The Pennsylvania State University

The Graduate School

Department of Psychology

BASE RATES AND CORRELATES OF ABNORMAL NEUROPSYCHOLOGICAL TEST PERFORMANCE ON A CONCUSSION ASSESSMENT BATTERY AMONG HEALTHY COLLEGIATE ATHLETES

A Dissertation in

Psychology

by

Fiona H. Barwick

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2011

The dissertation of Fiona H. Barwick was reviewed and approved* by the following:

Peter A. Arnett

Associate Professor of Psychology

Dissertation Adviser

Chair of Committee

Frank G. Hillary

Associate Professor of Psychology

Kristin A. Buss

Associate Professor of Psychology

Elana Farace

Associate Professor of Neurosurgery and Health Evaluation

Melvin Mark

Psychology Department Head

*Signatures are on file in the Graduate School.

ABSTRACT

Concussion in sport represents a significant source of disability, loss of time, and financial cost among collegiate athletes. The current paradigm for diagnosing and managing sport-related concussion requires baseline neuropsychological evaluation. Pre-injury performance levels are established against which post-injury recovery can be compared. The baseline testing paradigm assumes that healthy athletes will perform better than injured athletes on a battery of neuropsychological tests. However, this assumption has been challenged by recent research showing that abnormally low or discrepant scores on at least some tests in a neuropsychological assessment battery is normal for healthy individuals. To date, such research has focused on older adults and children rather than healthy younger adults such as collegiate athletes. The current study proposed to fill this gap by examining base rates and correlates of abnormally poor neuropsychological test performance on a baseline concussion assessment battery among a large sample of collegiate athletes. Consistent with studies looking at older adults and children, this study found that obtaining some abnormally low or discrepant scores on a multiple test battery is common among healthy collegiate athletes and is dependent upon the stringency of cutoff criteria. Factors correlated with abnormally poor neuropsychological test performance among healthy collegiate athletes at baseline assessment include gender, IQ, and time point in test battery. Information tables are provided, which include base rates for abnormally poor test performance across a multiple test battery among collegiate athletes by gender. Clinical examples illustrate how these information tables can be used in the interpretation of baseline and post-concussion test results. Incorporating such information will help to minimize misdiagnosis of cognitive impairment, refine clinical test inference, and improve return to play decisions for all collegiate athletes.

TABLE OF CONTENTS

APPENDIX F: LIST OF TABLES

APPENDIX G: LIST OF FIGURES

Chapter 1: Introduction

Overview of sport-related concussion

Athletes in any sport, especially those in high contact sports, run the risk of sustaining a concussion. Approximately 300,000 sport-related brain injuries occur in the United States each year, and over 3700 concussions are estimated to occur annually among collegiate athletes (Centers for Disease Control and Prevention: CDCP, 1997; Hootman, Dick, & Agel, 2007). These figures likely underestimate the true incidence rate of sports-related concussion because athletes and sports personnel sometimes fail to recognize or report their injuries (Collins et al., 1999; McCrea, Hammeke, Olsen, Leo, & Guskiewicz, 2004). Significantly, concussion rates among 15 collegiate sports have doubled over the past decade and a half (Hootman et al., 2007). Concussive injury thus represents a significant source of disability, loss of time, and financial cost among all athletes, including collegiate athletes.

International consensus guidelines have defined sport-related concussion as a temporary alteration in consciousness caused by complex pathophysiological processes, which result from traumatic biomechanical forces acting on the brain (Barth, Freeman, Broshek, & Varney, 2001; Giza & Hovda, 2004; McCrory et al., 2005). By definition, structural neuroimaging finds no abnormalities after concussive injury, but subsequent symptoms can be numerous and diverse: headache, dizziness, balance problems, sleep disturbance, fatigue, memory difficulties, attention deficits, slowed processing speed, loss of consciousness, posttraumatic amnesia, irritability, and mood changes (Aubry et al., 2001; Barth et al., 1989; Broglio & Puetz, 2008; Macciocchi, Barth, Alves, Rimel, & Jane, 1996; Moser et al., 2007). Although post-concussion symptoms typically resolve within two to fourteen days, some athletes recover more slowly, especially those who have sustained multiple concussions, suggesting that post-concussion effects might not always be as transient or reversible as believed (Barth et al., 1989; Belanger & Vanderploeg, 2005; Iverson, Brooks, Collins, & Lovell, 2005; McCrea et al., 2003).

The current paradigm for diagnosing and managing sports-related concussion among athletes, endorsed by the International Symposia on Concussion in Sport (Aubry et al., 2001; McCrory et al., 2005, 2009) and the National Athletic Trainers' Association (Guskiewicz et al., 2004), incorporates baseline neuropsychological testing as an integral part of any post-concussion evaluation. In this model, all athletes undergo neuropsychological testing prior to the start of practices or games in order to establish a pre-injury performance level against which post-injury recovery can be compared. After a concussion, the athlete undergoes serial reassessment until neuropsychological performance and reported symptoms return to baseline levels, at which point the athlete is permitted to return to play (Barth et al., 1989; Macciocchi et al., 1996). Baseline neuropsychological evaluation is thus crucial to the diagnosis and management of sport-related concussion.

Typically, baseline testing is designed to assess the functional domains that are most frequently affected in sports-related concussion, such as reaction time, information processing speed, attention, memory, executive functioning, symptoms, and mood. Baseline testing also helps to control for non-injury factors that can affect neuropsychological performance, such as gender, age, education level, intellectual ability, learning or attention deficits, and history of prior concussion (Binder, Iverson, & Brooks, 2009; Echemendia & Julian, 2001; Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008). Underlying the baseline testing paradigm is the assumption that healthy athletes will not perform as poorly as injured athletes on a battery of neuropsychological tests.

However, studies examining base rates of abnormally poor neuropsychological test performance among healthy older adults provide evidence that unusually low scores are not uncommon in these populations. No matter how abnormally poor performance is defined – as test scores that are 1, 1.5 or 2 standard deviations below the mean or as discrepancies between test scores of 2 or more standard deviations – healthy adults frequently obtain abnormally low test scores and demonstrate abnormally large discrepancies between tests. In fact, some researchers have concluded that abnormally poor performance on at least some tests in a neuropsychological assessment battery is normal for healthy adults (Binder et al., 2009; Schretlen et al., 2008).

However, most of the studies examining the frequency with which abnormally low test scores occur on a neuropsychological assessment battery in healthy populations have focused on older adults or children and adolescents. Although statistical models have been developed for estimating the probability that abnormally low scores will occur across a multiple test battery, the predictive accuracy of these models for younger adults has yet to be adequately evaluated (Ingraham & Aiken, 1996; Crawford, Garthwaite, & Gault, 2007). A better understanding of the base rates of abnormally poor test performance on a neuropsychological assessment battery administered to healthy young adults is needed, along with an examination of the predictive accuracy of statistical models for estimating the probability of abnormally poor test performance in this population.

This need is especially relevant to healthy collegiate athletes who undergo baseline neuropsychological testing as part of a program for diagnosing and managing sport-related concussion. Knowing how frequently abnormally low or discrepant test scores occur among healthy collegiate athletes on a baseline concussion assessment battery will aid in the interpretation of test results at both baseline and post-concussion evaluations. For example, knowledge of base rates could reduce the probability of inferring impairment in an athlete on the basis of one or two abnormally low test scores when such results might be common in the collegiate athlete population. The identification of possible factors associated with abnormally poor baseline performance among healthy collegiate athletes, including factors related to test administration or to athlete characteristics, might also help when interpreting results.

Thus, understanding base rates and correlates of abnormally low or discrepant test scores at baseline neuropsychological assessment among healthy collegiate athletes might lead to better clinical inferences and safer return-to-play decisions for all collegiate athletes. Because base rate information usually is not provided even for commonly used neuropsychological test batteries and because prior studies have focused on healthy older adults or children, this study calculates the base rates of abnormally poor or discrepant test performance on a baseline concussion assessment battery in a sample of healthy collegiate athletes. It then compares these base rates to statistical estimates of abnormally poor performance among collegiate athletes on such a battery.

Study results address three critical questions. First, information describing the maximum discrepancy between lowest and highest standardized test scores, along with the percentage of athletes showing test scores that are 1, 1.5, or 2 standard deviations (SD) below the mean, is used to determine the importance of taking into account base rates of abnormally low or discrepant test scores when interpreting the results of baseline concussion assessment among collegiate athletes. Second, information on those aspects of test administration or athlete characteristics – for example, time points in the test battery, types of neuropsychological tests, cognitive domains assessed, and athlete mood or personality variables – that are associated with abnormally low or discrepant test scores is used to ascertain whether such correlates might help in the interpretation of baseline performance among collegiate athletes. Finally, all study results are integrated to determine whether and how to incorporate information regarding base rates and correlates of abnormally poor baseline test performance into post-concussion evaluations and return to play decisions.

Epidemiology of sport-related concussion

Mild traumatic brain injury (mTBI), or concussion, represents a significant source of disability, loss of time, and financial cost among athletes at all levels of sports participation. Athletes in all sports run a greater risk of mTBI, with almost 300,000 athletes per year sustaining concussive injuries (CDCP, 1997, 2007). This rate likely underestimates the true incidence of sport-related concussion, as the definition on which the Center for Disease Control and Prevention bases its statistics requires loss of consciousness (LOC) when concussive injury typically involves no LOC (Moser et al., 2007). Incidence studies also rely upon sports participants and personnel to accurately identify and report all concussions, but athletes sometimes fail to recognize or report their injuries (Collins et al., 1999; Delaney, Lacroix, Leclerc, & Johnston, 2002; McCrea et al., 2004), and sports personnel may lack adequate training for detecting concussion (Macciocchi, 2006).

Data from the National Collegiate Athletic Association Injury Surveillance System (NCAA-ISS), which provides information on injury trends and exposure rates in intercollegiate athletics, show that in the 16-year period between 1988 and 2004 an average of 563 concussions occurred per year across 15 sports, and an estimated average of 3753 concussions occurred annually among athletes (Hootman et al., 2007). Although concussive injuries constitute approximately 5% of all injuries for most sports and competition levels (Covassin et al., 2003; Powell & Barber-Foss, 1999), they represent a larger percentage in men's football (6%), women's lacrosse (6.3%), men's ice hockey (7.9%), and women's ice hockey (18.3%).[1] These same sports show higher injury rates when the number of injuries is calculated per 1000 athlete exposures, where athlete exposure is defined as one athlete participating in one practice or game (Hootman et al., 2007). Other epidemiological studies have found similar percentages for these sports (Guskiewicz, Weaver, Padua, & Garrett, 2000; Macciocchi et al., 1996). Across all sports, concussive and other injuries are more likely to occur during games than during practice (Covassin, Swanik, & Sachs, 2003; Hootman et al., 2007).

According to reports (Hootman et al., 2007), concussion rates among 15 collegiate sports have increased significantly over the 16-year period between 1988 and 2004, roughly doubling from .17 per 1000 athlete exposures in 1988-1989 to .34 per 1000 athlete exposures in 2003-2004. Research groups caution, however, that this increase might reflect factors other than, or in addition to, an actual increase in the number of concussions per athlete exposure. For example, Title IX has increased sports

[1] Hootman and colleagues (2007) point out that the NCAA-ISS has collected data from women's ice hockey for only four years (2000-2004) and thus should be interpreted with caution.

participation among women by 80% as well as in sports participation overall. In addition, the number of certified athletic trainers working in the collegiate setting has increased by 86% between 1995 and 2005 (Covassin et al., 2003; Hootman et al., 2007; Macciocchi, 2006). These factors might increase the absolute number of athletes injured as well as the number of injuries reported without contributing to a proportionate increase in overall concussion rate.

Definition of sport-related concussion

According to consensus guidelines developed by international symposia, concussion is defined as “a complex pathophysiological process affecting the brain, induced by traumatic biomechanical forces” (Aubry et al., 2001; McCrory et al., 2005, 2009). Concussive injury results when a blow to the head or body generates a rapid change in the head’s velocity over a short period of time. The abrupt acceleration and deceleration forces produced by impact move the brain within and against the skull in a linear and rotational fashion. This movement causes the shearing of tissue and the stretching or tearing of axons, leading to diffuse axonal injury (Barth et al., 2001; Lampert & Hardman, 1984). Even for mild concussion, such damage leads to complex neurophysiological changes at the cellular and vascular level in the seconds, minutes, and hours following injury, a process described by Giza and Hovda as a “post-concussive neurometabolic cascade” involving changes in ionic equilibrium, neurotransmission, energy metabolism, and cerebral blood flow (Hovda et al., 1995; Giza & Hovda, 2004).

Symposia definitions emphasize that a concussion can result not only from a direct blow to the head but from a blow sustained elsewhere on the body. They also highlight the fact that concussive injury may or may not involve loss of consciousness and that subsequent symptoms are short-lived, usually resolving spontaneously within seven to ten days. The definition of concussive injury as a functional disturbance rather than a structural injury reflects the fact that standard structural neuroimaging techniques such as computerized tomography or magnetic resonance imaging typically reveal no abnormalities in brain tissue (Aubry et al., 2001; McCrory et al., 2005, 2009). Although more advanced neuroimaging techniques currently remain in the experimental stage for the investigation of sport-related concussion (Davis, Iverson, Guskiewicz, Ptito, & Johnston, 2009; Johnston, Ptito, Chankowsky, & Chen, 2001), functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) studies have identified possible neurophysiological abnormalities associated with clinical recovery (Chen, Johnston, Frey, Petrides, Worsely, & Ptito, 2004, 2008; Davis et al., 2009; Jantzen, Anderson, Steinberg, & Kelso, 2004; Lovell et al., 2007).

Symptoms and recovery in sport-related concussion

The temporary alteration in consciousness that defines concussive injury may be accompanied by loss of consciousness and posttraumatic amnesia, although the vast majority of sport-related concussions occur without either LOC or PTA (Guskiewicz et al., 2000; Macciocchi et al., 1996; McCrea et al., 2003; McCrory et al., 2005). Other common symptoms include headache, dizziness, nausea, balance problems, memory difficulties, sleep disturbance, attention and concentration deficits, fatigue, irritability, and mood changes (Aubry et al., 2001; Delaney et al., 2002; Erlanger, Kaushik, et al., 2003; Macciochi et al., 1996; McCrory & Johnston, 2002).

Athletes who sustain a concussive injury display cognitive decrements on neuropsychological tests which may last hours, days, or sometimes weeks after injury (Belanger & Vanderploeg, 2005; Broglio & Puetz, 2008; Erlanger, Feldman et al., 2003; Moser et al., 2007). The cognitive domains most affected include reaction time, information processing speed, attention/working memory, learning and longer-term memory, and executive functioning (Barth et al., 1989; Collie, Makdissi, Maruff, Bennell, & McCrory, 2006; Echemendia, Putukian, Mackin, Julian, & Shoss, 2001; Hinton-Bayre, Geffen, Geffen, McFarland, & Friis, 1999; Guskiewicz, Ross, & Marshall, 2001; Lovell & Collins, 1998; Maddocks & Saling, 1996; McCrea, Kelly, Randolph, Cisler, & Berger, 2002). Concussed athletes may also experience changes in mood, including increased symptoms of irritability, anxiety, and depression (Broshek & Fremman, 2005; Chen et al., 2008).

The symptoms and cognitive decrements that occur after sport-related concussion appear to be transient and reversible for the majority of injuries, typically resolving within two to fourteen days after injury (Barth et al., 1989; Belanger & Vanderploeg, 2005; Bleiberg et al., 2004; Iverson et al., 2005; Lovell, Collins, Iverson, Johnston, & Bradley, 2004; Macciocchi et al., 1996; McClincy, Lovell, Pardini, Collins, & Spore, 2006; McCrea et al., 2002, 2003). However, some athletes recover more slowly. Evidence points to a greater risk of adverse effects, including more severe and long-lasting symptoms as well as the possibility of cumulative and permanent neurodegenerative effects, for athletes who have sustained multiple concussions (Cantu, 2003; Collins, Lovell, Iverson, Cantu, Maroon, & Field, 2002; Covassin, Stearne, & Elbin, 2008; Delaney et al., 2002; Guskeiwicz et al., 2003, 2005, 2007; Iverson, Gaetz, Lovell, & Collins, 2004; Killam, Cautin, & Santucci, 2005; Moser, Schatz, & Jordan, 2005; Schwarz, 2009). Other factors that may be associated with differences in post-concussion symptom levels include gender (Broshek, Kaushik, Freeman, Erlanger, Webbe, & Barth, 2005; Dick, 2009; Gaetz & Iverson, 2009), type of sport (Schulz et al., 2004), presence or absence of diagnosed learning disability (Collins et al., 1999), and degree of motivation or effort exerted (Bailey, Echemendia, & Arnett, 2006).

Evaluation of sport-related concussion: Baseline neuropsychological assessment

Two decades ago, mild traumatic brain injury was attributed mostly to psychological factors and largely ignored by health care professionals. As research began to show that even mild injuries could impair cognitive functioning, however, neuropsychological evaluation became an important tool in diagnosing head injuries, including sport-related concussion (Barth, Macchiocchi, Giordani, Rimel, Jane, & Boll, 1983; Rimel, Giordani, Barth, Boll, & Jane, 1981). Barth and colleagues at the University of Virginia expanded the use of neuropsychological testing in sport-related concussion by developing a paradigm that incorporated baseline assessment as an integral part of the post-concussion evaluation (Barth et al., 1989; Macciocchi et al., 1996). In their Sports as Laboratory Assessment Model (SLAM), athletes undergo neuropsychological testing prior to the start of practices or games to establish a pre-injury performance level against which post-injury recovery can be compared. If athletes sustain a concussion, they undergo serial re-evaluation until post-concussion cognitive performance and reported symptoms return to baseline levels, at which point clinical practice deems it safe to return them to play.

The model developed by Barth and colleagues is now considered the "gold standard" for assessment of professional, collegiate, and high school athletes who have sustained a concussion (Barth, Broshek, & Freeman, 2006). It is endorsed by both the International Symposia on Concussion in Sport (Aubry et al., 2001; McCrory et al., 2005, 2009) and by the National Athletic Trainers' Association (Guskiewicz et al., 2004) as part of a multidisciplinary approach to managing sport-related concussion. Although not yet widely used at the high school and collegiate levels because of the significant resources required, the SLAM model has been adopted by some universities, including Pennsylvania State University's Concussion Program, which was established in 1995 (Echemendia & Julian, 2001).

Baseline neuropsychological evaluation is designed to assess the cognitive, mood, and behavioral domains that are most frequently affected in concussion. It also helps to control for non-injury factors that can affect cognitive performance or symptom report, such as gender, intellectual ability, and learning or attention deficits. Because it establishes a reference point against which post-injury changes can be compared, baseline neuropsychological testing allows the more accurate detection of cognitive deficits resulting from concussive injury. Such information is useful in diagnosing sport-related concussion, tracking recovery, and making safe of return to play decisions (Echemendia & Julian, 2001).

Abnormal neuropsychological test performance in sport-related concussion

But how are cognitive deficits best defined? In the context of a neuropsychological evaluation, cognitive deficits are seen as commensurate with abnormally low test scores. However, definitions of "abnormally low" abound both in the neuropsychological research literature and in clinical practice. Individual scores can be defined as abnormally low if they are 1, 1.5, or 2 SD below the mean, or if they fall in the "borderline" (below the 10th percentile i.e., 1.3 SD below the mean) or "impaired" (below the 5th percentile i.e., 1.7 SD below the mean) ranges. Abnormal performance can also be defined as the amount of "test scatter," or the discrepancy between highest and lowest standardized scores across a test battery (Strauss et al., 2006). However, few guidelines exist for determining what constitutes abnormally discrepant neuropsychological test scores, especially across a multiple test battery.

Furthermore, a test score that is abnormally low may appear not infrequently in the population, as the magnitude of statistical difference between two test scores in a battery yields no information about the frequency of that difference (Lezak et al., 2004). In other words, an abnormally poor neuropsychological test performance, as represented by a statistically significant deviation from the population mean, is not always clinically significant and does not always correlate with clinical impairment (Binder et al., 2009). For example, a 15-point absolute difference between Verbal and Performance IQ on the WAIS-III is statistically significant (i.e., $p < .05$), but it occurs in almost 20% of the WAIS-III normative sample.

Such considerations are relevant to the use of baseline testing with healthy collegiate athletes because it appears that healthy individuals not infrequently obtain abnormally low or discrepant test scores on common neuropsychological assessment batteries. For example, normative samples from the Wechsler Adult Intelligence Scale-III (WAIS-III) and Wechsler Adult Intelligence Scale-IV (WAIS-IV) had an average maximum discrepancy between highest and lowest scaled scores of over 2 SD when all tests were administered. On both batteries, more than three-fourths of the normative samples earned at least one test score that was 1 SD below the mean, and almost half the samples earned at least one test score that was 1.5 SD below the mean. More than half had three or more test scores that were 1 SD below the mean, and more than one-quarter had two or more scores that were 1.5 SD below the mean (Binder et al., 2009; Wechsler, 1997).

Similar findings appear for less commonly used assessment batteries, such as the Halstead-Reitan Battery. On at least one test from this 25-test battery, 87% of the normative sample obtained scores 1 SD below the mean, 59% obtained scores 1.5 SD below the mean, and 28% obtained scores 2 SD below the mean (Heaton, Miller, Taylor, & Grant, 2004). Studies examining test performance on flexible neuropsychological assessment batteries administered to diverse samples find results similar to performance on fixed batteries. For example, 65% of 197 adults aged 20 to 92 years showed a discrepancy greater than 3 SD between their highest and lowest standardized test scores on a 15-test neuropsychological battery (Schretlen et al., 2003).

Studies focusing on older adults also find high rates of abnormally low or discrepant scores on multiple test batteries. In the WMS-III standardization sample, which consisted of 550 adults aged 55 to 87 years, 64% obtained one or more scores at or below the 16th percentile, or 1 SD below the mean, and 26% obtained one or more scores at or below the 5th percentile, or 1.7 SD below the mean (Brooks et al., 2008). On ten memory scores derived from the Neuropsychological Assessment Battery (Stern & White, 2003), 56% of 742 adults from the standardization sample, aged 55 to 79 years, had one or more test scores 1 SD below the mean, and 31% had one or more test scores 1.7 SD below the mean (Brooks et al., 2007). In another study, 73% of 132 healthy older adults aged 50 to 79 years had at least one test score out of ten in the borderline range or 1.3 SD below the mean, while 48% had two or more such scores; 37% had at least one test score in the impaired range or 1.7 SD below the mean while 24% had two or more such scores (Palmer et al., 1998).

Similar findings also occur with children and adolescents. In the standardization sample for the Children's Memory Scale, which consisted of 1000 healthy children and adolescents between 5 and 16 years of age, 38% obtained at least one test score out of six that was 1 SD below the mean, and 22% obtained at least one test score out of six that was 1.7 SD below the mean (Brooks et al., 2009). Using the standardization sample for the NEPSY-II, which consisted of 1200 healthy children between the ages of 3 and 16 years, Brooks and colleagues (2010) found that 41% of 3-4 year olds, 37% of 5-6 year olds, and 52% of 7-16 year olds had at least one test score at or below the 10^{th} percentile (Brooks et al., 2010).

Whether the test battery is fixed or flexible and whether cut-off criteria are 1 SD below the mean or lower, abnormally low scores and unusually large discrepancies between scores appear frequently in samples of healthy children and adolescents (Brooks et al., 2009, 2010) as well as in healthy older adults (Brooks et al., 2007, 2008; Palmer et al., 1998). Such findings have led some of these authors to conclude that few participants demonstrate consistent performance across a neuropsychological assessment battery that includes multiple tests, where consistency is defined as all scores falling within 1 SD of each other (Schretlen et al., 2003). In fact, they assert that abnormal performance on some proportion of neuropsychological tests in a battery is psychometrically normal in healthy adults (Binder et al., 2009).

However, most studies examining base rates of abnormally poor neuropsychological test performance have focused on older adults or children. Those studies that have included younger adult participants have examined base rates of abnormally poor test performance across, but not within, age ranges. Although statistical models have been developed for estimating the probability that healthy adults will obtain abnormally low scores across multiple tests in an assessment battery (Ingraham & Aiken, 1996; Crawford et al., 2007), the predictive accuracy of these models for younger adults has been examined in only one study thus far, which was conducted with a non-athlete college sample (Axelrod & Wall, 2007). Furthermore, research has identified some correlates of abnormally poor neuropsychological test performance among healthy adults – for example, the frequency of abnormally low scores is positively correlated with the number of tests administered and negatively correlated with IQ and stringency of cutoff levels (Binder et al., 2009; Schretlen et al., 2008) – but such correlates have yet to be examined in a healthy collegiate athlete population.

The relevance of these findings to collegiate athletes is clear. Given the frequency with which healthy older adults and healthy children obtain abnormally low or discrepant test scores on neuropsychological assessment batteries, and given the little that is known about such frequencies among healthy younger adults, we need to investigate test score variability in the latter population. This need is even greater with collegiate athletes because of the role of baseline assessment in the diagnosis and management of sport-related concussion. In order to understand and accurately interpret abnormally poor performance at post-concussion evaluation, we need to characterize normal performance at baseline assessment among healthy collegiate athletes, including the extent of performance variability on a typical neuropsychological concussion test battery. To properly characterize normal performance, we need to establish base rates and identify possible correlates of abnormally low or discrepant test scores among collegiate athletes at baseline assessment.

Proposed Study

With these considerations in mind, the current study examines abnormally poor neuropsychological test performance on a baseline concussion assessment battery among a group of healthy collegiate athletes. To better characterize baseline neuropsychological test performance in this population, the investigation calculates the average maximum discrepancy between highest and lowest standardized test scores for each athlete as well as the percentage of athletes showing scores that are 1, 1.5, or 2 SD below the mean on one or more tests. The study also examines whether abnormally low test scores emerge at particular points in the battery (e.g., beginning, middle, or end), on specific modes of test administration (e.g., computer-administered or pencil-and-paper), or in specific cognitive domains (e.g., processing speed or memory)?

The study examines athlete demographic and psychological factors to discover whether any are correlated with abnormally poor neuropsychological test performance at baseline and to determine how much of the variance is accounted for by these factors. It also compares the percentage of athletes earning abnormally low test scores in the actual sample to percentage estimates derived from other statistical methods, such as binomial probability distribution or Monte Carlo simulation, to determine the accuracy of estimation methods. Finally, the athlete sample is divided into groups to determine whether differences in gender or intellectual functioning are associated with differences in abnormally poor neuropsychological test performance at baseline.

Study results address three critical questions. First, information describing the maximum discrepancy between lowest and highest standardized test scores, along with the percentage of athletes showing test scores that are 1, 1.5, or 2 standard deviations (SD) below the mean, is used to determine the importance of taking into account base rates of abnormally low or discrepant test scores when interpreting the results of baseline concussion assessment among collegiate athletes. Second, information on those aspects of test administration or athlete characteristics – for example, time points in the test battery, types of neuropsychological tests, cognitive domains assessed, and athlete mood or personality variables – that are associated with abnormally low or discrepant test scores is used to ascertain whether such correlates might help in the interpretation of baseline performance among collegiate athletes. Finally, all study results are integrated to determine whether and how to incorporate information regarding base rates and correlates of abnormally poor baseline test performance into post-concussion evaluations and return to play decisions.

Chapter 2: Methods

Penn State Concussion Program

The Penn State University Concussion program was established in 1995 (Echemendia & Julian, 2001) in order to implement the neuropsychological evaluation paradigm developed by Barth and colleagues at the University of Virginia (Barth et al., 1989, 2006). Procedures for conducting baseline assessment of incoming athletes and acute assessment of concussed athletes were established. The assessment battery, which is the same for both baseline and acute evaluations, was designed to be administered easily and to include tests with adequate psychometric properties. The tests evaluate functional domains most frequently affected in sport-related concussion: reaction time, information processing speed, attention, verbal and nonverbal learning and memory, executive functioning, and mood and symptom report. Both pencil-and-paper as well as computerized tests are included in the battery and, where possible, alternate test forms are used to conduct serial evaluations.

Participants

Participants were 612 student athletes who had in the past or were presently attending Penn State University and whose teams participate in the Penn State Concussion Program. The teams include men’s football, ice hockey, and wrestling, along with men’s and women’s soccer, lacrosse, and basketball. Although Penn State University also supports teams in baseball, softball, volleyball, gymnastics, cheerleading, and rugby, athletes from these sports do not participate consistently in the concussion testing program.

Procedures

Incoming freshman athletes typically undergo baseline testing prior to the start of practices and games. Testing was conducted on campus, either at the Penn State Psychological Clinic or at the Lasch Football Building. Athletes were usually tested individually in a quiet, well-lit room by graduate or undergraduate assessors who had been trained in administration of the concussion test battery. Testing began only after informed consent was obtained and clinical paperwork was completed. All testing and procedures were conducted in accordance with the requirements of Pennsylvania State University's Institutional Review Board and with the ethical guidelines of the American Psychological Association.

Measures

Contextual measures

Background Information Form. Participants were asked to provide information about sex, age, race/ethnicity, education, hand dominance, sport type, drug/alcohol use, and previous diagnoses of learning disability or attention deficit disorder.

Previous Head Injury Questionnaire: Participants were given a definition of concussive injury and its associated symptoms and then asked to indicate the number of concussions they had sustained over their lifetime. For each concussion, participants provided information about date, age, circumstances, symptoms, treatment, and recovery.

Wechsler Test of Adult Reading (WTAR: Holdnack, 2001). Participants were shown a list of 50 irregularly pronounced words which they were asked to say aloud (e.g., aisle, porpoise, exigency), with one point awarded for each word pronounced correctly. The raw score was then converted to an estimated WAIS-III Full Scale IQ using normative tables stratified by age, gender, ethnicity, and education. The WTAR provides an estimate of overall intellectual functioning for adults aged 16-89 years. Performance on the WTAR shows excellent internal consistency ($\alpha \geq .90$) and test-retest reliability ($r_{12} \geq .90$). It also shows good correlations with WAIS-III FSIQ scores for ages 18 to 74 years ($r = .70 - .80$).

Post-Concussive Symptom Scale (PCSS: Lovell & Collins, 1998; Lovell et al., 2006). Participants were asked to rate their current state on a 22-item, self-report inventory listing common post-concussion symptoms using a 6-point scale where 0 = none and 6 = severe. Ratings were totaled across all items to assess symptom severity. Normative data for the PCSS from a sample of 1,039 university athletes show excellent internal consistency (α =. 88).

Beck Depression Inventory-Fast Screen (BDI-FS: Beck et al., 2000). Participants were asked to rate common depression symptoms during the past two weeks using a 3-point scale. Ratings were totaled across all 7 items, with higher ratings indicating greater symptom severity. The BDI-FS shows good internal consistency (α = .84) and good concurrent validity with other depressive measures such as the BDI-II (r = .85) and CES-D (r = .86).

NEO-Five Factor Inventory (NEO-FFI: Costa & McCrae, 1992). Participants were asked to rate 60 self-statements (e.g., I am not worrier, I usually prefer to do things alone) using a 5-point scale ranging from “Strongly Agree” to “Strongly Disagree”. Subscale totals were derived for each of five personality factors: neuroticism, extraversion, openness, agreeableness, and conscientiousness. Factors on the NEO-FFI show good internal consistency (α = .86, .77, .73, .81, and .68 for neuroticism, extraversion, openness, conscientiousness, and agreeableness respectively) as well as high correlations with factors on the full NEO-PI measure (r = .75 - .89).

Cognitive measures

Digit Span from Wechsler Adult Intelligence Scale-III (WAIS-III: Weschler, 1997). Participants listened to sequences of numbers read aloud and were asked to repeat the sequences forward and backward. Sequences increased in length across trials. Performance was assessed using combined total correct. The forward condition is hypothesized to measure short-term auditory attention whereas the backward condition is hypothesized to measure auditory working memory.

Brief Visuospatial Memory Test-Revised (BVMT-R: Benedict, 1997). Participants viewed a set of six figures for ten seconds which they were then asked to draw from memory. Three trials were administered in succession, with the figures displayed each time, and a delayed recall trial was administered after twenty minutes. Figures were scored on accuracy and location, with performance assessed for total correct, yielding scores for immediate and delayed recall. The BVMT-R is intended to measure visual learning and memory. It has six alternate forms for conducting repeated assessments. It shows adequate to good test-retest reliability for the different forms (r_{12} = .65 to .80 for total recall), excellent interrater reliability (k = .90), and good concurrent validity with other memory measures (r_{xy} = .65 for HVLT and .80 for WMS-R Visual Reproduction).

Hopkins Verbal Learning Test-Revised (HVLT-R: Brandt, 1991; Brandt & Benedict, 2001). Participants were read aloud a list of twelve nouns grouped into three conceptual categories (e.g., animals, foods, and fuels) and asked to recall as many words as they were able. Three trials were administered in succession followed by a delayed trial administered after twenty minutes. Performance was assessed for total correctly recalled, yielding scores for immediate and delayed recall. The HVLT-R is intended to measure non-contextualized verbal learning and memory. It has six alternate forms for conducting repeated assessments. It demonstrates adequate test-retest reliability (r_{12} = .74 for total recall) and adequate concurrent validity with other verbal memory measures (r_{xy} = .72 for CVLT Total Learning and .65 for Wechsler Logical Memory).

PSU Cancellation Task (PSU-CX: Echemendia et al., 2001). Participants were asked to scan across rows of shapes, crossing out all those that matched a target shape given at the top of the page. Performance was assessed by the number of target shapes correctly crossed out in 90 seconds minus the number of errors. The PSU-CX has four equivalent forms for conducting repeated assessments and demonstrates adequate test-retest reliability (Echemendia et al., 2001). It is intended to measure simple and complex attention, visual scanning, and processing speed.

Symbol Digit Modalities Task (SDMT: Smith, 1991). Participants were asked to pair numbers with abstract symbols as rapidly as possible according to a key provided at the top of the page. Performance was assessed by tallying the total number correct after 90 seconds as well as the total number of errors. The SDMT is designed to measure simple and divided attention, visual scanning, and psychomotor processing speed. It has six alternate forms for conducting repeated assessments. It shows good test-retest reliability (r_{12} = .80) and excellent concurrent validity with other processing speed tasks (r_{xy} = .78 - 91 for Wechsler Digit Symbol Coding).

Vigil (Cegalis & Cegalis, 1994). Participants were presented with a computer-administered continuous performance task in which they were required to press the space bar as quickly as possible, first in response to a designated target letter and then only if the designated target letter was preceded by another specified letter. Performance was assessed for average response time as well as total number of omissions and commissions. The Vigil is intended to assess sustained attention, response time, and response inhibition. It shows good test-retest reliability and concurrent validity with other measures of reaction time and processing speed (Cegalis & Cegalis, 1994).

Stroop Color-Word Test (Stroop: Trenerry et al., 1989). Participants were first asked to read aloud a list of 112 color words as quickly as possible (Color-Naming condition). They were then asked to state aloud the ink color in which the words were printed as quickly as possible (Color-Word condition). Performance was assessed using the completion time for each condition along with the number of errors. The Stroop is hypothesized to measure attention, processing speed, mental tracking, response inhibition, response shifting, and response monitoring. It demonstrates good test-retest reliability for the Color-Naming condition (r_{12} = .84), adequate test-retest reliability for the Color-Word condition (r_{12} = .73), and only moderate correlation between Color-Naming and Color-Word conditions, suggesting that the conditions measure similar but not identical functions.

Trail Making Test (TMT: Partington & Leiter, 1949) and Comprehensive Trail-Making Test (Reynolds, Pearson, & Voress, 2002). Participants were first asked to draw a line as quickly as possible sequentially connecting a series of numbers. They were then asked to alternate between sequentially connecting numbers and letters. Empty and filled distracter circles increased task difficulty. Performance was assessed using the completion time for each condition along with the number of errors. The TMT is hypothesized to measure attention, processing speed, visual scanning, sequencing, and response inhibition. It shows adequate test-retest reliability ($r_{12} \geq .70$) for all subtests and excellent interrater reliability (k ≥ .90). Different conditions correlate modestly with each other, suggesting that they measure different functions.

Immediate Post-concussion Assessment and Cognitive Testing (ImPACT: Lovell, Collins, Podell, Powell, & Maroon, 2000). Participants completed a 25-minute computer-administered assessment battery consisting of three parts: demographic data, symptom checklists, and neuropsychological tests. The demographic section requests sports, medical history, and previous concussion information. The symptom section consists of the Post-Concussion Symptom Scale. The neuropsychological tests comprise modules targeting attention, memory, processing speed, and reaction time. From these tests, five composite indices are generated: Verbal Memory, Visual Memory, Visuomotor Speed, Reaction Time, and Impulse Control. The overall assessment battery has shown adequate to good reliability and validity (Iverson, Lovell, & Collins, 2003; Iverson, Gaetz, Lovell, & Collins, 2005) as well as excellent sensitivity and specificity (Schatz, Pardini, Lovell, Collins, & Podell, 2006).

Data Analysis

Preliminary analyses

All statistical analyses were conducted using PASW Statistics 18.0 software package. Demographic data were obtained for the overall athlete sample, including age, sex, self-reported race/ ethnicity, sports type, hand dominance, self-reported prior diagnosis of learning disability or attention-deficit/hyperactivity disorder, current substance use, and number of prior concussions. Data were examined to determine the extent and pattern of missing variables using Missing Values Analysis as well as group comparisons of cases with and without missing values on all neuropsychological outcome measures. Case deletion or multiple imputation were considered depending upon the amount of missing data, with the intention of performing subsequent analyses on both complete cases only as well as on complete and imputed cases. Data on all measures serving as criterion or predictor variables were examined and, where necessary, transformed to ensure that requisite assumptions of normality, linearity, and homogeneity or homoscedasticity of variance were met.

Primary analyses

The average maximum discrepancy between lowest and highest test scores across the 17-test battery was determined by first using the mean and standard deviation from the overall athlete group[2] to calculate z-scores for each individual athlete on each neuropsychological outcome measure (Digit Span Total, HVLT Total Recall, HVLT Delayed Recall, BVMT Total Recall, BVMT Delayed Recall, SDMT Total Correct, SDMT Incidental Memory, TMT 1 and 2 Times, Stroop Color-Word and Color-Naming Times, PSU-Cancellation Total Correct, Vigil Average Delay, and Impact Verbal Memory, Visual Memory, Motor Speed, and Reaction Time Composites) and then subtracting the lowest z-score from the highest z-score. Where lower z-scores indicated better performance (e.g., reaction time tasks), scores were reversed so that lower scores always indicated worse performance. The percentage of athletes who obtained abnormally low test scores across all 17 tests was determined by using z-scores to assign a single point for each test on which an athlete scored 1, 1.5, or 2 SD below the mean and then by tallying points across all tests to create an index of abnormal scores for each cutoff level. The frequency of index scores was then used to determine the percentage of athletes showing test scores at least 1, 1.5, or 2 SD below the mean on one or more tests across the battery.

[2] The mean and standard deviation from the overall athlete group was used instead of published normative data because several of the measures lack universally accepted normative data, especially for younger age groups such as collegiate athletes, and because normative samples are likely to show demographic differences relative to the current sample which might make interpretation of results more difficult.

To determine whether the total number of abnormally low test scores differed significantly among sections of the battery, the 12 neuropsychological tests not administered by computer [3] were divided into three sections of approximately equal length: the first section comprised the four tests administered during the initial 15- to 40-minute period, the second section comprised the four tests administered during the middle 40- to 65-minute period, and the third section comprised the four tests administered during the final 65- to 90-minute period (See Appendix A). Repeated-measures ANOVA, accompanied by appropriate post-hoc comparisons, were conducted at each of the three cutoff levels (1, 1.5, and 2 SD below the mean), with the total number of abnormally low scores at the beginning, middle, and end of the test battery serving as the within-subjects factor. Similar analyses were conducted comparing the total number of abnormally low test scores between computer-administered and pencil-and-paper modes of test administration (see Appendix B). With the latter analyses, the number of abnormally low test scores in each mode was calculated as a proportion of the total number of tests included in that mode.

To determine whether abnormally low test scores across the battery differed significantly among specific cognitive domains, neuropsychological outcome measures were divided into domains of cognitive functioning based on either theoretical assumptions or empirical considerations (see Appendix C). Based on theoretical assumptions, tests were assigned to one of four cognitive domains: (a) attention/information processing speed; (b) verbal learning and memory; (c) visual learning and memory; and (d) executive functioning. For empirical considerations, all tests were subject to factor analysis and cognitive domains were created based on factor loadings, with the number of abnormally low test scores in each domain calculated as a proportion of the total number of tests included in that domain. Repeated measures ANOVA, accompanied by appropriate post-hoc comparisons, were

[3] Computerized tests are excluded because the majority of these tests (4/5) are administered at the same point in the assessment battery as part of the ImPACT, thus confounding test mode and time of administration.

conducted at each of the three cutoff levels (1, 1.5, and 2 SD below the mean), with the total number of abnormally low scores in each cognitive domain serving as the within-subjects factor.

Demographic or psychological factors that might be associated with abnormally low scores across the test battery were identified by examining correlations between athlete characteristics – including IQ, gender, sport, mood, and personality – and the total number of test scores 1, 1.5 and 2 SD below the mean. Variables that were significantly correlated with neuropsychological test performance at each cutoff level were examined for conformity with statistical assumptions, transformed as necessary, and then included as independent predictors in standard multiple regression analyses, with total number of test scores at 1, 1.5, and 2 SD below the mean serving as dependent variables.

The possible impact of gender differences or differences in intellectual functioning[4] on baseline neuropsychological test performance was evaluated using several methods. Independent samples t-tests or one-way ANOVA were used to compare gender or IQ groups on the average maximum discrepancy between highest and lowest standardized test scores. Proportional significance testing was used to compare gender or IQ groups on the percentage of athletes showing test scores at 1, 1.5, and 2 SD below the mean. Repeated measures ANOVA, with gender or IQ group as a between-subjects factor, was used to determine whether group differences occur in the total number of abnormally low test scores in different sections of the test battery, different modes of test administration, or different cognitive domains. Finally, demographic, mood, and personality variables that were significantly correlated with athlete gender or IQ were included as independent predictors in standard multiple

[4] The athlete sample was divided into three groups according to WTAR-estimated premorbid IQ: low average (IQ < 90); average (90 ≤ IQ < 110); and high average (IQ ≥ 110).

regression analyses for each group, with total number of test scores at 1, 1.5, and 2 SD below the mean serving as dependent variables.

Finally, binomial probability distribution methods (Ingraham & Aiken, 1996), which assume that test scores are uncorrelated, and Monte Carlo simulation methods (Crawford et al., 2007), which assumes that test scores are correlated, were used to estimate the percentage of abnormally low test scores at each cutoff level in the overall athlete sample. These estimates were then compared to percentages derived from the actual sample to determine the accuracy of the predictions yielded by these different methods for base rates of abnormally low test scores in a healthy young athlete sample undergoing a multiple-test neuropsychological assessment battery.

Chapter 3: Results

Preliminary Analyses

Demographics

Basic demographic data for the overall athlete sample are shown in Table 1. The sample consisted of 612 athletes, 74% male and 26% female, with an average age of 18.5 years and an average estimated Full Scale IQ of 103. The majority of athletes were right-hand dominant. The two largest ethnic groups were Caucasian-Americans (75%) and African-Americans (19%). Football players comprised almost one-third of the athlete sample (30%), and lacrosse and soccer players slightly more than one-fifth of the sample (22% and 21% respectively). Football, ice hockey, and wrestling included only male athletes, whereas soccer, lacrosse, and basketball included male and female athletes. Sports categorized as “Other” included rugby, baseball, softball, volleyball, track and field, cheerleading, and women’s ice hockey. The average age at which most of the athletes had begun to participate in organized sports was 6.6 years, with a range of 2 to 16 years. The average number of prior head injuries was .56, with a range of 0 to 7 prior concussions.

Approximately 10% of athletes in the sample reported some previous psychological testing. Almost 6% of the athletes reported diagnosed or possible attention-deficit/hyperactivity disorder, and almost 5% reported a diagnosed or possible learning disability. Over 5% of athletes left blank a question asking about weekly alcohol consumption. Of those athletes who answered the question, the average number of drinks per week was 2.8, with a range of 0 to 35 drinks. Slightly more than 3% of athletes did not answer questions asking about IV drug use or marijuana use. Of those who answered the questions, 95% denied any IV drug use and 88% denied any marijuana use; however, almost 8% reported occasional marijuana use.

Missing Values

During the years in which the Penn State Concussion Program has been ongoing, the program director, administration conditions, and assessment battery have all changed, most notably in late summer 2004 when a new program director took over. As a result of these changes, not all athletes were administered all tests. A Missing Values Analysis (MVA) showed that less than 1% of values were missing on demographic variables, including age, sex, race/ethnicity, sport, hemispheric dominance, LD/ADHD diagnosis, age at first sports participation, and number of prior concussions. However, more than 5% of values were missing (range 6.7% to 14.9%) for all neuropsychological test scores that were to serve as dependent variables in subsequent analyses. Furthermore, Little's MCAR test was significant, χ^2 (637) = 814, $p < .001$, indicating that data were not missing completely at random.

As one check, groups were created using missing and non-missing values on all neuropsychological tests that were to serve as dependent variables in subsequent analyses. These groups were then compared to see if any significant differences emerged on all other dependent variables. Most missing groups had only 0 to 4 members, and they failed to show any significant differences on outcome measures, including other neuropsychological test scores, compared to non-missing groups. However, groups missing values on BDI-FS total score (20 members), WTAR FSIQ (30 members), NEO-FFI trait scores (75 members), or ImPACT (8 members) showed significant differences on several measures, including HVLT-R Total and Delayed Recall, SDMT Total Correct, TMT 1 and 2, Stroop Color-Word and Color-Naming, and ImPACT Visual Motor Speed and Reaction Time Composites.

MVA separate variance t-tests also showed that, for several variables (e.g., WTAR FSIQ, all ImPACT indices, TMT 1 and 2, and Stroop Color-Word and Color-Naming), missing and non-missing groups significantly differed. MVA analyses identified only one pattern of missing data with more than 5% of cases missing, however, where NEO-FFI data were missing for 46 athletes. This measure had not been administered during the initial years of the concussion program under the first program director.

Furthermore, athletes missing the greatest number of tests (12% to 100%) were those who had undergone baseline concussion assessment under this first director.

Given this information, groups were created for athletes tested under the first program director (n = 117) versus those tested under the second program director (n = 495). Independent samples t-tests showed that athletes tested under the first program director were significantly older (19.0 compared to 18.4 years) and were reporting a significantly greater number of prior concussions (0.7 compared to 0.5 concussions) than athletes tested under the second program director. The two groups were then compared on neuropsychological outcome measures which were to be included in subsequent analyses. Groups differed significantly, or showed a trend toward significant differences, on several variables: HVLT-R Total and Delayed Recall, SDMT Total Correct, TMT 1 and 2, Stroop Color-Word and Color-Naming, PCSS Total Score, and ImPACT Visual Motor Speed and Reaction Time Composite indices.

Several of these measures were the same as the ones which differed significantly between missing and non-missing groups. More importantly, either the form or the administration for several of these measures had changed with the change in directors (e.g., HVLT, SDMT, and TMT) or with updated versions (e.g., ImPACT). It appeared likely that the changes that occurred in administration protocol, testing conditions, and battery composition with the 2004 change in directorship led to differences in neuropsychological test performance among athletes. Excluding from the sample those athletes tested under the first program director erased all differences between missing and non-missing groups on neuropsychological tests scores. Furthermore, the amount of missing data was reduced to less than 5% for all variables of interest, no patterns of missing data were identified, and Little's MCAR test was rendered non-significant, χ^2 (447) = 482, p > .01. Given MVA results as well the likely explanation for these results, a decision was made to exclude from all subsequent analyses those athletes who had been tested under the first program director rather than to impute missing data points, leaving a final sample size of n = 495.

Statistical Assumptions

Several neuropsychological test scores that served as dependent variables required some form of transformation to reduce skewness and kurtosis and thus meet assumptions of statistical hypothesis testing. BVMT Delayed Recall, and Impact Verbal Memory, Visual Memory, and Reaction Time Composites were square-root transformed. BDI Total Score, PCSS Total Score, SDMT Incidental Memory Total Correct, Stroop Color-Naming Time, and TMT 1 and 2 Times were log10 transformed. Vigil Average Delay and Stroop Color-Word Time were inverse transformed. Whereas for some of these transformed variables the improvements were substantial (e.g., PCSS Total, BDI Total, TMT 1 and 2 Time, and Stroop Color-Word and Color-Naming Times), for several others the improvements were only modest (e.g., Vigil Average Delay, BVMT Delayed Recall, SDMT Incidental Memory Total Correct, and Impact Verbal Memory, Visual Memory, and Reaction Time Composites). As transformation can complicate interpretation of results, analyses were tried both with and without transformed variables. Where results were similar, non-transformed variables were used to facilitate interpretation. Although there were univariate and multivariate outliers on several of the variables, outliers were retained given the large sample size.

Primary analyses: main questions

As shown in Table 2, among a sample of 495 athletes administered a 17-test battery, the average maximum discrepancy between lowest and highest test scores was 3.2 SD. The average lowest z-score was 1.8 SD below the mean and the average highest z-score was 1.4 SD above the mean. The percentage of athletes showing test scores at least 1, 1.5, or 2 SD below the mean on one or more tests across the battery is reported in Tables 3 and 4 and shown in Figures 1 and 2. Over three-fourths of the athletes scored at least 1 SD below the mean, over half of the athletes scored at least 1.5 SD below the mean, and over one-quarter of the athletes scored at least 2 SD below the mean on at least one test.

Thus, less than one-quarter of the athletes showed no abnormal scores on any tests in the battery at 1 SD below the mean and less than one-half of the athletes showed no abnormal scores on any tests in the battery at 1.5 SD below the mean. Cumulative percentages remained high even as the number of tests on which athletes obtained abnormal scores increased. For example, over one-quarter of the athletes showed test scores 1. 5 SD below the mean on two or more tests, and a similar percentage showed test scores 1 SD below the mean on four or more tests. Not surprisingly, the number of tests on which athletes scored significantly below the mean decreased with more stringent cut off levels.

Significant differences in the total number of abnormally low test scores appearing at the beginning, middle, or end of the test battery for each cutoff level are reported in Table 5 and shown in Figure 3. Repeated measures ANOVA, with total number of abnormally low scores in different sections of the test battery serving as the within-subjects dependent variable, showed significant differences between time points in the average number of test scores obtained by athletes at 1.5 and 2 SD below the mean but not at 1 SD below the mean. Huynh-Feldt adjustments were used to correct for non-sphericity. Follow-up pair-wise comparisons using Bonferroni adjustment indicated that at 1.5 SD and 2 SD below the mean, athletes obtained significantly fewer test scores at the end of the test battery than at the beginning or middle of the test battery, all $p < .05$. Effect sizes were small, however, with Cohen's d ranging from .15 to .22.

As the order of test administration was not counterbalanced – in other words, the same four tests were always given at the same point in the battery – test type was confounded with time point. Hence, follow-up paired-sample t-tests were conducted to compare performance on the four individual tests at the end of the battery with performance on the four individual tests at the beginning and middle of the battery. No significant differences emerged in average z-score between any of the tests (all $t < .25$, all $p > .10$). Paired samples t-tests showed no significant differences in the total number of test scores appearing at 1, 1.5, or 2 SD below the mean on computer-administered versus pencil-and-paper tests (all $t \leq 1.0$, all $p > .10$). Even when tests were divided into three administrative modes rather than two, for those requiring a computerized, written, or oral response (see Appendix B), no significant differences appeared between modes of test administration in the total number of test scores at 1, 1.5, or 2 SD below the mean (all $F < 1.5$, all $p > .10$).

To determine whether abnormally low test scores across the battery differed significantly among theoretically or empirically derived cognitive domains, three neuropsychological test scores were assigned to one of four cognitive domains based on theoretical assumptions (see Appendix C). These twelve tests, along with three additional ones (SDMT Incidental Memory, Stroop Color Word, and IMPACT Visual Motor Speed Composite), were then subject to principal factors extraction with promax oblique rotation (kappa = 2). Principal components extraction was used prior to principal factors extraction to estimate number of factors, presence of outliers, absence of multicollinearity, and factorability of the correlation matrices.

Almost 500 cases provided a very good sample size for factor analysis. Examination of pairwise scatter plots showed no evidence of curvilinearity. With $\alpha = .001$ as a cutoff level, 15 athletes produced scores that identified them as outliers. These cases were deleted from principal factors extraction and all subsequent analyses were run on a reduced data set of 480 athletes. The largest squared multiple correlation between variables, where each in turn serves as a dependent variable for all others, was .79, indicating no threat of multicollinearity. Two measures (Digit Span Total and PSU Cancellation Task Total) were deleted from subsequent analyses as they had Kaiser's measures of sampling adequacy < .60, indicating poor linear correlation with other variables. All other variables had Kaiser's measures of sampling adequacy > .70 with an overall KMO = .80.

Three factors were extracted which accounted for 50% of the variance. As indicated by factor loadings on the pattern matrix, factors were internally consistent, although factor loadings for some items were < .40. Because oblique rotation allowed factors to correlate modestly with each other, variables were not as clearly defined by factors, with six of the variables (TMT 2, SDMT Total Correct and Incidental Memory, and Impact Verbal Memory, Visual Memory, and Visual Motor Speed Composite) loading on more than one factor. Loadings of variables on factors, correlations between factors, communalities, and percents of variance are shown in Table 6. Variables are ordered and grouped by size of loading to facilitate interpretation. Loadings under .35 are suppressed, and interpretative labels are suggested for the three factors: speeded processing, visual memory, and verbal memory.

Repeated measures ANOVA comparing the total number of test scores obtained at 1, 1.5, and 2 SD below the mean in the four theoretically derived cognitive domains showed no significant differences between speeded processing, verbal memory, visual memory, and executive functioning domains (all univariate $F < 2.0$, all $p > .10$). Similarly, for the empirically derived cognitive domains, repeated measures ANOVA comparing the total number of test scores, calculated as a proportion of the total number of tests included in each domain, obtained at 1, 1.5, and 2 SD below the mean showed no significant differences between speeded processing, verbal memory, and visual memory domains (all univariate $F < 2.0$, all $p > .10$).

To determine whether demographic or psychological factors might account for some of the variance in abnormally low scores across a concussion assessment battery, three standard multiple regression analyses were performed with total number of test scores 1, 1.5, and 2 SD below the mean as dependent variables and athlete gender, IQ, sport, BDI-FS Total score, PCSS Total score, and all five NEO-FFI personality factors (Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness) as independent variables. To prevent the confounding of athlete sex and sport, only sports teams with male and female athletes were included (soccer, lacrosse, and basketball), which reduced the number of cases that from 495 to 265 athletes. Despite this, the ratio of cases to independent variables exceeded the minimum requirement for testing both the overall correlation and individual predictors (Tabachnik & Fidell, 2007).

After examining bivariate correlations between all proposed predictor and criterion variables, two variables – PCSS Total and BDI-FS Total – were excluded because they were not significantly correlated with any of the criterion variables.[5] All other proposed predictor variables were retained. Variables were screened using residual scatterplots obtained through an initial regression run in order to identify any departure from underlying assumptions, and further screening of individual variables was conducted through examination of descriptives and histograms. All dependent variables were square root transformed to correct for moderate positive skewness and kurtotisis, reduce the number of outliers, improve the normality, linearity, and homoscedasticity of residuals, and increase the chance of finding possible relationships between predictor and criterion variables. Tolerances (all > .70) and collinearity diagnostics indicated no problems with multicollinearity for any of the variables. Only one multivariate outlier was identified using $p < .001$ as the criteria for Mahalanobis distance (< 29.6); given the sample size, it was decided to retain the outlier.

Results of the three standard multiple regression analyses are reported in Tables 7, 8, and 9, including zero-order correlations (r), semipartial correlations (sr^2), intercepts, unstandardized regression coefficients (*B*), standardized regression coefficients (β), R and R^2 values. The R value for the overall model was significantly different from zero for all three criterion variables: for (square root of) number of test scores 1 SD below the mean, $F (9, 255) = 5.5$, $p < .001$; for (square root of) number of test scores 1.5 SD below the mean, $F (9, 255) = 3.0$, $p < .01$; and for (square root of) number of test scores 2 SD below the mean, $F (9, 255) = 2.2$, $p < .05$. R^2 values indicate that 16% of the variance in number of test

---

[5] These two variables were significantly correlated with each other and with NEO-FFI Neuroticism (all r = .35 to .49, all $p < .01$), and with several of the other NEO-FFI factors (r = .11 to .25). Their shared variance, which is ignored in multiple regression, might have masked a significant relationship with the criterion variables.

scores 1 SD below the mean, 10% of the variance in test scores 1.5 SD below the mean, and 7% of the variance in number of test scores 2 SD below the mean was predicted by demographic and personality variables in a sample of collegiate athletes.

Athlete IQ was a significant predictor of number of test scores below the mean at 1, 1.5, and 2 SD and consistently accounted for the greatest amount of unique variance in abnormally low test scores at all cutoff levels. Regression coefficients indicated that the number of abnormally low test scores decreased as athlete IQ increased. Athlete sport was also a significant predictor of number of test scores below the mean at 1 and 1.5 SD, with lacrosse and basketball players significantly more likely than soccer players to earn abnormally low test scores. Athlete sex was a significant predictor of abnormally low test scores but only at 1 SD below the mean, with regression coefficients indicating that male athletes were significantly more likely than female athletes to earn abnormally test scores. A trend toward a significant relationship between NEO-Openness and number of test scores below the mean emerged at 1 and 1.5 SD, with athletes who reported greater openness less likely to earn test scores below the mean.

To examine whether demographic or performance differences exist between male and female athletes on a multiple test concussion assessment battery, demographic data were generated for both groups as shown in Table 10. Few differences appeared between gender groups except on diagnosed or possible ADHD or LD, where a smaller percentage of female athletes reported either condition (4% versus 6%), and on previous psychological testing, where a larger percentage of female athletes reported prior testing (16% versus 9%). Female athletes also reported lower alcohol consumptions (1.3 drinks per week versus 3.2 drinks per week) and lower marijuana use (.10 occasions of use versus .49).

For average maximum discrepancy between lowest and highest z-scores, an independent samples t-test revealed a trend toward a significant difference, $t\ (488) = 1.8$, $p = .08$, with male athletes showing a larger discrepancy than female athletes (3.2 SD versus 3.0 SD) as reported in Table 11.

Independent samples t-tests, corrected for unequal variances, also revealed significant differences between male and female athletes on average number of test scores below the mean at all cutoff levels: at 1 SD below the mean, $t(258) = 3.1$; at 1.5 SD below the mean, $t(309) = 3.2$; and at 2 SD below the mean, $t(349) = 3.1$, all $p < .001$. As reported in Table 12, male athletes obtained significantly more test scores below the mean than female athletes at all cutoff levels, and they obtained abnormally low scores on a greater number of tests.

Chi-square analyses, using separate categories for zero through four tests and collapsing five or more tests into a single category in order to ensure that all cells had an expected count greater than five, failed to show any significant differences in the percentage of male and female athletes with test scores at 1, 1.5, or 2 SD below the mean, all $\chi^2(5) = 8.0$ to 9.1, all $p > .10$. However, as reported in a Tables 13 and 14 and shown in Figures 6 and 7, significance tests of the difference in cumulative percentages[6] indicated that significantly more male athletes had abnormally low scores on one, two, and three or more tests at 1 and 1.5 SD below the mean, and significantly fewer male athletes had zero test scores below the mean at 1 and 1.5 SD cutoff levels.

A comparison of the total number of abnormally low test scores for male and female athletes at the beginning, middle, and end of the test battery yielded a significant main effect for time as described above for the overall athlete sample. At 1.5 and 2 SD below the mean, all athletes obtained significantly fewer test scores below the mean at the end of the test battery than at the beginning or middle of the test battery, with small effect sizes ranging from .15 to .22. A main effect for gender also emerged at 1, 1.5, and 2 SD below the mean, with male athletes obtaining significantly more test scores below the mean than female athletes at 1, 1.5, and 2 SD. Again, effect sizes were small, with Cohen's d ranging

---

[6] A significance test of the differences between two percentages was calculated by dividing the difference between percentages by the standard error of the difference, where SE (diff%) = Sqrt $[(P \times Q)/N_1 + (P \times Q)/N_2]$.

from .20 to .25. The interaction between athlete gender and time point in the test battery was not significant at any cutoff level. Results are reported in Table 15 and shown in Figure 8.

A comparison of the total number of abnormally low test scores for male and female athletes on pencil-and-paper versus computer-administered tests yielded a significant interaction, or a trend toward a significant interaction, between mode and gender, along with a significant main effect of gender at almost every cutoff level. The main effect for gender was as described above, with male athletes showing significantly more test scores at 1, 1.5, and 2 SD below the mean than female athletes. For the interaction effect, male athletes tended to obtain more scores below the mean on pencil-and paper tests and fewer scores below the mean on computer-administered tests, whereas female athletes tended to earn more scores below the mean on computer-administered tests and fewer score below the mean on pencil-and-paper tests. When male and female athletes were compared on total number of scores below the mean for tests requiring a computer, written, or oral response, only the main effect for athlete gender was significant. Results are reported in Table 16 and shown in Figure 9.

A comparison of the total number of abnormally low test scores for male and female athletes in different theoretically derived cognitive domains yielded only a main effect for athlete gender at all cutoff levels as described above (all $F > 6.0$, all $p < .01$). Neither the main effect for cognitive domain nor the interaction effect between athlete gender and cognitive domain was significant at any cutoff level (all $F < 1.5$, all $p > .10$). As with theoretically derived cognitive domains, a comparison of the total number of abnormally low test scores for male and female athletes in different empirically derived cognitive domains found only a significant main effect for athlete gender at all cutoff levels as described above (all $F > 7.0$, all $p < .01$), with no significant main effect for cognitive domain or interaction effect between athlete gender and cognitive domain (all $F < 1.0$, all $p > .10$).

Finally, standard multiple regression analyses were conducted to determine whether male and female athletes differ on the particular demographic and personality factors accounting for any significant variance in the number of test scores at 1, 1.5, or 2 SD below the mean across the concussion assessment battery. Predictor variables included in the regression analyses were the same as for the overall sample – athlete IQ, sport, and five NEO-FFI factors – minus athlete gender. Results of the three standard multiple regression analyses are reported in Table 17. The R value for the overall model was significantly different from zero only for female athletes at 1 SD and 1.5 SD cutoff levels, but it trended toward significance for male athletes at the 1 SD cutoff level. The (square root of) number of test scores at 1 SD below the mean yielded $F\ (8, 110) = 4.7$, $p < .001$, for female athletes and $F\ (8, 137) = 1.8$, $p = .09$ for male athletes; the (square root of) number of test scores 1.5 SD below the mean yielded $F\ (8, 110) = 2.3$, $p < .05$, for female athletes and $F\ (8, 137) = 1.1$, $p > .10$, for male athletes had. Values for R^2 indicated that, for female athletes, 25% of the variability in number of test scores 1 SD below the mean and 14% of the variability in number of test scores 1.5 SD below the mean was predicted by demographic and personality variables.

Athlete IQ was a significant predictor of number of test scores below the mean at 1 SD and 1.5 SD below the mean for both male and female athletes, but it was a significant predictor of number of test scores at 2 SD below the mean only for female athletes. At all cutoff levels, it accounted for the largest amount of unique variance in abnormally low test scores compared to other predictor variables. Regression coefficients indicated that the number of abnormally low test scores decreased as athlete IQ increased. For female athletes only, sport was a significant predictor of number of test scores below the mean at 1 SD, and a trend toward significance emerged at 1.5 SD below the mean. Regression coefficients indicated that female lacrosse and basketball players were significantly more likely than female soccer players to obtain abnormally low test scores. None of the NEO personality factors accounted for a significant amount of variance in number of test scores below the mean. For female athletes, however, NEO-Openness showed a trend toward significance at 1 and 2 SD below the mean, with regression coefficients indicating that female athletes reporting greater openness were significantly less likely to earn test scores 1 or 2 SD below the mean.

An attempt was made to examine whether differences in intellectual functioning are associated with differences in number of test scores at 1, 1.5, or 2 SD below the mean on a concussion assessment battery. Athletes were divided into groups based on WTAR estimated premorbid Full Scale IQ: Low Average (IQ < 90, n = 12); Average (90 ≤ IQ < 110, n = 401); and High Average (IQ ≥ 110, n = 64). Significant discrepancies in group sizes limited the utility of the originally planned analyses. Furthermore, although High Average and Low Average IQ groups were equivalent on demographic variables such as age and handedness, they differed significantly on a number of other variables, including sport, race/ ethnicity, alcohol or drug use, and incidence of LD or ADHD. Because IQ groups were confounded with several other variables, none of the proposed analyses were conducted because findings would have been inconclusive or invalid.

Lastly, binomial probability distribution (Ingraham & Aiken, 1996) and Monte Carlo simulation (Crawford et al., 2007) methods were used to estimate the percentage of abnormally low test scores at each cutoff level in the overall athlete sample. A binomial (or Bernoulli) probability distribution is based on a binomial statistical experiment, one which consists of *n* repeated trials where each trial has only two possible outcomes (event or non-event). The probability, P, of an event occurring is the same for each trial because trials are seen as independent of each other. In other words, the outcome on one trial does not affect the outcome on any other trial or, in our study, one test score is not correlated with any other test score. The classic binomial experiment is the coin toss (see Appendix D for formulae).

Cumulative binomial probability distributions were calculated[7] for a battery of 17 tests (n = 17) where the event was the number of test scores below the mean at 1, 1.5, or 2 SD (P = .159, .067, or

---

[7] An online Binomial Distribution Calculator was used from the StatTrak website: http://stattrek.com/Tables/Binomial.aspx.

.023). Results are shown in Figure 4. At all cutoff levels, the binomial probability distribution model significantly over-predicted the number of athletes earning scores below the mean on fewer tests (e.g., one or more and two or more tests), while it significantly under-predicted the number of athletes earning scores below the mean on more tests (e.g., five or more and six or more tests). Significant differences[8] between the percentage of athletes earning scores below the mean as estimated by the binomial probability distribution method and the percentage of athletes earning scores below the mean in the actual athlete sample emerged for all numbers of tests except for four of more tests at 1 SD, three or more tests at 1.5 SD, and one or more, five or more, and six or more tests at 2 SD (all $p < .01$).

Monte Carlo methods include an array of statistical simulation methods which use sequences of random numbers to simulate complex behavioral phenomena. These simulation methods only require that the behavioral phenomena be described using probability density functions. Simulation proceeds by random sampling from the probability density function, which quickly generates multiple random distributions for the phenomenon of interest with a mean of 0 and a standard deviation of 1. Many simulations are then performed and averaged to predict the desired outcome: simulation of a particular complex system. Such methods allow for correlation between outcomes.

Using MatLab, Monte Carlo simulation estimates were calculated for the cumulative percent of athletes obtaining test scores at 1, 1.5, and 2 SD below the mean across a 17-test battery (see Appendix E for MatLab program). Results for 500,000 randomly generated pseudo-athletes yielded estimates similar to the binomial probability distribution model, as shown in Figure 5. At 1 and 1.5 SD cutoff levels, the Monte Carlo simulation model significantly over-predicted the percent of athletes earning scores below the mean on fewer tests (e.g., one or more). At a 2 SD cutoff level, however, it accurately predicted the percent of athletes with low scores on one or more tests. It also accurately predicted the

[8] Differences between percents were tested for significant using an online z-test calculator for two proportions at http://www.dimensionresearch.com/resources/calculators/ztest.html.

percent of athletes with low scores on more tests at the 1 SD cutoff level (e.g., five or more and six or more) and at the 1.5 SD cutoff level (e.g., two or more and three or more). At 1.5 and 2 SD below the mean, however, the Monte Carlo simulation model significantly under-predicted the percentage of athletes earning scores below the mean on more tests (e.g., four or more and five or more).

Chapter 4: Discussion

Review

Results of the study address three critical questions. First, when interpreting the results of baseline concussion assessment among healthy collegiate athletes, how important is it to incorporate information about base rates of abnormally low or discrepant test scores? Second, do demographic or psychological characteristics or test administration factors that are associated with abnormally low or discrepant scores obtained by healthy collegiate athletes on a neuropsychological test battery help in the interpretation of results and in the understanding of performance at baseline assessment? Finally, how do we incorporate information about base rates and correlates of abnormally poor baseline test performance into post-concussion evaluations and return to play decisions?

To answer the first question, a brief review of the literature on the size and frequency of discrepancies between highest and lowest subtest scores on some common neuropsychological test batteries will be helpful (Binder et al., 2009). Results are summarized in Table 18. For the Wechsler test batteries (WAIS-R, WAIS-III, and WAIS-IV), all of which have large normative samples with an age range between 16 and 74 to 90 years and an education range between 8 and 16 or more years, the mean maximum discrepancy between highest and lowest subtest scaled scores is approximately 2.2 to 2.5 SD, with 18% to 30% of the sample showing a mean maximum discrepancy of 3 SD or greater (Matarazzo & Prifitera, 1989; Wechsler, 1997; Wechsler, 2008). In contrast, the smaller normative sample for the Aging, Brain Imaging, and Cognition study (ABC: Schretlen et al., 2003), in which the age range is 20 to 92 years and the average education level is 14 years, shows a mean maximum discrepancy between highest and lowest subtest scaled scores of over 3 SD, with approximately 65% of the sample showing a mean maximum discrepancy of 3 SD or greater. The differences between the ABC and the WAIS samples might be accounted for by the greater number of test scores examined in the latter sample.

In comparison to the above, the relatively large sample of healthy collegiate athletes from the Penn State Concussion Testing Program, with an age range of 18 to 21 years and an average education level of 12 years, shows a mean maximum discrepancy between highest and lowest subtest scores of over 3 SD. As reported in Table 2, the smallest maximum discrepancy is 1.8 SD and the largest is 8.8 SD. Furthermore, 50% of athletes show a mean maximum discrepancy of 3 SD or greater. Surprisingly, the size of the discrepancy displayed by healthy young collegiate athletes on a concussion assessment battery, as well as the percent of athletes showing a discrepancy of 3 SD or greater on the battery, is closer to that of the older normative sample from the ABC study than the younger normative sample from the WAIS standardization studies.

When comparing the cumulative percentage of collegiate athletes obtaining abnormally low scores on one or more tests across a concussion assessment battery to the cumulative percentage of subjects obtaining abnormally low scores on one or more tests across common neuropsychological test batteries, similar findings emerge as reported in Table 19. At the 1 SD cutoff level, 78% to 92% of subjects from other studies score below the mean on one or more tests as compared to 79% of collegiate athletes; 36% to 75% score below the mean on two or more tests as compared to 55% of athletes; and 51% to 58% score below the mean on three or more tests as compared to 38% of athletes. At the 1.5 SD cutoff level, 43% to 70% of subjects from other studies score below the mean on one or more tests as compared to 50% of collegiate athletes; 15% to 49% score below the mean on two or more tests as compared to 27% of athletes; and 16% to 25% score below the mean on three or more tests as compared to 15% of athletes. At the 2 SD cutoff level, 27% to 44% of subjects score below the mean on one or more tests as compared to 29% of collegiate athletes; 3% to 22% score below the mean on two or more tests as compared to 13% of athletes; and 3% to 12% score below the mean on three or more tests as compared to 7% of athletes.

Healthy collegiate athletes thus exhibit base rates of abnormally low or discrepant test scores that are comparable to those found for other subjects on other test batteries. Despite the athletes' youth and health, we can conclude that considerable variability across test scores is normal and expected for collegiate athletes. In fact, the average lowest score obtained by athletes across the 17-test concussion assessment battery was almost 2 SD below their estimated WTAR FSIQ, suggesting that no matter what their level of intellectual functioning, healthy collegiate athletes show significant variability in neuropsychological test performance. The results of the study support the conclusion reached for other groups on other test batteries: obtaining some unusually low or discrepant scores on a large battery of tests is the rule rather than the exception and is common even among healthy young adults (Binder et al., 2009). Incorporating information on base rates of abnormally poor neuropsychological test performance is thus critical when interpreting the results of baseline concussion assessment among healthy collegiate athletes.

In addressing the second question – are there demographic and psychological characteristics associated with collegiate athletes or contextual factors associated with neuropsychological test administration that might help us to better understand abnormally poor test performance at baseline assessment – one of the clearest and most consistent findings is the relatively poorer and more variable performance of male athletes as compared to female athletes. As noted in Tables 11 through 14, male athletes show lower best and worst test scores, on average, than female athletes, along with a greater discrepancy between best and worst test scores and a wider range across all test scores. They also perform more poorly than female athletes on a greater number of tests at all cutoff levels, and a greater percentage of them have test scores significantly below the mean. These findings are consistent with extant literature on gender differences in cognitive test performance (Sternberg, 2000). However, they also indicate that the any interpretation of abnormally poor or variable performance on a multiple test neuropsychological assessment battery among collegiate athletes can be improved by including gender.

Another clear and consistent finding from the study (Tables 7, 8, and 9) is the association between athlete IQ and low test scores. Athlete IQ is not only modestly but significantly correlated with total number of low test scores across a concussion assessment battery ($r = -.20$ to $-.28$); it also accounts for the greatest amount of unique variance at all cutoff levels ($sr^2 = .03$ to $.08$). Regression coefficients indicate that for every ten-point increase in athlete IQ, the number of test scores at 1 SD below the mean is reduced by 0.5, or 42% of 1.2 (the average number of test scores 1 SD below the mean for all athletes). A ten-point increase in athlete IQ drops the number of test scores at 1.5 SD below the mean by 0.3, or 48% of .63 (the average number of test scores 1.5 SD below the mean for all athletes). Finally, a ten-point increase in athlete IQ reduces the number of test scores at 2 SD below the mean by 0.2, or 67% of 0.3 (the average number of test scores 2 SD below the mean for all athletes).

Sport also accounts for a small but significant amount of unique variance at 1 and 1.5 SD cutoff levels in number of low test scores obtained by collegiate athletes on a concussion assessment battery ($sr^2 = .02$ to $.04$). Regression coefficients indicate that at the 1 SD cutoff level, lacrosse players obtain 0.33 more test scores below the mean and basketball players obtain 0.24 more test scores below the mean as compared to soccer players. At 1.5 SD cutoff, lacrosse players obtain 0.17 more test scores below the mean and basketball players obtain 0.25 more test scores below the mean as compared to soccer players. Thus, male and female soccer players are performing best overall at baseline concussion assessment, male and female lacrosse players are performing slightly worse, with more test scores 1 SD below the mean, and male and female basketball players are performing worst of all, with more test scores 1.5 SD below the mean.

Although regression models at all cutoff levels are statistically significant, they account for only a small amount of the total variance in abnormally poor baseline test performance among collegiate athletes: 16% at the 1 SD cutoff level, 10% at the 1.5 SD cutoff level, and 7% at the 2 SD cutoff level. Demographic characteristics such as IQ and sport contribute the greatest amount of variance, whereas psychological characteristics such as mood or personality factors contribute little to no variance. Possibly, any associations between low test scores and mood or symptom inventories such as the BDI-FS and the PCSS were affected by the limited range of these measures as well as by their moderate correlation with each other ($r = .43$). Similarly, several of the NEO-FFI personality factors showed small but significant correlations with number of low test scores at the three cutoff levels ($r = .10$ to $.14$) but larger and more significant correlations with each other ($r = .17$ to $.41$). Thus, their shared variance might have masked any unique contributions to the regression analyses. Notably, only NEO-Openness, which was not significantly correlated with other personality traits, showed a trend toward contributing unique variance at the 1 and 1.5 SD cutoff levels.

When examining demographic and psychological characteristics associated with poorer baseline test performance for male and female athletes separately, the picture described above holds but only for female athletes (Table 17). Athlete IQ continues to account for the greatest amount of unique variance in number of low test scores at all cutoff levels, but the association is stronger and more consistent for female athletes ($sr^2 = .02$ to $.07$ for male athletes and $sr^2 = .04$ to $.10$ for female athletes). Female lacrosse and basketball players, but not male lacrosse and basketball players, are significantly more likely to obtain test scores 1 or 1.5 SD below the mean on a concussion assessment battery. NEO-Openness shows a trend toward significance at 1 and 2 SD cutoff levels but only for female athletes ($sr^2 = .02$ to $.03$). Finally, the overall regression models account for a greater amount of variance in number of test scores at 1, 1.5, and 2 SD below the mean for female athletes (25%, 14%, and 11% respectively) as compared to those for male athletes (9%, 6%, and 8% respectively).

Contextual factors – such as mode of test administration, time point in test battery, and type of cognitive domain assessed – appear to have little impact on abnormally poor neuropsychological test performance among healthy collegiate athletes. Athletes perform neither better nor worse on computer-administered versus pencil-and-paper tasks, on measures requiring an oral or a written response versus a computer response, or on tests evaluating speeded processing, verbal memory, visual memory, or executive functioning domains. However, athletes did show significantly fewer low scores toward the end of the test battery as compared to the beginning and middle of the test battery at the 1.5 and 2 SD cutoff levels (Table 5 and Figure 3). This improvement in test performance over the course of the battery is consistent with habituation to the test environment and to testing demands (Lezak et al., 2004). Effect sizes are small (.15 to .22), however, indicating that increased adaptation has only a minimal impact on baseline test performance. Male and female athletes showed few gender differences, performing similarly across time points, test modes, and cognitive domains. Although a trend toward gender differences on computer-administered versus pencil-and-paper tasks emerged, with male athletes earning fewer low test scores on computer-administered tasks and female athletes earning fewer low test scores on pencil-and-paper tasks, the trend failed to reach significance.

In summary, the results of the current study allow us to draw several conclusions in response to the second question. First, male athletes tend to show poorer and more variable neuropsychological test performance, on average, than female athletes at baseline concussion assessment. Second, IQ exerts the strongest influence on number of low scores earned by athletes at baseline assessment, regardless of athlete gender or sport. Third, only for female athletes does sport, and possibly personality factors such as openness, impact baseline test performance. Fourth, performance on a concussion assessment battery improves for all athletes as they become habituated to the environment and to expectations associated with testing. Fifth, male athletes are likely to obtain slightly more test scores below the mean on pencil-and-paper tasks, whereas female athletes are likely to obtain slightly more test scores below the mean on computer-administered tasks. Finally, no test administration factors yield significant effect sizes nor do any athletes characteristics account for a large amount of variance, leaving the preponderance of variability in abnormally poor test performance at baseline assessment among collegiate athletes still to be explained.

The third major question is whether and how to incorporate information about base rates and correlates of abnormally poor baseline test performance among healthy collegiate athletes into post-concussion evaluations and return to play decisions. One of the most important issues in sport-related concussion is determining when concussed athletes are able to return to competition and practice and resume other activities safely. Making this determination requires interpreting post-injury test performance relative to pre-injury, or baseline, test performance. Without knowing how to understand or interpret abnormally low or discrepant test scores at baseline, we must use caution when interpreting similar results after a concussion, as poor performance may not be synonymous with cognitive impairment. Incorporating information about base rates and correlates of abnormally low or discrepant test performance at baseline assessment might improve our understanding and interpretation of such results at post-concussion evaluations.

But do we use statistical estimation methods or do we extrapolate from actual clinical samples when generating and incorporating base rate information for collegiate athletes? Results from the current study support the latter. Statistical estimation methods, such as binomial probability distribution or Monte Carlo simulation, minimize problems associated with generalizing results that might be specific to different athlete groups or test batteries. However, our study shows that both binomial probability distribution and Monte Carlo simulation methods typically overestimate the number of low scores obtained across a multiple test battery when compared to percentages found in the actual athlete sample, especially for one or more and two or more tests. They also significantly underestimate the number of low scores for five or more and six or more tests. The inaccuracy of the predictions generated by both methods renders them less suitable for establishing clinically useful base rates of abnormally low test scores among healthy collegiate athletes. Furthermore, neither binomial probability distribution nor Monte Carlo simulation methods provide estimates that take athlete gender into account.

Thus, using a large and representative sample to calculate the percentage of athletes earning abnormally low or discrepant test scores across a multiple test battery provides the most accurate and specific criteria for defining abnormally poor test performance among healthy male and female collegiate athletes on a concussion assessment battery. It generates base rates describing the prevalence of abnormally low test scores among healthy collegiate athletes who are administered a multiple test battery on which many test scores are interpreted simultaneously. It allows for intercorrelation between test scores as well as non-normal distribution among neuropsychological measures. It also allows for the calculation of separate normative data for male and female athletes. Tables which provide such information can reduce over-interpretation of low test scores, helping to refine the interpretation of overall test performance and improve clinical decision making.

Table 21 provides base rate information in a clinically useful format for a large sample of healthy collegiate male and female athletes on a multiple test concussion assessment battery. To facilitate the interpretation of test scores for this population, the number of abnormally low scores at ≤ 1 SD (16th percentile), ≤ 1.5 SD (7th percentile), and ≤ 2 SD (2nd percentile) is provided for the total sample and for male and female athletes separately. Cumulative percentages presented in bold italics mark an “uncommon” number of low scores, where uncommon is based on a prevalence rate of ≤ 10% of the sample (similar tables appear in Brooks et al., 2010). Information from Table 21 allows us to see the number of abnormally low test scores that optimally define uncommon performance across a multiple test battery among healthy collegiate athletes. Uncommon neuropsychological test performance provides greater evidence of cognitive impairment. This evidence, in turn, helps to refine the interpretation of overall test performance and improve clinical decision making, an assertion that is best illustrated by looking at individual cases.

Clinical Examples

Athlete 1

A 21-year-old, right-handed, male lacrosse player sustained a concussion during an evening game, when he was checked in the back of the head by an opposing player. He was tested 5 days and 9 days post-injury. Although he experienced dizziness, nausea, vomiting, and headache for a few hours after the injury and described feeling mentally foggy during the following two days, he reported no residual symptoms at post-concussion evaluations. At baseline assessment, the athlete had no test scores in the impaired range (i.e. ≥ 1. 5 SD below the mean) but three test scores ≥ 1 SD below the mean. Base rates of abnormally low scores among male collegiate athletes (Table 14) show that this latter finding is not unusual, as over 40% of male athletes obtain three or more test scores at least 1 SD below the mean at baseline assessment. Furthermore, although the athlete showed intellectual functioning in the average range (estimated WTAR FSIQ = 104), he reported a diagnosis of ADHD, which might account for some of his test performance variability.

At first post-concussion evaluation, the athlete showed performance decrements ≥ 1.5 SD on six neuropsychological tests. The information on base rates provided by Table 21 indicates that the number of tests on which the athlete was showing performance decrements, along with the magnitude of these decrements, is clinically unusual. The athlete had shown no performance decrements of this magnitude at baseline, and less than 6% of male athletes performed similarly at baseline. Thus, the athlete would appear to be experiencing sufficient residual cognitive dysfunction to justify withholding him from play. At second post-concussion evaluation, the athlete's performance improved to within 1 SD of baseline or better on all previously impaired tests except for one delayed memory task, which was still ≥ 2 SD below baseline. However, over 31% of male athletes had one or more tests at least 2 SD below the mean at baseline assessment, and the athlete was performing at or above baseline on ten other measures, suggesting that he had recovered sufficiently for a gradual return to play.

Athlete 2

An 18-year-old, right-handed female soccer player sustained a concussion during an afternoon game in head-to-head contact with an opposing player. She was tested 4 days and 8 days post-injury. She reported symptoms of disorientation, dizziness, balance problems, mental fogginess, difficulty with concentration, and headache, with the latter two symptoms lasting several days. At baseline assessment, the athlete had no test scores in the impaired range (i.e. ≥ 1. 5 SD below the mean) but two test scores ≥ 1 SD below the mean. Despite the athlete's average intellectual functioning (estimated WTAR FSIQ = 101), base rates of abnormally low test scores for female athletes show that this latter finding is not unusual, as over 44% of female athletes had two or more test scores at least 1 SD below the mean (Table 14).

At first post-concussion evaluation, the athlete showed performance decrements ≥ 1.5 SD on four neuropsychological tests. Information from Table 21 indicates that the number of tests on which the athlete was showing performance decrements, along with the magnitude of these decrements, was clinically unusual. The athlete had shown no performance decrements of this magnitude at baseline, and less than 6% of female athletes performed similarly at baseline. Thus, the cognitive deficits exhibited by this athlete would appear sufficient to warrant withholding her from play. At second post-concussion evaluation, the athlete's performance improved to within 1 SD of baseline or better on all previously impaired tests except for one, which was still ≥ 2 SD below baseline. She also showed slight decrements (≤ 0.5 SD) on three previously unimpaired tasks. However, almost 25% of female athletes obtained at least one test scores 2 SD below the mean at baseline, and over 60% earned scores 0.5 SD below the mean on three or more tests (not shown in Table 21). Furthermore, she was performing above baseline on ten other tests, suggesting that she had recovered sufficiently for a gradual return to play.

Athlete 3

A right-handed, male basketball player sustained a concussion during practice, when he dove after a loose ball and was accidentally kicked in the face by a teammate. He was tested 4 days post-injury. He described symptoms including disorientation, dizziness, nausea, balance problems, headache, and mental fogginess lasting approximately one hour, with the last persisting through post-concussion evaluation. At baseline assessment, the athlete had two test scores in the impaired range (i.e. ≥ 1. 5 SD below the mean) and two additional test scores ≥ 1 SD below the mean. Table 21 indicates that the athlete's number of impaired scores is not unusual, despite his high average intellectual functioning (estimated WTAR FSIQ = 110). At baseline, almost 30% of male athletes had four or more test scores 1 SD below the mean and two or more test scores 1.5 SD below the mean, as did over 20% of athletes with high average IQ (not shown in tables).

At first post-concussion evaluation, the athlete showed performance decrements ≥ 1.5 SD on three neuropsychological tests, which is borderline abnormal: according to Table 21, 18% of male athletes had three or more test scores 1.5 SD below the mean at baseline assessment. However, this athlete had obtained two test scores in the impaired range at baseline which, in itself, is not exceptional. Furthermore, he was performing at or above baseline on 13 other measures, including all except one on which he had scored 1 or 1.5 SD below the mean at baseline. Given this information, decrements displayed by this athlete at post-concussion evaluation do not appear sufficient to suggest a need for further recovery time, justifying the decision for a gradual return to play.

Athlete 4

A 20-year-old, left-handed, female basketball player sustained a concussion during afternoon practice in head-to-head contact. She was tested 2 days post-injury. At the time of injury, the athlete reported brief loss of consciousness and anterograde amnesia in addition to disorientation, dizziness, balance problems, visual disturbance, mental fogginess, pressure in the head, and headache. She was experiencing the last three symptoms at the time of post-concussion evaluation. At baseline assessment, the athlete had two test scores in the impaired range (i.e. ≥ 1. 5 SD below the mean) and three additional test scores ≥ 1 SD below the mean. Table 21 indicates that her number of impaired scores is borderline abnormal, especially given her average range of intellectual functioning (estimated WTAR FSIQ = 98): 19% of female athletes had two or more test scores 1.5 SD below the mean and only 12% had at least five test scores 1 SD below the mean.

At first post-concussion evaluation, the athlete showed performance decrements ≥ 1.5 SD on seven neuropsychological tests, with declines ≥ 2 SD on four computer tests. Table 21 indicates that the number of tests on which the athlete was showing performance decrements, along with the magnitude of these decrements, is highly unusual: at baseline assessment, less than 3% of female athletes had five or more test scores 1.5 SD below the mean and less than 2% had four or more test scores 2 SD below the mean. Even with a borderline abnormal baseline performance, the athlete’s post-concussion test scores were clinically unusual, as she had obtained only two test scores in the impaired range at baseline assessment versus seven test scores in the impaired range at post-concussion evaluation. The decrements would thus appear sufficient to suggest that the athlete abstain from practice and games until she had recovered further.

Athlete 5

An 18-year-old, right-handed, male football player sustained a concussion during an evening game, when he was hit in the back of the head after being tackled. He was tested 2 days and 5 days post-injury. He described symptoms including dizziness, balance problems, difficulties with concentration, mental fogginess, and headache lasting approximately two hours. Only the headache remained at post-concussion evaluation. At baseline assessment, the athlete had seven test scores in the impaired range (i.e. ≥ 1. 5 SD below the mean) along with four additional scores ≥ 1 SD below the mean. Table 21 indicate that this number of impaired scores is extremely unusual, especially given the athlete's average intellectual functioning (estimated WTAR FSIQ = 106). Less than 7% of male athletes had five or more test scores 1.5 SD below the mean and less than 6% had eight or more test scores 1 SD below the mean at baseline assessment.

At first post-concussion evaluation, the athlete showed performance decrements ≥ 1.5 SD on three measures. At the same time, he showed improvements ≥ 1.5 SD on several other tests, including all except two tests on which he performed abnormally poorly at baseline. Table 21 indicates that the three tests on which the athlete was showing performance decrements at post-concussion evaluation, along with the magnitude of these decrements, is borderline abnormal: 18% of male athletes obtained scores ≥ 1.5 SD below the mean on three or more tests at baseline. However, post-concussion test results must be interpreted in light of the athlete's highly abnormal baseline performance, which likely did not reflect his true level of functioning. The recognition that any decline from baseline might have been obscured by his highly abnormally baseline performance would suggest that the athlete be withheld from play for follow-up evaluation. At second post-concussion evaluation, the athlete's performance improved on two of the three measures on which he had shown impairment. In addition, his performance on six of the other measures on which he had shown improvement remained stable. Thus, the athlete was showing impairment on only one measure, a performance that represented a significant improvement over both his baseline and first post-concussion evaluations and one that was well within the range of clinical normality, as over 50% of male athletes had test scores 1.5 SD below the mean on at least one test at baseline. The improvement in the athlete's performance would justify allowing him to return to play.

As can be seen with the above examples, knowing base rates of abnormally low scores among healthy male and female collegiate athletes on a neuropsychological test battery at baseline assessment provides information that is relevant and helpful in the clinical interpretation of post-concussion test results. The information provided in Table 21 reduces the likelihood of mistakenly identifying a commonly occurring number of low scores as indicative of cognitive impairment. For example, if we know that healthy collegiate athletes not uncommonly obtain at least two test scores in the impaired range (1.5 SD below the mean) on a multiple test battery, as over 27% of our sample did, we may be less quick to assume that an injured athlete performing in the "impaired" range on two tests at post-concussion assessment indicates a clinically significant problem, as with athlete 3 above. Such information helps to optimally define poor performance across a multiple test battery among healthy collegiate athletes, thus refining overall test interpretation and clinical decision making.

Limitations

Although the overall sample size was large, it was notably skewed toward male Caucasian freshman athletes, who constituted 75% of the sample. Thus, findings from the current study might not generalize to athlete groups that differ markedly on age, gender, or racial/ethnic characteristics. Athlete IQ subgroups were also notably skewed toward male athletes and those with average intellectual functioning. This imbalance, coupled with confounds between IQ and other variables, made it impossible to conduct proposed analyses examining a possible relationship between level of intellectual functioning and abnormally poor baseline test performance.

Because all athletes from teams participating in the Penn State Concussion Testing Program undergo baseline assessment, no exclusionary criteria were used to screen out athletes prior to testing. Thus, it is possible that the frequency of abnormally low or discrepant test scores resulted at least in part from unidentified cognitive impairment among these healthy collegiate athletes. However, the reported frequency of LD (5%), ADHD (6%), chronic illness (7%), and marijuana use (7%) was low in the overall athlete sample. Almost half of the sample reported some alcohol consumption during the week, but over 80% of athletes in the sample identified themselves as moderate drinkers, consuming five or fewer drinks per week. Although over one-third of the athletes in the sample reported a history of prior head injury, 27% reported only one previous concussion, leaving just 10% reporting two or more previous concussions.

Furthermore, no correlations appeared between measures of abnormally variable test performance (maximum discrepancy across all tests) or abnormally low test scores (total number of tests 1, 1.5 or 2 SD below the mean) and ADHD diagnosis, chronic illness, alcohol or drug use, or number of prior concussions (all $p > .10$). Small but significant correlations appeared with LD diagnosis ($r = .10$ to .13), suggesting that learning difficulties might have influenced the size of the discrepancy between tests as well as the number of low scores across all tests. Correlations were small, however, as was the frequency of LD diagnosis, making it unlikely that the number of individuals with learning difficulties who were included in the analyses would have been large enough to account for the high maximum discrepancy between tests and the high base rates of abnormally low scores found in this study. Finally, the average Full Scale IQ for the overall sample was 103, further supporting the conclusion that any pre-existing cognitive impairment that in the sample did not unduly influence study findings.

The type of tests administered and the order of administration are specific to the neuropsychological battery employed by the Penn State Concussion Assessment Program. Different neuropsychological measures clearly entail different levels of cognitive ability, with more difficult tests increasing the possibility of obtaining low test scores. Prior research also indicates that the number of abnormally low test scores obtained on any neuropsychological test battery is positively correlated with the length of the battery: fewer tests decrease the likelihood of obtaining low test scores, whereas more tests increase the likelihood of obtaining low scores (Binder et al., 2009; Brooks & Iverson, 2010; Schretlen et al., 2008). The base rates of low test scores found in this study presumably reflect, at least in part, the number and type of tests administered. However, this hypothesis could not be explicitly examined because neuropsychological test batteries of different lengths and/or compositions were not administered.

Failure to counterbalance the order of test administration made it difficult to determine to what extent, if any, test order influenced the number of abnormally low scores earned across all tests in the battery. An indirect attempt was made to investigate this possibility by comparing performance on the four individual tests at the end of the battery with performance on the four individual tests at the beginning and middle of the battery. Although no significant differences emerged between individual tests administered at different time points in the battery, the lack of counterbalancing in this study makes it difficult to draw definitive conclusions, as test type was confounded with test administration.

Future Directions

Future studies need to determine the degree to which base rates calculated in the current study generalize to individuals from different demographic backgrounds. Determining whether athletes with different levels of intellectual functioning, diagnosed LD or ADHD, or a history of prior concussion show differences in base rates of abnormally low or discrepant test scores at baseline assessment would also provide useful clinical information. Additional variables that might account for a greater amount of variance in abnormally poor baseline test performance among healthy collegiate athletes need to be identified, as factors analyzed in the current study failed in account for more than 16% of performance variability. In addition, base rate information from the current study could usefully be expanded to include base rate information for test batteries of different lengths.

Developing base rate tables for athlete groups with different levels of intellectual functioning is especially important because previous studies have yielded conflicting results. Some studies have found that the total number of test scores below the mean increases with decreasing IQ and decreases with increasing IQ (Brooks et al., 2007, 2008, 2009; Schretlen et al., 2003). Such studies raise the possibility of misdiagnosing cognitive impairment in concussed athletes depending upon level of intellectual functioning, leading to higher rates of false positives for athletes with lower IQ, who tend to obtain more test scores below the mean, and higher rates of false negatives for athletes with higher IQ, who tend to obtain fewer test scores below the mean. Other studies have found that higher IQ is associated with greater neuropsychological test performance variability (Matarazzo & Prifitera, 1989; Schinka et al., 1994). These contrary findings suggest the need to clarify whether and how base rates of abnormally low or discrepant test scores among healthy collegiate athletes are contingent upon level of intellectual functioning as they are contingent upon gender.

Future studies should also compare base rates of abnormally low test scores or abnormally high discrepancies among healthy and concussed collegiate athletes at baseline and post-concussion assessments to determine the stringency of cutoff levels and the number of low or discrepant test scores that optimally define cognitive impairment among concussed athletes and to detect whether the pattern and frequency of abnormally low or discrepant test scores changes after injury or during recovery. For example, do injured athletes obtain abnormally low scores on the same or different tests at baseline versus post-concussion evaluation? Are particular patterns of abnormally low test scores after injury or changes in these patterns during recovery associated with higher levels or longer durations of specific symptoms? Future studies should also try to ascertain whether interpreting patterns of abnormally low test scores in relation to patterns of abnormally high test scores helps to refine clinical inferences for healthy and concussed collegiate athletes, as studies examining this relationship have indicated possible clinical relevance for children with moderate to severe cognitive disability (Brooks et al., 2008).

Summary and conclusions

The overall goal of the current study was to provide base rate information on the frequency and magnitude of abnormally low or discrepant test scores among healthy collegiate athletes on a multiple test concussion assessment battery so as to minimize the misidentification of cognitive impairment, improve diagnostic accuracy, and refine clinical decision making for injured athletes at post-concussion assessment. A secondary goal was to explore athlete characteristics or test administration factors that might be associated with abnormally poor or variable neuropsychological test performance so as to identify factors that might account for a significant amount of performance variability at baseline assessment.

Results of the current study augment the growing literature examining base rates of abnormally low or discrepant test scores among healthy individuals. These results are consistent with the general trends found in other studies examining older and younger subjects (Binder et al., 2009; Brooks & Iverson, 2010). Namely, obtaining some abnormally low or discrepant scores on a large battery of neuropsychological tests is common among healthy collegiate athletes and is dependent upon the stringency of cutoff criteria and the number of tests administered.

Knowing base rates of low or discrepant test scores among healthy collegiate athletes is critical to understanding and interpreting the results of baseline and post-concussion assessments, especially when interpreting multiple test scores simultaneously. For example, the assumptions of normal distribution suggest that less than 16% of athletes are likely to score 1 SD below the mean, less than 10% are likely to score 1.5 SD below the mean, and less than 2% are likely to score 2 SD below the mean on a single test. However, these estimates and distributions are not adjusted for the multiple comparisons required when interpreting a battery of test scores. In our sample of collegiate athletes who were administered 17 tests, 77% earned at least one test score 1 SD below the mean, 51% earned at least one test score 1.5 SD below the mean, and 29% earned at least one test score 2 SD below the mean.

The current study provided some evidence that abnormally poor neuropsychological test performance among healthy collegiate athletes across multiple tests varies by demographic characteristics such as gender and IQ, although additional research is needed to examine how level of intellectual functioning might affect base rates of abnormally low or discrepant test scores and thus the interpretation of test performance. As athlete gender and IQ are the characteristics most clearly and consistently correlated with abnormally poor neuropsychological test performance among healthy collegiate athletes, however, these factors should be integrated in the interpretation of baseline and post-concussion results in order to increase the accuracy of overall test interpretation.

Knowing the base rates of abnormally poor neuropsychological test scores on a multiple test concussion assessment battery administered to healthy collegiate athletes minimizes the over-interpretation of cognitive impairment. Tables which provide this information in a convenient and clinically relevant format, such as Table 21 in this study, make it easier to incorporate information on base rates and correlates such as gender when evaluating baseline and post-concussion test performance among collegiate athletes. Including this information, in turn, refine clinical test inference and improve diagnostic accuracy, thus leading to improved return to play decisions for all collegiate athletes.

REFERENCES

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB '97: Computerized Assessment of Response Bias: Manual.* Durham, NC: COgniSyst.

Army Individual Test Battery: Manual of Directions and Scoring. (1944). Washington, D.C.: War Department, Adjutant General's Office.

Aubry, M., Cantu, R. Dvorak, J., Graf-Baumann, T., Johnston, K., Kelly, et al. (2001). Summary and agreement statement of the 1st International Conference on Concussion in Sport, Vienna 2001. *British Journal of Sport Medicine*, *36*, 3-7.

Axelrod, B. N., & Wall, J. R. (2007). Expectancy of impaired neuropsychological test scores in a non-clinical sample. *International Journal of Neuroscience, 117*, 1591-1602.

Bailey, C. M., Echemendia, R. J., & Arnett, P. A. (2006). The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. *Journal of the International Neuropsychological Society, 12,* 475-84.

Barr, W.B., & McCrea, M. (2001). Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. *Journal of the International Neuropsychological Society, 7*, 693-702.

Barth, J. T., Alves, W. M., Ryan, T. V., Macciochi, S. N., Rimel, R. W., Jane, J. A., & Nelson, W. E. (1989). Mild head injury in sports: Neuropsychological sequelae and recovery of function. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Mild Head Injury*, pp. 257-75. New York: Oxford.

Barth, J. T., Broshek, D. K., & Freeman, J. R. (2006). Sports: A new frontier for neuropsychology. In R. J. Echemendia (Ed.), Sports Neuropsychology: Assessment and Management of Traumatic Brain Injury, pp. 3-16. New York: Guilford Press.

Barth, J. T., Freeman, J. R., Broshek, D. K., & Varney, R. N. (2001). Acceleration-deceleration sports-related concussion: the gravity of it all. *Journal of Athletic Training, 36*, 253-56.

Barth, J. T., Macciocchi, S. N., Giordani, B., Rimel, R. Jane, J. A., & Boll, T. J. (1983).Neuropsychological sequelae of minor head injury. *Neurosurgery, 13*, 529-33.

Beck, A. T., Steer, R. A., & Brown, G. K. (2000). *BDI-Fast Screen for Medical Patients: Manual*. San Antonio, TX: Psychological Corporation.

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 345-57.

Benedict, R. H. B. (1997). *Brief Visuospatial Memory Test-Revised*. Odessa, FL: Psychological Assessment Resources.

Benton, A. L., Hamsher, K., & Sivan, A. B. (1994). *Multilingual Aphasic Examination, 3rd Ed*. Iowa City: AJA Assocatiates.

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: “abnormal” neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology, 24*, 31-46.

Bleiberg, J. Cernich, A. N., Cameron, K. Sun, W., Peck, K., Ecklund, P. J. et al. (2004). Duration of cognitive impairment after sports concussion. *Neurosurgery, 54*, 1073-78.

Brandt, J. (1991). The Hopkins verbal learning test: Development of a new verbal memory test with six equivalent forms. *Clinical Neuropsychologist, 5,* 125-42.

Brandt, J. & Benedict, R. H. B. (2001). *Hopkins Verbal Learning Test-Revised*. Odessa, FL: Psychological Assessment Resources.

Broglio, S. P., & Puetz, T. W. (2008). The effect of sport concussion on neurocognitive function, self-report symptoms, and postural control: a meta-analysis. *Sports Medicine, 28*, 53-67.

Brooks, B. L., Iverson, G. L., Holdnack, J. A., & Feldman, H. H. (2008). The potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society, 14*, 463-78.

Brooks, B. L., Iverson, G. L., & White, T. (2007). Substantial risk of 'accidental MCI' in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society, 13,* 490-500.

Brooks, B. L., Iverson, G. L., Sherman, E. M. S., & Holdnack, J. A. (2009) . Healthy children and adolescents obtain some low scores across a battery of memory tests. *Journal of the International Neuropsychological Society, 15,* 613-17.

Brooks, B. L., Sherman, E. M. S., & Iverson, G. L. (2010). Healthy children get low scores too: prevalence of low scores on the NEPSY-II in preschoolers, children, and adolescents. *Archives of Clinical Neuropsychology, 25*, 182-90.

Broshek, D. K., & Freeman, J. R. (2005). Psychiatric and neuropsychological issues in sports medicine. *Clinical Sports Medicine, 24*, 663-79.

Broshek, D. K., Kaushik, T., Freeman, J. R., Erlanger, D., Webbe, F., & Barth, J. T. (2005). Sex differences in outcome following sports-related concussion. *Journal of Neurosurgery, 102*, 856-63.

Büki, A., & Povlishock, J. T. (2006). All roads lead to disconnection? Traumatic axonal injury revisited. *Acta Neurochirurgica, 148*, 181-94.

Cantu, R. C. (2003). Recurrent athletic head injury: risks and when to retire. *Clinical Sports Medicine, 22*, 593-603.

Cegalis, J. A., & Cegalis, S. (1994). *The Vigil/W Continuous Performance Test (manual)*. New York: ForThought.

Centers for Disease Control and Prevention. (1997). Sports-related recurrent brain injuries – United States. *Morbidity & Mortality Weekly Report, 46,* 224-27.

Centers for Disease Control and Prevention. (2007). Nonfatal traumatic brain injuries from sports and recreation activities – United States, 2001-2005. *Morbidity & Mortality Weekly Report, 56,* 733-37. Published online at http://www.cdc.gov/datastatistics/2008/brainInjuries/ .

Chen, J. K., Johston, K. M., Frey, S., Pterides, M., Worsely, K., & Ptito, A. (2004). Functional abnormalities in symptomatic concussed athletes: an fMRI study. *Neuroimage, 22,* 68-82.

Chen, J. K., Johston, K. M., Frey, S., Pterides, M., Worsely, K., & Ptito, A. (2008). Neural substrates of symptoms of depression following concussion in male athletes with persisting postconcussion symptoms. *Archives of General Psychiatry, 65,* 81-9.

Collie, A., Makdissi, M., Maruff, P., Bennell, K., & McCrory, P. (2006). Cognition in the days following concussion: comparison of symptomatic versus asymptomatic athletes. *Journal of Neurology, Neurosurgery, and Psychiatry, 77*, 241-45.

Collins, M. W., Grindel, S. H., Lovell, M. R., Dede, D. E., Moser, D. J., Phalin, B. R., et al. (1999). Relationship between concussion and neuropsychological performance in college football players. *Journal of the American Medical Association, 282*, 964-70.

Collins, M. W., Lovell, M. R., Iverson, G. L., Cantu, R. C., Maroon, J. C., & Field, M. (2002). Cumulative effects of concussion in high school athletes. *Neurosurgery, 51*, 1175-81.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R Professional Manual: Revised NEO Personality Inventory and NEO Five-Factor Inventory*. Odessa, FL: Psychological Assessment Resources, Inc.

Covassin, T., Stearne, D., & Elbin, R. (2008). Concussion history and postconcussion neurocognitive performance and symptoms in collegiate athletes. *Journal of Athletic Training, 43*, 119-24.

Covassin, T., Swanik, C. B., & Sachs, M. L. (2003). Epidemiological considerations of concussions among intercollegiate athletes. *Applied Neuropsychology, 10,* 12-22.

Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: a generic method with applications. *Neuropsychology, 21*, 419-30.

Davis, G. A., Iverson, G. L., Guskiewicz, K. M., Ptito, A., & Johnston, K. M. (2009). Contributions of neuroimaging, balance testing, electrophysiology and blood markers to the assessment of sport-related concussion. *British Journal of Sports Medicine, 43*, Supp I, i36-i45.

De Beaumont, L., Brisson, B., Lassonde, M., & Jolicoeur, P. (2007). Long-term electrophysiological changes in athletes with a history of multiple concussion. *Brain Injury, 21*, 631-44.

Delaney, J. S., Lacroix, V. J., Leclerc, S., & Johnston, K. M. (2002). Concussions among university football and soccer players. *Clinical Journal of Sports Medicine, 12*, 331-38.

Dick, R. W. (2009). Is there a gender difference in concussion incidence and outcomes? (2009). *British Journal of Sports Medicine, 43 Suppl*, i46-i50.

Echemendia, R. J. (1998). Neuropsychological assessment of the college athlete: The Penn State Concussion Program.

Echemendia, R. J., & Julian, L. J. (2001). Mild traumatic brain injury in sports: Neuropsychology's contribution to a developing field. *Neuropsychology Review, 11*, 69-88.

Echemendia, R. J., Putukian, M., Mackin, R. S., Julian, L., & Shoss, N. (2001). Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clinical Journal of Sport Medicine, 11,* 23-31.

Erlanger, D., Feldman, D., Kutner, K., Kaushik, T., Kroger, H., Festa, J., et al. (2003). Development and validation of a web-based neuropsychological test protocol for sports-related *return-to-play* decision-making. *Archives of Clinical Neuropsychology, 18*, 293-316.

Erlanger, D., Kaushik, T., Cantu, R., Barth, J. T., Broshek, D. K., Freeman, J. R., et al. (2003). Symptom-based assessment of the severity of a concussion. *Journal of Neurosurgery, 98*, 477-84.

Field, M., Collins, M. W., Lovell, M. R., & Maroon, J. (2003). Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. *Journal of Pediatrics, 142*, 546-53.

Gaetz, M., Goodman, D., & Weinberg, H. (2000). Electrophysiological evidence for the cumulative effects of concussion. *Brain Injury, 14*, 1077-88.

Gaetz, M. B., & Iverson, G. L. (2009). Sex difference in self-reported symptoms after aerobic exercise in non-injured athletes. Implications for concussion management programmes. *British Journal of Sports Medicine, 43*, 508-13.

Giza, C. C., & Hovda, D. A. (2004). The pathophysiology of traumatic brain injury. In M. R. Lovell, R. J. Echemendia, J. T. Barth, & M. W. Collins (Eds.), *Traumatic Brain Injury in Sports* (pp. 45-70). Lisse: Swets & Zeitlinger.

Guskiewicz, K. M., Bruce, S. L., Cantu, R. C., Ferrara, M. S., Kelly, J. P., McCrea, M., et al., (2004). Recommendations on management of sport-related concussion: summary of the National Athletic Trainers' Association position statement. *Neurosurgery, 55*, 891-95.

Guskiewicz, K. M., Marshall, S. W., Bailes, J., McCrea, M., Harding, H. P., Jr., Matthews, A., et al. (2007). Recurrent concussion and risk of depression in retired professional football players. *Medicine & Science in Sports & Exercise, 39*, 903-09.

Guskiewicz, K. M., Marshall, S. W., Bailes, J., McCrea, M., Cantu, R. C., Randolph, C., et al. (2005). Association between recurrent concussion and late-life cognitive impairment. *Neurosurgery, 57*, 719-26.

Guskiewicz, K. M., McCrea, M., Marshall, S. W., Cantu, R. C., Randolph, C. Barr, W., et al. (2003). Cumulative effects associated with recurrent concussion in collegiate football players: The NCAA concussion study. *Journal of the American Medical Association, 290*, 2549-555.

Guskiewicz, K. M., Ross, S. E., & Marshall, S. W. (2001). Postural stability and neuropsychological deficits after concussion in collegiate athletes. *Journal of Athletic Training, 36*, 263-73.

Guskiewicz, K. M., Weaver, N. L., Padua, D. A. & Garrett, W. E., Jr. (2000). Epidemiology of concussion in collegiate and high school football players. *American Journal of Sports Medicine, 28*, 643-50.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults professional manual. Lutz, FL: Psychological Assessment Resources.

Hinton-Bayre, A., Geffen, G., Geffen, L., McFarland, K., & Friis, P. (1999). Concussion in contact sports: Reliable change indices of impairment and recovery. *Journal of Clinical and Experimental Neuropsychology, 21*, 70-86.

Holdnack, J. A. (2001). WTAR: Wechsler Test of Adult Reading Manual. San Antonio, TX: Psychological Corporation.

Hootman, J. M., Dick, R., & Agel, J. (2007). Epidemiology of collegiate injuries for 15 sports: Summary and recommendations for injury prevention initiative. Journal of Athletic Training, 42, 311-19. Published online at http://www.nata.org/jat/readers/archives/42.2/i1062-6050-42-2-311.pdf.

Hovda, D. A., Lee, S. M., Smith, M. L., Von Stuck, S., Bergsneider, B., Kelly, D., et al. (1995). The neurochemical and metabolic cascade following brain injury: Moving from animal models to man. *Journal of Neurotrauma*, 12, 903-906.

Ingraham, L. J., & Aiken, C. B. (1996). An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology, 10*, 120-4.

Iverson, G. L. (2005). Outcome from mild traumatic brain injury. *Current Opinion in Psychiatry, 18*, 301-17.

Iverson, G. L., Brooks, B., L., Collins, M. W., & Lovell, M. R. (2005). Tracking neuropsychological recovery following concussion in sport. *Brain Injury, 20*, 245-52.

Iverson, G. L., Brooks, B. L., Lovell, M. R., & Collins, M. W. (2006). No cumulative effects for one or two previous concussion. *British Journal of Sports Medicine, 40*, 72-5.

Iverson, G. L., Gaetz, M., Lovell, M. R., & Collins, M. W. (2004). Cumulative effects of concussion in amateur athletes. *Brain Injury, 1*8, 433-43.

Iverson, G. L., Gaetz, M., Lovell, M. R., & Collins, M. W. (2005). Validity of ImPACT for measuring processing speed following sports-related concussion. *Journal of Clinical and Experimental Neuropsychology, 27,* 683-89.

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2003). Interpreting change on ImPACT following sport concussion. *Clinical Neuropsychologist, 17*, 460-7.

Jantzen, K. J., Anderson, B., Steinberg, F. L., & Kelso, J. A. (2004). A prospective functional MR imaging study of mild traumatic brain injury in college football players. *American Journal of Radiology, 25*, 738-45.

Johnston, K. M., Ptito, A., Chankowsky, J., & Chen, J. K. (2001). New frontiers in diagnostic imaging in concussive head injury. *Clinical Journal of Sport Medicine, 11*, 166-75.

Kant, R., Smith-Seemiller, L, Isaac, G., & Duffy, J. (1997). Tc-HMPAO SPECT in persistent post-concussion syndrome after mild head injury: comparison with MRI/CT. *Brain Injury, 11*, 115-24.

Killam, C., Cautin, R. L., & Santucci, A. C. (2005). Assessing the enduring residual neuropsychological effects of head trauma in college athletes who participate in contact sports. *Archives of Clinical Neuropsychology, 20*, 599-611.

Koh, J. O., Cassidy, J. D., & Watkinson, E. J. (2003). Incidence of concussion in contact sports: A systematic review of the evidence. *Brain Injury, 17*, 901-17.

Lampert, P. W., & Hardman, J. M. (1984). Morphological changes in brains of boxers. *Journal of the American Medical Association, 251*(20), 2676-2679.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer J. S. (2004). *Neuropsychological Assessement, 4th edition*. Oxford: Oxford University Press.

Lovell, M. R., & Collins, M. W. (1998). Neuropsychological assessment of the college football player. *Journal of Head Trauma Rehabilitation, 13*, 9-26.

Lovell, M. R., Collins, M. W., Iverson, G. L., Johnston, K. M., & Bradley, J. P. (2004). Grade 1 or "ding" concussions in high school athletes. *American Journal of Sports Medicine, 32*, 47-54.

Lovell, M. R., Collins, M. W., Podell, K., Powell, J., & Maroon, J. (2000). *ImPACT: Immediate post-concussion assessment and cognitive testing*. Pittsburgh, PA: NeuroHealth Systems, LLC.

Lovell, M. R., Iverson, G. L., Collins, M. W., Podell, K., Johnston, K. M., Pardini, D. et al. (2006). Measurement of symptoms following sports-related concussion: reliability and normative data for the post-concussion scale. *Applied Neuropsychology, 13*, 166-74.

Lovell, M. R., Pardini, J. E., Welling, J., Collins, M. W., Bakal, J., Lazar, N., et al. (2007). Functional brain abnormalities are related to clinical recovery and time to return-to-play in athletes. *Neurosurgery, 61*, 253-9.

Macciocchi, S. N. (2006). Epidemiology of cerebral concussion: the extent of the problem.

Macciocchi, S. N., Barth, J. T., Alves, W., Rimel, R. W. & Jane, J. A. (1996). Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery, 39*, 510-14.

Macciocchi, S. N., Barth, J. T., Littlefield, L., & Cantu, R. C. (2001). Multiple concussion and neuropsychological functioning in collegiate football players. *Journal of Athletic Training, 36,* 303-06.

Matarazzo, J. D., & Prifitera, A. (1989). Subtest scatter and premorbid intelligence: Lessons from the WAIS-R standardization sample*. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 186-91.

Maddocks, D., & Saling, M. (1996). Neuropsychological deficits following concussion. *Brain Injury, 12*, 99-103.

McClincy, M. P., Lovell, M. R., Pardini, J., Collins, M. W., & Spore, M. K. (2006). Recovery from sports concussion in high school and collegiate athletes. *Brain Injury, 20*, 33-9.

McCrea, M., Guskiewicz, K. M., Marshall, S. W., Barr, W., Randolph, C., Cantu, R. C., et al. (2003). Acute effects and recovery time following concussion in collegiate football players: The NCAA concussion study. *Journal of the American Medical Association, 290*, 2556-563.

McCrea, M., Hammeke, T., Olsen, G., Leo, P, & Guskiewicz, K. (2004). Unreported concussion in high school football players: Implications for prevention. *Clinical Journal of Sport Medicine, 14*, 13-17.

McCrea, M. Kelly, J. P., Randolph, C., Cisler, R., & Berger, L. (2002). Immediate neurocognitive effects of concussion. *Neurosurgery, 50*, 1032-42.

McCrory, P., Johnston, K., Meeuwisse, W., Aubry, M., Cantu, R., Dvorak, J., et al. (2005). Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague, 2004. *British Journal of Sports Medicine, 39*, 196-204.

McCrory, P., & Johnston, K. (2002). Acute clinical symptoms of concussion. *The Physician and Sportsmedicine, 30*, 43-7.

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Malloy, M., & Cantu, R. (2009). Consensus statement on concussion in sport – the 3rd International Conference on concussion in sport, held in Zurich, 2008. *Journal of Clinical Neuroscience, 16*, 755=763.

Moser, R. S., Inverson, G. L., Echemendia, R. J., Lovell, M. R., Schatz, P., Webbe, F. M., et al. (2007). Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Archives of Clinical Neuropsychology, 22*, 909-16.

Moser, R. S., Schatz, P. & Jordan, B. D. (2005). Prolonged effects of concussion in high school athletes. *Neurosurgery, 57*, 300-06.

Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1998). Base rates of "impaired" neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology, 13,* 503-11.

Partington, J. E., & Leiter, R. G. (1949). Partington's pathway test. *Psychological Service Center Bulletin, 1* 9-20.

Pellman, E. J. Lovell, M. R., Viano, D. C., & Casson, I. R. (2006). Concussion in professional football: Recovery of NFL and high school athletes assessed by computerized neuropsychological testing – part 12. *Neurosurgery, 58,* 263-74.

Powell, J. W., & Barber-Foss, K. D. (1999). Traumatic brain injury in high school athletes. *Journal of the American Medical Association, 282*, 985-63.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery*. Tucson, AZ: Neuropsychology Press.

Reynolds, C. R., Pearson, N. A., & Voress, J. K. (2002). *Developmental Test of Visual Perception-Adolescent and Adult*. Austin, TX: Pro-Ed.

Rimel, R. W., Giordani, B., Barth, J. T., Boll, T. J., & Jane, J. A. (1981). Disability caused by minor head injury. *Neurosurgery, 9*, 221-28.

Schatz, P., Pardini, J. E., Lovell, M. R., Collins, M. W., & Podell, K. (2006). Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Archives of Clinical Neuropsychology, 21*, 91-9.

Schinka, J. A., Vanderploeg, R. D., & Curtiss, G. (1994). Wechsler Adult Intelligence Scale-Revised subtest scatter as a function of maximum subtest scaled score. *Psychological Assessment, 6*, 364-67.

Schretlen, D. J., Munro, C. A., Anthony, J. C., & Pearlson, G. D. (2003). Examining the range of normal intraindividual variability in neuropsychological test performance. *Journal of the International Neuropsychological Society, 9*, 864-70.

Schretlen, D. J., Testa, S., M., Winicki, J. J., Pearlson, D. G., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society, 14*, 436-45.

Schulz, M. R., Marshall, S. W., Mueller, F. O., Yang, J., Weaver, N. L., Kalsbeek, W. D., & Bowling, J. M. (2004). Incidence and risk factors for concussion in high school athletes, North Carolina, 1996-1999. *American Journal of Epidemiology, 160*, 937-44.

Schwarz, A. (2009). New sign of brain damage in NFL. Published online at the New York Times, January 27, at http://www.nytimes.com/2009/01/28/sports/28iht-footbrain28.19747479.html.

Schwarz, A. (2009). Dementia risk seen in players in NFL study. Published online at the New York Times. September 9, at http://www.nytimes.com/2009/09/30/sports/football/30dementia.html?_r=1.

Smith, A. (1991). *Symbol Digit Modalities Test.* Los Angeles, CA: Western Psychological Services.

Stern, R. A., & White, T. (2003). *Neuropsychological Assessment Battery*. Lutz, FL: Psychological Assessment Battery.

Sternberg, R. J. (2000). *Handbook of Intelligence*. London: Cambridge University Press.

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 3rd Ed*. Oxford: Oxford University Press.

Tabachnik, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics: Fifth Edition*. San Francisco: Pearson Education, Inc.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *Stroop Neuropsychological Screening Test*. Odessa, FL: Psychological Assessment Resources.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale:Third Edition: Administration and Scoring Manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *WAIS-IV Wechsler Adult Intelligence Scale: Fourth Edition: Administration and Scoring Manual.* San Antonio, TX: NCS Pearson, Inc.

Wilson, B. A., Cockburn, J., & Baddely, A. (2001). The Rivermead Behavioral Memory Test. Bury St. Edmunds, England: Thames Valley Test Company.

Zemper, E. D. (2003). Two-year prospective study of relative risk of second cerebral concussion. *American Journal of Physical Medicine Rehabilitation, 82*, 653-659.

APPENDIX A

Neuropsychological tests: Time of administration[9]

First Section (15-40 minutes)

Brief Visuospatial Memory Test- Immediate Recall

Hopkins Verbal Learning Test- Immediate Recall

Symbol Digit Modalities Test

Symbol Digit Modalities Test- Incidental Memory

Second Section (40-65 minutes)

Trail Making Test-Condition 1

Trail Making Test-Condition 2

Hopkins Verbal Learning Test- Delayed Recall

Brief Visuospatial Memory Test- Delayed Recall

Third Section (65-90 minutes)

Digit Span Test- Total

Stroop Color-Naming Test

Stroop Color-Word Test

PSU Cancellation Test

[9]Computerized tests are excluded because the majority of these tests (4/5) are administered at the same point in the assessment battery as part of the ImPACT, thus confounding test mode and time of administration.

APPENDIX B

Neuropsychological tests: Mode of administration

2 Administrative Modes

Pencil-and-paper (7)

- HVLT Total Recall
- BVMT Total Recall
- SDMT Total Correct
- TMT 1 Time
- TMT 2 Time
- Stroop Color-Naming
- Stroop Color-Word

Computer-administered (5)

- Vigil Average Delay
- ImPACT-Verbal Memory Index
- ImPACT-Visual Memory Index
- ImPACT-Visual Motor Speed Composite
- ImPACT-Reaction Time Composite

3 Response Modes

Oral (5)

- Digit Span
- HVLT Total Recall
- HVLT Delayed Recall
- Stroop Color-Naming
- Stroop Color-Word

Written (5)

- BVMT Total Recall
- BVMT Delayed Recall
- SDMT Total Correct
- TMT 1 Time
- TMT 2 Time

Computer (5)

- Vigil Average Delay
- ImPACT-Verbal Memory Index
- ImPACT-Visual Memory Index
- ImPACT-Visual Motor Speed Composite
- ImPACT-Reaction Time Composite

APPENDIX C

Neuropsychological tests: Cognitive domains

Theoretical Cognitive Domains

Attention/speeded processing (3)

SDMT Total Correct

Vigil Average Delay

ImPACT-Reaction Time Composite

Verbal learning/memory (3)

HVLT Total Recall

HVLT Delayed Recall

ImPACT-Verbal Memory Composite

Visual learning/memory (3)

BVMT Total Recall

BVMT Delayed Recall

ImPACT-Visual Memory Composite

Executive functioning (3)

Stroop Color-Word

TMT 1 Time

TMT 2 Time

Empirical Cognitive Domains

Speeded processing (8)

SDMT Total Correct

Vigil Average Delay

ImPACT-Reaction Time Composite

ImPACT-Visual Memory Composite

TMT 1 Time

TMT 2 Time

Stroop Color-Word

Stroop Color-Naming

Verbal memory (3)

HVLT Total Recall

HVLT Delayed Recall

ImPACT-Verbal Memory Composite

Visual memory (4)

BVMT Total Recall

BVMT Delayed Recall

ImPACT-Visual Memory Composite

SDMT Incidental Memory

APPENDIX D

Binomial probability distribution formulae

The binomial probability distribution is the probability distribution of a binomial random variable. The formula for calculating the binomial probability for a specific event or events is

$$b\,(x;\, n,\, P) = {}_{n}C_{x} * P^{x} * (1 - P)^{\,n - x}$$

where

- x: the number of events i.e., number of test scores 1, 1.5, or 2 SD below the mean (1, 2, 3, etc.).
- n: the number of trials in the experiment e.g., the number of tests in the battery (17).
- P: the probability of an event occurring on an individual trial i.e., the probability of obtaining a test score 1, 1.5, or 2 SD below the mean (.159, .091, .023).
- 1 – P: the probability of a non-event occurring on an individual trial i.e., the probability of obtaining a test score less than 1, 1.5, or 2 SD below the mean (.841, .909, .977).
- b (x; n, P): binomial probability, or the probability of exactly *x* events occurring across a n-trial experiment when the probability of an event on an individual trial is P.
- ${}_{n}C_{x}$: the number of combinations of *n* things taken *x* at a time i.e., the number of 17 test scores taken 1, 2, 3, etc. at a time.

Cumulative binomial probability refers to the probability that a binomial random variable falls within a specified range i.e., is greater than or equal to a specified lower limit or is less than or equal to a specified upper limit i.e., is greater than or equal to 2 test scores that are 1.5 SD below the mean. The formulae for cumulative binomial probability is

$$b_{c}\,(x \geq \text{lower limit};\, n,\, P) = \sum_{LL}^{n} [{}_{n}C_{x} * P^{x} * (1 - P)^{n - x}] \;\; \textit{or} \;\; b_{c}\,(x \leq \text{upper limit};\, n,\, P) = \sum_{0}^{UL} [{}_{n}C_{x} * P^{x} * (1 - P)^{n - x}]$$

where, in addition to the variables defined above, ∑ refers to the sum of all individual binomial distribution probabilities from a lower limit to n or from 0 to an upper limit.

APPENDIX E

MatLab program for Monte Carlo simulation

Performance of n subjects on a 17-test battery

1. *Generate correlation matrix(a) for 17-test battery and check matrix size (17x17).*

```
a=[1.0 -.184 -.117 -.232 .308 -.044 -.068 -.019 -.088 -.216 -.065 .126 .124 .205 .192 .065 -.173;-
.184 1.0 .484 .408 -.238 .323 .281 .316 .315 .408 .359 -.228 -.228 -.148 -.229 .182 .084;-.117 .484
1.0 .308 -.236 .336 .312 .208 .226 .320 .252 -.199 -.178 -.103 -.154 .183 .095;-.232 .408 .308 1.0 -
.374 .241 .204 .242 .294 .459 .211 -.268 -.245 -.229 -.381 .229 .265;.308 -.238 -.236 -.374 1.0 -
.178 -.190 -.159 -.219 -.358 -.160 .284 .249 .183 .333 -.143 -.190;-.044 .323 .336 .241 -.178 1.0
.758 .277 .252 .327 .402 -.196 -.251 -.049 -.177 .158 -.046;-.068 .281 .312 .204 -.190 .758 1.0
.216 .219 .280 .423 -.185 -.238 -.034 -.141 .082 -.115;-.019 .316 .208 .242 -.159 .277 .216 1.0
.688 .343 .240 -.151 -.193 -.085 -.217 .121 .076;-.088 .315 .226 .294 -.219 .252 .219 .688 1.0 .363
.240 -.160 -.193 -.037 -.253 .094 .073;-.216 .408 .320 .459 -.358 .327 .280 .343 .363 1.0 .403 -
.278 -.234 -.224 -.380 .192 .278;-.065 .359 .252 .211 -.160 .402 .423 .240 .240 .403 1.0 -.162 -
.188 -.071 -.116 .125 -.028;.126 -.228 -.199 -.268 .284 -.196 -.185 -.151 -.160 -.278 -.162 1.0 .645
.191 .184 -.159 -.025;.124 -.228 -.178 -.245 .249 -.251 -.238 -.193 -.193 -.234 -.188 .645 1.0 .160
.198 -.207 .243;.205 -.148 -.103 -.229 .183 -.049 -.034 -.085 -.037 -.224 -.071 .191 .160 1.0 .514 -
.209 -.167;.192 -.229 -.154 -.381 .333 -.177 -.141 -.217 -.253 -.380 -.116 .184 .198 .514 1.0 -.205 -
.191;.065 .182 .183 .229 -.143 .158 .082 .121 .094 .192 .125 -.159 -.207 -.209 -.205 1.0 .074;-.173
.084 .095 .265 -.190 -.046 -.115 .076 .073 .278 -.028 -.025 .243 -.167 -.191 .074 1.0];
size(a);
```

2. *Generate Lower Choleski decomposition of correlation matrix(a), where a = CDa*CDa', and define size of matrix CDa [rows,columns].*

```
CDa=chol(a,'lower');
[rows columns] = size(CDa);
```

3. *Generate number of pseudo-subjects (as many as you want e.g., 5 to 500,000).*

```
n=500000;
```

4. *Generate a set of numbers or 17 "test scores" (D) for each pseudo-subject 1 through n by multiplying the Choleski decomposition (CDa) by n rows of random normal variables.*

```
DD=CDa*randn(rows,n);
```

5. *Clear previous [rows, columns] for matrix CDa and define size of matrix DD [rows,columns] which will be 17 rows x n columns*

```
clear rows columns
[rows columns] = size(DD);
DD_size=rows*columns;
```

6. *Determine how many numbers per column (i.e., per pseudo-subject) going down the row of "test scores" (i.e., out of 17) are less than each of three defined cutoff levels(≤ -1.0, ≤ -1.5, ≤ 2.0) and sum the total for each level (number_below_cutoff)*

```
cutoff=[-1.99 -1.49 -.99];
num_cutoff=numel(cutoff);
AtLeastFailuresInCutoff=zeros(num_cutoff,columns);
for i=1:num_cutoff
  AtLeastFailuresInCutoff(i,:)=sum(DD<=cutoff(i));
end;
number_below_cutoff=sum(AtLeastFailuresInCutoff,2);
```

In addition, number_below_cutoff17 (a matrix 17 x 3) is the fraction of 1 or more tests below each cutoff, 2 or more tests below each cutoff, etc.

```
number_below_cutoff17=zeros(rows,num_cutoff);
for i=1:num_cutoff
  tmp=zeros(rows,columns);
  for j=1:rows
    tmp(j,:)= AtLeastFailuresInCutoff(i,:)>=j;
  end;
  number_below_cutoff17(:,i)=mean(tmp,2);
end;
```

7. *Display total number of test scores at 1, 1.5, and 2 SD below mean (number_below_cutoff) and percent of these scores out of total test scores for all randomly generated pseudo-subjects (DD_size = n x 17) for 1 or more, 2 or more, 3 or more, etc, .tests up to 17 (AtLeastFailuresInCutoff).*

```
format bank
display('Total number below cutoff 1, 1.5, and 2')
number_below_cutoff
display('Total number of tests')
DD_size
display('Fraction below cutoff 1, 1.5, and 2')
number_below_cutoff./DD_size
display('Each column refers to a cutoff value, each row is the fraction of patients with n or more
tests below cutoff')
number_below_cutoff17
```

APPENDIX F

Tables

Table 1

Demographics for Overall Athlete Sample

N = 612	M	SD
Age (years)	18.5	1.0
Full Scale IQ	103.0	5.9
	Frequency	Percent
Sex		
Male	454	74.2
Female	158	25.8
Handedness		
Right	529	86.4
Left	66	10.8
Bilateral	11	1.8
Ethnicity		
White	457	74.7
Black	114	18.6
Multiracial	15	2.5
Hispanic	7	1.1
Asian	6	1.0
Other	11	1.8
Sport		
Football	184	30.1
Lacrosse	136	22.2
Soccer	127	20.8
Basketball	66	10.8
Ice Hockey	56	9.2
Wrestling	18	2.9
Other	25	4.1

Table 2

Lowest Z-score, Highest Z-score, and Maximum Discrepancy Between Scores on a Concussion Assessment Battery

N = 495	Mean	Minimum	Maximum	Range
Lowest z-score	-1.8	-7.7	0.2	7.9
Highest z-score	1.4	-0.3	3.8	4.1
Maximum discrepancy	3.2	1.3	8.8	7.5

Table 3

Frequency and Percent of Athletes with Test Scores 1, 1.5, and 2 SD Below the Mean on a Concussion Assessment Battery

	1 SD Below Mean		1.5 SD Below Mean		2 SD Below Mean	
# Low Test Scores	Frequency	Percent	Frequency	Percent	Frequency	Percent
0	106	21.4	245	49.5	350	70.7
1	117	23.6	116	23.4	80	16.2
2	85	17.2	58	11.7	31	6.3
3	56	11.3	28	5.7	15	3.0
4	45	9.1	17	3.4	10	2.0
5	26	5.3	13	2.6	5	1.0
6	20	4.0	4	.8	3	.6
7	11	2.2	4	.8	1	.2
8	11	2.2	5	1.0		
9	6	1.2	4	.8		
10	4	.8	--	--		
11	3	.6	1	.2		
12	3	.6				
13	2	.4				
Total	495	100.0	495	100.0	495	100.0

Table 4

Cumulative Percent of Athletes with Test Scores 1, 1.5, and 2 SD Below the Mean on a Concussion Assessment Battery

	1 SD Below Mean		1.5 SD Below Mean		2 SD Below Mean	
# Low Test Scores	Frequency	Cumulative Percent	Frequency	Cumulative Percent	Frequency	Cumulative Percent
5+	86	17.4	31	6.3	9	1.8
4+	45	26.5	17	9.7	10	3.8
3+	56	37.8	28	15.4	15	6.9
2+	85	54.9	58	27.1	31	13.1
1+	117	78.6	116	50.5	80	29.3
0	106	21.4	245	49.5	350	70.7
Total	495		495		495	

Table 5

Significant Differences in Number of Test Scores 1, 1.5, and 2 SD Below the Mean at the Beginning, Middle, and End of a Concussion Assessment Battery

	Within-Subject Effect of Time					
R-M ANOVA	Total # Low Test Scores M (SD)	95% Confidence	df Huynh-Feldt	F Univariate	p	Cohen's d cf. to End
1 SD Below Mean			1.9, 930.3	0.1	.93	
Beginning	.60 (.9)	.52 - .68				--
Middle	.59 (.9)	.51 - .67				--
End	.58 (.8)	.51 - .65				--
1.5 SD Below Mean			1.9, 948.9	8.5	.00	
Beginning	.34 (.7)	.28 - .40				.22
Middle	.30 (.6)	.24 - .35				.16
End	.21 (.5)	.16 - .23				--
2 SD Below Mean			1.9, 960.6	5.2	.01	
Beginning	.15 (.5)	.11 - .19				.15
Middle	.15 (.4)	.11 - .19				.17
End	.09 (.3)	.06 - .10				--

Table 6

Factor Correlations, Factor Loadings, Communalities (h^2), and Percents of Variance for Principal Factors Extraction with Oblique Rotation on Neuropsychological Tests

	Factor Correlations			
Factor 1[a]	1.0	--	--	
Factor 2[a]	-.21	1.0	--	
Factor 3[a]	-.32	.30	1.0	
Test		Factor Loadings		Comm
	1	2	3	h^2
Stroop 1 Time	.57			.32
Stroop 2 Time	.56			.40
IMPACT Motor Speed	-.54			.35
IMPACT Reaction Time	.53			.28
TMT 1 Time	.43			.43
Vigil Average Delay	.43			.16
SDMT Total Correct	-.41			.41
TMT 2 Time	.36			.44
BVMT Delayed Recall		.86		.57
BVMT Total Recall		.81		.58
SDMT Memory Correct		.39		.26
IMPACT Visual Memory		.38		.30
HVLT Delayed Recall			.82	.49
HVLT Total Recall			.77	.48
IMPACT Verbal Memory			.41	.37
% of Variance	30.15	11.42	8.45	T=50.02

[a]Suggested factor labels: 1 = Speeded Processing, 2 = Visual Memory, 3 = Verbal Memory

Table 7

Standard Multiple Regression of Athlete Demographic and Personality Variables on Number of Test Scores 1 SD Below the Mean

N=265				Correlation (r)							Regression		
Variables	DV[a]	Sex	IQ	Sport$_1$[b]	Sport$_2$[b]	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C	*B*	β	sr^2
Sex	.14**	--									.22*	.14	.01
IQ	-.28**	.14**	--								-.05**	-.30	.08
Sport$_1$	.13*	.08	.15**	--							.33**	.20	.03
Sport$_2$	.10*	.01	-.25**	-.44**	--						.24^	.12	.01
NEO-N	.05	-.14**	-.10*	.05	.00	--					.00	.01	.00
NEO-E	-.04	-.15**	-.02	.07	-.02	-.32**	--				.00	.00	.00
NEO-O	-.14**	-.16**	.02	-.08	.01	.03	.02	--			-.02^	-.11	.01
NEO-A	-.10*	-.30**	.02	-.01	-.06	-.21**	.22**	.05	--		.00	.00	.00
NEO-C	-.10*	-.14**	-.03	-.03	.01	-.29**	.31**	-.17**	.41**	--	-.01	-.09	.00
					Mean and Standard Deviation						Intercept	R	R^2
M	1.23	0.55	103.00	0.44	0.20	17.07	32.68	24.20	32.70	33.15	6.47	.40**	.16[c]
SD	0.82	0.50	5.28	0.50	0.40	6.70	4.55	6.01	5.46	6.14			

**p < .01; *p < .05; ^p < .10.

[a]DV is the number of tests 1 SD below the mean, square root transformed.

[b]Soccer is the reference group, Sport$_1$ = Lacrosse, Sport$_2$ = Basketball.

[c]Unique variance = .14; shared variance = .02.

Table 8

Standard Multiple Regression of Athlete Demographic and Personality Variables on Number of Test Scores 1.5 SD Below the Mean

N=265	Correlation (r)										Regression		
Variables	DV[a]	Sex	IQ	$Sport_1$[b]	$Sport_2$[b]	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C	*B*	β	sr^2
Sex	.10*	--									.14	.10	.00
IQ	-.20**	.14**	--								-.03**	-.20	.03
$Sport_1$	.05	.08	.15**	--							.17^	.12	.01
$Sport_2$	.13*	.01	-.25**	-.44**	--						.25*	.13	.01
NEO-N	.06	-.14**	-.10*	.05	.00	--					.00	.02	.00
NEO-E	-.06	-.15**	-.02	.07	-.02	-.32**	--				.00	-.02	.00
NEO-O	-.11*	-.16**	.02	-.08	.01	.03	.02	--			-.01^	-.10	.01
NEO-A	-.08	-.30**	.02	-.01	-.06	-.21**	.22**	.05	--		.00	.03	.00
NEO-C	-.11*	-.14**	-.03	-.03	.01	-.29**	.31**	-.17**	.41**	--	-.01	-.11	.00
	Mean and Standard Deviation										Intercept	R	R^2
M	0.63	0.55	103.00	0.44	0.20	17.07	32.68	24.20	32.70	33.15	3.90	.31**	.10[c]
SD	0.73	0.50	5.28	0.50	0.40	6.71	4.55	6.00	5.46	6.14			

**p < .01; *p < .05; ^p < .10.

[a]DV is the number of tests 1.5 SD below the mean, square root transformed.

[b]Soccer is the reference group, $Sport_1$ = Lacrosse, $Sport_2$ = Basketball.

[c]Unique variance = .06; shared variance = .04.

Table 9

Standard Multiple Regression of Athlete Demographic and Personality Variables on Number of Test Scores 2 SD Below the Mean

N=265	Correlation (r)										Regression		
Variables	DV[a]	Sex	IQ	$Sport_1$[b]	$Sport_2$[b]	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C	*B*	β	sr^2
Sex	.07	--									.09	.08	.00
IQ	-.20**	.14**	--								-.02**	-.18	.03
$Sport_1$	-.11*	.08	.15**	--							-.07	-.06	.00
$Sport_2$	.17**	.01	-.25**	-.44**	--						.14	.10	.00
NEO-N	.02	-.14**	-.10*	.05	.00	--					.00	.00	.00
NEO-E	-.04	-.15**	-.02	.07	-.02	-.32**	--				.00	.00	.00
NEO-O	-.05	-.16**	.02	-.08	.01	.03	.02	--			.00	-.05	.00
NEO-A	-.07	-.30**	.02	-.01	-.06	-.21**	.22**	.05	--		.00	.00	.00
NEO-C	-.07	-.14**	-.03	-.03	.01	-.29**	.31**	-.17**	.41**	--	.00	-.07	.00
	Mean and Standard Deviation										Intercept	R	R^2
M	0.31	0.55	103.00	0.44	0.20	17.07	32.68	24.20	32.70	33.15	2.57	.27*	.07[c]
SD	0.57	0.50	5.28	0.50	0.40	6.70	4.55	6.01	5.46	6.14			

**p < .01; *p < .05; ^p < .10.

[a]DV is the number of tests 2 SD below the mean, square root transformed.

[b]Soccer is the reference group, $Sport_1$ = Lacrosse, $Sport_2$ = Basketball.

[c]Unique variance = .04; shared variance = .03.

Table 10

Demographics for Male and Female Athlete Groups

	Male (n=371)	Female (n=124)
	M (SD)	M (SD)
Age (years)	18.5 (0.9)	18.1 (0.7)
Full Scale IQ	103.2 (6.2)	102.2 (4.8)
	Frequency (pct)	Frequency (pct)
Handedness		
Right	321 (86.5)	110 (88.7)
Left	41 (11.1)	12 (9.7)
Bilateral	7 (1.9)	2 (1.6)
Ethnicity		
White	261 (70.4)	101 (81.5)
Black	84 (22.6)	15 (12.1)
Multiracial	10 (2.7)	5 (4.0)
Hispanic	4 (1.1)	2 (1.6)
Asian	3 (0.8)	0
Other	9 (2.4)	1 (0.8)
Other		
ADHD	23 (6.2)	5 (4.0)
LD	22 (5.9)	5 (4.0)
Prior Testing	35 (9.4)	20 (16.1)

Table 11

Lowest Zscore, Highest Zscore, and Maximum Discrepancy between Scores for Male and Female Athletes on a Concussion Assessment Battery

	Sex	N	Mean	Minimum	Maximum	Range	Cohen's d
Lowest z-score**	Male	371	-1.8	-7.7	-0.9	7.6	.30
	Female	124	-1.5	-4.6	0.2	4.8	
Highest z-score^	Male	371	1.4	-0.3	3.5	3.9	.18
	Female	124	1.5	0.5	3.8	3.3	
Maximum discrepancy^	Male	371	3.2	1.3	8.8	7.5	.20
	Female	124	3.0	1.3	6.0	4.7	

**p < .01; ^p < .10.

Table 12

Total Number of Test Scores 1, 1.5, and 2 SD Below the Mean for Male and Female Athletes on a Concussion Assessment Battery

			Average # Low Test Scores			
	Sex	N	Mean	SD	Range	Cohen's d
1 SD below mean**	Male	371	2.7	2.7	13	.28
	Female	124	2.0	2.1	9	
1.5 SD below mean**	Male	371	1.3	1.9	11	.28
	Female	124	0.8	1.3	7	
2 SD below mean**	Male	371	0.6	1.2	7	.27
	Female	124	0.3	0.7	4	

**p < .01.

Table 13

Frequency and Percent of Male and Female Athletes with Test Scores 1, 1.5, and 2 SD Below the Mean on a Concussion Assessment Battery

		1 SD Below Mean		1.5 SD Below Mean		2 SD Below Mean	
Sex	# Low Test Scores	Frequency	Percent	Frequency	Percent	Frequency	Percent
Male	0	71	19.1	174	46.9	255	68.7
	1	83	22.4	87	23.5	61	16.4
	2	67	18.1	44	11.9	23	6.2
	3	44	11.9	25	6.7	15	4.0
	4	35	9.4	14	3.8	8	2.2
	5	19	5.1	11	3.0	5	1.3
	6	18	4.9	3	0.8	3	0.8
	7	10	2.7	3	0.8	1	0.3
	8	9	2.4	5	1.3		
	9	3	0.8	4	1.1		
	10	4	1.1	--	--		
	11	3	0.8	1	0.3		
	12	3	0.8				
	13	2	0.5				
	Total	371	100.0	371	100.0	371	100.0
Female	0	35	28.2	71	57.3	95	76.6
	1	34	27.4	29	23.4	19	15.3
	2	18	14.5	14	11.3	8	6.5
	3	12	9.7	3	2.4	--	--
	4	10	8.1	3	2.4	2	1.6
	5	7	5.6	2	1.6		
	6	2	1.6	1	0.8		
	7	1	0.8	1	0.8		
	8	2	1.6				
	9	3	2.4				
	Total	124	100.0	124	100.0	124	100.0

Table 14

Cumulative Percent of Male and Female Athletes with Test Scores 1, 1.5, and 2 SD Below the Mean on a Concussion Assessment Battery

		1 SD Below Mean		1.5 SD Below Mean		2 SD Below Mean	
Sex	# Low Test Scores	Frequency	Cumulative Percent	Frequency	Cumulative Percent	Frequency	Cumulative Percent
Male	5+	71	19.1	27	7.3	9	2.4
	4+	35	28.6	14	11.1	8	4.6
	3+	44	40.4*	25	17.8*	15	8.6
	2+	67	58.5*	44	29.6*	23	14.8
	1+	83	80.9*	87	53.1*	61	31.3
	0	71	19.1*	174	46.9*	255	68.7
	Total	371		371		371	
Female	5+	15	12.1	4	3.2	--	--
	4+	10	20.2	3	5.6	2	1.6
	3+	12	29.8*	3	8.1*	--	--
	2+	18	44.4*	14	19.4*	8	8.1
	1+	34	71.8*	29	42.7*	19	23.4
	0	35	28.2*	71	57.3*	95	76.6
	Total	124		124		124	

* Percentages different between male and female athletes significantly at $p < .05$.

Table 15

Significant Differences Between Male and Female Athletes in Number of Test Scores 1, 1.5, and 2 SD Below the Mean at Beginning, Middle, and End of a Concussion Assessment Battery

R-M ANOVA	Total # Low Test Scores M (SD)	Marginal M: Sex	Univariate F Test (Huynh-Feldt corrected) Effects of Time, Sex, and Time X Sex		Cohen's d
1 SD Below Mean					
Male (n=371)		.64			.25
Beginning	.67 (1.0)				
Middle	.65 (0.9)				
End	.60 (0.8)		Time	F (2, 931) = 0.3, p >.10	
Female (n=124)		.43	*Sex	F (1, 493) = 10.2, p < .01	.25
Beginning	.37 (0.6)		Time X Sex	F (2, 931) = 2.1, p > .10	
Middle	.41 (0.7)				
End	.52 (0.7)				
Marginal M: Time	Beginning = .52, Middle = .53, End = .56				--
1.5 SD Below Mean					
Male (n=371)		.31			.20
Beginning	.39 (0.8)				
Middle	.33 (0.7)				
End	.23 (0.6)		*Time	F (2, 949) = 4.6, p <.05	
Female (n=124)		.18	*Sex	F (1, 493) = 7.7, p < .01	.20
Beginning	.20 (0.4)		Time X Sex	F (2, 949) = 0.8, p > .10	
Middle	.21 (0.5)				
End	.14 (0.4)				
Marginal M: Time	Beginning = .30, Middle = .27, end = .18				.22
2 SD Below Mean					
Male (n=371)		.15			.20
Beginning	.17 (0.5)				
Middle	.17 (0.5)				
End	.11 (0.4)		*Time	F (2, 961) = 3.6, p <.05	
Female (n=124)		.07	*Sex	F (1, 493) = 6.9, p < .01	.20
Beginning	.09 (0.3)		Time X Sex	F (2, 961) = 0.2, p > .10	
Middle	.09 (0.3)				
End	.03 (0.2)				
Marginal M: Time	Beginning = .13, Middle = .13, end = .07				.15

*p < .05.

Table 16

Significant Differences Between Male and Female Athletes in Number of Test Scores 1, 1.5, and 2 SD Below the Mean on Pencil-and-Paper Versus Computer-Administered Tests on a Concussion Assessment Battery

R-M ANOVA	Total # Low Test Scores M (SD)	Marginal M: Sex	Univariate F Test Effects of Mode, Sex, and Mode X Sex		Cohen's d
1 SD Below Mean					
Male (n=371)		.16			.27
Pencil-Paper	.16 (0.1)				
Computer	.15 (0.2)		Mode	F (1, 493) = 0.5, p > .10	
Female (n=124)		.12	*Sex	F (1, 493) = 5.3, p < .05	.27
Pencil-Paper	.10 (0.1)		*Mode X Sex	F (1, 493) = 3.9, p < .05	
Computer	.13 (0.2)				
Marginal M: Mode	Pencil-Paper = .13, Computer = .14				--
1.5 SD Below Mean					
Male (n=371)		.08			.30
Pencil-Paper	.09 (0.1)				
Computer	.07 (0.1)		Mode	F (1, 493) = 0.1, p > .10	
Female (n=124)		.05	*Sex	F (1, 493) = 6.6, p < .05	.30
Pencil-Paper	.04 (0.1)		^Mode X Sex	F (1, 493) = 3.3, p = .06	
Computer	.05 (0.1)				
Marginal M: Mode	Pencil-Paper = .07, Computer = .06				--
2 SD Below Mean					
Male (n=371)		.04			.20
Pencil-Paper	.04 (0.1)				
Computer	.04 (0.1)		Mode	F (1, 493) = 0.1, p > .10	
Female (n=124)		.02	*Sex	F (1, 493) = 5.0, p < .05	.20
Pencil-Paper	.02 (0.1)		^Mode X Sex	F (1, 493) = 3.1, p = .07	
Computer	.03 (0.1)				
Marginal M: Mode	Pencil-Paper = .03, Computer = .03				--

*p < .05; ^p < .10.

Table 17

Standard Multiple Regression of Athlete Demographic and Personality Variables on Number of Test Scores 1, 1.5, and 2 SD Below the Mean for Male and Female Athletes on a Concussion Assessment Battery

	1 SD Below Mean						1.5 SD Below Mean						2 SD Below Mean					
Variables	*B*	β	r	sr^2	R	R^2	*B*	β	r	sr^2	R	R^2	*B*	β	r	sr^2	R	R^2
Male (n=146)																		
Int	5.50				.31^	.09[c]	3.45				.25	.06[d]	2.30				.27	.08[e]
IQ	-.04**	-.27	-.28**	.07			-.02*	-.18	-.20**	.03			-.02^	-.15	-.19*	.02		
$Sport_1$ [b]	.15	.09	.04	.00			.13	.08	.02	.00			-.17	-.14	-.20**	.01		
$Sport_2$ [b]	.11	.06	.08	.00			.18	.09	.09	.00			.13	.08	.18*	.00		
NEO-N	.00	.02	.07	.00			.00	.01	.06	.00			.00	.01	.05	.00		
NEO-E	.01	.03	.03	.00			.00	.01	.00	.00			.00	-.02	-.05	.00		
NEO-O	-.01	-.06	-.07	.00			-.01	-.08	-.07	.00			.00	.02	.03	.00		
NEO-A	.00	.01	-.03	.00			.01	.07	.00	.00			.00	.00	-.02	.00		
NEO-C	-.01	-.07	-.05	.00			-.01	-.12	-.08	.00			.00	-.05	-.06	.00		
Female (n=119)																		
Int	8.48				.50**	.25[f]	4.91				.37*	.14[g]	3.10				.33	.11[h]
IQ	-.06**	-.33	-.34**	.10			-.03*	-.22	-.25**	.04			-.02*	-.21	-.25**	.04		
$Sport_1$ [b]	.56**	.33	.21**	.09			.24^	.17	.07	.02			.07	.07	.00	.00		
$Sport_2$ [b]	.35^	.17	.14^	.02			.30^	.17	.18*	.02			.16	.13	.15*	.01		
NEO-N	.00	-.02	.08	.00			.00	.02	.11	.00			.00	-.04	-.01	.00		
NEO-E	-.01	-.02	-.07	.00			-.01	-.05	-.12	.00			.01	.06	.01	.00		
NEO-O	-.02^	-.16	-.19*	.02			-.02	-.12	-.14^	.01			-.02^	-.17	-.16*	.03		
NEO-A	-.01	-.04	-.11	.00			-.01	-.04	-.13	.00			.00	-.02	-.09	.00		
NEO-C	-.02	-.14	-.11	.01			-.01	-.12	-.12	.01			-.01	-.10	-.06	.00		

***p < .01; p < .05; ^p < .10.

[b]Soccer is the reference group, $Sport_1$ = Lacrosse, $Sport_2$ = Basketball.

[c]Unique variance = .07 and shared variance = .02; [d]Unique variance = .03 and shared variance = .03; [e]Unique variance = .03 and shared variance = .05.

[f]Unique variance = .24 and shared variance = .01; [g]Unique variance = .10 and shared variance = .04; [h]Unique variance = .08 and shared variance = .03.

Table 18

Mean Maximum Discrepancy Between Highest and Lowest Standardized Scores on Common Neuropsychological Test Batteries Along with Percent of Sample Showing Discrepancy ≥ 3 SD

	Normative Sample					
Test Battery	N	Age (years) Range	Education (years) Range	# Subtests/ # Scores	Mean Maximum Discrepancy	% with Maximum Discrepancy ≥ 3 SD
WAIS-R	1,880	16-74	8-16+	11/11	2.2 SD	18%
WAIS-III	2.450	16-89	8-16+	13/13	2.3 SD	22%
WAIS-IV	2,200	16-90	8-16+	15/15	2.5 SD	30%
Aging, Brain Imaging, & Cognition (ABC)	197	20-92	14	15/32	3.4 SD	65%
PSU-Concussion Testing Program	495	18-21	12-15	15/17	3.2 SD	50%

Note: Information from Matarazzo & Prifitera, 1989; *WAIS-III Administration and Scoring Manual*, Table B.5, 1997; *WAIS-IV Administration and Scoring Manual*, Table B.6, 2008; Schretlen et al., 2003.

Table 19

Cumulative Percent of Subjects Scoring 1, 1.5 or 2 SD Below the Mean for 0 to 5 Tests on Common Neuropsychological Test Batteries

Author	Battery	# Test scores	Sample size	# Tests > 1 SD below mean	# Tests > 1.5 SD below mean	# Tests > 2 SD below mean
Heaton et al., 2004	E-HRNB	25	1,189	13% = 0	41% = 0	72% = 0
				87% ≥ 1	59% ≥ 1	28% ≥ 1
				72% ≥ 2	37% ≥ 2	10% ≥ 2
				58% ≥ 3	25% ≥ 3	5% ≥ 3
				45% ≥ 4	15% ≥ 4	2% ≥ 4
				36% ≥ 5	10% ≥ 5	1% ≥ 5
Schretlen et al. (2008)	ABC study	10	220-327	-- = 0	-- = 0	-- = 0
				-- ≥ 1	-- ≥ 1	-- ≥ 1
				36% ≥ 2	15% ≥ 2	3% ≥ 2
Schretlen et al. (2008)	ABC study	25	220-327	-- = 0	-- = 0	-- = 0
				-- ≥ 1	-- ≥ 1	-- ≥ 1
				75% ≥ 2	40% ≥ 2	14% ≥ 2
				53% ≥ 3	16% ≥ 3	3% ≥ 3
Iverson et al. (2008)	NAB	36	1,269	8% = 0	30% = 0	56% = 0
				92% ≥ 1	70% ≥ 1	44% ≥ 1
				--	49% ≥ 2	22% ≥ 2
				44% ≥ 5	16% ≥ 5	12% ≥ 3
Iverson et al. (2008)	WAIS-III/ WMS-III	20	1,250	22% = 0	57% = 0	73% = 0
				78% ≥ 1	43% ≥ 1	27% ≥ 1
				51% ≥ 3	28% ≥ 2	14% ≥ 2
				34% ≥ 5	18% ≥ 3	6% ≥ 3
Unpublished	PSU-CX Program	17	495	21% = 0	50% = 0	71% = 0
				79% ≥ 1	50% ≥ 1	29% ≥ 1
				55% ≥ 2	27% ≥ 2	13% ≥ 2
				38% ≥ 3	15% ≥ 3	7% ≥ 3
				27% ≥ 4	10% ≥ 4	4% ≥ 4
				17% ≥ 5	6% ≥ 5	2% ≥ 5

Note: E-HRNB, Expanded Halstead-Reitan Neuropsychological Battery; ABC, Aging, Brain Imaging, and Cognition; NAB, Neuropsychological Assessment Battery; WAIS-III, Wechsler Adult Intelligence Scale, 3rd Edition.

Source: Reproduced from Binder et al., 2009,Table 6

Table 20

Number of Tests at 1, 1.5, and 2 SD Below the Mean on Tests Requiring a Computer, Oral, or Written Response for Male and Female Athletes on a Concussion Assessment Battery

	Total # of Tests Below Mean		
	1 SD	1.5 SD	2 SD
Male (n=371)	M (SD)	M (SD)	M (SD)
Computer Response	.76 (1.0)	.37 (0.8)	.18 (0.5)
Oral Response	.79 (1.0)	.38 (0.7)	.19 (0.5)
Written Response	.81 (1.2)	.42 (0.8)	.21 (0.6)
Female (n=124)			
Computer Response	.66 (1.0)	.27 (0.7)	.14 (0.5)
Oral Response	.61 (1.0)	.19 (0.5)	.05 (0.3)
Written Response	.47 (0.8)	.24 (0.6	.10 (0.4)

Table 21

Base Rates of Low Test Scores on a 17-Test Concussion Assessment Battery in Collegiate Athletes

Number Low Test Scores	Total Athlete Sample	Male Athletes	Female Athletes
≤ 1.0 SD/16th Percentile			
8 or more	5.9	6.5	4.0
7 or more	***8.1***	***9.2***	4.8
6 or more	12.1	14.0	***6.5***
5 or more	17.4	19.1	12.1
4 or more	26.5	28.6	20.2
3 or more	37.8	40.4	29.8
2 or more	54.9	58.5	44.4
1 or more	78.6	80.9	71.8
None	21.4	19.1	28.2
≤ 1.5 SD/7th Percentile			
5 or more	6.3	***7.3***	3.2
4 or more	***9.7***	11.1	5.6
3 or more	15.4	17.8	***8.1***
2 or more	27.1	29.6	19.4
1 or more	50.5	53.1	42.7
None	49.5	46.9	57.3
≤ 2.0 SD/2nd Percentile			
4 or more	3.8	4.6	1.6
3 or more	***6.9***	***8.6***	1.6
2 or more	13.1	14.8	***8.1***
1 or more	29.3	31.3	23.4
None	70.7	68.7	76.6

APPENDIX G

Figures

Figure 1

Percent of Athlete Sample with Test Scores 1, 1.5, or 2 SD Below the Mean on a Concussion Assessment Battery

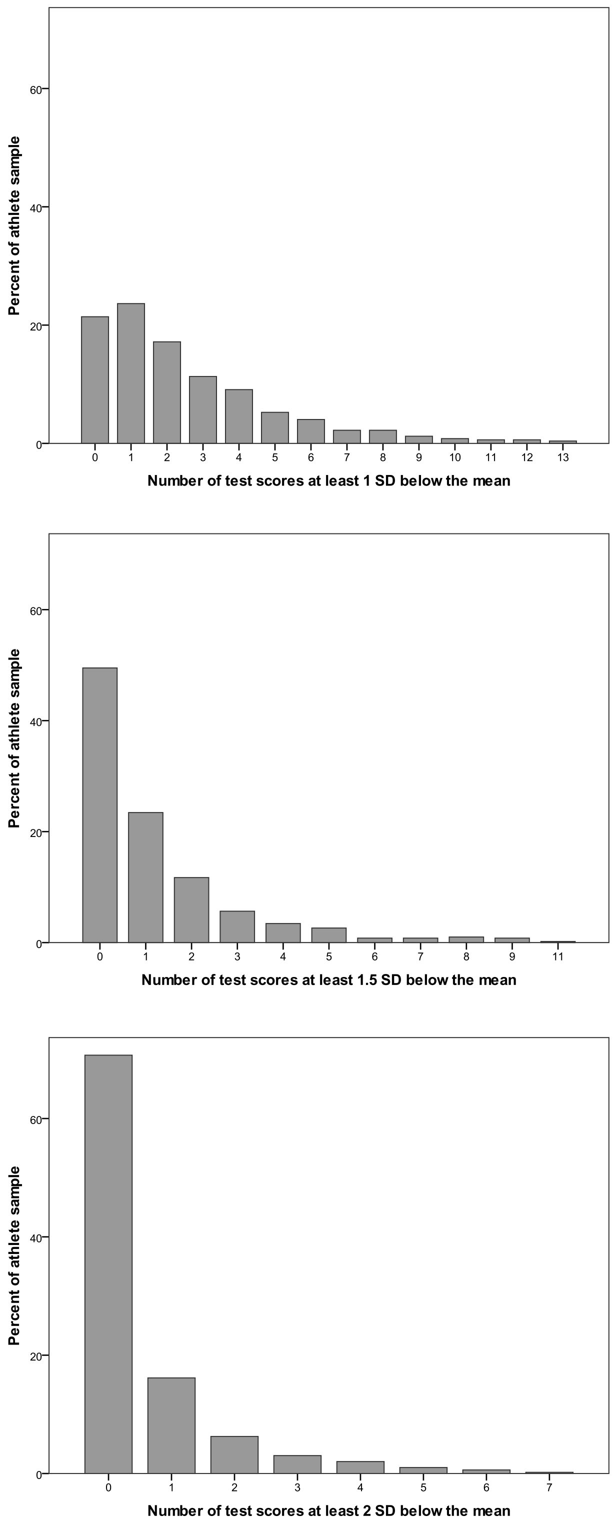


Figure 2

Cumulative Percent of Athletes with Test Scores 1, 1.5, or 2 SD Below the Mean on a Concussion Assessment Battery

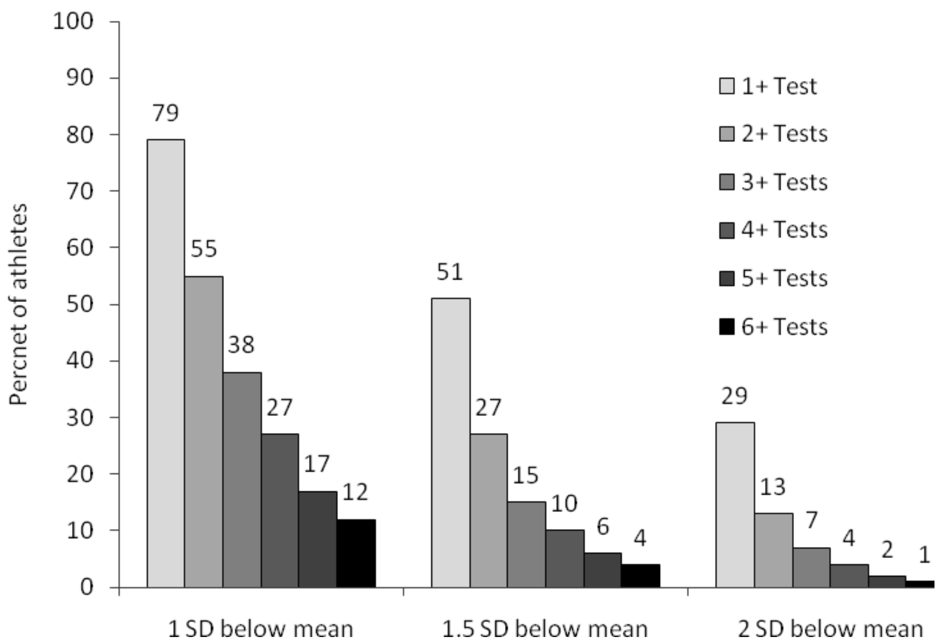


Figure 3

Number of Test Scores 1.5 and 2 SD below the Mean at Beginning, Middle, and End of a Concussion Assessment Battery

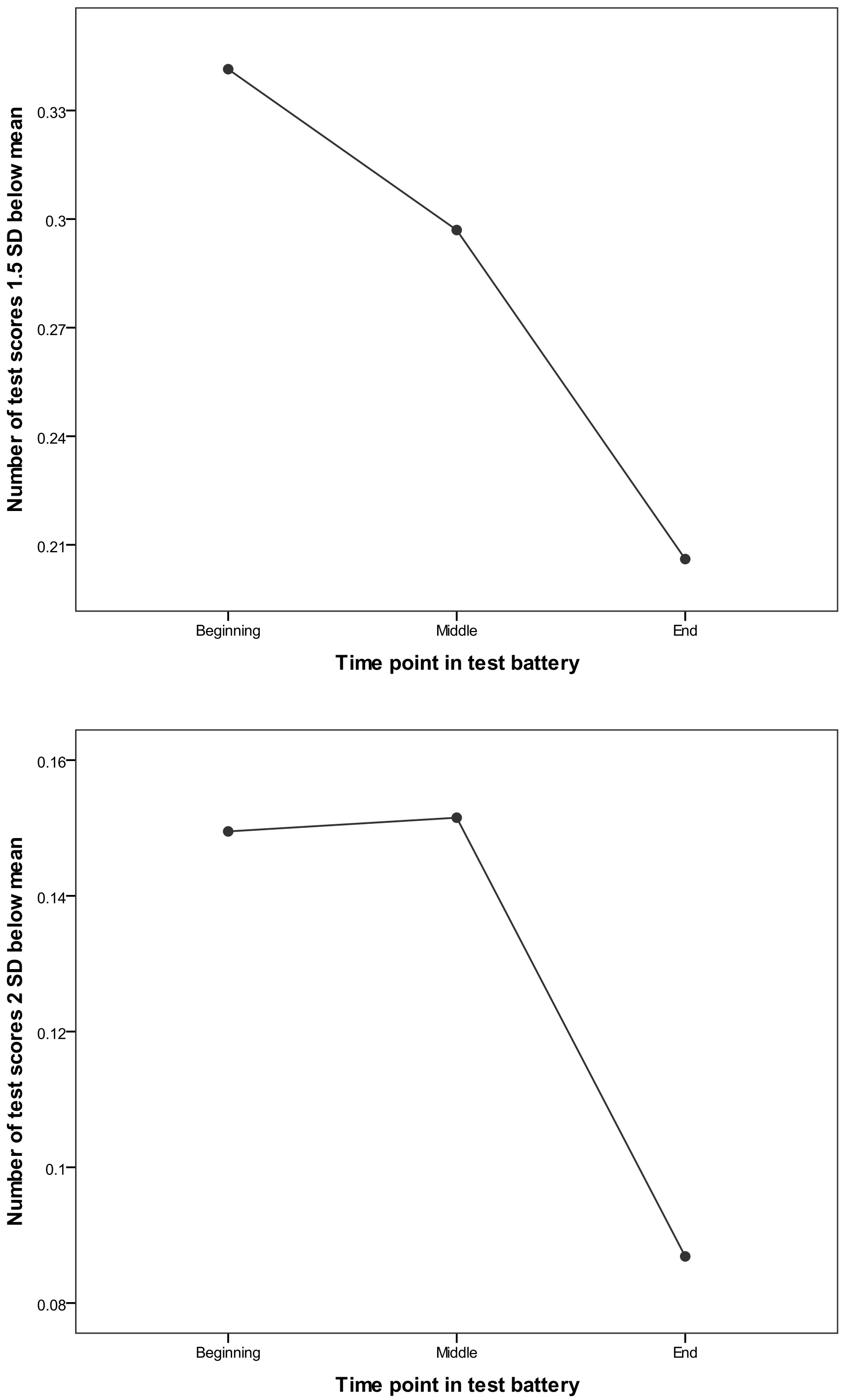


Figure 4

Binomial Probability Distribution Predictions: Cumulative Percent of Athletes Obtaining Test Scores at Least 1, 1.5, or 2 SD Below the Mean on a Concussion Assessment Battery

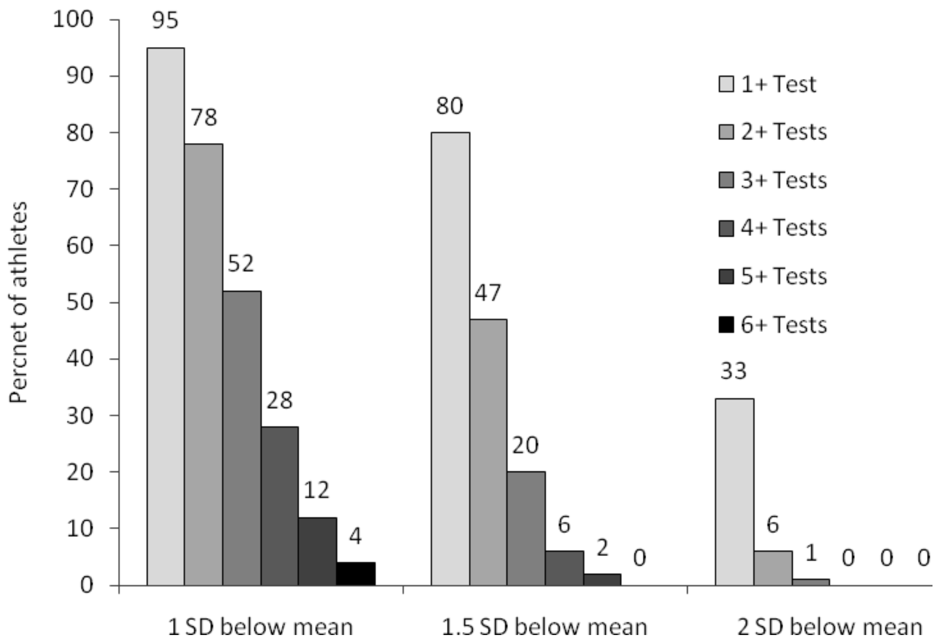


Figure 5

Monte Carlo Simulation Probability Distribution Predictions: Cumulative Percent of Athletes Obtaining Test Scores at Least 1, 1.5, or 2 SD Below the Mean on a Concussion Assessment Battery

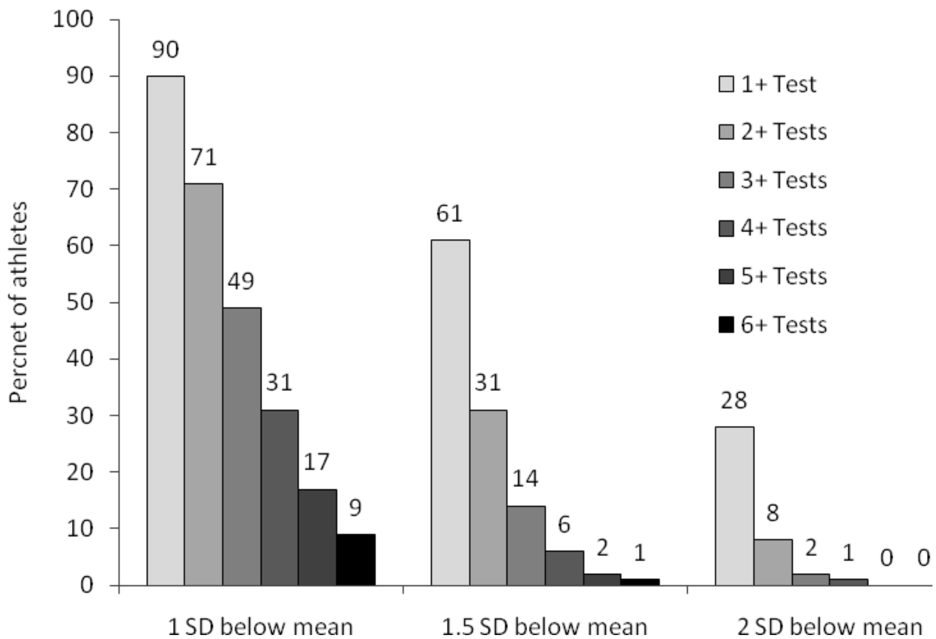


Figure 6

Percent of Male and Female Athletes with Test Scores 1, 1.5, or 2 SD Below the Mean on Concussion Battery

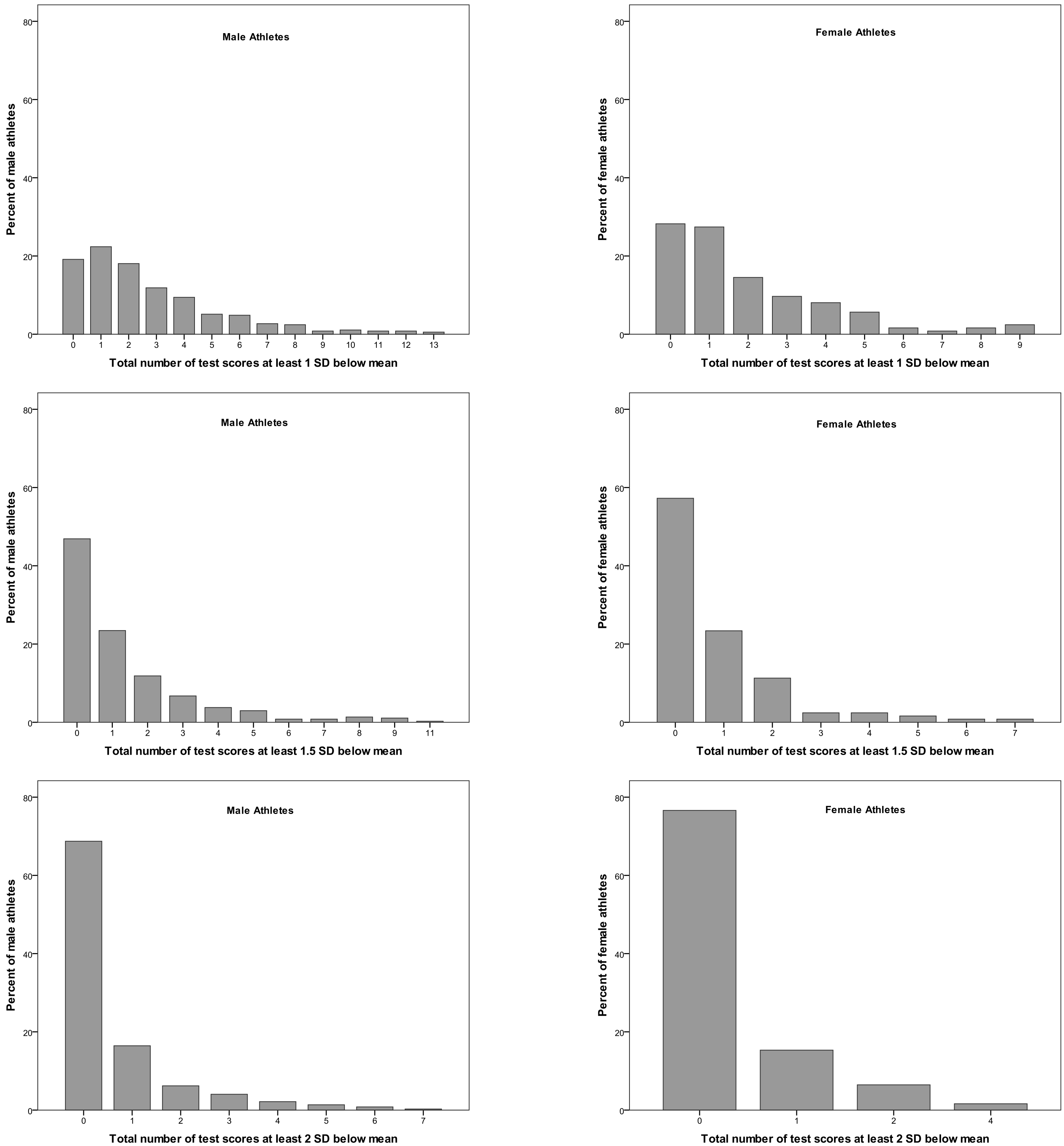


Figure 7

Cumulative Percent of Male and Female Athletes with Test Scores 1, 1.5, or 2 SD Below the Mean on a Concussion Assessment Battery

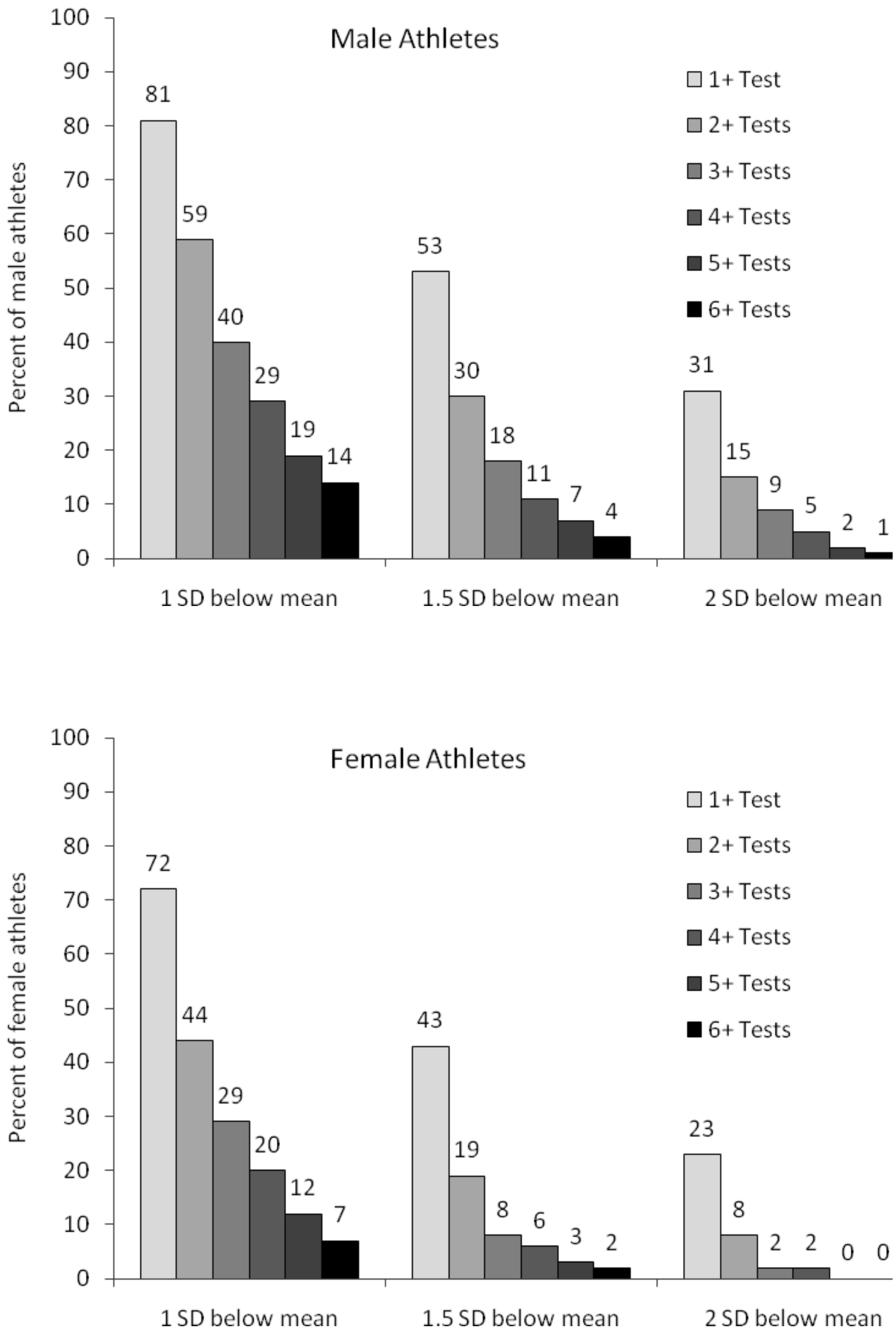


Figure 8

Number of Test Scores 1.5 and 2 SD below the Mean at Beginning, Middle, and End of a Concussion Assessment Battery for Male and Female Athletes

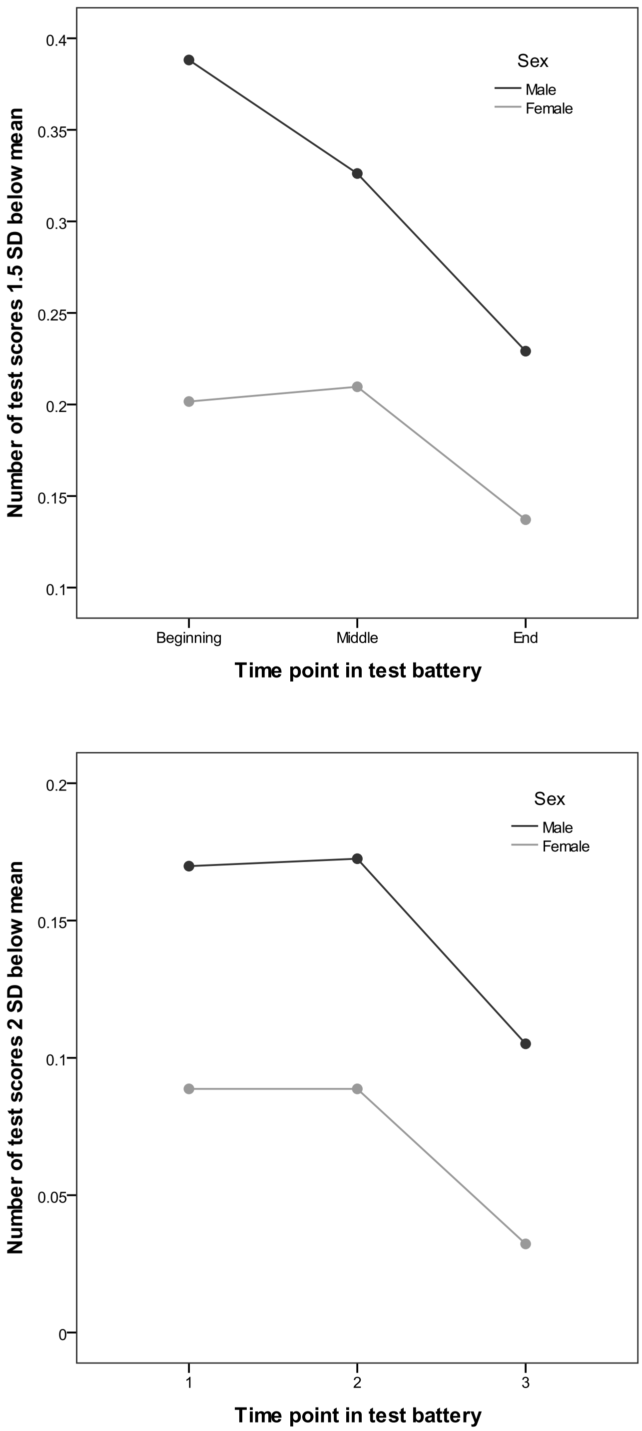
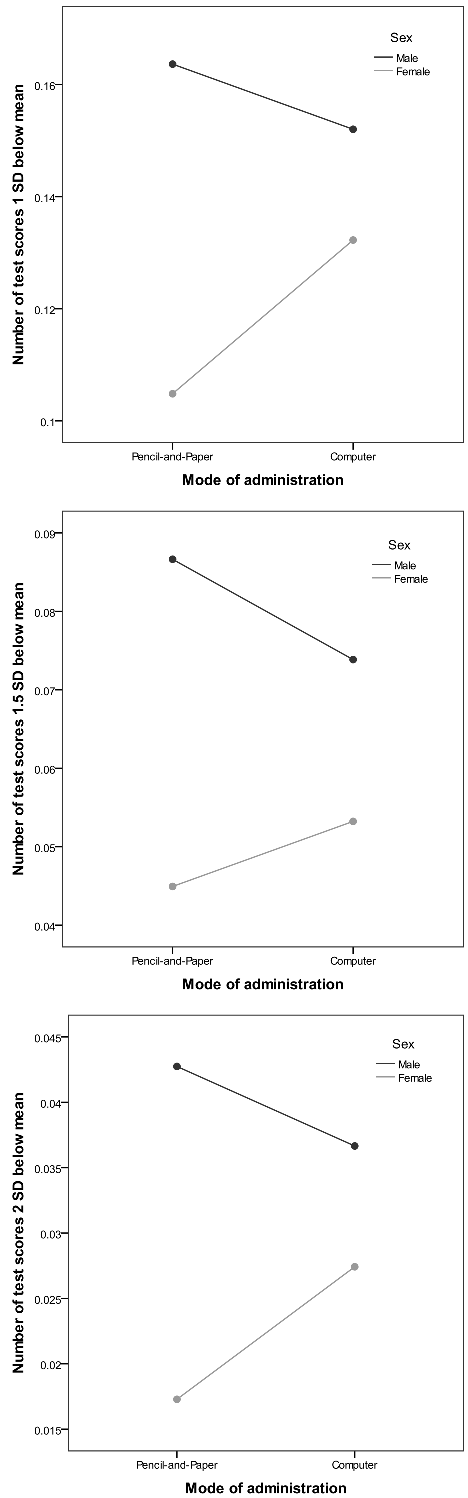


Figure 9

Number of Test Scores 1.5 and 2 SD below the Mean on Pencil-and-Paper Versus Computer-Administered Tests for Male and Female Athletes on a Concussion Assessment Battery



VITA

FIONA H. BARWICK

Education and Degrees:

2010 – 2011	Internship, Clinical Psychology (APA-accredited program) Veterans Affairs Palo Alto Health Care System Neuropsychology Track
2008 – 2011	Ph.D., Clinical Psychology (APA-accredited program) Pennsylvania State University, State College, PA Neuropsychology Track Ph.D. Dissertation: Defended July 2010
2004 – 2007	M.S., Clinical Psychology (APA-accredited program) Pennsylvania State University, State College, PA Neuropsychology Track
2000 – 2003	B.S., Psychology/Biology (CUNY BA) Hunter College, City of New York, NY