

The Pennsylvania State University

The Graduate School

College of the Liberal Arts

**PREMORBID MEASURES AND SPORTS-RELATED MILD TRAUMATIC**

**BRAIN INJURY: APPROPRIATENESS AND ACCURACY**

A Thesis in

Psychology

by

Christopher M. Bailey

© 2007 Christopher M. Bailey

Submitted in Partial Fulfillment  
of the Requirements  
for the Degree of

Doctor of Philosophy

August 2007

The thesis of Christopher M. Bailey has been reviewed and approved\* by the following:.

Peter A. Arnett  
Associate Professor of Psychology  
Thesis Adviser  
Chair of Committee

Frank G. Hillary  
Assistant Professor of Psychology

William J. Ray  
Professor of Psychology

Clarence B. Blair Jr.  
Associate Professor of Human Development and Family Studies

Melvin M. Mark  
Professor of Psychology  
Head of the Department of Psychology

\*Signatures are on file in the Graduate School.

## ABSTRACT

The current study examined the accuracy of estimating baseline neuropsychological performance using a variety of methods (Wechsler Test of Adult Reading, WTAR; demographically-based methods; and empirically constructed methods).

103 collegiate athletes were administered baseline mild traumatic brain injury (MTBI) neuropsychological batteries. 19 of these participants sustained concussions and were used in post-MTBI analyses. Initial baseline estimates were developed according to demographics [WTAR demographics-only and Barona et al. (1984) methods], WTAR reading performance (WTAR-P), and WTAR performance plus demographics (WTAR-PD). Correlations between each estimated baseline and observed baseline measure were conducted. Also, separate repeated measures ANOVAs were conducted comparing observed baseline performance and each estimation method with significant demographic groups entered as between-subject factors. Stepwise regressions predicting the observed baseline performance were also conducted by entering the above estimation methods and significant demographic variables (age, sex, race, previous concussion status, and sport). The proportions of accurately estimated baselines were also compared using Binomial Tests for each measure. Difference scores between observed and estimated baseline standard scores and post-injury standard scores were created and compared for the post-MTBI sample with repeated measure ANOVAs. The proportions of declined participants when utilizing observed or estimated baselines were compared using Binomial Tests for each measure.

Mild to moderate effects were observed on correlations and few mean differences were observed on the baseline repeated measure ANOVAs. Stepwise regressions showed a pattern of selecting demographic factors (especially race) and demonstrated generally moderate effects. Few accuracy differences were observed on the Binomial Tests, though some measures demonstrated differential accuracy for demographic-based or word-reading-based estimates. No decrement in sensitivity to the cognitive change associated with MTBI was observed in the post-MTBI analyses. Increased sensitivity over observed baseline performance was noted for some estimated baselines on some post-MTBI measures.

Baselines were estimated moderately well by the estimation methods. The stepwise regressions (which were largely based on race and other demographics) likely provided the best overall estimate for most measures with the WTAR-P providing the best baseline estimate for the DST and Stroop. Neuropsychologists should consider using multiple estimates for baseline performance in clinical situations.

## TABLE OF CONTENTS

List of Tables.....	v
List of Figures.....	vi
Introduction.....	1
Biological Basis of Mild Traumatic Brain Injury.....	4
Sports-Related Neuropsychology.....	8
Issues Associated with Baseline Testing.....	12
Premorbid Estimates.....	15
Questions to be Addressed.....	22
Methodology.....	24
Participants.....	24
Measures.....	25
Procedure.....	32
Analyses.....	33
Results.....	39
Discussion.....	52
Race, Sex, and Sport as Predictors of Baseline Performance.....	67
Estimating Digit Span and Stroop Baseline Performance.....	69
Limitations.....	71
Final Conclusions.....	72
References.....	73
Appendix A. Concussion Grading System Criteria .....	104
Appendix B. Barona Baseline Scatterplots.....	105
Appendix C. WTAR-Demographics Only Baseline Scatterplots.....	110
Appendix D. WTAR-Performance Only Baseline Scatterplots.....	115
Appendix E. WTAR-Performance & Demographics Baseline Scatterplots.....	120

## LIST OF TABLES

Table 1. Group Demographic Information.....	40
Table 2. Sport and Census Region Information.....	40
Table 3. Means and Standard Deviations of the Estimates and Measures.....	42
Table 4. Baseline Correlations.....	42
Table 5. Baseline Univariate Comparisons.....	44
Table 6. Repeated Measures ANOVA Estimated Marginal Means for Baseline Performance and Premorbid Estimates.....	45
Table 7. Stepwise Regressions Predicting Baseline Measures.....	47
Table 8. Stepwise Regression Equations for Baseline Measures.....	47
Table 9. Frequency of Participant Estimated Baseline Scores that Did Not Differ from the Observed Baseline by One Standard Deviation or More.....	49
Table 10. Repeated Measures ANOVA Estimated Marginal Means for Post-MTBI Difference Scores.....	50
Table 11. Frequency of Participants who were Identified as Declined Post-MTBI.....	51
Table 12. Best Baseline Predictors for Sport-Related MTBI.....	66

## LIST OF FIGURES

Figure 1. BVMT-R Baseline and Premorbid Estimates Estimated Marginal Means by Sex.....	46
Figure 2. SDMT Baseline and Premorbid Estimates Estimated Marginal Means by Sex.....	46
Figure 3. SDMT Baseline and Premorbid Estimates Estimated Marginal Means by Previous Concussion Status.....	46
Figure 4. Stroop-W Baseline and Premorbid Estimates Estimated Marginal Means by Sex.....	46

## **Introduction**

Historically, there has been much debate as to what specific features characterize cerebral concussion or, as is known within the field of neuropsychology, mild traumatic brain injury (MTBI). However, the definition that is generally accepted by the neuropsychological community reflects the variability of events and symptoms that can be associated with MTBI. It states that “a concussion is a trauma-induced alteration in mental status that may or may not involve loss of consciousness (Kelly and Rosenberg, 1997).” This “trauma-induced alteration in mental status” may or may not be accompanied by cognitive, physical, and emotional deficits as well. Recent meta-analytic studies have demonstrated that multiple domains of cognitive functioning, such as memory acquisition, delayed memory, processing speed, and fluency, are common deficits after MTBI in both non-athletic (Berlanger, Curtiss, Demery, et al., 2005) and athletic populations (Berlanger & Vanderploeg, 2005). Alves, Macchiocchi, & Barth (1993) provide evidence that headache and headache combined with dizziness are the most frequently documented physical symptoms; however, they suggest that other symptoms are also commonly seen including neurasthenia (fatigue), hyperesthesia (sensitivity to light and sound), and emotional lability. No long-term effects were thought to result from MTBI possibly because of the rapid recovery of symptoms associated with MTBI (Barth et al., 1989; Alves, Macchiocchi, & Barth, 1993; Vanderploeg, Curtiss, & Belanger, 2005) as well as the previous lack of empirically supported diagnostic instrumentation (Erlanger et al., 1999; Echemendia & Julian, 2001). However, some research has shown that MTBI can be associated with pervasive and persistent physical and cognitive deficits in a minority of cases (Stiller and Weinberger, 1985; Binder,

Rohling, & Larrabee, 1997) and it has also been recognized as the cause of death in some rare cases (Cantu & Voy, 1995).

These findings as well as the media's concentration on the injuries of professional athletes have led to a greater awareness of MTBI and its potential impact on the sports arena. Those athletes and coaches who might have been encouraged to ignore the effects of MTBI in the past are now aware of the possible long-term effects that such a course of action may produce. Echemendia and Julian (2001) stated:

Historically, sports-related MTBIs have been dismissed as “bell ringers” that are simply “part of the game” with no cause for concern. Recently, the highly publicized cases of professionals such as Steve Young, Troy Aikman, and Merrill Hoge in football and Pat LaFontaine, Eric Lindros, and Brett Lindros in ice hockey, as well as others, have helped underscore the potentially serious consequences of MTBI. (p.69)

This newly emerging interest related to the effects of MTBI on athletes produced several concerns that need to be addressed. Among them are the questions: what *exactly* is a MTBI?; how can the effects of MTBI be detected?; and when is an athlete safe to return to play after suffering from a MTBI? Clinical neuropsychology has been one area of science that has attempted to answer some of these questions.

“Clinical neuropsychology is an applied science concerned with the behavioral expression of brain dysfunction” (Lezak, Howieson, & Loring, 2005, p.3). Through the use of tests that concentrate on functions such as memory, attention, and speed of information processing, neuropsychologists are able to evaluate how specific organic deficiencies influence daily function. Neuropsychologists can also use tests to develop

profiles of functional deficits that are typical for a specific organic deficiency. Given the expectation of cognitive repercussions following a specific deficiency, neuropsychological testing may be used for the detection and diagnosis of organic dysfunction and some researchers have suggested it to be more sensitive to the effects of certain organic phenomena, such as MTBI, than diagnostic techniques such as computerized tomography (CT) and magnetic resonance imaging (MRI; Povlishock & Coburn, 1989; Hayes & Dixen, 1994). Because of their sensitivity to associated sequelae, neuropsychological instruments have been considered crucial in the diagnosis, detection, and monitoring of the cognitive repercussions of MTBI (Lezak, Howieson, & Loring, 2005). However, this sensitivity may be improved and, because of the inconsistent evidence regarding long-term outcome of those who suffer MTBI, there is an increased need to further enhance the sensitivity of the neuropsychological instruments to MTBI. One way to achieve this goal might be through the use of pre-injury estimations of cognitive function.

This paper will provide a review of the recent response to many of the concerns mentioned above as well as attempt to further clarify some of the present concerns regarding sports-related MTBI that have yet to be addressed. Specifically, the present paper will focus on the use of premorbid estimates of intelligence as a replacement for baseline testing. First, a review of the literature on the biological basis of MTBI that makes the application of neuropsychological instruments appropriate will be presented. The use of neuropsychological measures in making “return to play” decisions will then be discussed. Next, a discussion of premorbid estimates will be presented. A methodology for the evaluation of the appropriateness of premorbid estimates in the

diagnosis of MTBI will then follow. Finally, the results of the analyses and the implications of those results will be presented.

### *Biological Basis of Mild Traumatic Brain Injury*

Although originally attributed to a shock to the head resulting in loss of consciousness, MTBI is now thought to be associated with the acceleration or deceleration forces that act on the brain and cause stretching and the shearing of nerve cells that result in axonal damage (Povlishock & Coburn, 1989; Giza & Hovda, 2001). Though the stretching/shearing leaves the nerve cell alive, it also leaves the nerve cell temporarily dysfunctional. This temporary dysfunction has been posited by researchers to result in part from a metabolic cascade within the brain that begins during the first hour after impact (Hovda et al., 1999; Giza & Hovda, 2001). Based on animal research, Giza and Hovda (2001) suggest that this cascade begins with a massive increase in the extracellular concentration of potassium ions resulting from the indiscriminate firing of neurotransmitters (most likely glutamate). This indiscriminate firing is thought to occur as a result of the deformation of cells during the stretching/shearing from the concussive impact. The massive excitation not only renders the cell unable to produce action potentials (known as the spreading depression), but it also increases the cellular need for glucose to fuel the now over-activated sodium-potassium pump, which attempts to return sodium-potassium concentration to normal levels. Junger et al. (1997) indicate that these radical shifts in neuronal activity are also accompanied by dysfunction in the autoregulation of cerebral blood flow (the ability to maintain cerebral blood flow to the

brain to provide needed glucose among other nutrients). Therefore, the increased need for glucose to support cells' need to return ionic levels to normal is coupled with a decrease in cerebral blood flow which results in a potentially damaging energy crisis. Hovda et al. (1999) suggest that this metabolic cascade coupled with minimized cerebral blood flow may not only result in the symptoms associated with MTBI, but the associated vulnerable state of the brain may also explain why individuals who sustain one MTBI could be more likely to sustain another.

This model has been supported by recent electroencephalograph (EEG) research by Nuwer, Hovda, Schrader, & Vespa (2005). These authors describe animal research using EEG which demonstrates epileptiform activity occurring immediately following injury and is in turn followed by suppressed cortical activity for seconds to minutes. Nuwer et al. also suggest that in more severe cases, vascular injury results in ischemia and toxicity from exposure to blood and may result in significant cell damage. They suggest that such trauma may result in compromises in the integrity of the blood-brain barrier and lead to the formation of free radicals. Nuwer et al. suggest that these transient break-downs of the blood-brain barrier may be what results in the symptoms associated with what has been called Post-Concussive Syndrome.

To summarize, most of the pathophysiological evidence associated with MTBI suggests that there is typically only transient physiological damage. This conclusion is consistent with the meta-analytic reviews by Berlinger et al. (2005) and Berlinger & Vanderploeg (2005) which suggest that most individuals experience complete or near complete recovery in seven to ten days. However, a minority of cases may experience permanent dysfunction and sometimes fatal damage may occur as a result of MTBI.

What is described as Postconcussional Disorder (PCD) in the most recent edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR, 2000) is perhaps the most frequently occurring, long-lasting phenomenon resulting from MTBI. Though listed as a diagnosis that is in need of further study by the DSM-IV-TR, PCD is characterized by sustained difficulty in attention and memory. PCD may include any of the following symptoms which occur shortly after trauma and last at least three months: becoming easily fatigued, disordered sleep, headache, vertigo or dizziness, irritability or aggression with little or no provocation, affective lability, changes in personality, apathy, or lack of spontaneity. Though this disorder may be the most prevalent, the long-lasting effects of MTBI are not limited to PCD. Two other severe conditions may result from MTBI: Dementia Pugilistica and Second Impact Syndrome.

Dementia Pugilistica, also known as chronic traumatic encephalopathy or the “punch-drunk” syndrome, refers to the persistent, Parkinson’s-like, neurologic symptoms experienced by individuals who have sustained repetitive blows to the head. Jordan (1987) described Dementia Pugilistica in the context of boxing and suggested that individuals may suffer from symptoms such as headache, dizziness, imbalance, irritability, fatigue, poor memory, and dysarthria, as well as symptoms as serious as chronic amnesia, psychosis, and rage reactions. Autopsies of individuals having clinical histories compatible with traumatic encephalopathy were synthesized by Stiller and Weinberger (1985) and were described as having one of the following five brain irregularities: (1) abnormalities of the septum pellucidum; (2) cerebellar abnormalities; (3) cerebral scarring and atrophy; (4) degeneration of specific nuclear groups (substantia nigra, locus ceruleus, and nucleus basalis of Meynert); or (5) regional occurrence of

neurofibrillary tangles. These findings suggest that, though initial symptomatology may dissipate shortly after a MTBI, the accumulation of MTBIs over time may produce a condition where an individual experiences chronic post-concussion-like symptoms caused by resulting neurological abnormalities.

Second Impact Syndrome (SIS) is a condition where a neurological crisis following MTBI produces coma and respiratory failure and often leads to death. This disorder occurs when an individual sustains an original blow that produces typical postconcussional symptoms such as difficulty with thought or memory. The individual then receives a second blow before the symptoms of the original injury resolve. What follows in the next several minutes is described by Cantu and Voy (1995) as a disruption of the brain's blood autoregulatory system. "Vascular engorgement within the cranium markedly increases intracranial pressure, leading to herniation of the medial surface (uncus) of the temporal lobe or lobes below the tentorium or of the cerebellar tonsils through the foramen magnum (Cantu and Voy, 1995, p.31)."

SIS is a rare condition typically seen in young athletes from sports where successive blows to the head occur (such as football, hockey, and boxing). However, because of its speed of onset and its lethality, SIS is one of the driving forces (along with PCD and Dementia Pugilistica) behind the need to accurately diagnose MTBI and to devise a methodology for determining when athletes are safe to return to play after MTBI (a subject that will be touched on later). Neuropsychological assessment has stepped into the role of diagnosis of MTBI as well as the documentation of the recovery of functioning.

As discussed earlier, clinical neuropsychology measures behavioral deficits to study organic dysfunction. This process occurs through the use of psychometric tests that are able to detect functional deficits associated with neurological phenomena while other diagnostic techniques may not. Though Nuwer et al. (2005) suggest that EEG has been sensitive to immediate changes in animal models, no consistent evidence has been observed in the use of neurodiagnostic techniques in the identification of damage associated with MTBI in human patients. Also, given the ambiguous nature of the definition of MTBI, its diagnosis may be less important than characterizing the impact that the injury has had on the individual. Therefore, neuropsychology has emerged as the necessary procedure for evaluating the presence, magnitude, and symptom resolution associated with MTBI (Lezak, Howieson, & Loring, 2005).

### *Sports-Related Neuropsychology*

Much of the concern regarding MTBI has been focused on athletic populations, and with good reason. Sports-related head injuries represent approximately 20% of the 1.54 million head injuries estimated to occur annually in the United States (Erlanger et al., 1999). Gerberich and colleagues (1983) reported that 20% of all high school football players sustained cerebral concussions each year. However, this phenomenon is not limited to football alone nor is it limited to male dominated sports. Echemendia (1997), using a multisport college sample, found that 29.8% of all football players reported a history of at least one MTBI prior to participation in college sports. He also found that 55.8% of ice hockey players, 41.2% of male soccer players, 42.2% of female soccer players, 36.8% of male basketball players, and 31.3% of female basketball players

reported having a similar history of MTBI. These incidence rates contrast with the 2-10% incidence in matched samples of non-athletes according to Echemendia (1997). Because of the prevalence in athletic populations, there have been many attempts not only to identify the symptoms and consequences of MTBI within the athletes, but there also has been a heavy emphasis on determining at what point it would be safe to allow an athlete to return competitive play after having sustained such a head injury.

Athletes rigorously train for competitive play. An athlete's ability to perform publicly against opponents not only is the reward for such demanding training, but the ability (or inability) to play may determine whether or not he or she will advance to the next level of competition (an endeavor that may dictate how such individuals will maintain a livelihood). Not only is it unnecessary to hold athletes out of competitive play when they are in no serious danger, but it may also be quite costly to the athlete. However, as noted earlier, it may be dangerous and even life-threatening if an athlete is returned to play before a full recovery. Many methods for determining whether or not an athlete is ready for return to play after MTBI have been developed, but few have been empirically supported.

Traditionally, return to play was determined by the use of guidelines developed primarily by neurologists. Different standards for returning athletes to play were based on the grade (severity) of the concussion as well as the number of concussions that an athlete had sustained per season. An example of a widely used grading system is the one developed by Robert Cantu. In his system, Cantu (1998) divided what he described as cerebral concussion into three grades based on the presence and duration of certain symptomatology such as post-traumatic amnesia (PTA) as well as the presence and

duration of loss of consciousness (LOC). In Cantu's system, an individual with a grade I cerebral concussion (the least severe injury defined by no LOC and PTA lasting less than 30 minutes) should be allowed to return to play if asymptomatic for one week. If an athlete sustains three grade I concussions during the same season, Cantu's system recommends that the player terminate his or her season and return to play the next year if asymptomatic. At the other end of the spectrum, the athlete who sustains a grade III concussion (the most severe injury defined by more than five minutes of LOC and greater than twenty-four hours of PTA) is recommended to return to play after a minimum of one month rest followed by a week when the athlete is asymptomatic. After a second grade III concussion, it is suggested that the athlete terminate the season and return the following year if asymptomatic.

Although grading systems such as Cantu's attempt to standardize definitions of the severity of MTBI as well as focus attention on athletes' safety when returning them to play, there are several problems with the clinical application of such systems. One major problem with grading systems is that there are many different systems other than the one described above (recent examples include: Colorado Medical Society, 1991 and American Academy of Neurology, 1997). Each of these systems has its own unique grades, which not only vary in number, but also vary in the symptomatology necessary to determine grade. The specific criteria for three grading systems mentioned above (Cantu's revised system, the Colorado Medical Society's system, and the American Academy of Neurology's system) are listed in Appendix A. Another major problem with the grading systems is that they are based on self-report of symptoms from athletes. Such reports may be influenced by several factors, one of which is the strong motivation that

many athletes feel to return to play. Finally, it has been noted by Echemendia and Julian (2001) that little or no empirical evidence exists to support the grading systems and their associated guidelines. With these problems in mind, alternative means are necessary for evaluating when athletes are ready to return to play.

In the context of college football, Barth et al. (1989) performed the pioneering work which first utilized neuropsychological testing with sports-related MTBI. Preseason baseline data were gathered from college football players on measures that tested attention, concentration, speed of information processing, and memory. Those players who received a MTBI were retested at twenty-four hours post-injury, five days post-injury, and ten days post-injury. To control for practice effects, Barth et al. also gathered data from athletes who had not experienced a MTBI. They found that the athletes who had a MTBI did not show the expected practice effect shown by the control group of athletes on several measures, and they interpreted this as evidence of cognitive impairment. Also, cognitive impairment diminished over the ten day period before leveling off. The researchers concluded that neuropsychological measures are sensitive to the effects of MTBI at twenty-four hours following injury and that recovery of functioning is generally seen within a five to ten day period.

This idea of establishing a baseline for each player as a point of reference for their results post-injury created by Barth et al. (1989) has been the foundation for using neuropsychological measures in returning athletes to play after MTBI. Since the Barth et al. study, the serial testing protocol has been expanded in interval of testing as well as the types of athletes being tested (Echemendia & Julian, 2001; Erlanger et al., 1999). Echemendia and his colleagues at The Pennsylvania State University used a serial testing

protocol that tests athletes at baseline, two hours after injury, twenty-four to forty-eight hours after injury, one week after injury, and one month after injury (Echemendia and Julian, 2001). The Penn State program has expanded to include many sports other than football and has been described as a model program in this regard (Barth et al., 1997).

Though Barth et al.'s original model has demonstrated sensitivity to the cognitive repercussions of MTBI, there have been notable problems with his methodology since its development. First and foremost has been the issue of identifying clinically meaningful change in cognitive functioning. Second, Bailey, Echemendia, & Arnett (2006) have found that some athletes approach the multiple testing evaluations with differential motivation, which increases the error of testing and possibly obscures the cognitive repercussions of MTBI. Finally, given the more recent trend of the multi-sport, multi-evaluation method (Barth et al., 1997), many pragmatic, clinical, and economic issues are associated with the need to perform baselines on a multitude of incoming athletes including the question of what to do clinically when an athlete sustains a concussion and has not received a baseline evaluation. Each of these issues will be discussed in turn.

### *Issues Associated with Baseline Testing*

The need to document changes in cognitive performance over time is not new to clinical neuropsychology. Lezak, Howieson, & Loring (2005) have identified that the most common methodology among clinical neuropsychologists in attempting to identify meaningful change from time 1 to time 2 is the use of the one to two standard deviation(s) criterion. This criterion assumes that any change with a magnitude of one

standard deviation or more would likely be of clinical significance. Therefore, when applied to the MTBI literature, one might assume that any decline post-injury of 1 standard deviation or more from baseline would be indicative of the impact of MTBI. However, many authors object to this methodology given the impact of multiple confounding factors. First, Jacobson and Truax (1991) identify the need to take into account the test-retest reliability of the neuropsychological measures. That is, to ensure that the differences that are observed are solely due to the individual's true performance and not error inherent in the measure, one has to rule out the variability that is associated with limits in test reliability. The reliable change index (RCI) has been proposed as a solution to this problem. RCI is based upon the test-retest reliability of the instruments used and provides the amount of change necessary to identify that the difference between time points is not the result of error (Jacobson & Truax, 1991). However, other confounding factors have been identified as well, resulting in further adjustments to the RCI formula. These include the need to account for practice effects (Chelune et al., 1993), given the use of the same tests from one evaluation to the next, as well as the need to account for regression to the mean when extreme scores occur at baseline (Speer, 1992; Speer & Greenbaum, 1995). Therefore, the result is a comparison of post-injury data to the estimate of an estimate with each step of estimation being associated with a certain level of error.

Another issue associated with the use of baseline testing in the identification of post-MTBI cognitive repercussions is differential motivation. Bailey, Echemendia, and Arnett (2006) have demonstrated the existence of significant *improvements* in performance post-injury from baseline in individuals with suspect motivation at baseline.

This was in comparison to decline or no change from baseline in individuals with high motivation at baseline. These examiners suggest that increased motivation post-injury makes the post-injury testing sessions likely to be a more accurate reflection of the athlete's true cognitive functioning. However, during the baseline testing, those motivating factors that are associated with the post-injury testing (awareness of the importance of testing in making an RTP decision, resistance to cognitive change as a result of injury, etc.) are not present. In fact, they suggest that there may be factors with regard to cognitive performance which may work against similar levels of motivation being present at the baseline testing compared with the post-injury testing. Some of these factors may include limited awareness of the importance of baseline testing, personality style, approach tendencies for all academic and cognitive tasks, and differing levels of self-efficacy.

Finally, there are significant pragmatic, economic, and clinical issues associated with the trend of the multi-sport, multi-testing methodology endorsed by Barth et al. (1997). Such testing formats require that each athlete for each team undergo an initial baseline from a clinical neuropsychologist or trained assistants. These evaluations are both financially expensive and time consuming for both the athletes and the neuropsychologists. These financial and pragmatic issues have been the driving force behind the development of computerized testing procedures such as the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT; Lovell et al., 2002). Such tests can be completed in less time and do not require administration by a neuropsychologist. However, these computerized programs have not been adequately compared to paper and pencil testing by a neuropsychologist with regard to sensitivity to the repercussions of

MTBI. Given the burden of testing a large number of athletes and the possible hesitancy of the athletes themselves to undergo testing, it is likely that at least some of the athletes who sustain a MTBI over the time of their sports participation will not have undergone baseline testing. This leads to a dilemma for the neuropsychologist who may now be required to estimate whether an athlete continues to suffer from cognitive dysfunction as a result of the sustained MTBI without having an adequate reference from which to make such a determination. One possible way to resolve this dilemma, and each of the problems associated with the measurement of change from baseline and differential motivation at baseline versus post-injury, would be to use premorbid estimates of cognitive functioning.

#### *Premorbid Estimates*

Because baseline pre-injury test scores are typically not available, it is common practice within neuropsychology for clinicians to need to determine whether an obtained test score differs from what would be expected had the individual not suffered from any brain injury or dysfunction. This necessary reference of the level of premorbid functioning allows neuropsychologists to answer many of the referral questions that were originally posed for the evaluation. These include the establishment that the loss of some skill has occurred in civil litigation, the tracking of cognitive decline in cases such as dementia, and the tracking of progress toward original levels of functioning in other cases of neurologic disorders (Reynolds, 1997). Given the importance of the identification of premorbid functioning, in the absence of having prior test data, neuropsychologists have

been required to rely on estimating the level of functioning from a variety of factors including clinical judgment, demographic data, current test performance, and testing of so-called “hold skills” (Wechsler, 1958). Testing of “hold skills” involves the evaluation of skills that are thought not to be greatly impacted by cognitive dysfunction. Lezak, Howieson, & Loring (2005) indicate that this method was based on the observation that many cognitively deteriorating patients retained old, well-established verbal skills long after other skills such as memory, reasoning, and arithmetic ability were severely compromised. The specifics of such testing will be discussed further below. However, given the inconsistent findings regarding the effectiveness of these methods (Kareken, 1997; Orme, Ree, & Rioux, 2001; Griffin, Mindt, Rankin, et al., 2002; Powell, Bosart, & Reynolds, 2003; Harnett, Godfrey, & Knight, 2004; Skeel, Sitzer, Fogal, et al., 2004), no “gold standard” has been identified. The strengths and weaknesses of these different estimation methods will also be briefly described.

Lezak, Howieson, & Loring (2005) described the use of indirect methods for estimating premorbid functioning through the use of a combination of clinical judgment, anecdotal information, and demographic characteristics. Kareken (1997) suggests that this style of premorbid estimation is subjective and inherently biased and empirical evidence suggests that actuarial data and statistical formulas often provide decisions of accuracy equal to or exceeding that of diagnosing clinicians (Dawes, Faust, & Meehl, 1989). It was with this in mind, that multiple regression formulas for the estimation of premorbid intelligence/functioning were established using actuarial and demographic information such as age, race, education, occupation, and region of the country. Probably the most widely used of these models was the one developed by Barona, Reynolds, and

Chastain (1984) which uses age, sex, race, education, occupation, and region to estimate premorbid Full Scale IQ. However, Orme et al. (2001) describe a consistent effect of regression-based formulae like the Barona et al. (1984) method where high IQ scores are underpredicted and low IQ scores are overpredicted. Along with the demographic equations, the use of an individual's highest achieved performance on a battery of tests has been used as a measure of premorbid functioning, assuming that the highest level is likely consistent with the individual's general functioning prior to suffering from a specific organic condition. However, research associated with this approach has consistently demonstrated an overestimation of actual IQ due to an inability to account for factors such as regression to the mean and individual areas of strength prior to suffering from a particular neurological insult (Krull, Sherer, & Adams, 1995). Finally, the use of "hold tests" that are based on over-learned skills, such as word knowledge and the memory of school-aculturated declarative knowledge, have been suggested as the basis of premorbid functioning as well (Wechsler, 1958). Since then, several tests have been designed to measure over-learned skills, such as reading aloud or the application of semantic knowledge, because these are skills thought to be relatively resistant to the effects of brain injury in the non-aphasic adult (Harnett et al., 2004). Examples of these tests are the National Adult Reading Test (NART; Nelson & Willison, 1991) and the Wechsler Test of Adult Reading (WTAR; The Psychological Corporation, 2001), both of which have received some empirical support (Crawford, Parker, & Besson, 1988; Moss & Dowd, 1991; The Psychological Corporation, 2001)

There have been several recent studies which compared these methodologies for estimating premorbid functioning. As noted above, Griffin et al. (2002) evaluated the

accuracy of the North American Adult Reading Test (NAART), Wide Range Achievement Test – 3 (WRAT-3), the Barona method, and the Oklahoma Premorbid Intelligence Estimate (OPIE) on a group of 64 chronic pain patients across intelligence ranges. They found that the OPIE method demonstrated the best overall accuracy, but that each of the methods had similar accuracies in classifying people in the average range of intelligence. The differences in accuracy resulted in classifying people in the above or below average premorbid IQ ranges. McCarthy, Sellers, Burns, et al. (2003) created a regression equation to predict premorbid intelligence from the Mayo Clinic’s normative sample of older adults. They then compared their regression equation to the Barona method and found that the Barona method was at least as accurate as their equation and, given the relative popularity of the Barona method, suggested its continued use.

Powell et al. (2003) also compared the premorbid estimates of both the Barona method and the OPIE method among samples of non-impaired individuals and a sample of clinically-impaired individuals. Powell et al. drew their data from the WAIS-R normative sample, providing a non-impaired group of over 1800 participants. The clinically-impaired group was composed of participants referred for neuropsychological assessment and had biomedical data (CT, MRI, EEG data) that clearly demonstrated brain injury with unilateral right hemisphere damage, those with unilateral left hemisphere damage, and those with multifocal, bilateral damage. Then, the OPIE and Barona methods were compared in both the non-impaired and impaired groups by establishing the premorbid estimates (based on the equations) and comparing the average absolute residual between the estimated Full Scale IQ and the observed WAIS-R Full Scale IQ scores. Powell et al. indicated that the most accurate method for the non-

impaired group was identified as the method with the smallest average absolute residual; however, in the clinically-impaired group, Powell et al. indicated that the method with the largest absolute residual would be considered the most accurate (given that the WAIS-R data were obtained post-injury). Though they found that the OPIE method (mean residual = 6.16, standard deviation = 4.84) demonstrated more accurate measurement in the non-impaired group than the Barona method (mean residual = 12.96, standard deviation = 9.85), they found that the Barona method was more accurate in the clinically impaired group (mean residual = 13.65, standard deviation = 10.46) than the OPIE method (mean residual = 8.88, standard deviation = 5.40). They suggested that the Barona method may be more sensitive to the clinical decline in the impaired group than the OPIE method and also suggested its continued use. However, it is important to note that some authors have objected to the methodology of Powell et al. stating that just because Barona demonstrated a greater difference between the estimated premorbid Full Scale IQ and current functioning does not necessarily mean that the method was more accurate (Schoenburg, Scott, Ruwe, et al., 2004).

Skeel et al. (2004) also compared several different methods of detecting cognitive decline from premorbid levels including the likelihood based on the statistical frequency within the clinical population, the Barona method, and premorbid estimation based on the WRAT-3 reading test. These authors found a high level of agreement between each of the methods and suggested that the use of multiple methods would be the most optimal means of establishing a premorbid level of cognitive functioning.

There has been a particular focus on premorbid intelligence estimation with schizophrenia due to the degenerative nature of the disorder (Allen, Kelley, Miyatake, et

al., 2001; Morrison, Sharkey, Allardyce, et al. 2000; Kremen, Seidman, Faraone, et al., 1996; Tracy, McGrory, Josiassen, & Monaco, 1996). However, research on the estimation of premorbid intelligence with schizophrenia has also not identified a true “gold standard.” Morrison et al. (2000) followed a sample of psychotic patients who had experienced only a single psychotic episode over a period of seven years. They found that, despite observed decline in other measures, the NART remained consistent with a mean difference score from the first administration to the administration at 7 years follow-up of 1.4 and a test-retest reliability of .91. Kremen et al. (1996) also found that reading measures were strong estimators of premorbid ability in a group of schizophrenic patients of varied disability, even above some demographic factors. Kremen et al. found that the Wide Range Achievement Test, 3<sup>rd</sup> Edition (WRAT-III) Reading test was a better estimator of WAIS-R full scale IQ than parental education (with WRAT-III effect sizes more than double those of parental education). However, some authors have found that demographic factors are better premorbid estimators in schizophrenic patients than reading performance. Tracy et al. (1996) found that the NART was significantly related to a measure of cognitive decline in a sample of chronic schizophrenics and patients with schizoaffective disorder. However, the Barona et al. (1984) demographically-based estimates were not related to cognitive decline. Tracy et al. concluded that the Barona et al. estimate may be a better premorbid estimate than the NART because it may be less impacted by the cognitive deterioration associated with the psychiatric illness. It should be noted that in each of the articles reviewed, no true measure of intelligence was obtained prior to the first psychotic episode and, therefore, no true comparisons between the premorbid estimates and premorbid intelligence could be obtained.

The Wechsler Test of Adult Reading (WTAR) has also received empirical support for estimating premorbid functioning. The measure, which is similar in form to word recognition tests such as the NART, was administered and normed along with the Wechsler Adult Intelligence Scale, 3<sup>rd</sup> Edition (WAIS-III) and demonstrated significant correlations when estimating WAIS-III Full Scale IQ's that ranged from .70 to .80 for the ages in the ranges of 18-74 (The Psychological Corporation, 2001). Also, the WTAR demonstrated significant correlations to the Wechsler Memory Scale, 3<sup>rd</sup> Edition (WMS-III) providing evidence that the premorbid estimate could be used in identifying premorbid levels of cognitive domains other than global functioning like the IQ (The Psychological Corporation, 2001). These correlations ranged from .40 to .71 in the age ranges of 18-74. Finally, the WTAR was shown to be insensitive to neurological impairment (traumatic brain injury, mild and moderate Alzheimer's disease, Parkinson's disease, Huntington's Chorea, and Korsakoff's syndrome) by comparing WTAR estimates in clinical samples with matched controls. .

In the case of the sports-related MTBI, the model developed by Barth et al. (1989) eliminates the need for premorbid estimates because baseline measures were collected prior to injury. However, as noted in the previous section above, there have been multiple issues associated with the collection and use of baseline data including the required use of multiple estimations for establishing reliable change, problems associated with differential motivation pre-injury versus post-injury, and multiple other practical, financial, and clinical issues. One possible method for resolution of these problems associated with using baseline measures would be the use of an accurate premorbid estimate that could be administered post-injury. This would eliminate the need for the

administration of a baseline evaluation, thereby reducing familiarity with tests and practice effects. Additionally, since they could be administered post-injury, it would not be necessary to contend with issues of motivation identified in some baseline participants by Bailey, Echemendia, and Arnett (2006).

The above considerations aside, it is important to note that the research presented above focuses on the premorbid estimate's ability to measure global cognitive functioning (with the exception of the WTAR which demonstrated some relationship to the memory measures of the WMS-III) while the battery of tests administered for MTBI typically only measure the specific cognitive domains of attention, memory, executive functioning, and speed of information processing (Echemendia & Julian, 2001).

Therefore, it is unclear whether premorbid estimates would be accurate in identifying the baseline level of these specific measures, something that is a necessary assumption for the identifying meaningful cognitive change following MTBI. However, given the currently used methodology in the MTBI literature where baseline measures are collected, MTBI evaluations provide an ideal means for evaluating the accuracy of premorbid measures. This concern, among others, was the focus of the current project.

### *Questions to be Addressed*

1. How well do premorbid measures estimate actual MTBI baseline performance on specific measures of cognitive functioning? Specifically, how related and accurate are the scores that premorbid estimates generate compared to those that are

- actually collected from baseline measures of attention, memory, and speeded information processing?
2. Is one method of premorbid estimation a more accurate measure of MTBI baseline cognitive functioning? Specifically, are the scores that premorbid estimates generate more related and accurate, compared to actual baseline performance, for one method of premorbid estimation over another? If so, is this true for a particular measure or cognitive domain (attention, memory, and speeded information processing)?
  3. Are there differences in the accuracy of the premorbid estimation methods based on the demographic factors? Specifically, does the age, race, sex, previous head injury status, or sport of an athlete impact the accuracy of the scores generated by the premorbid estimates compared to actual baseline measures of attention, memory, and speeded information processing?
  4. What is the most accurate and parsimonious method or combination of methods for predicting the baseline performance for a specific measure or cognitive domain?
  5. Are the methods of premorbid estimation as sensitive to the cognitive repercussions of MTBI as the obtained baseline performance? Specifically, if used as a reference in calculating the cognitive change post-MTBI, does the

amount of post-MTBI change from premorbid measures correspond to the amount of post-MTBI change observed from the baseline measures of attention, memory, and speeded information processing?

## **Methods**

### *Participants*

Data were drawn from the Penn State Concussion Program, a multi-sport program that assesses college athletes at risk for concussion prior to and following MTBI. Athletes underwent a baseline neuropsychological battery before or during their freshman year at the Pennsylvania State University (PSU). During the baseline evaluation, athletes completed a questionnaire detailing any previous head injury information as well as detailed background information including age, race, sex, home address, etc. All athletes who underwent baseline testing were included as participants for the current project. Then, a sub-division of the baseline group who had undergone baseline testing and suffered a MTBI during the course of their career at PSU were included in the post-MTBI analyses. Specifically, 106 total athletes were included as participants in the baseline sample and 19 of those 106 (18%) were used in post-MTBI sample. The participants in both the baseline sample and post-MTBI sub-sample were mostly male, Caucasian, freshman, athletes who were 19-years old or younger. These athletes participated in a variety of sports including: Football, Men's and Women's Soccer, Hockey, Men's and Women's Basketball, Gymnastics, Men's and Women's Lacrosse, and Wrestling.

However, only 5 sports had 10 or more participants in the baseline sample: Football, Women's Soccer, Hockey, Men's Lacrosse, and Women's Lacrosse.

### *Measures*

The Penn State Concussion test battery is an eclectic group of tests that focuses on a large variety of cognitive, physiological, and affective symptoms. Most of the measures are either oral or paper and pencil tests; however, one test (the VIGIL/W test) is computerized. The measures include: the Hopkins Verbal Learning Test-Revised (HVLTR; Benedict, Schretlen, et al., 1998), the Brief Visuospatial Memory Test-Revised (BVMT-R; Benedict, 1997), the Vigil/W test (Cegalis & Cegalis, 1994), the Symbol-Digit Modalities Test (SDMT; Smith, 1982), the Trail Making Test (TMT; Reynolds, 2002), the Digit Span Test (DST; Wechsler, 1997), the Stroop Color-Word Test (SCWT; Trener et al., 1989), and the Wechsler Test of Adult Reading (WTAR; The Psychological Corporation, 2001).

The Hopkins Verbal Learning Test – Revised (HVLTR; Benedict, Schretlen, et al., 1998), is a measure of verbal memory and learning in which 12 words are presented over three separate trials. The words are semantically related into three categories (such as sports, professions, birds, clothes, etc.). The participant is presented the words at a rate of one word per 1 ½ seconds. Then, the participant is asked to repeat the words from the list in any order. After three trials, a fourth, delayed-recall trial is administered after a 20-25 minute delay where a participant is asked to repeat the list of words without being provided with the original list of stimuli words. The primary indices from the HVLTR

are the Total Recall Learning (the number of words correctly remembered across the three learning trials) and the Delayed Recall (the number of words remembered after the delay). The HVLT-R has six alternative forms that were equivalent for recall across the trials (Benedict et al., 1998). The HVLT-R has been demonstrated to be sensitive to the effects of Alzheimer's disease and vascular dementia (Shapiro et al., 1999; Frank & Bryne, 2000; Hogervorst et al., 2002). It also has demonstrated significant sensitivity to the effect of MTBI as well (Bazarian et al. 1999; Echemendia et al., 2001; Bruce & Echemendia, 2003). Also, the test has been shown reliable by Benedict et al. (1998) with a test-retest correlation of .74.

The Brief Visuospatial Memory Test – Revised (BVMT-R; Benedict, 1997) is a test of visual memory that requires a subject to memorize both the form and location of geometric figures. The participant is presented with a sheet on which 6 geometric figures are printed in a 2 x 3 array. The figures are presented for ten seconds and then the participant is asked to draw the display from memory on a blank sheet. There are three learning trials which follow the same procedure and a 25-30 minute delay trial where the participant is not presented with the figures. The BVMT-R is a relatively brief instrument and has five alternate forms. The primary indices for the BVMT-R are the Recall Total (the number of figures correctly remembered across the three learning trials) and the Delayed Recall (the number of figures remembered after the delay). Benedict, Schretlen, et al., 1996) demonstrated that the BVMT-R was highly correlated with previously validate instruments (Rey Complex Figure Test, Hopkins Verbal Learning Test, and Visual Reproduction from the WMS-III) when testing a mixed sample of individuals with a wide range of brain dysfunction (head trauma, psychiatric disorder, adult learning

disability, mental retardation, stroke, vascular dementia, progressive basal ganglia disease, and chronic alcoholism). Correlations ranged from .62-.78. Also, the BVMT-R has demonstrated significant incremental sensitivity to the effect of MTBI, even over measures sensitive to the verbal memory deficits associated with MTBI (Bailey, Echemendia, & Arnett, 2006). With regard to test-retest reliability, Benedict et al. (1996) demonstrated that, when combined across forms, the test-retest coefficient for the BVMT-R was .80 for a 56-day interval.

The Vigil/W is a computerized continuous performance test that examines attention/concentration (Cegalis & Cegalis, 1994). The K portion of this test, which was used by the Penn State Concussion Program, requires a participant to identify the target letter *K* from among other serially and randomly presented non-target letters by pressing the spacebar each time the target is presented. The test takes five minutes and important indices include the average delay (the average response speed to the targets), the number of omissions (the missed target letters), and the number of commissions (the non-target letters responded to). Cegalis and Cegalis (1994) demonstrated that the Vigil/W has good convergent validity with other tests of attention and adequate discriminant validity by being able to distinguish schizophrenics from controls, schizophrenics from depressives, bipolars from schizophrenics, and brain damaged patients from controls. Cegalis and Cegalis also provide a three-month, test-retest reliability coefficient of .74.

The Symbol Digit Modalities Test (SDMT; Smith, 1982) has been described as a test of the speed of information processing as well as complex scanning and visual tracking, which can be given in either a written or oral format (Lezak, Howieson, & Loring, 2005). Because of the time constraints, the only the written form of the SDMT

was collected by the Penn State Concussion Program. The specific design of the test is similar to the Digit Symbol subtest of the Wechsler Adult Intelligence Scale in that it requires a person to pair a familiar stimulus (number) with an unfamiliar stimulus (symbol). In the SDMT, the participant is provided with several rows of symbols and asked to write under the symbols the more familiar numbers that are paired in the key at the top of the page, as opposed to providing the individual with a number and asking him/her to draw the symbol (as in the Digit Symbol subtest). There are 110 symbols in total and the subject is given 90 seconds to pair as many numbers with the given symbols as possible. The primary score for this test includes recording the total number of correct responses provided within the 90 second period. The SDMT has been shown to be sensitive to a variety of pathologies and organic deficiencies (Pfeffer et al., 1981; Starkstein et al., 1988). Ponsford and Kinsella (1992) showed that severe head trauma patients scored more than a standard deviation lower than controls on the written format thus demonstrating the sensitivity of the test to head trauma. Also, the retest reliability of the written format of the SDMT has been shown to be .80 at one month (Levene et al., 2004).

Participants were also administered forms of the Trail Making Test (TMT; 1944) or the recent extension of the TMT by Reynolds (2002): the Comprehensive Trail Making Test. The TMT tests are a set of measures visual search and sequencing tasks that are heavily influenced by attention, concentration, resistance to distraction, and cognitive flexibility. The basic task of the TMT is to connect a series of stimuli (numbers, expressed as numerals or in written form, and/or letters) in a specified order as rapidly as possible. The primary score derived from each of the TMT tasks is the number of seconds

required to complete each task. The participants in the current project were administered either Trail 2 and Trail 4 or Trail 3 and Trail 5 from the Comprehensive Trail Making Test or forms A and B from the original test. Trails 2, 3, and A are composed of plain circles with the numbers 1-25 in them with 19 distractor circles that are either empty or filled with irrelevant line drawings. In Trails 4, 5, and B, the participant is required to shift sets. In Trail 4, the participant draws a line to connect in order the numbers 1-20, where 11 of the numbers are presented as Arabic numerals and the remaining numbers are spelled out in English. In Trail 5 and B, the participant draws a line to connect, in alternating sequence, the numbers 1-13 and the letters A-L. This is done by first connecting a number and then a letter (1-A-2-B-3-C, etc.). Trail-making tasks in general are considered highly sensitive to the effects of a variety of brain injuries and organic problems and are among the most popular (in terms of frequency of use) of all neuropsychological tests (Mitrushina, Boone, & D'Elia, 1999; Lezak, Howieson, & Loring, 2005). These tests have also demonstrated significant sensitivity to MTBI as well (Leininger et al., 1990; Vanderploeg, Curtiss, & Belanger, 2005). The Comprehensive Trail Making Test itself has demonstrated sensitivity to the effects of a cerebrovascular accident and learning disability (Reynolds & Fletcher-Janzen, 2000). Also, the TMT tasks have been established as a reliable test by Reynolds (2001) with a composite test-retest coefficient of .84.

The Digit Span Test (DST; Wechsler, 1997) is a test from the Wechsler batteries (the WAIS-III and the WMS-III) which measure span of immediate verbal recall. It is comprised of two separate tests – Digit Span Forward (DSF) and Backward (DSB). Both tests consist of eight pairs of random number sequences (each successive pair consisting

of one more number; 2-9 numbers) that the test administrator reads aloud at the rate of one word per second. For the DSF component, the participant is asked to repeat the number sequence exactly as it was heard. For the purposes of the current project, the test administrator read the first of the pair of number sequences (though the participant was awarded credit for both sequences) unless the participant did not recall the first sequence correctly, at which point the administrator marked that sequence as incorrect and read the second sequence. When the participant incorrectly recalled two sequences of the same pair (or completed the list), the test administrator then administered the DSB component. The procedure for this task was exactly the same as the DSF task with the exception that the participant was required to repeat the sequence in reverse order from which the administrator read the numbers. The primary indices of the DST are the total number of sequences correctly recalled for each trial (DSF, DSB, and Total). Lezak, Howieson, & Loring (2005) described the DST as one of the most frequently used tests in neuropsychology and notes its sensitivity to multiple neurological conditions. DST has also been identified as a memory measure that is sensitive to MTBI by Vanderploeg, Curtiss, and Belanger (2005). Levene et al. (2004) demonstrated a test-retest reliability coefficient for the DST using the same procedure as .81.

The Stroop Color-Word Test (SCWT; Trenergy et al., 1989) was developed as a test of attention, concentration, information processing, and response inhibition (Lezak, Howieson, & Loring, 2005). There are several forms of this test, but the one used in this study consisted of 112 words arranged in four columns. For the word only trial (Stroop-W), participants were asked to read aloud, down the columns of words (ignoring the ink color) as fast as they could without making a mistake. Participants were told that if they

made a mistake, to correct themselves and continue until all four columns had been read. Then, for the color-word trial (Stroop-CW), participants were asked to look down the list and say aloud the color of ink that each word was printed in as fast as they could without making a mistake. Again, participants were told that if they made a mistake, to correct the mistake and continue on until all four columns had been read. The assessor scored this test by recording the amount of time that the individual took to complete the list as well as by recording the number and type of errors that the individual made while reading (self-corrected and uncorrected). The SCWT has also been shown to be sensitive to the effects of mild head injury (Bohnen et al., 1992) and closed head injury (Stuss et al., 1985). Dikmen et al. (1999) also indicate that both trials of the SCWT task have a test-retest coefficient of .84.

As noted above, the Wechsler Test of Adult Reading (WTAR; The Psychological Corporation, 2001) is a test of reading recognition that was designed for premorbid estimation. The WTAR was developed and co-normed with the WAIS-III and WMS-III in both the US and the United Kingdom using the same large, representative sample of normally functioning adults. The WTAR consists of asking a participant to pronounce 50 words of increasing difficulty. The words are arranged in two columns of 25 words on a single sheet of paper. The paper is placed in front of participants and they are asked to read the words aloud, even if they are unsure or are not familiar with the word. Words that are correctly pronounced, including correct emphasis on word syllables, are awarded one point and the test is discontinued after twelve incorrect pronunciations in a row. As noted above, the WTAR has demonstrated strong correlations (.70-.80) to the WAIS-III Full Scale IQ for a wide age range (18-74) and also demonstrated correlations to the

memory tests of the WMS-III with correlations to the General Memory index of .40-.71 for the same age range (The Psychological Corporation, 2001). Also, as is consistent with the “hold test” literature, the WTAR was observed to be relatively resistant to the effects of traumatic brain injury, mild and moderate Alzheimer’s disease, Parkinson’s disease, Huntington’s Chorea, and Korsakoff’s syndrome (The Psychological Corporation, 2001). Also, the Psychological Corporation (2001) demonstrated strong correlations between the WTAR and other accepted premorbid measures including the American National Adult Reading Test (.90), National Adult Reading Test (.78), and the Wide Range Achievement Test – Revised Reading Test (.73). Also, the test-retest reliability was reported as being .92 for the 16-29 age group, .93 for the 30-54 age group, and .94 for the 55-74 age group (The Psychological Corporation, 2001).

### *Procedure*

As described above, all participants underwent baseline testing. The post-MTBI group was tested up to 4 times post-injury, with the first post-injury testing being utilized for the current analyses. The vast majority of the post-injury testing occurred within 1 week after having sustained an MTBI. The original baseline testing as well as all post-injury evaluations was administered by a Ph.D.-level clinical neuropsychologist, or graduate and undergraduate assistants who were trained by the Ph.D.-level clinical neuropsychologist. The order of tests administered were the same for both groups: HVLRT-R, BVMT-R, Vigil/W, SDMT, TMT, DST, Stroop, and WTAR. The testing took approximately 1 hour per evaluation, but also required approximately ½ hour for

paperwork, debriefing, and administration of other instruments not included in the current project.

### *Analyses*

Four separate methods (two demographic-based, one reading-based, and one demographic/reading-based method) for estimating premorbid functioning were utilized for the current project. First, the regression formula identified from Barona et al. (1984) was calculated for each participant. The formula is as follows:

$$\text{Estimated Full Scale IQ} = 54.96 + 0.47(\text{Age}) + 1.76(\text{Sex}) + 4.71(\text{Race}) + 5.02(\text{Education}) + 1.89(\text{Occupation}) + 0.59(\text{Region}).$$

For each of the nominal variables within the Barona equation, numerical values were entered depending on the demographics of each of the participants. Barona et al. set the following values for each variable which were used: sex – female = 1, male = 2; race – African American = 1, other ethnicity = 2, Caucasian = 3; region – Southern region = 1, North Central region = 2, Western region = 3, Northeast region = 4. Given that all participants were full-time college students, each participant's occupation score was entered as a 3 which was established by Barona et al. to fit those that were either students or not currently in the work force.

Next, three other Full Scale IQ estimates were computed using methods outlined in the WTAR manual. First, the use of WTAR demographic-only predicted estimates

(WTAR-D) were computed using the age, race (African American, Caucasian, Hispanic), and education of each participant. Next, the use of the WTAR performance alone (WTAR-P) was used to predict Full Scale IQ. Finally, a method that combined WTAR performance with demographic information was used to compute the WTAR performance/demographic predicted Full Scale IQ estimates (WTAR-PD).

Several analyses were executed to answer question 1 regarding how accurately the premorbid estimates measure true baseline performance. After the computation of the four separate estimates, zero-order correlations were computed between each of the estimates and the corresponding baseline measures. These zero-order correlations provide a measure of relationship between the premorbid estimates and the actual baseline data and were used to establish an effect size ( $r^2$  = the amount of variance accounted for) as a method for comparison of the accuracy of the premorbid estimates.

The initial correlations were followed by univariate analyses to identify whether baseline participants differed significantly in the cognitive performance based on categorical demographic variables that might modify the accuracy of the premorbid estimates. Variables analyzed included: age (19 years-old or younger versus 20 years-old and older), race (African American versus Caucasian), sex (female versus male), previous head injury status (no previous head injury versus having previous head injuries), and sport (football, women's soccer, men's lacrosse, women's lacrosse, and hockey). When significant differences were observed, these groups were used in the analyses below.

To further evaluate the accuracy of the premorbid estimates, each of the baseline measures were converted into standard scores (using the means and standard deviations for each measure from the entire sample of baseline data). A single factor, repeated

measures ANOVA was then conducted for each baseline instrument comparing the methods of premorbid estimation to the actual baseline level. Therefore, the repeated measures factor of premorbid/baseline performance consisted of 5 levels: the observed baseline performance for the given measure, the Barona estimate, the WTAR-D estimate, the WTAR-P estimate, and the WTAR-PD estimate. A significant main effect for premorbid/baseline performance would suggest that at least one of the methods significantly differed from the other (either a premorbid measure differed from the actual baseline measure and/or the premorbid measures differed from each other) and simple contrasts were used to compare each estimate to the baseline specifically. If the above univariate tests identified significant differences in baseline demographic factors (age, race, sex, previous head injury status, or sport) on any of the baseline measures, each significant factor would be entered into the ANOVA as a between-subjects factor. Then, to determine which levels differed within all factors and interactions, a Tukey's Honestly Significant Difference test was performed post-hoc. These analyses evaluated the premorbid measures' ability to accurately estimate the observed baseline performance as well as evaluated if the estimated premorbid levels differed for the levels (or combinations of the levels) of demographic variables.

For example, the repeated measures ANOVA for a specific measure would compare the mean standard scores of the actual baseline testing, the Barona, the WTAR-D, the WTAR-P, and the WTAR-PD for that measure. However, if it was found that men had a significantly lower mean score than women on one of the baseline measures (e.g. the HVLTR), then sex was entered as a between-subjects factor and the repeated measures ANOVA became a 2 (sex) x 5 (premorbid/baseline performance) mixed factor

repeated measures ANOVA for that measure. If other demographics factors were found to be significantly different in the univariate analyses, then they too were entered into the single ANOVA for that measure as between-subjects factors. This method addressed each of questions 1 (How well do premorbid measures estimate baseline performance?), 2 (Is one premorbid estimate more accurate than another?), and 3 (Do certain demographic factors moderate accuracy?).

The premorbid estimates and the demographic factors were then entered into stepwise univariate regressions predicting each of the specific baseline measures. The standard scores of each baseline measure were entered as the dependent variable in the univariate stepwise regressions using each of the four premorbid estimation methods and any demographic factors that were identified as significant predictors of baseline performance. The regression procedures then entered the variables that accounted for the most variance possible in baseline performance while excluding those variables that did not add any unique contribution to the model. The resulting predictors and coefficients addressed question 4 (What is the most accurate and parsimonious method for predicting baseline performance?).

To further address both questions 1 (How well do premorbid measures estimate baseline performance?) and 2 (Is one premorbid estimation method more accurate than another?), the accuracy of the premorbid estimates was also examined in a nonparametric manner. This was accomplished through the establishment of accuracy groups based on the difference of the premorbid estimates (including the predicted baseline performance according to the regression described above) from the observed baseline performance. Difference scores between each baseline measure and each premorbid estimate were

calculated and groups were created by dividing the sample into those whose premorbid estimate (Barona, WTAR-D, WTAR-P, WTAR-PD, regression predicted baseline) did not differ by one standard deviation (15 standard score points above or below) and those who did differ by one standard deviation (greater than or equal to 15 standard score points above or greater than or equal to 15 standard score points below). The proportion of those participants whose baseline performance was accurately estimated (i.e. premorbid estimate within one standard deviation of the observed baseline) for a particular premorbid estimate were then compared to the highest proportion of accurately estimated participants by a premorbid measure for the given measure using a binomial test. For instance, if the Barona estimate accurately estimated the baseline HVLT-R performance for 75% of the participants (and this was the highest proportion of the other premorbid methods), then each of the other premorbid methods' proportion of accurately estimated HVLT-R participants would be compared to .75 by the binomial test.

For those participants who sustained a MTBI, the post-injury test scores were converted to standard scores based on the means and standard deviations from the baseline sample. Using the calculated post-injury standard scores, difference scores were computed between the baseline testing and the post-injury testing as well as between each of the premorbid estimates (including the newly calculated regression estimates of baseline performance) and post-injury testing. These were used as measures of cognitive change post-MTBI. A single factor, repeated measures ANOVA similar to the one above was conducted for each of the post-MTBI measures as well. The single factor of post-MTBI cognitive change had six levels: change from actual baseline, change from Barona, change from WTAR-D, change from WTAR-P, change from WTAR-PD, and change

from the predicted baseline according to the obtained regression formula of a given measure. A significant main effect for post-MTBI cognitive change would suggest that at least one of the methods significantly differed from the other (either the amount of change from a premorbid measure differed from the change from the actual baseline measure and/or the change from the premorbid measures differed from each other). Again, simple contrasts were conducted that compared each change score to the change observed by the true baseline. Due to the limited size of the post-MTBI sample, no demographic factors were entered into the post-MTBI analyses. This method addressed question 5 (Are premorbid measures as sensitive to the cognitive repercussions of MTBI as observed baseline performance?).

To further address question 5 (Are premorbid measures as sensitive to the cognitive repercussions of MTBI as observed baseline performance?), the post-MTBI difference scores were also examined in a nonparametric manner. This was accomplished through the establishment of cognitive decline groups based on the difference scores used in the post-MTBI repeated measures ANOVAs (including the difference scores between the regression-predicted baseline performance). The decline groups were created by identifying those with observed cognitive decline post-MTBI (participants whose post-MTBI performance was 15 points or more below the observed, estimated, or predicted baseline) and those without observed cognitive decline (participants whose post-MTBI performance was less than 15 points below the observed, estimated, or predicted baseline). The proportion of those participants whose post-MTBI performance showed decline for a particular premorbid estimate were then compared to the baseline proportion of declined participants obtained by using the baseline performance. For instance, if the

observed baseline identified decline in the HVLTR performance for 50% of the post-MTBI participants, then each of the other premorbid methods' proportion of declined HVLTR participants would be compared to .50 by a binomial test.

## **Results**

Several steps were taken in an effort to reduce measurement and coding errors and to ensure the reliability of the data to be presented. First, all measures were scored and then rescored by two separate undergraduate research assistants prior to being entered in the database that was used to examine the data. Also, after the data were entered, all baseline measures underwent a reliability check where 10% of all the cases were randomly selected and all data were compared back to original measures. No errors were identified in the baseline reliability check. For the post-MTBI data, every case underwent a similar reliability check whereby all data entered into the database used for these analyses were compared back to the original measures. Also, any baseline data that were identified as outliers (3 standard deviations or greater from the total baseline mean) were also individually scrutinized to determine the accuracy of data. Two transpositions in data entry were identified when examining the outlier data (one set in the HVLTR data and one set in the BVMT-R data). The identified transposed data were corrected and all other outlier data (which were consistent with actual performance) were excluded when calculating the baseline means and standard deviations to be used in the standardization of the data; however, the outliers were included in the analyses to maintain consistency with the true nature of the observed data. No more than two

outlying cases were observed on any given measure and no outliers were identified on the Vigil, BVMT-R, HVLIT-R, SDMT, TMT 1 (Form A), TMT 2 (Form 4 and B), and Digit Span.

The first analyses were undertaken to identify demographic information for both the baseline and post-MTBI samples. Both samples were observed to have similar distributions of sex, race, and previous concussion history in that the vast majority of the sample was male, Caucasian (all participants identified as either African American or Caucasian), and had no previous concussions. Similar distributions of sport membership and region of the US in which they lived were also observed between both the samples. The vast majority of both groups (82% of the baseline sample and 90% of the MTBI sample) were comprised of athletes from the following sports: Football, Women's Soccer, Hockey, Men's Lacrosse, and Women's Lacrosse. Most of the two samples also reported home addresses in the Northeast US census region, followed by the South, and then the Midwest with no report of Western addresses. As one might expect given the timing of the baseline vs. post-MTBI testing, differences were observed in age and education level between the baseline and post-MTBI group with 73% of the baseline and only 32% of the post-MTBI group being 19 years-old or younger while 76% of the baseline and only 47% of the post-MTBI group were freshmen in college. All demographic information for the two samples is displayed in tables 1 and 2.

Insert Table 1.

Insert Table 2.

The observed baseline and post-MTBI data were then standardized using the means and standard deviations from the baseline measures. Again, all outlying data (data that were 3 standard deviations from the mean or greater) were removed from the means and standard deviations used to standardize the data. Kolmogorov-Smirnov (K-S) tests of normality were conducted on the baseline data as well. No significant K-S statistics were obtained on the following measures: Vigil (K-S = 0.07,  $p > .05$ ), SDMT (K-S = 0.09,  $p > .05$ ), TMT 1, Form 2 (K-S = 0.12,  $p > .05$ ), TMT 1, Form A (K-S = 0.15,  $p > .05$ ), TMT 2, Form B (K-S = 0.17,  $p > .05$ ), Stroop-W (K-S = 0.06,  $p > .05$ ), and Stroop-CW (K-S = 0.06,  $p > .05$ ). However, significant K-S statistics were observed on the BVMT-R (K-S = 0.10,  $p < .05$ ), HVLTR (K-S = 0.10,  $p < .05$ ), TMT 1, Form 3 (K-S = 0.17,  $p < .01$ ), TMT 2, Form 4 (K-S = 0.14,  $p < .05$ ), TMT 2, Form 5 (K-S = 0.14,  $p < .05$ ), and Digit Span (K-S = 0.10,  $p < .05$ ). BVMT-R, HVLTR, and Digit Span all showed negative skews in their distribution which is consistent with the ceiling effect observed in each when administered to non-injured participants. Each of the TMT 1 and TMT 2 distributions demonstrated a positive skew. The K-S tests were also conducted on each of the premorbid estimates. No significant K-S statistic was observed on the WTAR raw score (K-S = .10,  $p > .05$ ). However, significant K-S statistics were observed for all of the estimated Full Scale IQs provided by each measure: Barona (K-S = 0.16,  $p < .05$ ), WTAR-D (K-S = 0.26,  $p < .05$ ), WTAR-P (K-S = 0.09,  $p < .05$ ), and WTAR-PD (K-S = 0.09,  $p < .05$ ). In each case, the premorbid estimates demonstrated a negative skew despite a strong central tendency and mode. Though several of these measures were statistically non-normal which violates the assumption of many of the following analyses, the analyses were conducted due to the robustness of the procedures to violations of

normality (Achen, 1982; Girden, 1992). All descriptive information for both the baseline and post-MTBI samples is provided in Table 3.

Insert Table 3

Correlations between the premorbid estimates and each of the standardized baseline measures were then conducted. The significant correlations between the standardized baseline measures and the premorbid estimates generally showed a small to moderate effect. The Barona estimate was significantly correlated to the HVLT-R ( $r = .24, p < .05$ ), TMT 1 ( $r = .24, p < .05$ ), TMT 2 ( $r = .24, p < .05$ ), Digit Span ( $r = .20, p < .05$ ), and Stroop-W ( $r = .28, p < .01$ ). The WTAR-D estimate was significantly correlated to the HVLT-R ( $r = .21, p < .05$ ), TMT 1 ( $r = .29, p < .01$ ), TMT 2 ( $r = .26, p < .01$ ), and Stroop-W ( $r = .27, p < .01$ ). The WTAR-P estimate was significantly correlated to the BVMT-R ( $r = .20, p < .05$ ), HVLT-R ( $r = .21, p < .05$ ), SDMT ( $r = .19, p < .05$ ), TMT 2 ( $r = .20, p < .05$ ), Digit Span ( $r = .42, p < .001$ ), Stroop-W ( $r = .42, p < .001$ ), and Stroop-CW ( $r = .30, p < .001$ ). The WTAR-PD estimate was significantly correlated to the HVLT-R ( $r = .25, p < .05$ ), TMT 2 ( $r = .25, p < .05$ ), Digit Span ( $r = .40, p < .001$ ), Stroop-W ( $r = .41, p < .001$ ), and Stroop-CW ( $r = .24, p < .001$ ). No premorbid estimate was related to the Vigil. The correlations are presented in Table 4 and scatterplots for the correlations are presented in Appendices B, C, D, and E.

Insert Table 4.

Univariate analyses were conducted to identify baseline differences in demographics . Independent sample t-tests were employed to examine any possible differences in age, sex, race, and previous concussion status. The age groups (19 years and younger vs. 20 years and older) were not significantly different on any of the baseline measures. Sex differences were observed on the SDMT (  $t(104) = 2.46, p < .05$ ) and a trend was observed on both the Stroop-W (  $t(102) = 1.77, p < .10$ ) and Stroop-CW(  $t(102) = 1.90, p < .10$ ). In each case, female participants were observed with higher standard scores than male participants. With the exception of the Vigil, significant differences were observed between the African American and Caucasian participants on each of the baseline measures: BVMT-R (  $t(104) = -2.71, p < .01$ ), HVLT-R (  $t(104) = -3.49, p < .01$ ), SDMT (  $t(104) = -2.94, p < .01$ ), TMT 1 (  $t(104) = -2.47, p < .05$ ), TMT 2 (  $t(104) = -3.77, p < .001$ ), Digit Span (  $t(104) = -3.10, p < .01$ ), Stroop-W (  $t(104) = -4.09, p < .001$ ), and Stroop-CW (  $t(104) = -2.83, p < .01$ ). On each measure, the Caucasian participants were observed with higher standard scores. A significant difference was observed in previous concussion status (no concussions vs. one or more concussions) on the HVLT-R (  $t(104) = -2.13, p < .05$ ) and trends toward significance were observed on both the BVMT-R (  $t(104) = -1.70, p < .10$ ) and the SDMT(  $t(104) = -1.68, p < .10$ ). Surprisingly, the one or more concussion previous head injury group demonstrated greater baseline standard scores on these measures. A univariate ANOVA was employed to examine the effect of sport (football, women's soccer, hockey, men's lacrosse, and women's lacrosse) on each of the baseline measures. The only significant difference observed was noted on the HVLT-R (  $F(4, 87) = 4.12, p < .01$ ). A Tukey's Honestly Significant Difference post-hoc test identified that football performed significantly lower

on the HVLT-R than men's lacrosse and football trended toward being significantly lower than both women's soccer and women's lacrosse. All univariate means are presented in Table 5.

Insert Table 5.

One single factor, repeated measures ANOVAs and seven mixed factor, repeated measures ANOVA were conducted to compare the premorbid estimates to the observed baseline performance. Given that no differences were observed on the univariate tests for the Vigil baseline performance, a single factor, repeated measures ANOVA was conducted entering Vigil baseline performance and each of the four premorbid estimates as levels of the factor. For each of the other measures, significant differences were found on at least one of the univariate tests which were entered as between subject factors in the repeated measures ANOVA model resulting in a mixed factor design. If a univariate test was significant at the  $p < .10$  level, the demographic factor was entered into the ANOVA model as a between subject factor to identify any factor that had a meaningful influence on the accuracy of the premorbid estimates.

Significant within-subjects main effects for the repeated measures factor were observed on each of the following measures: Vigil ( $F(4, 416) = 3.25, p < .05, \text{partial eta squared} = .03$ ), SDMT ( $F(4, 392) = 2.83, p < .05, \text{partial eta squared} = .03$ ), TMT 1 ( $F(4, 416) = 3.43, p < .05, \text{partial eta squared} = .03$ ), TMT 2 ( $F(4, 416) = 6.15, p < .01, \text{partial eta squared} = .06$ ), and Digit Span ( $F(4, 408) = 4.14, p < .05, \text{partial eta squared} = .04$ ). Trends toward significance were observed on the Stroop-W ( $F(4, 400) = 2.87, p <$

.10, partial eta squared = .03) and Stroop-CW ( $F(4, 392) = 2.71, p < .10$ , partial eta squared = .03) as well. Simple contrasts comparing each premorbid measure to the observed baseline performance indicated that only the WTAR-D differed significantly from the observed baseline of the following measures: Vigil, TMT 1, TMT 2, Digit Span and Barona for the SDMT. The contrasts also identified the WTAR-D as trending toward being significantly different from the baseline of the Stroop-W and Stroop-CW. The estimated marginal means (predicted population means based on the sample means having controlled for error) are presented in Table 6.

Insert Table 6.

Several interactions were also observed on the mixed factor, repeated measure ANOVAs. A significant interaction between the repeated factor and sex ( $F(4, 392) = 2.67, p < .05$ , partial eta squared = .03) was observed for the BVMT-R. Significant interactions between the repeated factor and both sex ( $F(4, 392) = 5.11, p < .05$ , partial eta squared = .05) and previous head injury status ( $F(4, 392) = 2.66, p < .05$ , partial eta squared = .03) were observed on the SDMT. Also, a significant interaction between the repeated factor and sex ( $F(4, 408) = 4.14, p < .05$ , partial eta squared = .04) was observed for the Stroop-W ( $F(4, 400) = 4.90, p < .05$ , partial eta squared = .05). Simple contrasts were also employed to compare the premorbid estimates to the observed baseline performance for the significant interactions. The contrast for the BVMT-R repeated measures factor by sex interaction indicated that trends toward differences in estimating baseline BVMT-R performance by sex were observed in the Barona, WTAR-

D, and WTAR-PD. The contrast for SDMT repeated factor by sex interaction indicated that significant differences in estimating baseline SDMT performance were observed in each of the Barona, WTAR-D, and WTAR-PD with only a trend toward significance in the WTAR-P. The contrast for the SDMT repeated factor by previous concussion status interaction indicated that a significant difference in estimating baseline SDMT performance by previous concussion status was observed in the WTAR-D with a trend toward significance in the Barona. Finally, the contrast for the Stroop-W repeated factor by sex interaction indicated that a significant difference in estimating baseline Stroop-W performance by sex was observed in each of the Barona, WTAR-D, and WTAR-PD indices. The estimated marginal means for these interactions are displayed in Figures 1, 2, 3, and 4.

Insert Figure 1.

Insert Figure 2.

Insert Figure 3.

Insert Figure 4.

To further evaluate the relationship between the premorbid estimations and the observed baseline performance, several stepwise regression analyses were also conducted. Each stepwise regression predicted the observed baseline performance and entered or excluded each of the premorbid estimations (Barona, WTAR-D, WTAR-P, and WTAR-PD) and any significantly different demographic factors based on the univariate

analyses above. The entered and excluded variables as well as the associated effect sizes for the full model stepwise regressions are presented in table 7.

Insert Table 7.

For the Vigil, no correlations were observed between any of the premorbid estimators and no differences were observed on any of the demographic factors entered in the univariate analyses. Therefore, no variables were entered into the Vigil stepwise regression. Moderate effects were observed in the models that were created using the stepwise methodology for each of the other baseline measure regressions. For the BVMT-R stepwise regression, both race and the Barona estimate were included in the regression model (Full Model  $R^2 = .14$ ). However, it is important to note that a negative coefficient was found with the Barona given its semi-partial correlation with the BVMT-R standard score of  $-.27$ . For the HVLТ-R regression analyses, sport and race were included in the regression model (Full Model  $R^2 = .21$ ). Race and sex were included in the stepwise regression for the SDMT (Full Model  $R^2 = .12$ ) while the WTAR-D was the only predictor included by the TMT 1 regression (Full Model  $R^2 = .09$ ). The WTAR-P was the only predictor included in the Digit Span (Full Model  $R^2 = .18$ ) and Stroop-CW (Full Model  $R^2 = .09$ ) regressions. While the WTAR-P and race were the retained predictors for the Stroop-W (Full Model  $R^2 = .21$ ) regression. The unstandardized full model regression equations are presented in table 8.

Insert Table 8.

For the final analyses with the entire baseline sample, difference scores from each premorbid estimator (including the newly created regression-predicted baseline performance) and the observed baseline performance were calculated. Then the participants were divided into accuracy groups (those individuals who were estimated/predicted less than 15 points from the baseline standard score versus those that were greater than or equal to 15 points away from the baseline). Frequency descriptive information was then obtained and the estimation with the highest raw score proportion of participants with accurately estimated/predicted baseline performance for a given measure was identified. This highest proportion was then used as a reference point in a nonparametric, binomial test to identify if some of the premorbid estimators or regression-predicted baselines were significantly more accurate than others.

The highest proportions of accurately estimated/predicted baseline scores ranged from .65-.76 for the baseline measures. On the Vigil, the WTAR-P (.55,  $p < .05$ ) significantly differed from the highest proportion of accurately estimated baseline scores provided by the Barona (.65), while the WTAR-PD (.58,  $p < .10$ ) demonstrated a trend. The Barona (.59,  $p < .05$ ) was the only baseline estimator that significantly differed from the highest proportion of accurately estimated/predicted baseline BVMT-R performance provided by the WTAR-P (.69). On the HVLTR, the regression formula (.75) provided the most accurately categorized participants which was significantly higher than the WTAR-P (.64,  $p < .01$ ), WTAR-D (.67,  $p < .05$ ) with trends in difference observed by the Barona (.68,  $p < .10$ ) and the WTAR-PD (.68,  $p < .10$ ). On the SDMT, the regression formula (.74) provided the most accurately categorized participants which was

significantly higher than the WTAR-PD (.66,  $p < .05$ ) with a trend in difference observed by the WTAR-D (.67,  $p < .10$ ). No significant differences were observed on the TMT 1. On the TMT 2, the WTAR-D (.70) provided the most accurately estimated participants which was significantly higher than the WTAR-P (.59,  $p < .05$ ), with a trend in difference observed by the WTAR-PD (.63,  $p < .10$ ). Both the regression formula and WTAR-P demonstrated the highest number of accurately estimated/predicted baseline Digit Span performance (.71) which only showed a trend in being significantly higher than the WTAR-D (.64,  $p < .10$ ). No significant differences were observed on the Stroop-C. Only a trend toward being significantly different from the largest proportion of accurately estimated Stroop-CW baselines (WTAR-P; .76) was observed on the WTAR-D (.69,  $p < .10$ ). All percentages of accurately estimated/predicted baseline performance are presented in table 9.

Insert Table 9.

The post-MTBI data were also analyzed in the current project. As described above, difference scores were created by subtracting the baseline performance or premorbid estimation from the post-MTBI performance on each measure. This provided a measure of cognitive change (or estimated/predicted cognitive change) post-MTBI. Then, a single factor, repeated measures ANOVA was conducted for each measure and involved entering each of the post-MTBI difference scores (baseline, Barona, WTAR-D, WTAR-P, WTAR-PD, and regression difference scores) as the six levels of the repeated factor. No between subjects factors were entered due to the reduced sample size of the

post-MTBI sample (19 participants). Significant within subjects main effects were observed on the Vigil ( $F(4, 68) = 4.08, p < .05, \text{partial } \eta^2 = .19$ ) and Stroop-W ( $F(5, 85) = 3.56, p < .05, \text{partial } \eta^2 = .17$ ). Simple contrasts were then employed to compare each of the estimated/predicted cognitive change post-MTBI to the cognitive change post-MTBI identified by the observed baseline. On the Vigil, each of the premorbid estimators (no regression formula was created for the Vigil) resulted in significantly more decline post-injury than the observed baseline with the exception of the WTAR-PD, which only showed a trend toward a larger amount of decline. On the Stroop-W, the Barona and the regression formula showed significantly less decline than the observed baseline performance, while the WTAR-D demonstrated a trend toward showing less decline post-MTBI. The estimated marginal means for the post-MTBI difference scores are displayed in table 10.

Insert Table 10.

For the final analyses with the post-MTBI sub-sample, the difference scores utilized in the above repeated measure ANOVAs were used to divide the post-MTBI sample into cognitive decline groups for each of the observed, estimated, or predicted baseline performance. The decline groups were established by identifying those with observed cognitive decline post-MTBI (participants whose post-MTBI performance was 15 points or more below the observed, estimated, or predicted baseline) and those without observed cognitive decline (participants whose post-MTBI performance was less than 15 points below the observed, estimated, or predicted baseline). Then, frequency information

was obtained for the number of post-MTBI participants that were identified as declined by the observed baseline performance. The proportion of declined participants (as identified by the observed-baseline-post-MTBI difference score) was then used as the standard in binomial tests for each measure.

The proportions of declined participants as identified by the observed baseline difference ranged from .21-.42. No differences from the baseline proportion were observed by any of the premorbid estimators or the regression proportions on the BVMT-R, HVL-T-R, TMT 1, TMT 2, and Stroop-CW. In each case where significant differences were observed in the proportion of declined participants, the estimated/predicted baseline declined groups were observed with a significantly *larger* proportion of declined than the observed baseline. On the Vigil, both the WTAR-P (.42,  $p < .05$ ) and WTAR-PD (.42,  $p < .05$ ) significantly differed from the baseline proportion of declined post-MTBI participants (.22). The proportion of declined participants for the WTAR-D (.53,  $p < .05$ ) significantly differed from the baseline declined participants (.37) on the SDMT. On the Digit Span, each of the Barona (.58,  $p < .01$ ), WTAR-D (.53,  $p < .01$ ), WTAR-P (.47,  $p < .05$ ), WTAR-PD (.47,  $p < .05$ ), and the regression (.53,  $p < .01$ ) significantly differed from the baseline proportion of declined post-MTBI participants (.24). On the Stroop-W, again each of the Barona (.61,  $p < .001$ ), WTAR-D (.67,  $p < .001$ ), WTAR-P (.50,  $p < .01$ ), WTAR-PD (.55,  $p < .01$ ), and the regression (.44,  $p < .05$ ) significantly differed from the baseline proportion of declined post-MTBI participants (.21). All percentages of declined post-MBTI performance are presented in table 11.

Insert Table 11.

## **Discussion**

Premorbid estimation is considered an important and necessary function of the neuropsychologist for both the identification and tracking of the cognitive deficits associated with a variety of pathologies (Lezak, Howieson, & Loring, 2005). A variety of clinical methods have been used to estimate premorbid level of functioning, including those based on demographic factors (such as race, sex, education level, etc.), the use of “hold skills” (such as reading performance, which are less impacted by the presence of neurologic dysfunction), and the combination of the two (Lezak, Howieson, & Loring, 2005). However, the premorbid estimation literature has been inconsistent in the identification of a gold standard with a variety of disorders including chronic pain, dementia, traumatic brain injury, and schizophrenia (Allen, Kelley, Miyatake, et al., 2001; Griffin, Mindt, Rankin, et al., 2002; Harnett, Godfrey, & Knight, 2004; Kareken, 1997; Kremen, Seidman, Faraone, et al., 1996; Morrison, Sharkey, Allardyce, et al. 2000; Orme, Ree, & Rioux, 2001; Powell, Bosart, & Reynolds, 2003; Skeel, Sitzer, Fogal, et al., 2004; Tracy, McGrory, Josiassen, & Monaco, 1996). Neuropsychological assessment following sports-related MTBI could benefit from the accurate estimation of premorbid level of functioning for a variety of reasons. The present study was designed and conducted to evaluate the accuracy and appropriateness of commonly-used methods of premorbid estimation (demographic-based, reading-based, and a combination of demographic and reading-based premorbid estimation methods) with a sample of college athletes.

Five questions regarding the relationship among common premorbid estimation methods and the baseline performance of college athletes in MTBI testing were posed: 1.) How well do premorbid measures estimate baseline measures (i.e. what is the relationship between premorbid measures and observed baseline performance)?, 2.) Are there differences in the accuracy of some premorbid measures when compared to others? 3.) Do demographic factors influence the accuracy of premorbid measures? 4.) What are the most meaningful and parsimonious and methods for predicting baseline performance on a given measure? and 5.) Are premorbid measures sensitive to the cognitive change post-MTBI if used as an estimate of baseline performance? Four separate sets of analyses were employed to answer the first four questions using the baseline sample of 106 college athlete participants. These included zero-order correlations, repeated measures ANOVAs, stepwise regressions, and nonparametric binomial tests. Two separate sets of analyses were employed to answer the final question regarding sensitivity of premorbid estimators post-MTBI using the post-MTBI sub-sample of 19 college athlete participants who had sustained a MTBI. These included repeated measures ANOVAs and nonparametric binomial analyses. Each of these questions will be addressed individually based on the results of the above analyses. This will be followed by a discussion integrating the current findings with regard to the clinical question of: what measure/method should be employed when attempting to estimate baseline performance? Such broad conclusions will be made by evaluating the specific sets of analyses in relation to each other and attempting to find converging evidence for the use of a specific measure or method of premorbid estimation for a given baseline measure. Finally, a discussion of what factors likely led to the obtained results will conclude the current thesis.

The first question to be addressed was: *How well do the premorbid estimators measure observed baseline performance?* Based on the obtained results, the premorbid measures' estimation of observed baseline performance was generally moderate. This question was at least partially addressed by each of the following analyses: the zero-order correlations (which can be used to establish an effect size), the baseline repeated measures ANOVAs, and the baseline accuracy group binomial tests. When the significant zero-order correlations were examined, each of the premorbid measures was observed with small to moderate effect sizes ( $R^2$  ranging from .04 - .18) in estimating baseline performance. However, notable variation was observed by measure. For instance, no premorbid estimator was significantly related to Vigil performance and only the WTAR-P showed a marginal relationship with the BVMT-R and the SDMT; while each of the HVLTR, TMT 2, and Stroop-W tests demonstrated small to moderate relationships with all of the premorbid estimators. When the repeated measures ANOVA analyses were examined, few mean differences relative to observed baseline performance were observed using the Barona, WTAR-P, and WTAR-PD. This is consistent with the literature regarding premorbid estimators' ability to accurately estimate the average performance of a group (Griffen et al., 2002). Also, when the results for the binomial tests were examined, it was observed that the baseline accuracy categorization rates ranged from 55% - 76%. Again, these rates of categorization varied notably by measure. When these results are considered in combination, a pattern of accurate assessment of baseline performance was observed on the majority of the premorbid measures for the majority of the participants; however, a notable minority of participants was not accurately estimated according to analyses such as the accuracy group binomial tests.

This suggests that, although the measures were generally accurate, there is significant room for increased accuracy.

The second question to be addressed was: *Are there differences in the accuracy of some premorbid measures over others?* The answer to this question is *yes*; however, the differences observed varied by the baseline measure being estimated and were typically not associated with large effects. The analyses that were utilized in answering this question again included the zero-order correlations, the baseline repeated measures ANOVAs, and the accuracy group binomial tests.

Based on the zero-order correlations, the WTAR-P was generally more related to observed baseline performance relative to other premorbid estimators. The WTAR-P demonstrated both the most frequent significant relationships to the baseline measures (7 of the 9 measures were significantly related to the WTAR-P) and the largest effect sizes (both the Digit Span and Stroop-W were correlated at .42). Also, the correlational data suggested that the combination of demographic information and reading performance in the WTAR-PD estimate was less accurate than the estimates that either took into account only one of the components of demographics or reading performance. With the exception of the HVLIT-R, the WTAR-PD's correlation was not larger than that of the WTAR-D or WTAR-P. For most baseline measures (BVMT-R, SDMT, Digit Span, Stroop-W, and Stroop-CW), the correlation of the WTAR-P was reduced by combining WTAR performance and demographics (WTAR-PD). However, the WTAR-D correlation with both the TMT 1 and TMT 2 was reduced when WTAR performance and demographics were combined to provide a premorbid estimate (WTAR-PD).

The baseline repeated measures ANOVA analyses suggested less accuracy in estimation of baseline measures by the WTAR-D with some evidence for better accuracy in the WTAR-P. A consistent pattern of over-estimation of baseline standard scores by the WTAR-D was observed in that significant mean differences (relative to actual baseline performance) were obtained on four (Vigil, TMT 1, TMT 2, and Digit Span) of the nine baseline measures with trends toward over-estimation on two other measures (Stroop-W and Stroop-CW). Also, the TMT-2 demonstrated the most difficulty in baseline estimation by the premorbid measures with two of the four premorbid estimates (WTAR-D and WTAR-P) significantly over-estimating the observed baseline performance. However, it is important to note that the vast majority of the significant mean differences were only consistent with small within-subject effects (partial eta squared statistics). In fact, even when the largest within-subject effect was examined (TMT 2), only an effect at the low end of the moderate range was observed (.06), with a mean difference of seven-points between WTAR-D and baseline. Such a difference is well within one standard deviation (fifteen standard score points) and may not represent a clinically meaningful difference relative to traditional standards employed by neuropsychologists (Lezak, Howieson, and Loring, 2005).

Finally, the nonparametric binomial tests were conducted partially to address whether differences existed in the accuracy of the premorbid measures. Generally, few differences were also observed between premorbid estimates and regressions (with the exception of the HVLTR) and no differences at all were observed in TMT 1 and Stroop-W accuracy categorization rates. A possible pattern was observed (in measures with differing estimates/regressions) where either demographic estimates/regressions had

higher rates of accurately categorized participants or estimates/regressions relying on WTAR performance had higher rates of accurately categorized participants.

Demographic measures of premorbid ability or regressions that relied on demographic measures were accurate for the highest number of participants for the Vigil, HVLTR, and TMT-2. At least slight evidence or trends for higher numbers of participants with accurately estimated baselines by using WTAR performance over demographics were observed on the BVMT-R, Digit Span, and Stroop-CW. The SDMT was an exception to this pattern where methods that relied on either demographic or WTAR performance were among the highest rates of participants with accurately estimated baselines.

The third question which was to be addressed asked: *Do demographic factors influence the accuracy of premorbid measures?* The obtained results suggested that some demographic factors for certain premorbid measures did, indeed, have some impact on the accuracy of the estimation of observed baseline performance. Again, limited evidence for the greater accuracy of premorbid estimation of the WTAR-P (regardless of the demographic category of the participant) was obtained. The analyses that most directly addressed this question were the univariate analyses and the interaction analyses for the baseline repeated measure ANOVAs. The univariate tests identified a significant and consistent trend of differences in race across measures (with the exception of the Vigil), while the HVLTR, SDMT, Stroop-W, and Stroop-CW each had at least one other demographic factor that trended toward significance. Each of these factors was entered into the baseline repeated measure ANOVAs as between-subject factors. When the demographic by repeated measure interactions for the baseline measures were evaluated, notable differences were observed on the BVMT-R, SDMT, and Stroop-W interactions.

Differences in the premorbid estimates' ability to accurately estimate both male and female baseline performance were observed on the SDMT and Stroop-W, with trends toward differences in sex observed on the BVMT-R. Significant differences in previous concussion status were observed on the SDMT as well. For each of these interactions, fewer differences relative to baseline were observed for the WTAR-P than other premorbid estimates. Specifically, the WTAR-P was better able to estimate female baseline performance on average than other measures and the WTAR-P was more accurate in estimating those individuals with previous concussions as well. As with the baseline repeated measures ANOVA mean effects, only small to moderate (.03-.05) effects were observed, with only the sex differences on the SDMT approaching one standard deviation. Again, this calls into question the clinical meaning of such statistical differences.

Stepwise regression analyses were conducted to answer the question of: *What are the most meaningful and parsimonious methods for predicting observed baseline performance?* The results above suggested that the use of demographic factors provided the most meaningful method of predicting baseline performance for the majority of the measures while the use of the WTAR-P was the most meaningful and parsimonious method for the remaining baseline measures. Stepwise regression which entered the premorbid estimators and any significantly different demographic factors for each measure was the analysis that most directly addressed this question. For the stepwise regressions that were completed, moderate effects ( $R^2 = .09-.21$ ) in the final full regression models were observed. These effects were notably larger than the majority of the effects obtained in the zero-order correlations. As noted above, a pattern of including

demographic factors (especially race) while excluding most premorbid estimates was identified across the analyses. Race was an included factor on five (BVMT-R, HVLTR, SDMT, TMT 2, and Stroop-W) of the eight regressions conducted and the WTAR-D (which relies heavily on race in premorbid estimation) was included in one other regression (TMT 1). In fact, the influence of race as a predictor was illustrated in the BVMT-R stepwise regression where, after having removed the variance associated with race, the Barona estimate was negatively associated with BVMT-R baseline performance. Despite the pattern of the strength of prediction of demographic factors in the regression analyses, the WTAR-P was included on three of the eight regressions (Digit Span, Stroop-W, and Stroop-CW) and was the only included predictor for the Digit Span and the Stroop-CW.

The final question that was to be addressed asked: *Are premorbid estimates sensitive to the cognitive change post-MTBI if used as an estimate of baseline performance?* The above results suggest that, generally, the premorbid estimates are at least as sensitive to cognitive change post-MTBI relative to observed baseline performance, though again, this sensitivity varied by baseline measure. Two sets of analyses most directly addressed this final question. The first of these analyses was a single factor, repeated measures ANOVA that compared cognitive change difference scores in the post-MTBI sample. Most of the mean differences between the premorbid estimates/regressions for the measures did not differ significantly, suggesting similar sensitivity to post-MTBI cognitive change. However, significant differences of moderate effect were observed on the Vigil and the Stroop-W. For the Vigil, all of the estimates demonstrated significantly larger cognitive decline (with the WTAR-PD demonstrating

only a trend in difference) than the observed baseline performance. For the Stroop-W, when the Barona and regression formulas were used to calculate cognitive decline, significantly less decline than the observed baseline was observed while a trend was observed for the WTAR-D.

There were several issues that might call into question the validity of the post-MTBI repeated measures ANOVA results. First, given the small sample size ( $n = 19$ ), statistical power could have been an issue. Based on the effect sizes observed in the partial eta squared statistics from the non-significant analyses which ranged from .05 - .12, if a larger sample had been obtained, these results may have been significant. However, it is also important to note that large standard deviations in the observed baseline difference scores existed. These standard deviations which ranged from 14.16 – 26.66 (in standard score units) on the observed baseline post-MTBI differences suggest that some athletes likely saw decreases post-MTBI while others saw increases. This is consistent with the MTBI literature (Echemendia & Julian, 2001; Vanderploeg, Curtiss, & Belanger, 2005) which suggests that the cognitive repercussions of sports-related MTBI may be limited to a specific domain of cognitive performance of a given athlete. Therefore, those athletes who did not demonstrate cognitive decline in a given area likely could have experienced a practice effect (Rosenbaum, Arnett, Echemendia & Bailey, in press) and/or increased motivation post-injury (Bailey, Echemendia, & Arnett, 2006). These differences may diminish the meaningfulness of the mean differences post-MTBI.

The second set of analyses which directly addressed the final question were the post-MTBI binomial analyses. The frequency of post-MTBI participants who were identified as declined by the observed baseline on each of the measures ranged from 22%

to 42% and, as with the baseline measures, few differences in categorization rates were observed with no differences on the following measures: BVMT-R, HVLT-R, TMT 1, TMT 2, and Stroop-CW. However, significant differences were observed on the Vigil, SDMT, Digit Span, and Stroop-W. For each of the measures with significant differences in categorization, the estimates/regressions identified larger proportions of declined participants than did observed baseline performance. The most consistent effect of larger proportions being identified by the premorbid estimates/regressions was observed on the Digit Span and Stroop-W where all of the estimates/regressions significantly differed from baseline. Also, the WTAR-P and WTAR-PD significantly differed from baseline on the Vigil and the WTAR-D significantly differed from the baseline on the SDMT.

The obtained results on the post-MTBI binomial analyses were unexpected in that greater sensitivity to cognitive decline was found when premorbid estimates were utilized than when observed baseline performance was utilized for several measures. One of two possibilities are likely driving the results. First, the premorbid estimates could be over-identifying those participants with cognitive decline on these measures. As noted above, despite the fact that each of the post-MTBI participants had suffered a MTBI, it is likely that not each of them would experience cognitive decline over all areas of functioning (Echemendia & Julian, 2001; Vanderploeg, Curtiss, & Belanger, 2005). Therefore, only a portion of the post-MTBI athletes would likely experience cognitive decline on a given measure and, using the significant premorbid estimates/regressions instead of the observed baseline when calculating cognitive change post-MTBI on the Vigil, SDMT, Digit Span, and Stroop-W, may result in an inaccurate inflation of that proportion. However, it is also possible that the observed baseline performance may not have been an

accurate reflection of the athlete's optimal ability for a variety of factors (measurement error, fatigue, motivation difficulties, self-efficacy, etc.), and real change may have been attenuated because of practice effects (Rosenbaum et al., in press). In this case, it might be that the observed baseline performance underestimated the true proportion of post-MTBI participants that experienced true cognitive decline and the baseline estimate/regression (which may not be impacted by the above factors) provided a more accurate estimate of true pre-MTBI ability. The premorbid estimates/regressions for the Vigil, SDMT, Digit Span, and Stroop-W demonstrated increased sensitivity to MTBI; however, it is unknown if they were able to achieve this sensitivity at the cost of specificity. Unfortunately, without control subjects who had received the same measures on two administrations without injury, it is impossible to know the diagnostic utility of the premorbid measures or regressions over observed baseline performance. However, as Schoenberg et al. (2004) have noted in previous analyses, it cannot be assumed that because larger differences between a premorbid estimate and observed post-injury performance were observed that the premorbid estimate is a more accurate measure of premorbid functioning.

Based on these analyses, one might ask what a neuropsychologist should use when making a clinical decision in situations involving sports-related MTBI. In answering this question, I attempted to find the most converging overall evidence from the analyses both globally and for the measures specifically. First, it should be reiterated that no differences with large effect sizes were observed throughout any of the analyses and, as noted above, some of the sets of analyses may not have had clinically meaningful differences despite the statistical significance. This suggests that each measure may lend

some clinical utility in baseline estimation and that multiple estimates in an attempt to find convergent validity may be a clinician's most prudent option. However, if a clinician wanted to rely on one of the original measures examined (Barona, WTAR-D, WTAR-P, WTAR-PD) to provide an estimate for global baseline performance (an estimate for all nine baseline measures), the WTAR-P was observed to have the best general association and accuracy with global baseline performance. The WTAR-P had the most significant correlations with actual baseline performance and the largest effect sizes. It only differed significantly from the TMT 2 on the baseline repeated measures ANOVAs and was better able to estimate female participant baseline performance and the baseline performance of participants with previous head injury on measures where differences were observed. The WTAR-P was also within the group of measures with the highest rates of participants with accurately estimated baselines on six of the nine measures evaluated. Finally, the WTAR-P identified at least as many participants who declined post-MTBI as observed baseline performance across all of the measures according to the post-MTBI binomial tests.

Globally, the WTAR-P could be used, but the observed data also suggest that the best premorbid estimate varied by baseline instrument and, when including the stepwise regression formulae which were developed, reliance on demographic data for baseline prediction/estimation demonstrated increased utility over other measures. For three of the baseline measures (Digit Span, Stroop-W, Stroop-CW), the convergent evidence suggested that the WTAR-P was the most accurate estimate, while convergent evidence suggested that regression equations based on demographic measures were the most accurate estimate of baseline performance for all other measures with the exception of the

Vigil (which was still best estimated by the demographic-based estimate of the WTAR-D). It is also important to note that, across analyses, a consistent finding was that an estimate/regression that was based on a combination of WTAR performance and demographic information rarely led to the most accurate estimation of baseline performance.

There were several of the baseline measures where the results suggested that the best baseline estimates were those that relied solely on demographic factors. These measures include the Vigil, BVMT-R, HVLT-R, SDMT, TMT 1, and TMT 2. The results suggest that the Barona might be the baseline estimate of choice for the Vigil. None of the measures evaluated had significant correlations with the Vigil and no differences were observed on the univariate analyses all of which suggest that the Vigil might be the measure of attention/concentration least confounded that was least confounded by measurement error within the battery. However, a significant mean difference between the WTAR-D and baseline performance was observed on the repeated measures ANOVA analyses and only the WTAR-D and Barona were among the measures that demonstrated the highest proportion of participants with accurately estimated baselines. Also, the Barona identified at least as many participants who declined post-MTBI as the observed Vigil baseline performance across all of the measures according to the post-MTBI binomial tests. For each of the BVMT-R, HVLT-R, SDMT, TMT 1, and TMT 2, the best baseline estimate was the obtained stepwise regression for the measure, each of which relied on demographic factors such as race, sex, and sport. In each case, the overall effect sizes for the full model regressions were larger than the effects based on the zero-order correlations of the premorbid estimates (full model  $R^2$  statistics for the regressions ranged

from .09 to .21). Also, there were significant differences between the premorbid estimates and baseline performance on the repeated measures ANOVAs for these measures. The stepwise regressions for the BVMT-R, HVLTR, SDMT, TMT 1, and TMT 2 were always among the measures with the highest proportion of accurately estimated baselines according to the binomial tests. Finally, the post-MTBI binomial tests suggested no decrease in sensitivity to cognitive decline (relative to baseline) by using the stepwise regressions.

There were three baseline measures for which the results suggest that the best baseline estimate was that which relied on WTAR-P performance. These baseline measures included the Digit Span, Stroop-W, and Stroop-CW. For the Digit Span, the WTAR-P demonstrated a moderate effect size based on the zero-order correlation and was the only included predictor in the stepwise regression. The ANOVA suggested that the WTAR-D may overestimate Digit Span performance and the WTAR-P remained in the measure that was among the highest in proportion of accurately estimated baselines on the baseline binomial test. The WTAR-P was also significantly higher than baseline in the number of participants identified as declined post-MTBI. For the Stroop-W, the WTAR-P demonstrated a moderate effect size based on the zero-order correlation consistent with the stepwise regression that retained the WTAR-P along with race. The Stroop-W ANOVA suggested that the WTAR-P was not significantly different from the observed baseline in estimating sex differences while all other premorbid measures were significantly different. Also, the WTAR-P remained in the measures that were among the highest in proportion of accurately estimated Stroop-W baselines on the baseline binomial test. Also, the Stroop-W regression's mean post-MTBI difference score

demonstrated significantly less decline and the WTAR-P was also significantly higher than the Stroop-W baseline in the number of participants identified as declined post-MTBI. Finally, the WTAR-P was also identified as the best estimate for Stroop-CW performance as well. The WTAR-P demonstrated a moderate effect size on the zero-order correlations and was the only significant predictor retained by the stepwise regression. The Stroop-CW repeated measures ANOVA suggested few differences between measures with only a trend in overestimation by the WTAR-D. The WTAR-P was among the measures with the highest proportion of accurately estimated Stroop-CW baselines and no significant post-MTBI differences were observed in sensitivity.

Table 12 provides a summary of these best baseline estimators. First, when considering only the premorbid estimates which were evaluated in the current project, the best overall estimate of baseline ability across the measures was the WTAR-P. Next, when evaluating the baseline measures individually, the best estimates were those obtained when using the stepwise regression equations (based on demographics) or those obtained by using the WTAR-P. The stepwise regressions provided the best overall method for estimating baseline performance on the BVMT-R, HVLIT-R, SDMT, TMT 1, and TMT 2. These regression equations most often included race, though other demographic factors such as sex, sport, and the WTAR-D were also included. Baseline Vigil performance was best estimated by the WTAR-D alone. The WTAR-P was the best estimate for the Digit Span, Stroop-W, and Stroop-CW measures.

Insert Table 12.

## *Race, Sex, and Sport as Predictors of Baseline Performance*

As described above, several demographic factors were examined through univariate tests, mixed factor repeated measure ANOVAs, stepwise regressions, and binomial tests. The demographic factors, especially race, demonstrated important differences and predictive ability at baseline. Racial categories in neuropsychology, and clinical psychology in general, have been a topic of much controversy (Okazaki & Sue, 2003; Schaefer, 1998). Research in this area has been confounded by problems that range from the misunderstanding of the terms and categories being studied to general disinterest or apathy in approaching the topic at all (Lezak, Howieson, & Loring, 2005; Okazaki & Sue, 2003; Schaefer, 1998). Despite this controversy, differences in cognitive performance have been reported between racial groups. Norman et al. (2000) showed that, in a normative sample used to produce demographically-corrected norms, Caucasians significantly outperformed African Americans on the California Verbal Learning Test; however, the Caucasians had significantly more education than the African-American participants. Significant differences have also been identified on global measures of cognitive ability such as the WAIS-III and WAIS-R, with Caucasians demonstrating around a score exceeding that of African Americans by 2 scaled score points for each subtest up through participants in their mid-thirties (Kaufman, McLean, & Reynolds, 1988; Wechsler, 1997). Also, many measures provide separate norms by racial categories, something that also highlights performance differences between races (Heaton, Miller, Taylor, & Grant, 2004; Lezak, Howieson, & Loring, 2005). Like the current study, sex differences have also been identified in the neuropsychological

literature. In fact, a consistent pattern of women outperforming men on tests of psychomotor speed and accuracy has been relatively consistent (Majeres, 1990; Schmidt et al., 2000). The female advantage in psychomotor speed has also been observed on the SDMT, though not believed to be large enough to warrant separate norms (Heaton, Taylor, & Manly, 2003; Kaufman, McLean, and Reynolds, 1988). These findings are consistent with those noted in the above analyses. However, it is important not to deduce, based on these findings, that any certain category of participants have significantly lower cognitive capacities despite these consistent results. Though neural/biological differences between groups may have some impact (Lezak, Howieson, & Loring, 2005), other unmeasured factors may be important in influencing differences between demographic groups on neuropsychological tests. In fact, evidence for the influence of other explanatory factors such as socio-economic status (SES) exists within the current data as well.

Given the participant membership in sport, the SES of the individuals may be inferred. The likelihood of participation in sports such as soccer, ice hockey, and lacrosse in high school has been shown to be more frequent in areas of high SES than areas of low SES (Eitzen & Sage, 1991). Participation in contact sporting events in particular (such as football and boxing) has been particularly associated with households of low income and working class positions (Eitzen & Sage, 1991; Wilson, 2002). Therefore, one might expect athletes who engage in sports such as soccer, ice hockey, and lacrosse (suggesting higher SES) to demonstrate higher performance on cognitive measures. This was observed in the current data with the significant differences in sport observed in the HVLT-R, where football was significantly lower than men's lacrosse and trends were

observed in women's soccer, and women's lacrosse. When the race groups are broken down by sport participation, 87% of the African Americans engaged in sports associated with lower SES (football and basketball) while only 22% of Caucasian athletes participated in the same sports (football and basketball). The frequency of football and basketball participation relative to sex indicated that males (47%) were more likely to participate in these lower SES sports than females (13%). To further buffer these claims, the unexpected finding that having had previous concussions was associated with higher performance on the HVLN-R (with trends toward higher performance on the BVMT-R and the SDMT), could also be influenced by the same factor. The previous concussion group had fewer participants who engaged in football and basketball (24%) than those that did not have previous concussions (40%).

In the end, it is likely that the demographic differences that were observed in the current analyses and provided the best estimation/prediction of baseline performance on several measures are likely due to multiple factors. These may include, but are likely not limited to: race, sex, sport, culture, and SES. This alternative and confounding factors should be considered before sweeping conclusions regarding the meaning of demographic differences can be made.

#### *Estimating Digit Span and Stroop Baseline Performance*

Despite the demographic factors being the best estimate for the majority of the baseline measures, the Digit Span and both the Stroop-W and Stroop-CW were best estimated by the WTAR-P estimate. The explanation for the accuracy of the WTAR-P

over demographic and other estimators may be entirely psychometric. The WTAR-P was developed with and normed to estimate the WAIS-III (The Psychological Corporation, 2001). The Digit Span used in these analyses is a subtest of the WAIS-III and is correlated with the WAIS-III Full Scale IQ (which the WTAR was designed to estimate) at .57 (Wechsler, 1997). The accuracy of the WTAR-P with the Digit Span may, therefore, be based, in part, on the Digit Span test's relationship to the WAIS-III. With regard to the Stroop and the WTAR-P, one cannot help but note the similarity in test demands (especially between the Stroop-W and WTAR-P). Both measures require a participant to read words that are presented in columns. The over-learned construct of reading ability is a necessary component of both the WTAR-P and the Stroop-W. The addition of the interference component of the Stroop task (having to inhibit reading to identify ink color) in the Stroop-CW is likely the cause of the reduction in correlation between the Stroop-CW and WTAR-P (.30) from what was observed between the Stroop-W and WTAR-P (.42).

Another possible reason for the accuracy of the WTAR-P (though not directly tested in this sample), is the possible impact of motivation. Bailey, Echemendia, & Arnett (2006) demonstrated that baseline effort in MTBI testing may be diminished in some athletes. Tests that were included in those that were shown to be influenced by apathy/disinterest of athletes being tested at baseline were the Digit Span and the Stroop-CW. If the WTAR-P was also a measure that was impacted by mild fluctuations of motivation at baseline, it might explain the relationship between the Digit Span and Stroop-CW. Further evidence for this might be seen in the post-MTBI binomial tests for the Digit Span and both Stroop trials in that the premorbid estimates were more likely to

identify athletes as declined post-MTBI. However, given the test-retest reliability and resistance to other confounding factors (The Psychological Corporation, 2001), the hypothesis that the WTAR-P is influenced by baseline motivation may be unlikely.

### *Limitations*

There were at least three significant limitations to this project which should be addressed in future research. First, the OPIE and “best performance” methods of premorbid estimation were not evaluated in the current project. Several measures necessary for OPIE calculation were not administered as a part of the PSU Concussion Program baseline and therefore, the estimate could not be included in the current project. Also, due to the limited number of individuals in the post-MTBI sample, it was thought that the best-performance estimate would not be adequately evaluated relative to the other measures. A related second limitation to the current project was the limited post-MTBI sample. The nineteen participants who comprised the post-MTBI sample may not have yielded adequate statistical power and may have led to post-injury means that could be misleading. A final limitation of the current project was the lack of control participants who received both baseline and a second evaluation (consistent with the post-MTBI sample). Control data would have allowed for a more sophisticated evaluation of the diagnostic utility of the premorbid measures versus the diagnostic utility of baseline performance.

## *Final Conclusions*

Estimation of baseline performance is a crucial aspect of the clinical decisions made by neuropsychologists following sport-related MTBI. The convergent evidence from all of the analyses presented in this project suggests that the WTAR-P was the best premorbid estimate of global baseline performance relative to those examined. The stepwise regressions also revealed that demographic factors, especially those including race, typically accounted for the most variance of baseline performance. According to the convergent evidence from all of the analyses, the obtained stepwise regressions provided the best estimates for the majority of the individual baseline measures. The WTAR-P provided the best estimate for the remaining individual baseline measures. Limited evidence was identified for the possible increased sensitivity of premorbid measures over some observed baseline performance, but the validity for this claim would need to be the subject of future research before any firm conclusions could be made regarding the diagnostic utility of premorbid measures over baseline measures. In the end, due to the small effects of the differences between the estimation methods, these results provide limited evidence for the use of a single measure of premorbid estimation. Instead, like Skeel et al. (2004), further support was obtained for the use of multiple estimates of baseline performance in conjunction (both separate demographic-based measures and reading-based measures if possible) in an attempt to identify converging evidence toward the true baseline ability on a given measure.

## REFERENCES

- Achen, C. H. (1982). *Interpreting and Using Regression*. London: Sage Publications.
- Allen, L.M., Conder, R.L., Green, P., Cox, D.R. (1997). *CARB '97: Computerized Assessment of Response Bias. Manual*. Durham, NC: CogniSyst.
- Allen, D.N., Kelley, M.E., Miyatake, R.K., Gurklis, J.A., & van Kammen, D.P. (2001). Confirmation of a two-factor model of premorbid adjustment in males with schizophrenia. *Schizophrenia Bulletin*, 27, 39-46.
- Alves, W.A., Macchiocchi, S.N., & Barth, J.T. (1993). Postconcussive symptoms after mild head injury. *Journal of Head Trauma Rehabilitation*, 8, 48-59.
- American Academy of Neurology (1997). Practice parameter: The management of concussion in sports (summary statement). *Neurology*, 48, 581-585.
- American Psychiatric Association (2000). *Diagnostic and Statistic Manual IV – Test Revised*. Washington, DC: American Psychiatric Association.
- Bailey, C.M. & Arnett, P.A. (2004). *Response speed on the Computerized Assessment of Response Bias and its relationship to motivation in multiple sclerosis*. Poster presented at the National Academy of Neuropsychologists (NAN) conference, Seattle, Washington.

- Bailey, C.M. & Arnett, P.A. (2005). *Assessment of effort in baseline MTBI neuropsychological testing*. Poster to be presented at the National Academy of Neuropsychologists (NAN) conference, Tampa, Florida.
- 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*, 1-10.
- Bailey, C.M., Echemendia, R.J., Arnett, P.A. (2005). *Visual memory predicts concussion status independent of verbal memory and post-mild traumatic brain injury symptoms in college athletes*. Paper presented at the International Neuropsychological Society (INS) conference, St. Louis, Missouri.
- Barth, J.T., Alves, W.M., Ryan, T.V., Macciocchi, S.N., Rimel, R.W., Jane, J.A., & Nelson, W.E. (1989). Mild head injury in sports: Neuropsychological sequelae and recovery of function. In Levin, H.S., Eisenberg, H.M., and Benton, A.L. (eds.), *Mild Head Injury*. New York: Oxford University Press, 257-275.
- Barth, J.T., Bartolic, E., Echemendia, R., Frazen, M., Heilbronner, R., Kay, T., Koffler, S., Kutner, K., Lovell, M., & Parker, E. (1997). *Recommendations for the use of neuropsychological assessment procedures in sports*. Paper presented at the 2<sup>nd</sup> Annual Conference on Sports-Related Concussion on Spinal Injuries, Orlando, Fl.

Barona, A., Reynolds, C.R., and Chastain, R.L. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52*, 885-887.

Benedict, R.H.B., Schretlen, D., Groninger, L., Brandt, J. (1998). Hopkins Verbal Learning Test-Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist, 12*, 43-55.

Benedict, R.H.B., Schretlen, D., Groninger, L., Dobraski, M., Sphritz, B. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. *Psychological Assessment, 8*, 145-153.

Berlanger, H.G., Curtiss, G., Demery, J.A., Lebowitz, B.K., Vanderploeg, R.D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 215-227.

Berlanger, H.G. & Vanderploeg, R.D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 345-357.

- Bigler, E. E., & Snyder, J. L. (1995). Neuropsychological outcome and quantitative neuroimaging in mild head injury. *Archives of Clinical Neuropsychology, 10*, 159-174
- Binder, L.M. (1997). A Review of Mild Head Trauma. Part II: Clinical Implications. *Journal of Clinical and Experimental Neuropsychology, 19*, 432-457.
- Binder, L.M., & Rohling, M.L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry, 153*, 5-8.
- Bohnen, N., Jolles, J., & Twijnstra, A. (1992). Modification of the Stroop Color Word Test improves differentiation between patients with mild head injury and matched controls. *The Clinical Neuropsychologist, 6*, 178-184.
- Cantu, R.C. (1998). Return to play guidelines after a head injury. *Clinical Sports Medicine, 17*, 121-126.
- Cantu, R.C. (2001). Posttraumatic Retrograde and Anterograde Amnesia: Pathophysiology and Implications in Grading And Safe Return To Play. *Journal of Athletic Training, 36*, 244-248.

Cantu, R.C., & Voy, R. (1995). Second impact syndrome: A risk in any contact sport. *The Physician and Sports Medicine, 23*, 27-34.

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

Chelune, G. J., Naugle, R. I., Luders, H., Sedlak, J. & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*, 41-52.

Collins, M. W., Grindell, H., Lovell, M. R., Dede, D., Moser, D., Phalin, B., Nogle, S., Wasik, M., Cordy, D., Daugherty, M., Sears, S., Nicolette, G., Indelicato, P., & McKeag, D. (1999). Relationship between concussion and neuropsychological performance in college football players. *Journal of the American Medical Association, 282*, 964-970.

Colorado Medical Society (1991). *Report of the sports medicine committee: Guidelines for the management of concussion in sports (revised)*. Colorado Medical Society, Denver.

Crawford, J.R., Parker, D.M., & Besson, J.A.O, (1988). Estimation of premorbid intelligence in organic conditions. *British Journal of Psychiatry, 153*, 178-181.

- Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.
- Echemendia, R.J. (1997). *Neuropsychological assessment of college athletes: The Penn State Concussion Program*. Paper presented at the meeting of the National Academy of Neuropsychology, Las Vegas, NV.
- 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.
- Eitzen, S.D. & Sage, G.H. (1991). *Sociology of North American Sport (5<sup>th</sup> Ed.)*. Madison, WI: Brown & Benchmark.
- Erlanger, D.M., Kutner, K.C., Barth, J.T., & Barnes, R. (1999). Neuropsychology of sports-related head injury: Dementia Pugilistica to Post Concussive Syndrome. *The Clinical Neuropsychologist*, 13, 193-209.
- Gerberich, S.G., Priest, J.D., Boen, J.R., Straub, C.P., & Maxwell, R.E. (1983). Concussion incidences and severity in secondary school varsity football players. *American Journal of Public Health*, 73, 1370-1375.
- Girden, E. R. (1992). *ANOVA: Repeated Measures*. London: Sage Publications.

- Giza, C.C. & Hovda, D.A. (2001). The neurometabolic cascade of concussion. *Journal of Athletic Training, 36*, 228-235.
- Griffin, S.L., Mindt, M.R., Rankin, E.J., Ritchie, A.J., Scott, J.G. (2002). Estimating premorbid intelligence: Comparison of traditional and contemporary methods across the intelligence continuum. *Archives of Clinical Neuropsychology, 17*, 497-507.
- Harnett, M.A., Godfrey, H.P., & Knight, R.G. (2004). Regression equations for predicting premorbid performance on executive test measures by persons with traumatic brain injuries. *New Zealand Journal of Psychology, 33*, 78-87.
- Hayes, R. L., & Dixen, C.E. (1994). Neurochemical changes in mild head injury. *Seminars in Neurology, 14*, 25-31.
- 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*. Lutz, FL: Psychological Assessment Resources, Inc.

- Heaton, R.K., Taylor, M.J., & Manly, J. (2003). Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III. In Tulskey et al. (Eds.), *Clinical interpretation of the WAIS-III and WMS-III*. San Diego, CA: Academic Press.
- Hovda, D.A., Prins, M., Becker, D.P., Lee, S., Bergsneider, M., & Martin, N.A. (1999) Neurobiology of concussion. In Bailes, J.E., Lovell, M.R., & Maroon, J.C. (eds.) *Sports-Related Concussion*. St. Louis, Missouri: Quality Medical Publishing, Inc.
- Jacobson, N.S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Jordan, B.D. (1987). Neurologic aspects of boxing. *Archives of Neurology*, 44, 453-459.
- Junger, E.C., Newell, D.W., Grant, G.A., Avellino, A.M., Ghatan, S., Douville, C.M., Lam, A.M., Aaslid, R., & Winn, H.R. (1997). Cerebral autoregulation following minor head injury. *Journal of Neurosurgery*, 86, 425-432.
- Kaufman, A.S., McLean, G.E., & Reynolds, C.R., (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology*, 44, 231-248.

- Kareken, D.A (1997). Judgment pitfalls in estimating premorbid intellectual function. *Archives of Clinical Neuropsychology, 12*, 701-709.
- Kelly, J.P., & Rosenberg, J.H. (1997). The diagnosis and management of concussion in sports. *Neurology, 48*, 575-580.
- Kremen, W.S., Seidman, L.J., Faraone, S.V., Pepple, J.R., Lyons, M.J., & Tsuang, M.T. (1996). The “3 R’s” of neuropsychological function in schizophrenia: An empirical test of the matching fallacy. *Neuropsychology, 10*, 22-31.
- Krull, K., Sherer, M., & Adams, R. (1995). A comparison of indices of premorbid intelligence in clinical populations. *Applied Neuropsychology, 2*, 35-38.
- Lezak, M.D., Howieson, D.B, & Loring (2005) *Neuropsychological assessment (4<sup>th</sup> Edition)*. New York: Oxford University Press.
- Leininger, B.E., Gramling, S.E., & Farrell, A.D., et al. (1990). Neuropsychological deficits in symptomatic minor head injury patients after concussion and mild concussion. *Journal of Neurology, Neurosurgery, and Psychiatry, 53*, 293-296.
- Lovell, M.R., & Collins, M.W. (1998). Neuropsychological assessment of the college football player. *Journal of Head Trauma Rehabilitation, 12*, 9-26.

- Lovell, M. R., Collins, M. W., Podell, K., Powell, J., & Maroon, J. (2002). *ImPACT: Immediate post-concussion assessment and cognitive testing*. Pittsburgh, PA: NeuroHealth Systems, LLC.
- Majeres, R.L. (1990). Sex differences in comparison and decision processing when matching strings of symbols. *Intelligence, 14*, 357-370.
- McCarthy, F.M., Sellers, A.H., Burns, W.J., Smith, G., Ivnik, R.J., and Malec, J.F. (2003). Prediction of IQ in the Mayo older adult normative sample using multiple methods. *Journal of Clinical Psychology, 59*, 457-463.
- Mesulam, M.M. (1983). The functional anatomy and hemispheric specialization for directed attention: The role of the parietal lobe and its connectivity. *Trends in Neuroscience, Sept.*, 384-387.
- Mitrushina, M.N., Boone, K. B., & D'Elia, L.F. (1999). *Handbook of normative data for neuropsychological assessment*. New York: Oxford University Press.
- Morrison, G., Sharkey, V., Allardyce, J., Kelly, R.C., McCreadie, R.G. (2000). Nithsdale schizophrenia surveys 21: A longitudinal study of National Adult Reading Test stability. *Psychological Medicine, 30*, 717-720.
- Moss, A.R., & Dowd, T. (1991). Does the NART hold after head injury? A case report. *British Journal of Clinical Psychology, 30*, 179-180.

- Nelson, H.E., & Willison, J.R. (1991). *National Adult Reading Test (NART): Test Manual* (2<sup>nd</sup> ed.). Windsor, UK: NFER-Nelson.
- Nuwer, M.R., Hovda, D.A., Schrader, L.M., & Vespa, P.M. (2005). Routine and quantitative EEG in mild traumatic brain injury. *Clinical Neurophysiology*, in press.
- Okazaki, S. & Sue, S. (2003). Methodological issues in assessment research with ethnic minorities. In Kazdin, A.E. (ed.) *Methodological issues and strategies in clinical research* (3<sup>rd</sup> Edition). Washington, DC: American Psychological Association.
- Orme, D., Ree, M.J., & Rioux, P. (2001). Premorbid IQ estimates from a multiple aptitude battery: Regression vs. equating. *Archives of Clinical Neuropsychology*, 16, 679-688.
- Powell, B.D., Bossart, D.F., & Reynolds, C.R. (2003). Evaluation of the accuracy of two regression-based methods for estimating IQ. *Archives of Clinical Neuropsychology*, 18, 277-292.
- Povlishock, J.T., & Coburn, T.H. (1989) Morphopathological change associated with mild head injury. In Levin, H.S., Eisenberg, H.M., and Benton, A.L. (eds.), *Mild Head Injury*. New York: Oxford University Press, 37-53.

Reynolds, C.R. (1997). Postscripts on premorbid ability estimation: Conceptual addenda and a few words on data. *Clinical Neuropsychology, 1*, 29-31.

Reynolds, C.R. (2002). *CTMT: Comprehensive Trail-Making Test. Manual*. Austin, TX: Pro-Ed, Inc.

Rosenbaum, A.M., Arnett, P.A., Bailey, C.M., & Echemendia, R.J. (in press). Neuropsychological Assessment of Sports-Related Concussion: Measuring Clinically Significant Change. In Slobounov, S. & Sebastainelli, W. (eds.) *Foundation of Sports-Related Concussion*. New York: Springer Publishing Co.

Schaefer, R.T. (1998). *Racial and Ethnic Groups*. New York: Longman.

Schmidt, S.L., Oliveira, R.M., Rocha, F.R., & Abreu-Villaca, Y. (2000). Influences of handedness and gender on the grooved pegboard test. *Brain and Cognition, 44*, 445-454.

Schoenburg, M.R., Scott, J.G., Ruwe, W., Patton, D., Adams, R.L. (2004). Discussion – Assumptions that underlie predicting IQ: A comment on the “Evaluation of the accuracy of two regression-based methods for estimating premorbid IQ.” *Archives of Clinical Neuropsychology, 19*, 1103-1106.

- Skeel, R.L., Sitzer, D., Fogal, R., Wells, J., Johnstone, B. (2004). Comparison of predicted-difference, simple-difference, and premorbid estimation methodologies for evaluating IQ and memory score discrepancies. *Archives of Clinical Neuropsychology, 19*, 363-374.
- Smith, A. (1982) *Symbol Digit Modalities Test (SDMT). Manual (Revised)*. Los Angeles: Western Psychological Services.
- Snow, W.G., Tierney, M.C., & Zorzitto, M.L., (1988). One-year test-retest reliability of selected neuropsychological tests in older adults. *Journal of Clinical and Experimental Neuropsychology, 11*, 423-428.
- Speer, D.C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60*, 402-408.
- Speer, D.C. & Greenbaum, P.E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology, 63*, 1044-1048.
- Spreen, O., & Straus, E. (1991). *A compendium of neuropsychological tests*. New York: Oxford University Press.

Stiller, J.W., & Weinberger, D.R. (1985). Boxing and chronic brain damage. *Psychiatric Clinics of North America*, 8, 339-356.

Stuss, D.T., Ely, P., & Hugenholtz, H. (1985). Subtle neuropsychological deficits in patients with good recovery after closed head injury. *Neurosurgery*, 17, 41-47.

Tracy, J.I., McGrory, A.C., Josiassen, R.C. & Monaco, C.A. (1996). A comparison of reading and demographically-based estimates of premorbid intelligence in schizophrenia. *Schizophrenia Bulletin*, 22, 103-109.

Vanderploeg, R.D., Curtiss, G., & Belanger, H.G. (2005). Long-term neuropsychological outcomes following mild traumatic brain injury. *Journal of the International Neuropsychological Society*, 11, 228-236.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence (4<sup>th</sup> ed.)*. Baltimore, MD: Williams & Wilkins.

Wechsler, D.A. (1997). *Wechsler Adult Intelligence Scale-III*. San Antonio, TX: The Psychological Corporation.

Wilson, T.C. (2002). The paradox of social class and sports involvement. *International Review for the Sociology of Sport*, 37, 5-16.

The Psychological Corporation (2001). *The Wechsler Test of Adult Reading (WTAR): Test Manual*. San Antonio, TX: The Psychological Corporation.

## TABLES

Table 1.

Group Demographic Information

Sample	n	% Male	Age	% With Age < 20	Ed. Level	% College Freshmen	% Caucasian	% With PCs
Baseline Sample	106	70%	18.83	73%	12.36	76%	78%	27%
Post-MTBI Sample	19	68%	19.95	32%	13.00	47%	79%	37%

Age and Education Level were dichotomized into the groups described above; Ed. = Education; PCs = Previous Concussions; Freshmen = freshmen in college; MTBI = mild traumatic brain injury

Table 2.

Sport and Census Region Information

	Baseline Sample	Post-MTBI Sample
<b>Sport</b>		
Football*	26%	31%
Men's Soccer	5%	5%
Women's Soccer*	17%	11%
Hockey*	14%	21%
Men's Basketball	6%	--
Women's Basketball	4%	5%
Gymnastics	1%	--
Men's Lacrosse*	11%	11%
Women's Lacrosse*	14%	16%
Wrestling	2%	--
<b>Region</b>		
Northeast	73%	68%
Midwest	9%	11%
South	18%	21%

\* = Sports that were selected to be used in analyses; MTBI = mild traumatic brain injury

Table 3.

## Means and Standard Deviations for the Estimates and Measures

Measure	Baseline		Post-Injury	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
WTAR	34.22	7.02	--	--
WTAR-D	103.05	4.02	--	--
WTAR-P	102.02	8.80	--	--
WTAR-PD	100.22	7.29	--	--
Barona	99.66	4.85	--	--
Vigil	410.90	34.35	428.62	52.69
BVMT-R	28.22	4.86	28.42	4.39
HVLT-R	27.32	3.72	28.00	4.52
SDMT	59.49	10.15	61.84	13.67
TMT Trial 1				
Form 2	34.25	10.20	34.21	13.28
Form 3	42.55	14.08	38.44	14.45
Form A	20.94	8.08	--	--
TMT Trial 2				
Form 4	26.54	6.79	30.26	9.83
Form 5	49.01	14.36	44.86	16.02
Form B	47.33	13.99	--	--
Digit Span	18.84	3.51	21.05	4.68
Stroop-W	53.32	8.21	56.65	16.08
Stroop-CW	112.29	17.56	107.04	22.98

It is important to note that there were multiple forms of the TMT which were not designed to be equivalent; the TMT forms were standardized separately, but combined into the appropriate trial for analysis; *M* = Mean; *SD* = Standard Deviation; WTAR = Wechsler Test of Adult Reading; WTAR-D = Full Scale IQ as predicted by demographics only according to the WTAR; WTAR-P = Full Scale IQ as predicted by WTAR performance; WTAR-PD = Full Scale IQ as predicted according to WTAR performance and demographics; Barona = Full Scale IQ as predicted by the Barona et al. (1984) equation; Vigil = Vigil Reaction Time in milliseconds; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test Time in seconds; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time in seconds; Stroop-CW = Stroop Color-Word Trial Time in seconds

Table 4.

## Baseline Correlations

Measure	Vigil	BVMT-R	HVLT-R	SDMT	TMT 1	TMT 2	Digit Span	Stroop-W	Stroop-CW
Barona	-.03	.02	.24*	.12	.24*	.25*	.20*	.28**	.17
WTAR-D	.10	.12	.21*	.16	.29**	.26**	.17	.27**	.16
WTAR-P	-.17	.20*	.21*	.19*	.06	.20*	.42***	.42***	.30***
WTAR-PD	-.15	.18	.25*	.14	.18	.25*	.40***	.41***	.24*

All correlations displayed are between the premorbid estimations and the baseline standard scores for each measure; \* = correlation significantly different from zero at  $p < .05$ ; \*\* = correlation significantly different from zero at  $p < .01$ ; \*\*\* = correlation significantly different from zero at  $p < .001$ ; WTAR = Wechsler Test of Adult Reading; WTAR-D = Full Scale IQ as predicted by demographics only according to the WTAR; WTAR-P = Full Scale IQ as predicted by WTAR performance; WTAR-PD = Full Scale IQ as predicted according to WTAR performance and demographics; Barona = Full Scale IQ as predicted by the Barona et al. (1984) equation; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial Time

Table 5.

## Baseline Univariate Comparisons

Factor	Vigil	BVMT-R	HVLT-R	SDMT	TMT 1	TMT 2	Digit Span	Stroop-W	Stroop-CW
Age	< 20 years	101.45	99.34	101.02	98.57	98.66	100.13	98.31	99.06
	20 + years	96.20	101.77	97.89	99.93	98.19	99.56	100.37	98.71
Sex	Female	100.13	103.25	105.33*	95.60	101.73	97.35	103.22 <sup>t</sup>	103.52 <sup>t</sup>
	Male	99.94	98.60	97.70	100.39	97.15	101.16	96.93	96.89
Race	Afr. American	99.91	90.82	92.14	90.60	86.89	91.77	86.63	89.89
	Caucasian	100.02	102.04**	102.18**	101.25*	101.76***	102.32**	102.15***	101.19**
Previous Concussion Status	None	100.47	98.13	98.51	97.91	97.75	99.43	98.47	97.41
	1+	98.70	104.00 <sup>t</sup>	103.96 <sup>t</sup>	101.68	100.61	101.42	99.98	103.11
Sport	Football	102.12	91.80 <sup>C</sup>	100.49	98.43	93.85	98.85	96.76	95.97
	Women's Soccer	101.21	103.64 <sup>B</sup>	105.02	105.67	104.14	102.35	101.90	106.07
	Hockey	102.79	98.08	96.71	102.13	99.57	102.21	106.18	98.08
	Men's Lacrosse	95.89	99.32	94.44	99.37	96.72	101.07	93.95	93.05
	Women's Lacrosse	98.61	101.17	104.62 <sup>B</sup>	106.47	88.89	102.36	96.41	104.02

All measures are presented as standard scores; <sup>t</sup> = mean significantly larger at p < .10; \* = mean significantly larger at p < .05; \*\* = mean significantly larger at p < .01; \*\*\* = mean significantly larger at p < .001; <sup>A</sup> = mean significantly larger than <sup>B</sup> at p < .05; <sup>B</sup> = mean significantly larger than <sup>C</sup> at p < .10; <sup>C</sup> = mean significantly smaller than <sup>A</sup> at p < .05 and smaller than <sup>B</sup> at p < .10; Afr. = African; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial Time

Table 6.

Repeated Measures ANOVA Estimated Marginal Means for Baseline Performance and Premorbid Estimates

Factor	Vigil	BVMT-R	HVLT-R	SDMT	TMT 1	TMT 2	Digit Span	Stroop-W	Stroop-CW
Baseline	100.00	100.35	100.65	101.50	95.93	94.33	97.04	96.49	96.68
Barona	99.65	96.52	98.91	96.52*	97.14	97.14	97.08	96.58	96.39
WTAR-D	103.03*	99.98	102.96	99.98	101.00*	101.00**	100.97*	100.39 <sup>†</sup>	100.55 <sup>†</sup>
WTAR-P	101.91	100.43	101.58	100.43	99.01	99.01*	98.98	99.33	99.12
WTAR-PD	100.14	97.27	99.62	97.27	97.15	97.15	97.06	96.78	96.70

All estimated marginal means are displayed as standard scores; \* = mean significantly different from baseline at  $p < .10$ ; \*\* = mean significantly different from baseline at  $p < .01$ ; \*\*\* = mean significantly different from baseline at  $p < .001$ ; † = mean significantly different from baseline at  $p < .05$ ; \*\* = mean significantly different from baseline at  $p < .01$ ; \*\*\* = mean significantly different from baseline at  $p < .001$ ; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial

Table 7.

Stepwise Regressions Predicting Baseline Measures

	Vigil	BVMT-R	HVLT-R	SDMT	TMT 1	TMT 2	Digit Span	Stroop-W	Stroop-CW
Variables Entered	None	Race Barona	Sport Race	Race Sex	WTAR-D	Race	WTAR-P	WTAR-P Race	WTAR-P
Variables Excluded	Barona WTAR-D WTAR-P WTAR-PD PCS	WTAR-D WTAR-P WTAR-PD Sex PCS	Barona WTAR-D WTAR-P WTAR-PD Race	Barona WTAR-D WTAR-P WTAR-PD PCS	Barona WTAR-P WTAR-PD Race WTAR-PD Sex	Barona WTAR-D WTAR-P WTAR-PD Race Sex	Barona WTAR-D WTAR-PD Race	Barona WTAR-D WTAR-PD Sex	Barona WTAR-D WTAR-PD Race Sex
Full Model R <sup>2</sup>	.00	.14	.21	.12	.09	.12	.18	.21	.09

WTAR = Wechsler Test of Adult Reading; WTAR-D = Full Scale IQ as predicted by demographics only according to the WTAR; WTAR-P = Full Scale IQ as predicted by WTAR performance; WTAR-PD = Full Scale IQ as predicted according to WTAR performance and demographics; Barona = Full Scale IQ as predicted by the Barona et al. (1984) equation; PCS = Previous Concussion Status; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial

Table 8.

## Stepwise Regression Equations for Baseline Measures

Measure	Equation
BVMT-R	BVMT-R Baseline Standard Score = $213.89 + 20.93(\text{race}) + (-1.31)(\text{Barona})$
HVLT-R	HVLT-R Baseline Standard Score = $78.99 + (6.99)(\text{sport}) + (10.62)(\text{race})$
SDMT	SDMT Baseline Standard Score = $97.48 + (9.29)(\text{race}) + (-6.82)(\text{sex})$
TMT 1	TMT 1 Baseline Standard Score = $-41.98 + (1.37)(\text{WTAR-D})$
TMT 2	TMT 2 Baseline Standard Score = $86.90 + (14.86)(\text{race})$
Digit Span	Digit Span Baseline Standard Score = $27.56 + (0.71)(\text{WTAR-P})$
Stroop-W	Stroop-W Baseline Standard Score = $30.87 + (0.60)(\text{WTAR-P}) + (9.23)(\text{race})$
Stroop-CW	Stroop-CW Baseline Standard Score = $42.10 + (0.56)(\text{WTAR-P})$

It is important to note that the demographic variables for the regression analyses were recoded according to the univariate analysis results. Therefore, the following codes were utilized: race (0=African American, 1 = Caucasian), sex (0 = female, 1 = male), and sport (1 = football, 2 = women's soccer, hockey, women's lacrosse, 3 = men's lacrosse). WTAR = Wechsler Test of Adult Reading; WTAR-D = Full Scale IQ as predicted by demographics only according to the WTAR; WTAR-P = Full Scale IQ as predicted by WTAR performance; Barona = Full Scale IQ as predicted by the Barona et al. (1984) equation; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial.

Table 9.

Frequency of Participant Estimated Baseline Scores that Did Not Differ from the Observed Baseline by One Standard Deviation or More

Measure	% of Participants within 1 SD		% of Participants within 1 SD
Vigil		TMT 2	
Barona	65 <sup>A</sup>	Barona	67
WTAR-D	64	WTAR-D	70 <sup>A</sup>
WTAR-P	55*	WTAR-P	59*
WTAR-PD	58 <sup>t</sup>	WTAR-PD	63 <sup>t</sup>
		Regression	68
BVMT-R		Digit Span	
Barona	59*	Barona	68
WTAR-D	66	WTAR-D	64 <sup>t</sup>
WTAR-P	69 <sup>A</sup>	WTAR-P	71 <sup>A</sup>
WTAR-PD	67	WTAR-PD	69
Regression	66	Regression	71 <sup>A</sup>
HVLT-R		Stroop-W	
Barona	68 <sup>t</sup>	Barona	70
WTAR-D	67*	WTAR-D	72 <sup>A</sup>
WTAR-P	64**	WTAR-P	68
WTAR-PD	68 <sup>t</sup>	WTAR-PD	71
Regression	75 <sup>A</sup>	Regression	70
SDMT		Stroop-CW	
Barona	72	Barona	72
WTAR-D	67 <sup>t</sup>	WTAR-D	69 <sup>t</sup>
WTAR-P	69	WTAR-P	76 <sup>A</sup>
WTAR-PD	66*	WTAR-PD	73
Regression	74 <sup>A</sup>	Regression	73
TMT 1			
Barona	70 <sup>A</sup>		
WTAR-D	70 <sup>A</sup>		
WTAR-P	65		
WTAR-PD	70 <sup>A</sup>		
Regression	68		

<sup>t</sup> = proportion significantly different from reference at  $p < .10$ ; \* = proportion significantly different from reference at  $p < .05$ ; \*\* = proportion significantly different from reference at  $p < .01$ ; <sup>A</sup> = highest proportion of participants which was used as reference within analyses; WTAR = Wechsler Test of Adult Reading; WTAR-D = Full Scale IQ as predicted by demographics only according to the WTAR; WTAR-P = Full Scale IQ as predicted by WTAR performance; WTAR-PD = Full Scale IQ as predicted according to WTAR performance and demographics; Barona = Full Scale IQ as predicted by the Barona et al. (1984) equation; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial

Table 10.

Repeated Measures ANOVA Estimated Marginal Means for Post-MTBI Difference Scores

Factor	Vigil	BVMT-R	HVLT-R	SDMT	TMT 1	TMT 2	Digit Span	Stroop-W	Stroop-CW
Baseline	-1.27	-1.30	2.14	4.22	7.18	3.87	9.30	-14.54	8.58
Barona	-9.01*	-0.97	-0.47	1.89	3.41	-1.22	6.47	-7.98*	2.59
WTAR-D	-9.98*	-2.06	-1.86	0.79	2.31	-2.32	5.30	-9.31 <sup>†</sup>	1.27
WTAR-P	-10.71*	-3.22	-2.27	-0.37	1.15	-3.48	3.94	-10.42	0.57
WTAR-PD	-9.29 <sup>†</sup>	-2.17	-1.39	0.69	2.20	-2.42	5.53	-9.53	1.04
Regression	--	2.99	1.48	3.32	6.50	1.74	6.30	-6.72*	4.28

All difference score estimated marginal means are in standard score units; <sup>†</sup> = mean significantly different from baseline at  $p < .10$ ; \* = mean significantly different from baseline at  $p < .05$ ; MTBI = mild traumatic brain injury; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial

Table 11.

## Frequency of Participants who were Identified as Declined Post-MTBI

Measure	% Declined		% Declined
Vigil		TMT 2	
Baseline	22 <sup>A</sup>	Baseline	42 <sup>A</sup>
Barona	32	Barona	26
WTAR-D	37	WTAR-D	26
WTAR-P	42*	WTAR-P*	32
WTAR-PD	42*	WTAR-PD <sup>t</sup>	26
		Regression	42
BVMT-R		Digit Span	
Baseline	26 <sup>A</sup>	Baseline	24 <sup>A</sup>
Barona	32	Barona	58**
WTAR-D	26	WTAR-D	53**
WTAR-P	26	WTAR-P	47*
WTAR-PD	37	WTAR-PD	47*
Regression	16	Regression	53**
HVLT-R		Stroop-W	
Baseline	37 <sup>A</sup>	Baseline	21 <sup>A</sup>
Barona	37	Barona	61***
WTAR-D	37	WTAR-D	67***
WTAR-P	21	WTAR-P	50**
WTAR-PD	21	WTAR-PD	55**
Regression	41	Regression	44*
SDMT		Stroop-CW	
Baseline	32 <sup>A</sup>	Baseline	33 <sup>A</sup>
Barona	37	Barona	39
WTAR-D	53*	WTAR-D	44
WTAR-P	42	WTAR-P	39
WTAR-PD	32	WTAR-PD	50
Regression	42	Regression	44
TMT 1			
Baseline	37 <sup>A</sup>		
Barona	21		
WTAR-D	26		
WTAR-P	32		
WTAR-PD	21		
Regression	47		

<sup>t</sup> = proportion significantly different from reference at  $p < .10$ ; \* = proportion significantly different from reference at  $p < .05$ ; \*\* = proportion significantly different from reference at  $p < .01$ ; <sup>A</sup> = highest proportion of participants which was used as reference within analyses; MTBI = mild traumatic brain injury; WTAR = Wechsler Test of Adult Reading; WTAR-D = Full Scale IQ as predicted by demographics only according to the WTAR; WTAR-P = Full Scale IQ as predicted by WTAR performance; WTAR-PD = Full Scale IQ as predicted according to WTAR performance and demographics; Barona = Full Scale IQ as predicted by the Barona et al. (1984) equation; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial

Table 12.

## The Best Baseline Estimators for Sport-Related MTBI

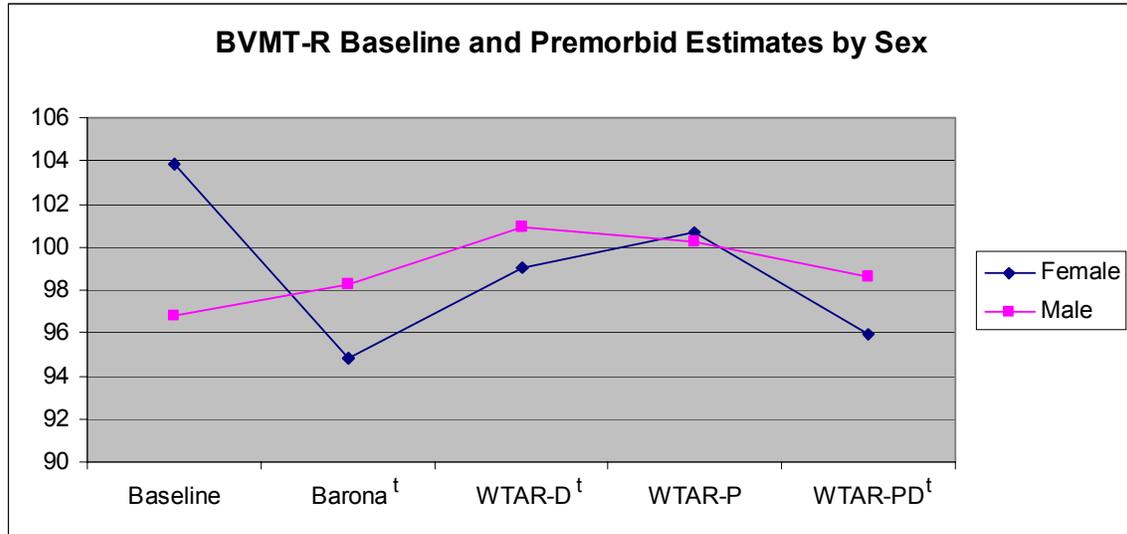
Measure	Best Estimate	Estimation Basis
All Measures (Using Original Premorbid Estimates)	WTAR-P	Reading-Based
Individual Baseline Measures		
Vigil	Barona	Demographic-Based
BVMT-R	Stepwise Regression (Race and Barona)	Demographic-Based
HVLT-R	Stepwise Regression (Sport and Race)	Demographic-Based
SDMT	Stepwise Regression (Race and Sex)	Demographic-Based
TMT 1	Stepwise Regression (WTAR-D)	Demographic-Based
TMT 2	Stepwise Regression (Race)	Demographic-Based
Digit Span	WTAR-P	Reading-Based
Stroop-W	WTAR-P	Reading-Based
Stroop-CW	WTAR-P	Reading-Based

WTAR = Wechsler Test of Adult Reading; WTAR-D = Full Scale IQ as predicted by demographics only according to the WTAR; WTAR-P = Full Scale IQ as predicted by WTAR performance; Barona = Full Scale IQ as predicted by the Barona et al. (1984) equation; Vigil = Vigil Reaction Time; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning; HVLT-R = Hopkins Verbal Learning Test-Revised Recall Total Learning; SDMT = Symbol Digit Modalities Test Total Correct; TMT = Trail Making Test; Digit Span = Digit Span Test Total Correct; Stroop-W = Stroop World Only Trial Time; Stroop-CW = Stroop Color-Word Trial.

## FIGURES

Figure 1.

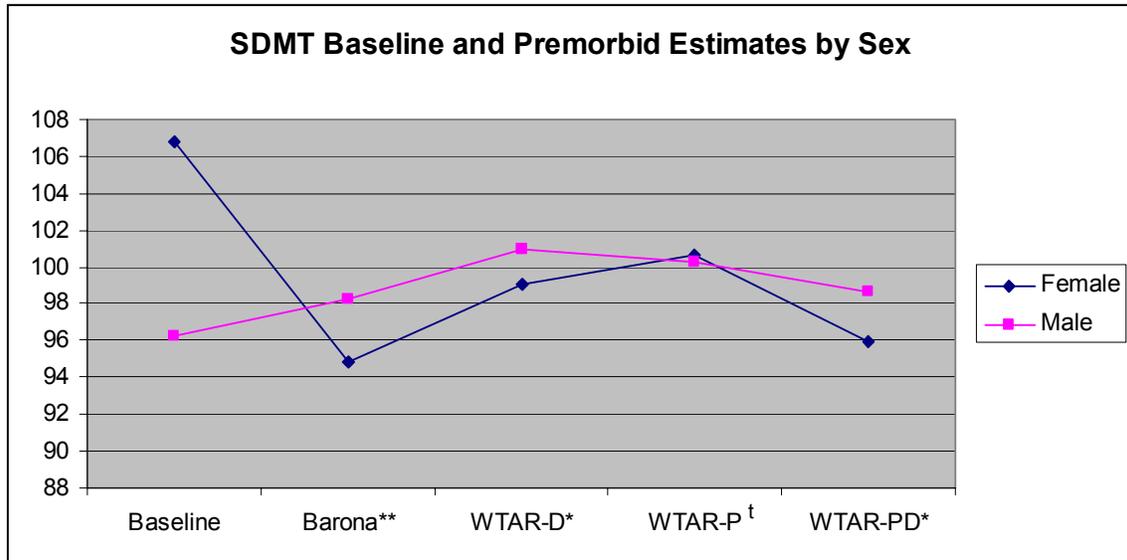
BVMT-R Baseline and Premorbid Estimates Estimated Marginal Means by Sex



All estimated marginal means are in standard score units; <sup>t</sup> = mean significantly different from baseline at  $p < .10$ ; BVMT-R = Brief Visuospatial Memory Test-Revised Total Learning

Figure 2.

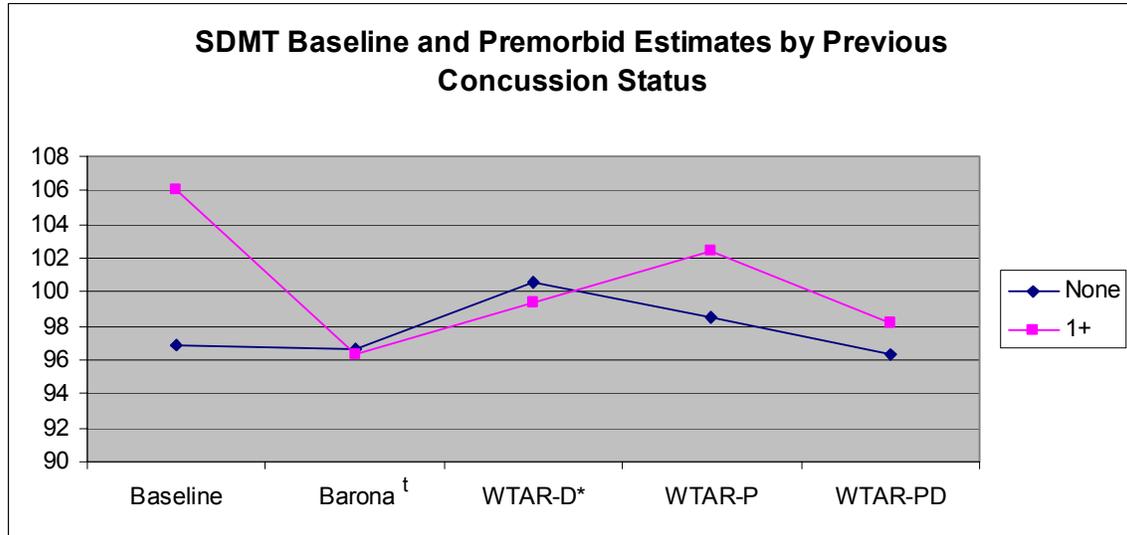
SDMT Baseline and Premorbid Estimates Estimated Marginal Means by Sex



All estimated marginal means are in standard score units; † = mean significantly different from baseline at  $p < .10$ ; \* = mean significantly different from baseline at  $p < .05$ ; \*\* = mean significantly different from baseline at  $p < .01$ ; SDMT = Symbol Digit Modalities Test Total Correct

Figure 3.

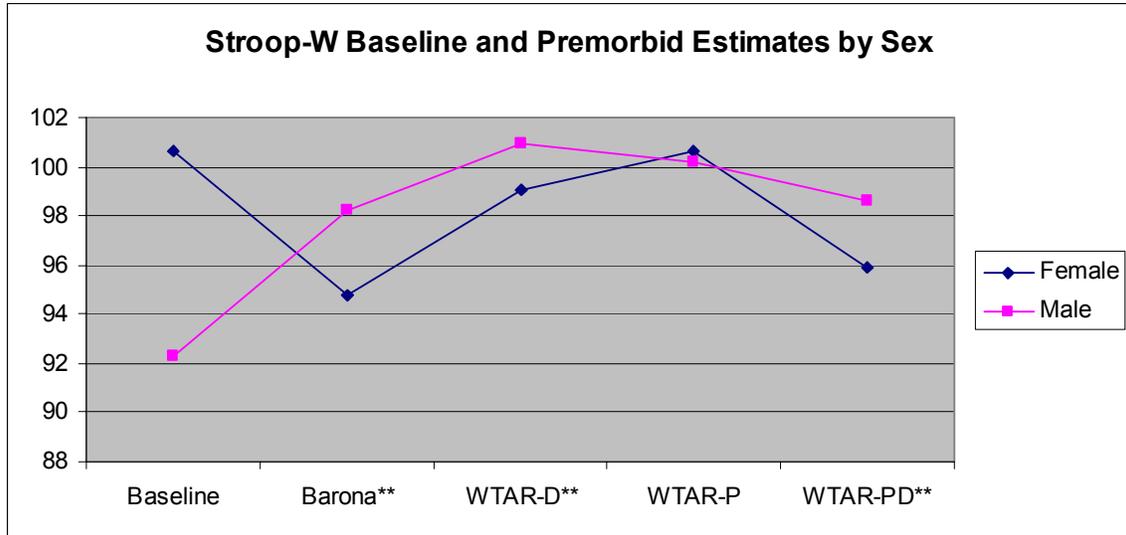
SDMT Baseline and Premorbid Estimates Estimated Marginal Means by Previous Concussion Status



All estimated marginal means are in standard score units; <sup>t</sup> = mean significantly different from baseline at  $p < .10$ ; \* = mean significantly different from baseline at  $p < .05$ ; SDMT = Symbol Digit Modalities Test Total Correct

Figure 4.

Stroop-W Baseline and Premorbid Estimates Estimated Marginal Means by Sex



All estimated marginal means are in standard score units; \*\* = mean significantly different from baseline at  $p < .01$ ; Stroop-W = Stroop World Only Trial Time

## Appendix A.

### Concussion Grading System Criteria

Concussion Grade	Cantu (2001)	Colorado Medical Society (1991)	American Academy of Neurology (1997)
Grade 1 (Mild)	<ul style="list-style-type: none"> <li>- No LOC</li> <li>- Either PTA or Post-concussion signs that last less than 30 minutes</li> </ul>	<ul style="list-style-type: none"> <li>- Transient mental confusion</li> <li>- No LOC</li> <li>- No PTA</li> </ul>	<ul style="list-style-type: none"> <li>- Transient confusion</li> <li>- No LOC</li> <li>- Symptoms or abnormalities clear in less than 15 minutes</li> </ul>
Grade 2 (Moderate)	<ul style="list-style-type: none"> <li>- LOC of less than 1 minute and PTA <i>or</i></li> <li>- Post-concussion signs or symptoms that last longer than 30 minutes but less than 24 hours</li> </ul>	<ul style="list-style-type: none"> <li>- No LOC</li> <li>- Confusion with PTA</li> </ul>	<ul style="list-style-type: none"> <li>- No LOC</li> <li>- Symptoms or abnormalities that last longer than 15 minutes</li> </ul>
Grade 3 (Severe)	<ul style="list-style-type: none"> <li>- LOC of greater than 1 minute <i>or</i></li> <li>- PTA lasting longer than 24 hours <i>or</i></li> <li>- Post-concussion signs or symptoms that last longer than 7 days</li> </ul>	<ul style="list-style-type: none"> <li>- Any LOC, no matter how brief</li> </ul>	<ul style="list-style-type: none"> <li>- Any LOC, whether brief (seconds) or long (minutes)</li> </ul>

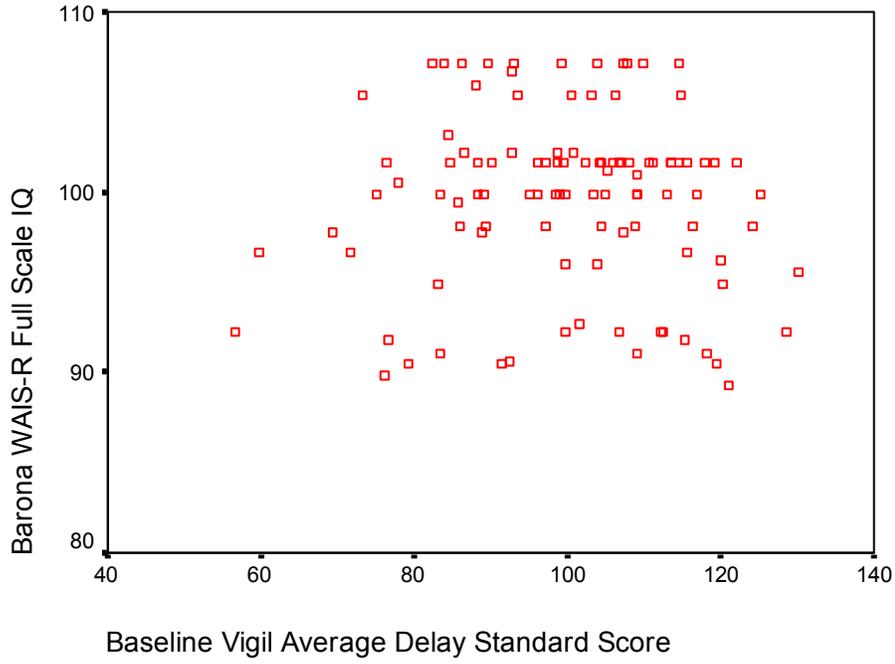
---

LOC = Loss of Consciousness; PTA = Posttraumatic amnesia; American Academy of Neurology (1997). Practice parameter: The management of concussion in sports (summary statement). *Neurology*, 48, 581-585; Cantu, R.C. (2001). Posttraumatic Retrograde and Anterograde Amnesia: Pathophysiology and Implications in Grading And Safe Return To Play. *Journal of Athletic Training*. 36, 244-248; Colorado Medical Society (1991). *Report of the sports medicine committee: Guidelines for the management of concussion in sports (revised)*. Colorado Medical Society, Denver.

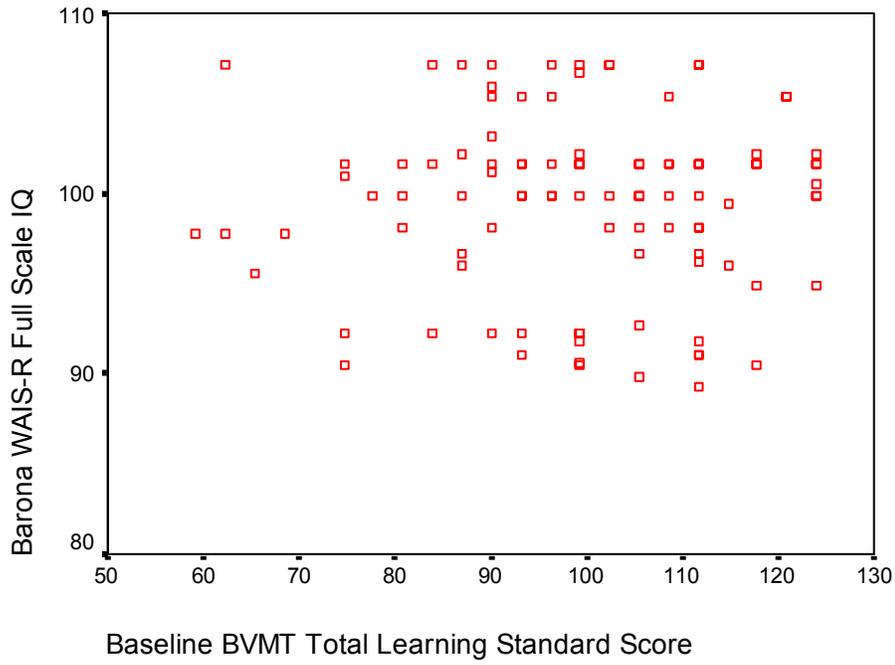
Appendix B.

Barona Baseline Scatterplots

Barona IQ and Vigil Scatterplot



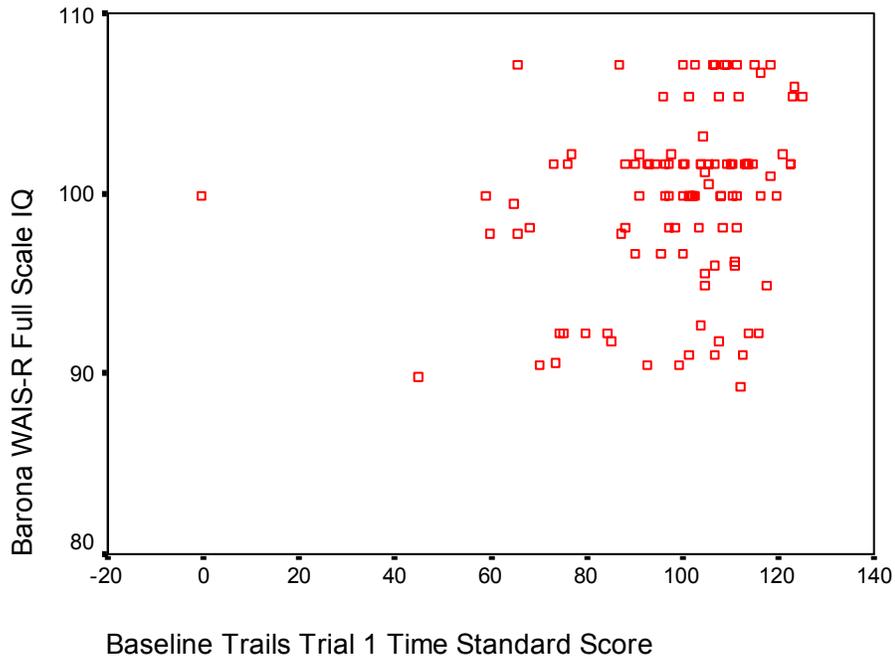
### Barona IQ and BVMT Scatterplot



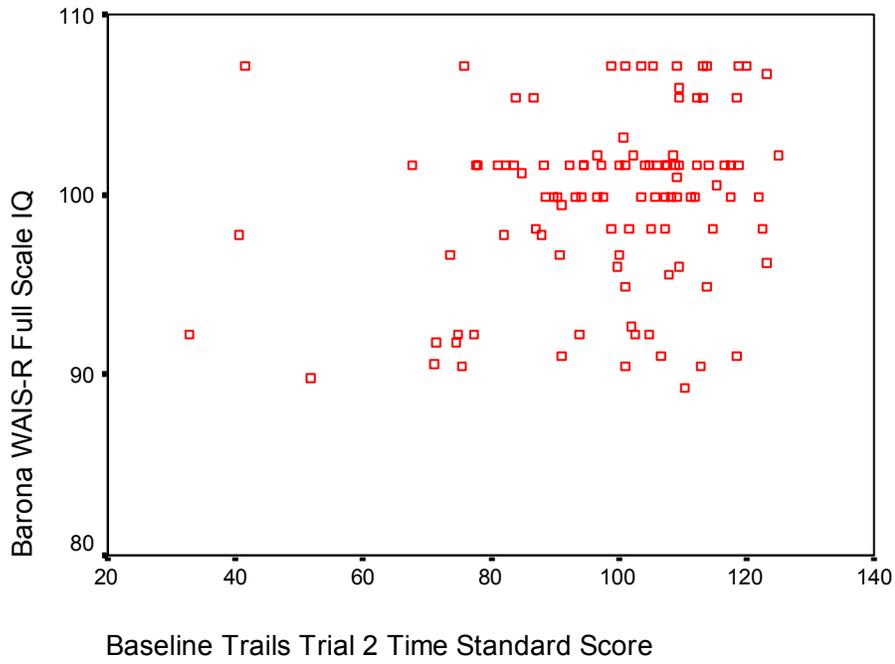
### Barona IQ and SDMT Scatterplot



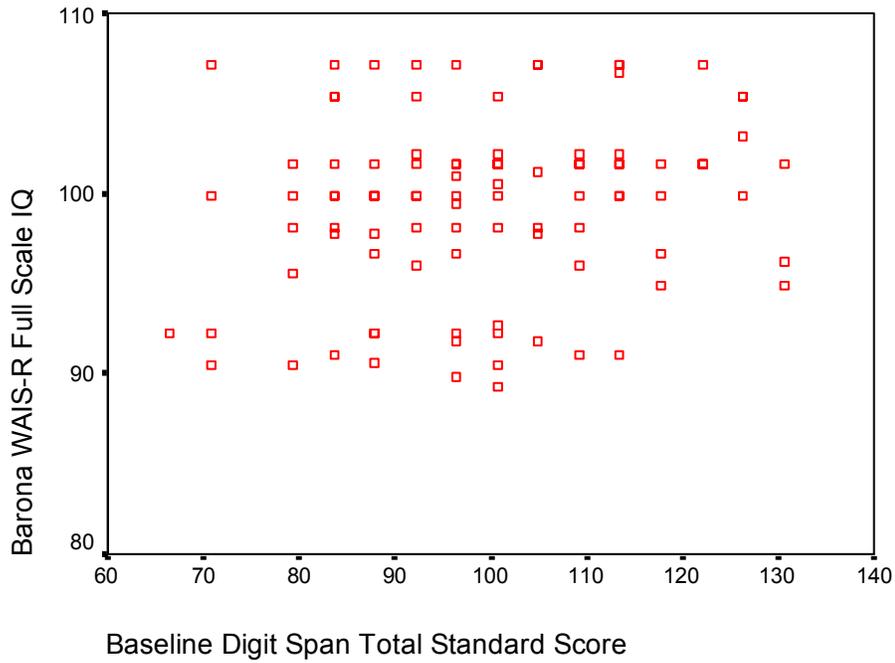
### Barona IQ and TMT Trial 1 Scatterplot



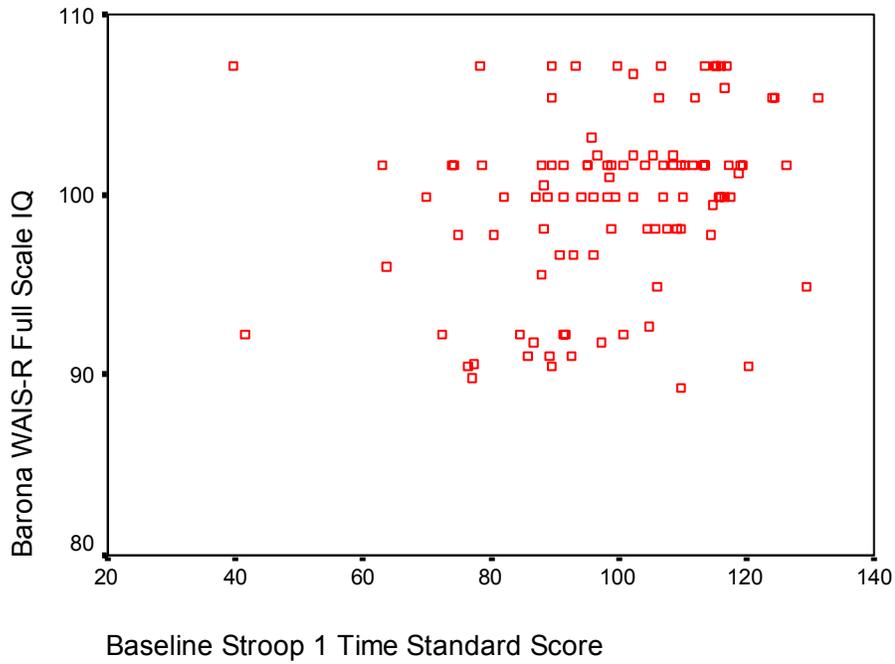
### Barona IQ and TMT Trial 2 Scatterplot



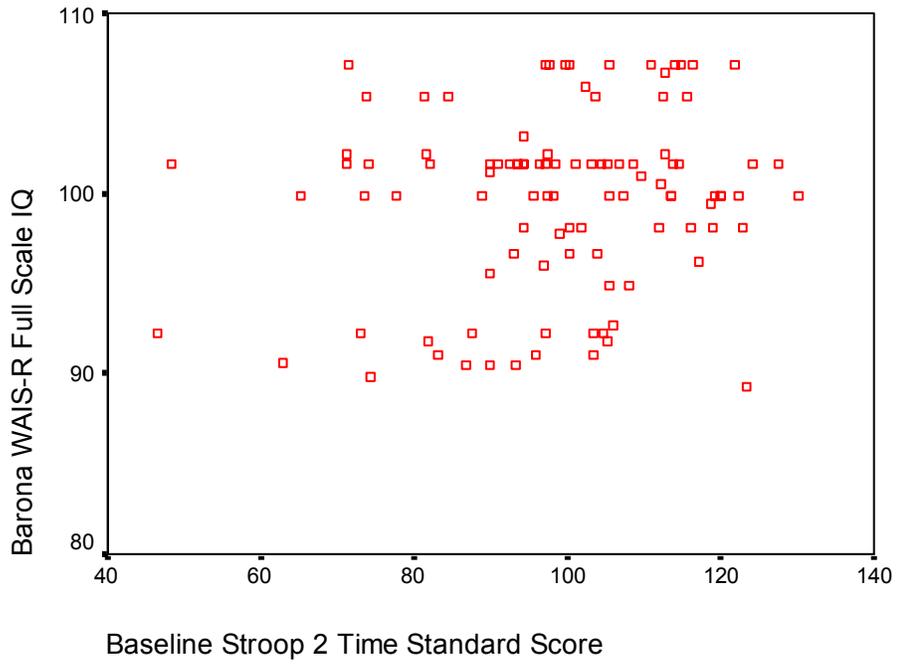
### Barona IQ and Digit Span Total Scatterplot



### Barona IQ and Stroop 1 Scatterplot



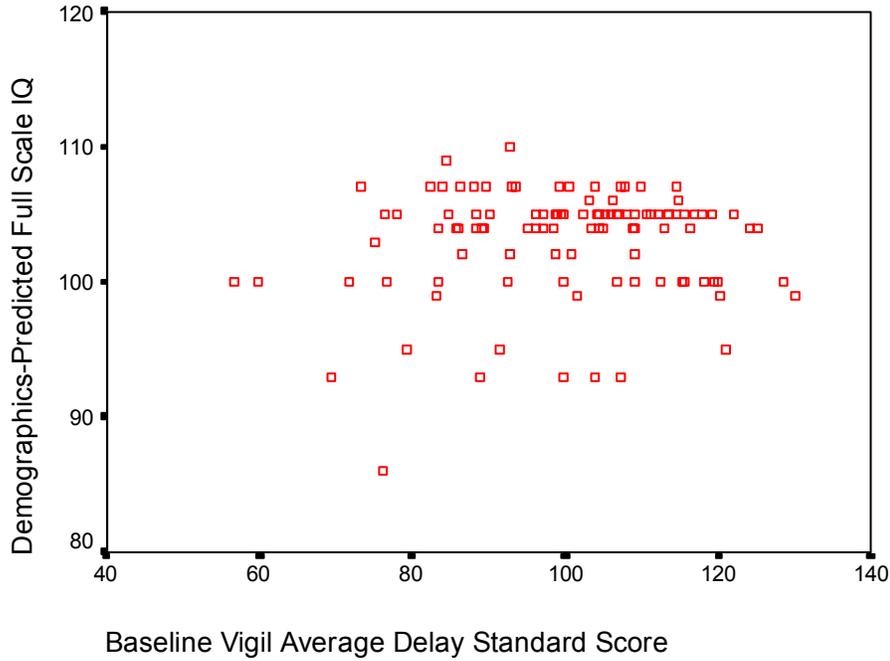
Barona IQ and Stroop 2 Scatterplot



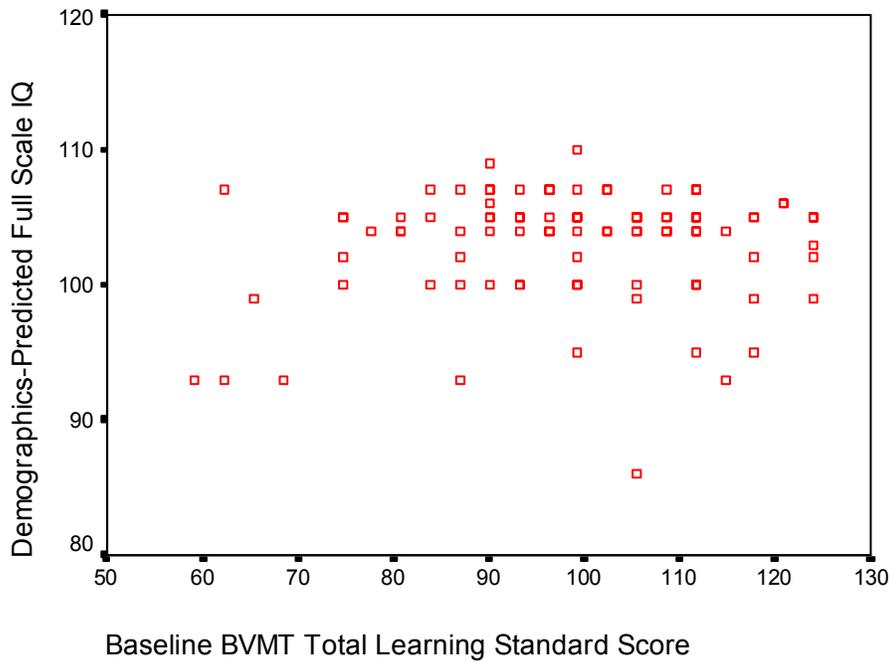
Appendix C.

WTAR-Demographics Only Baseline Scatterplots

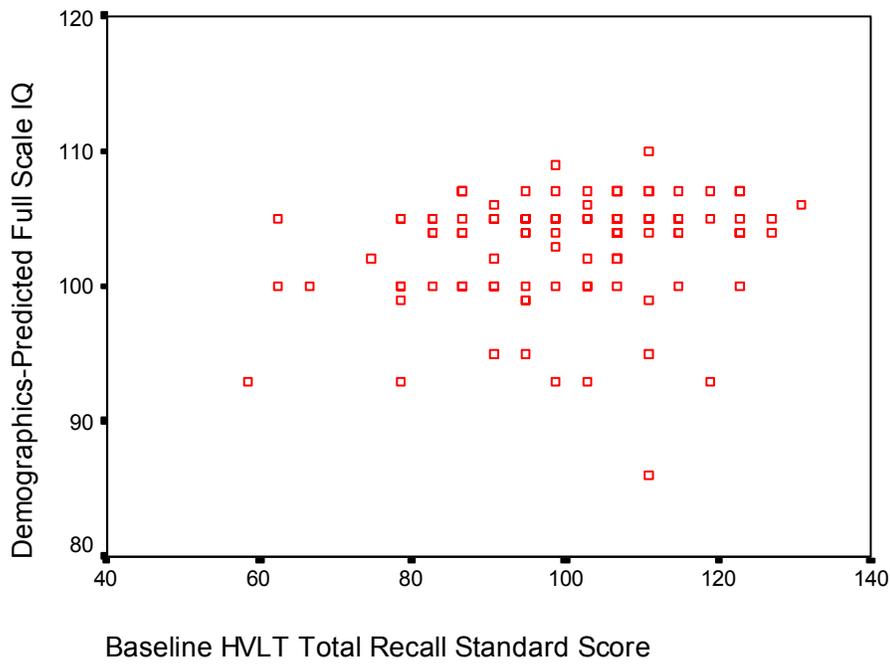
WTAR-D IQ and Vigil Scatterplot



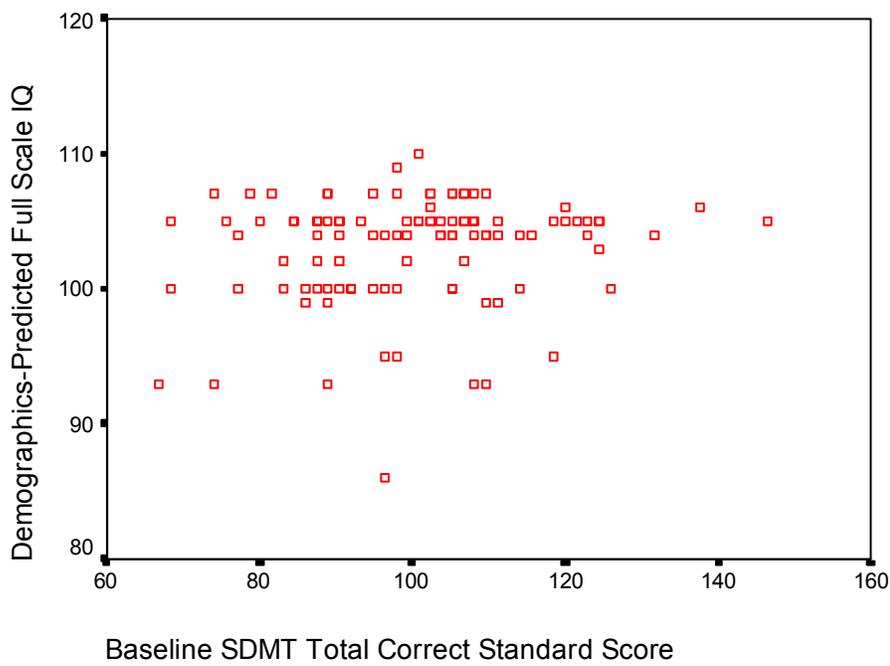
WTAR-D IQ and BVMT Scatterplot



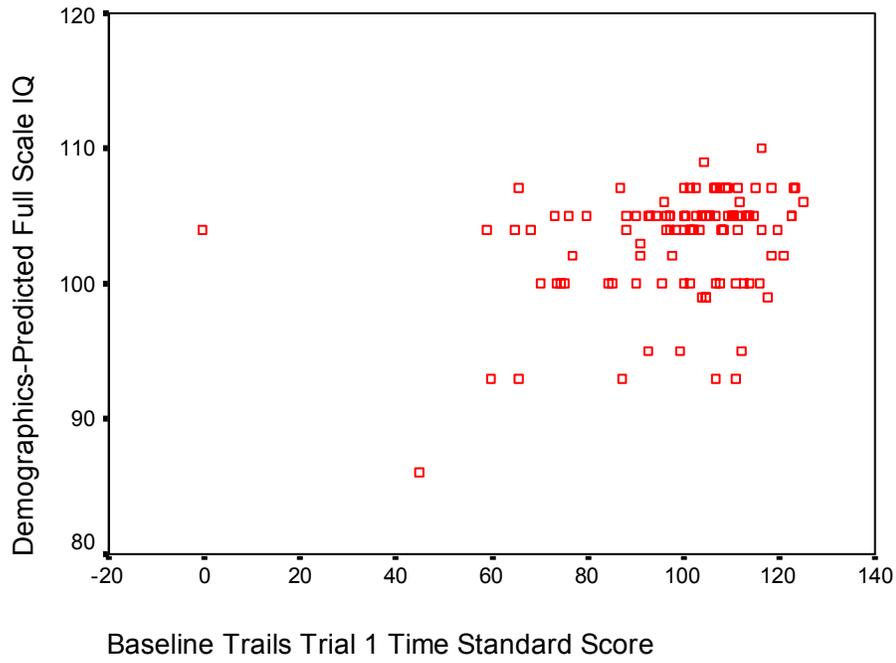
### WTAR-D IQ and HVLT Scatterplot



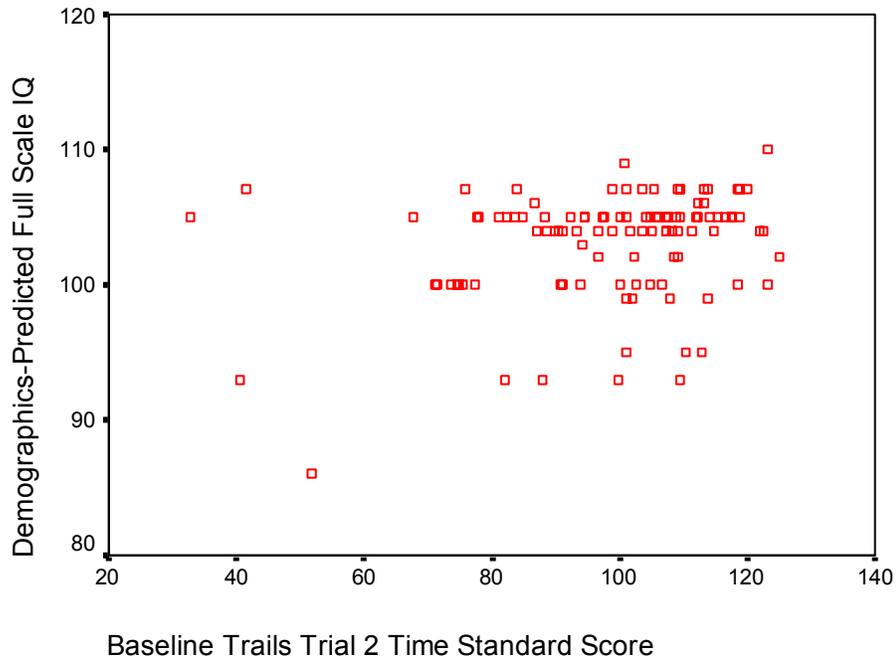
### WTAR-D IQ and SDMT Scatterplot



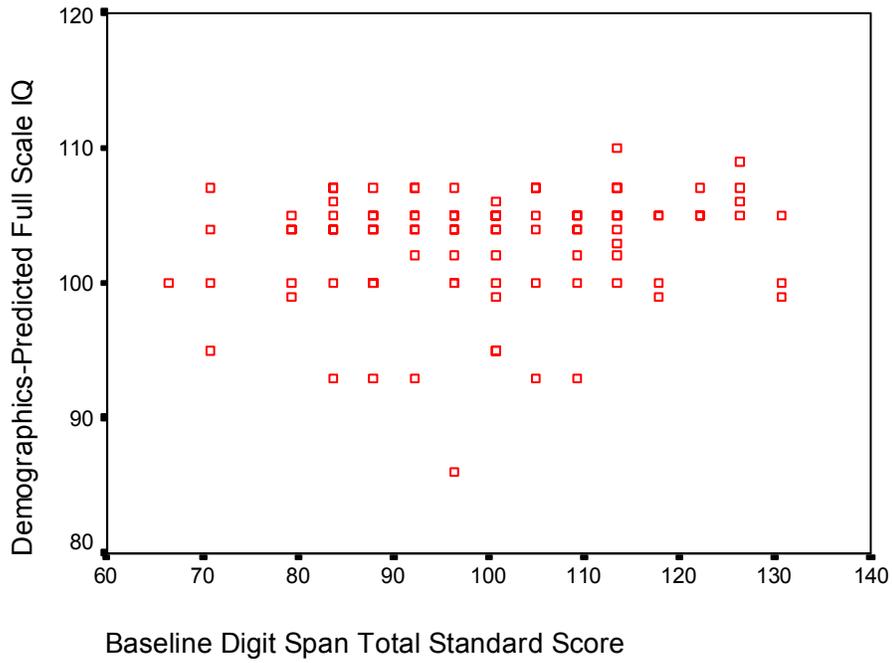
### WTAR-D IQ and TMT Trial 1 Scatterplot



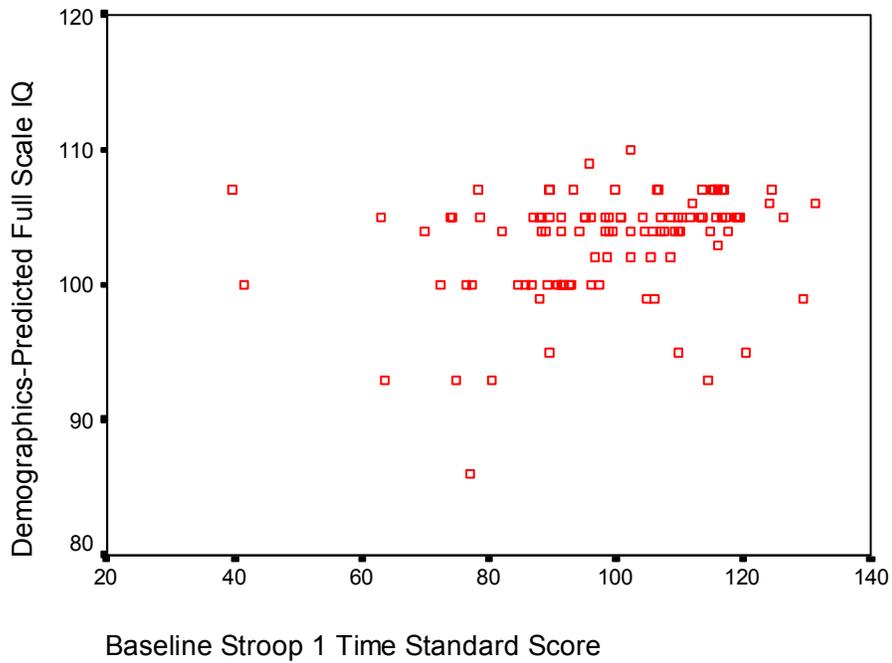
### WTAR-D IQ and TMT Trial 2 Scatterplot



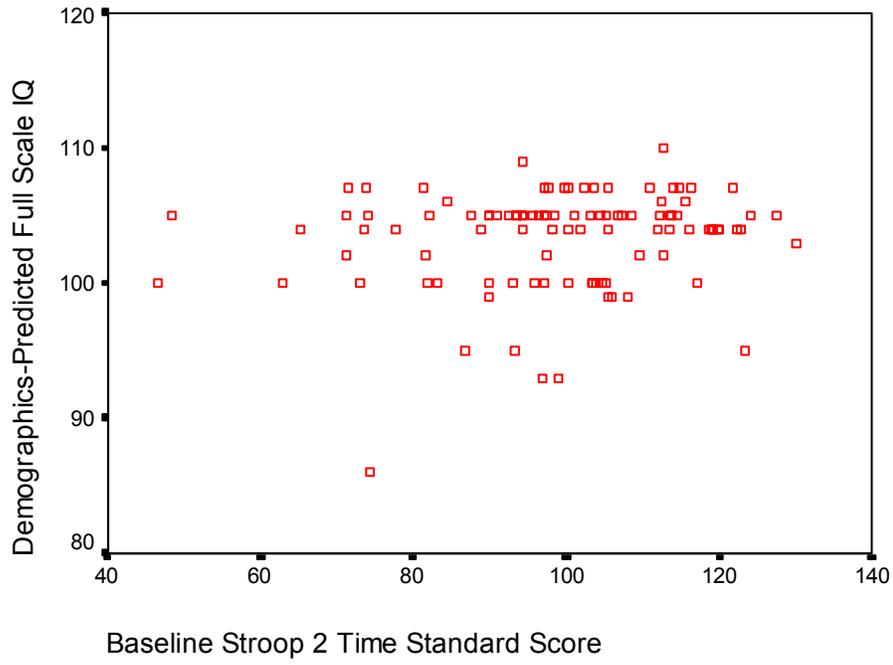
### WTAR-D IQ and Digit Span Total Scatterplot



### WTAR-D IQ and Stroop 1 Scatterplot



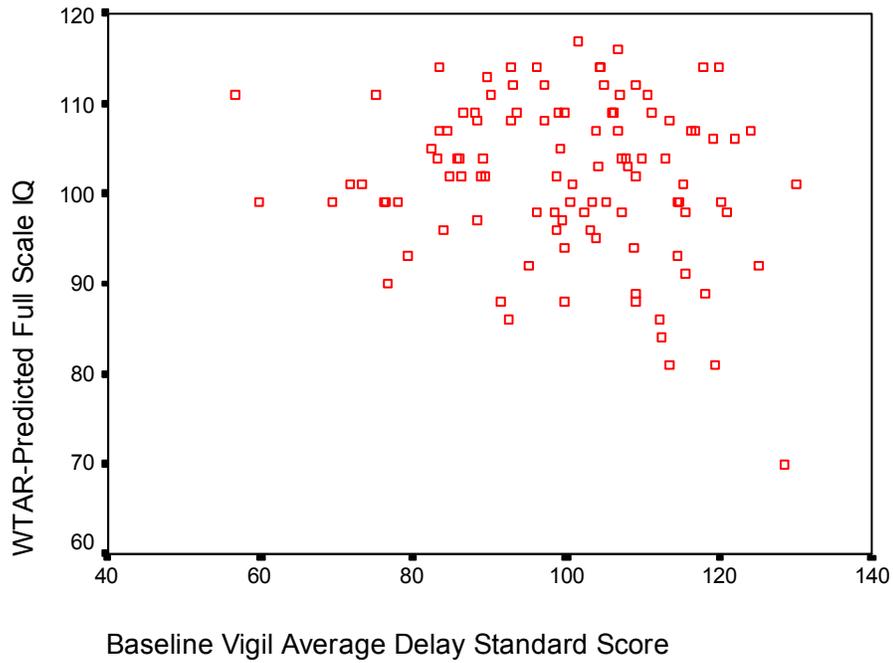
WTAR-D IQ and Stroop 2 Scatterplot



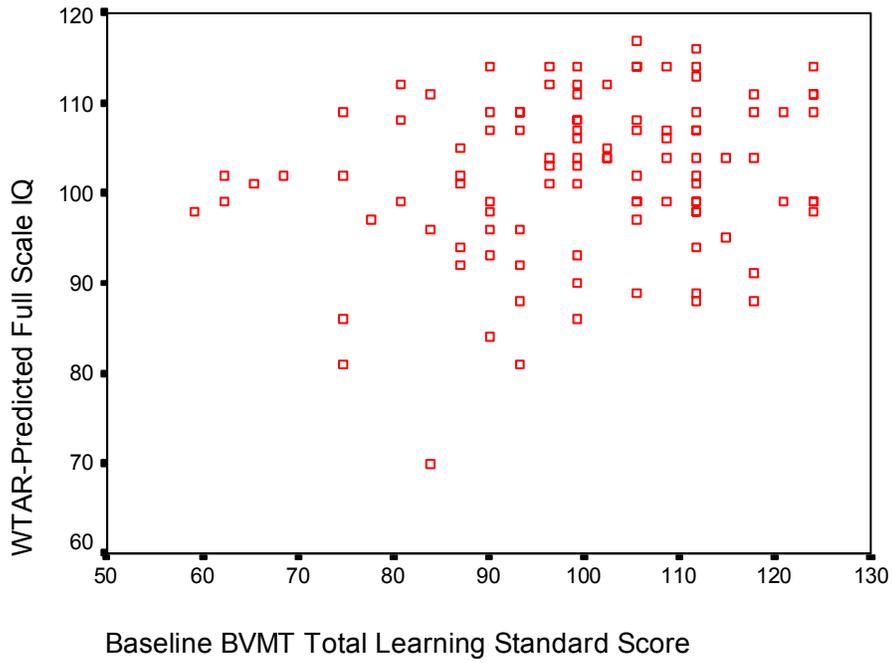
Appendix D.

WTAR-Performance Only Baseline Scatterplots

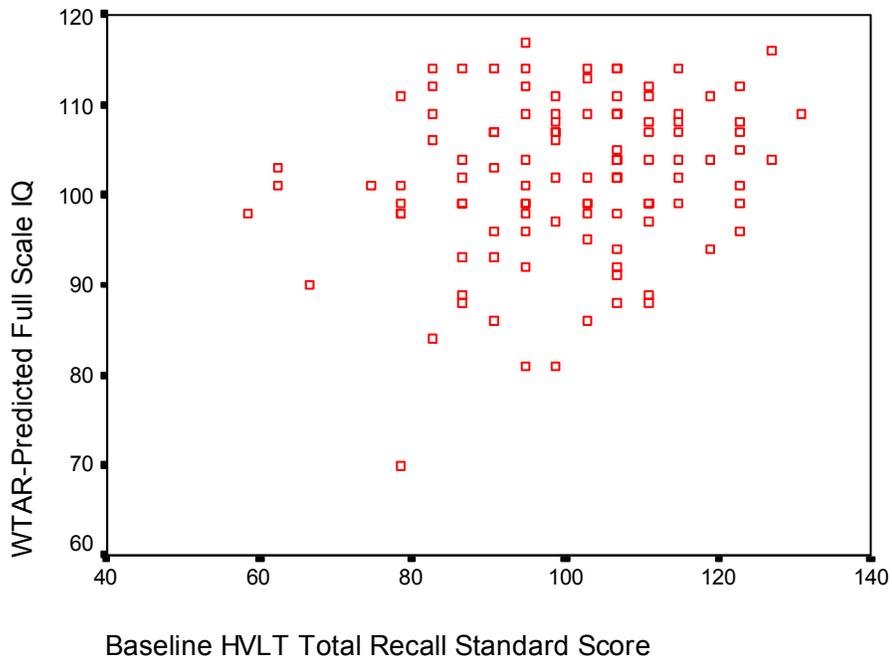
WTAR-P IQ and Vigil Scatterplot



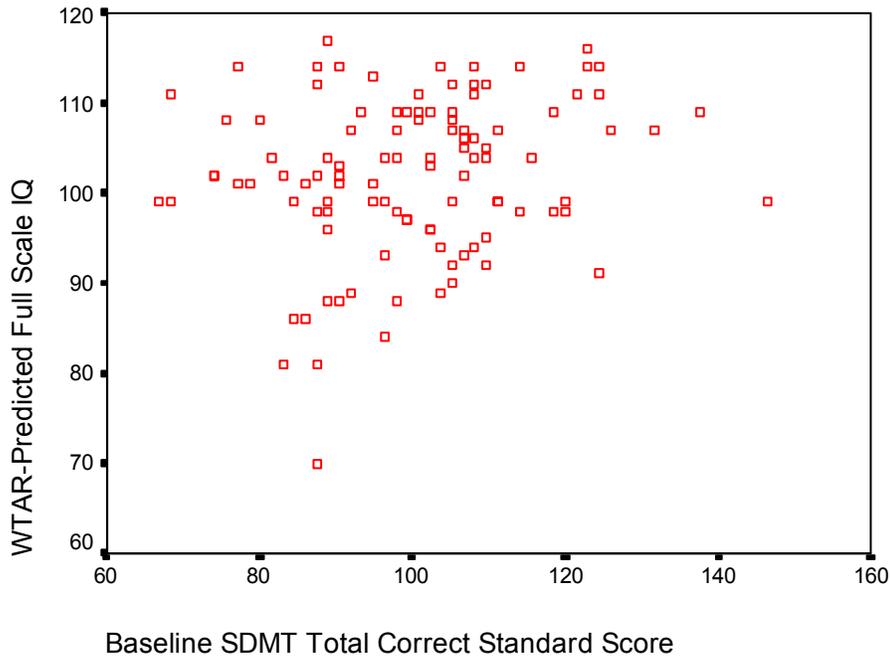
### WTAR-P IQ and BVMT Scatterplot



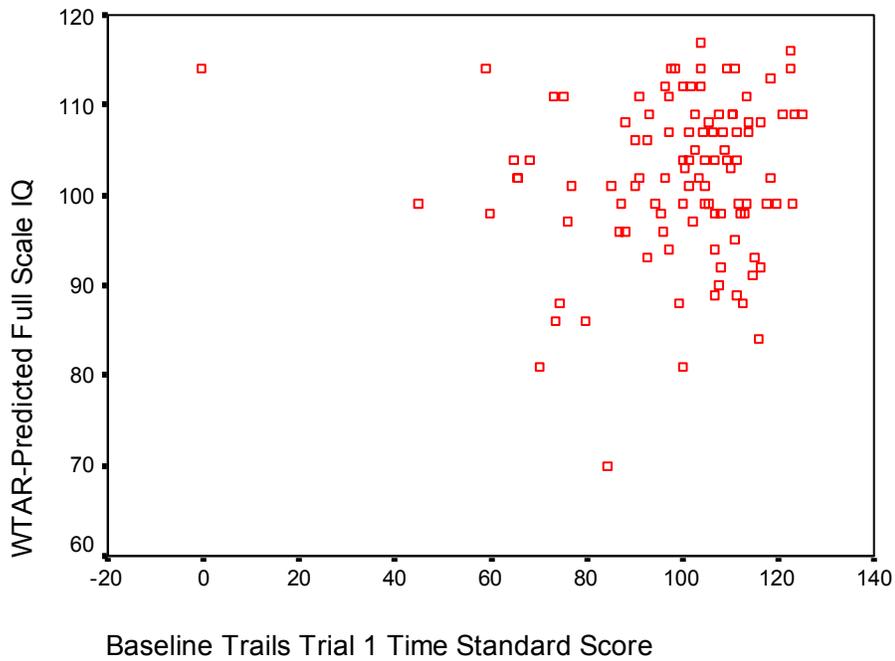
### WTAR-P IQ and HVLN Scatterplot



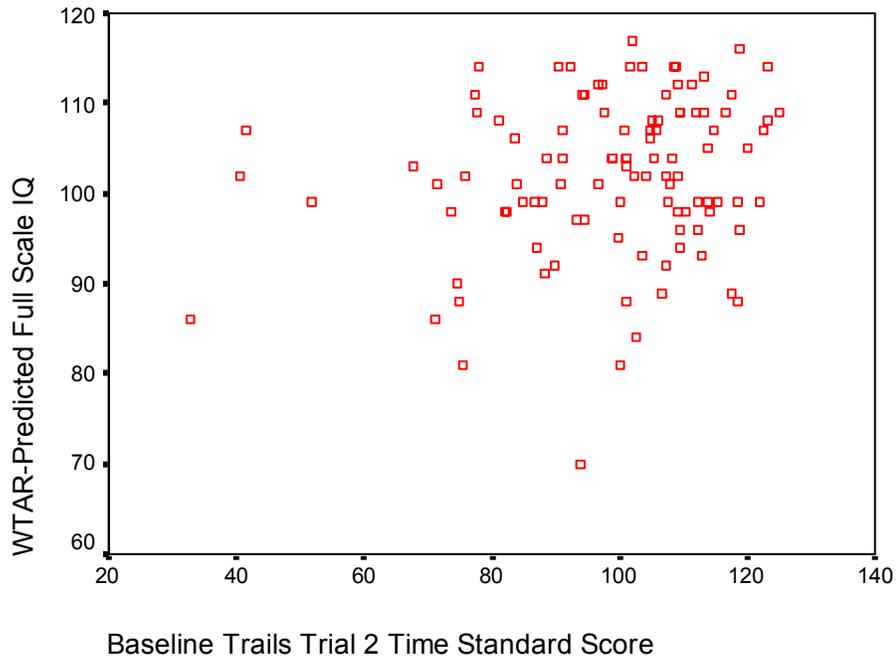
### WTAR-P IQ and SDMT Scatterplot



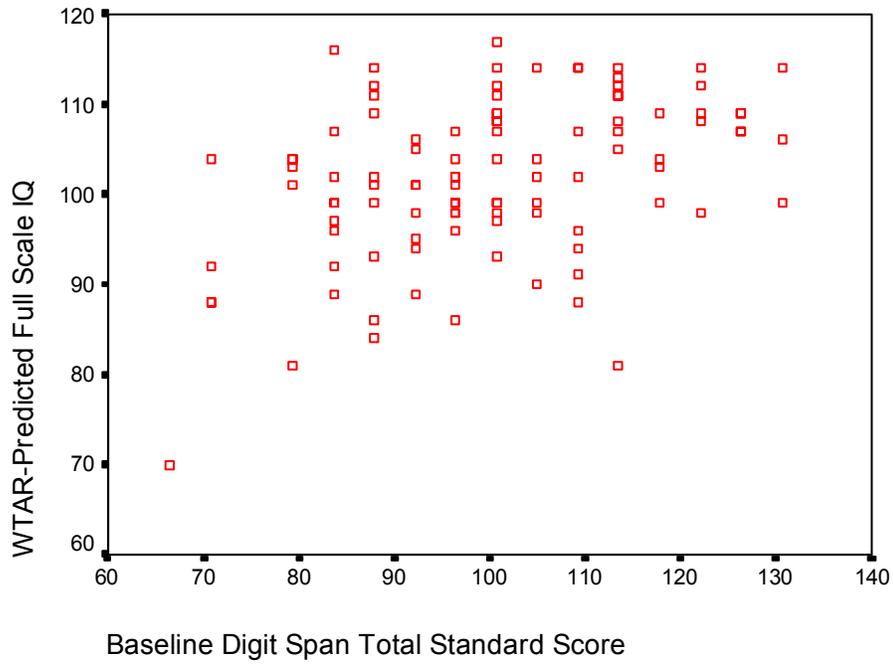
### WTAR-P IQ and TMT Trial 1 Scatterplot



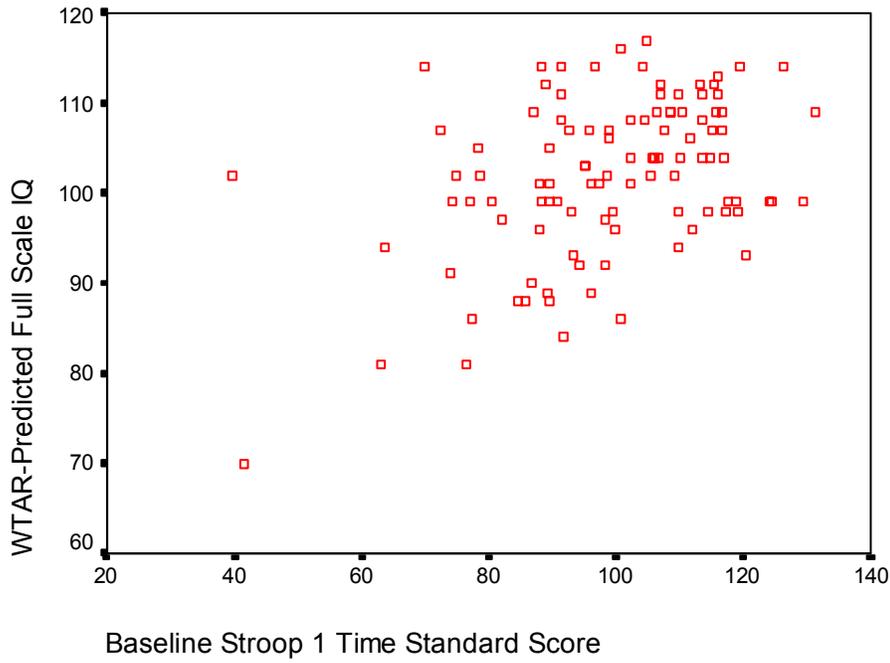
### WTAR-P IQ and TMT Trial 2 Scatterplot



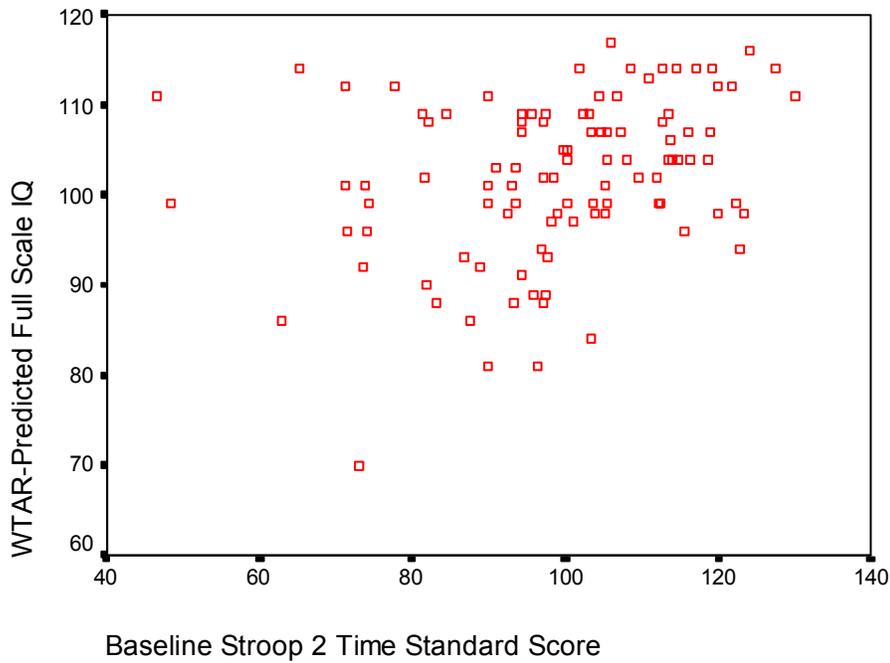
### WTAR-P IQ and Digit Span Total Scatterplot



### WTAR-P IQ and Stroop 1 Scatterplot

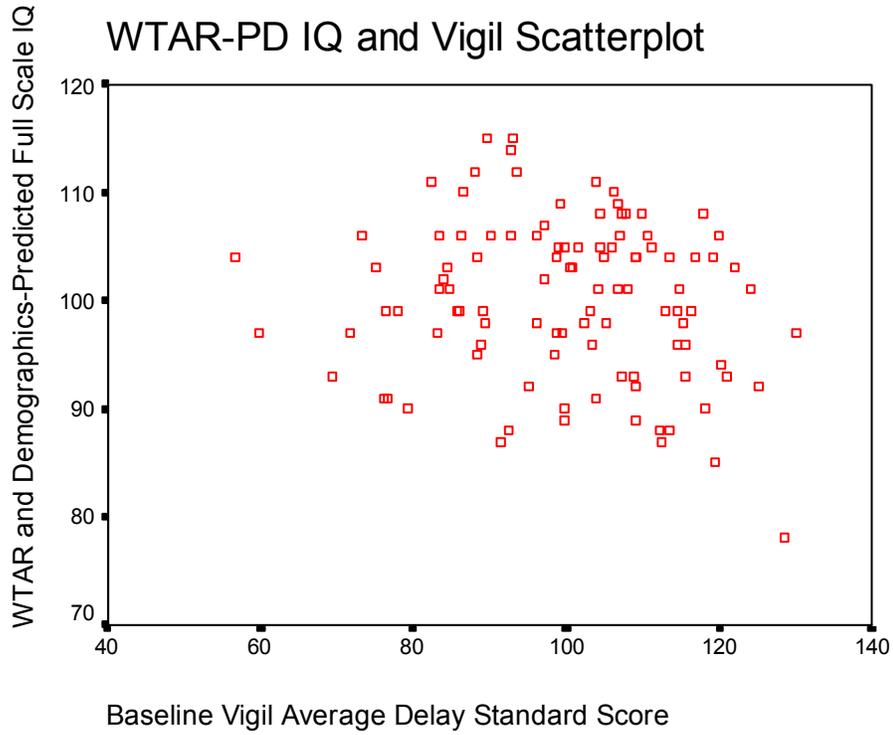


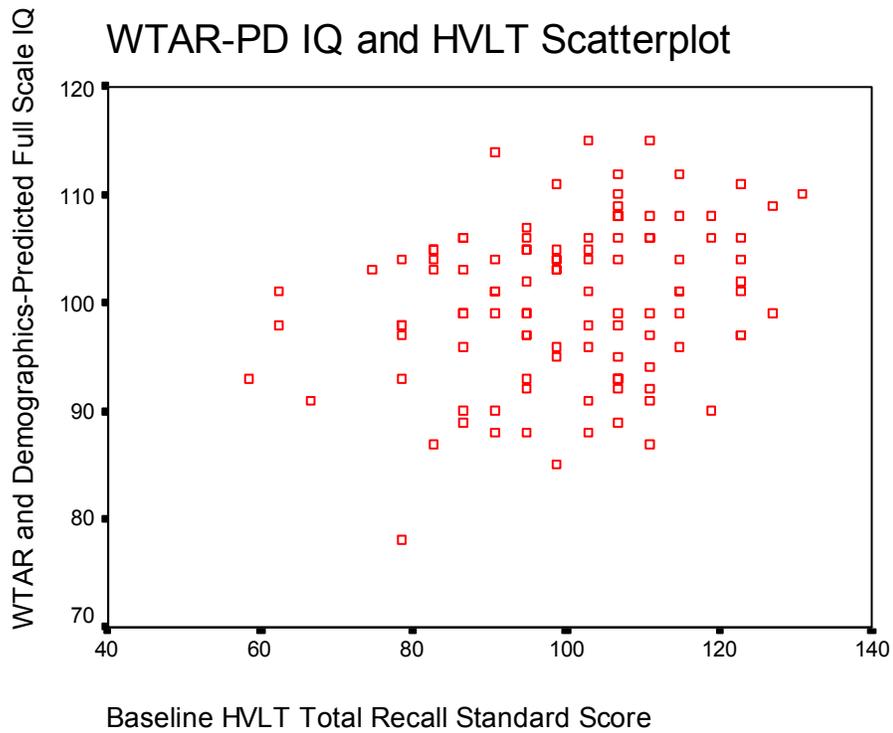
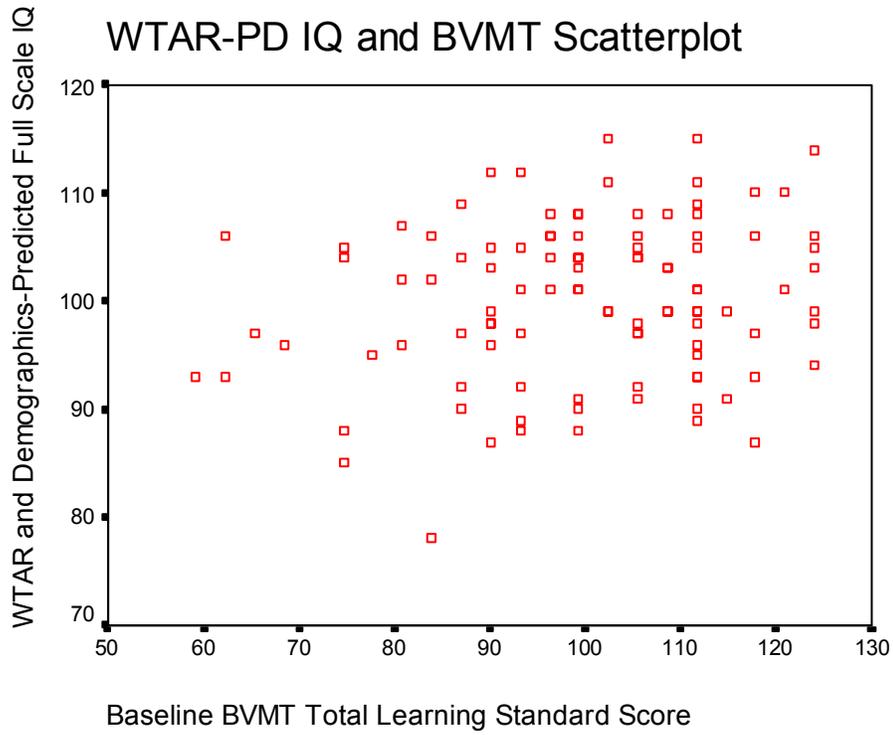
### WTAR-P IQ and Stroop 2 Scatterplot

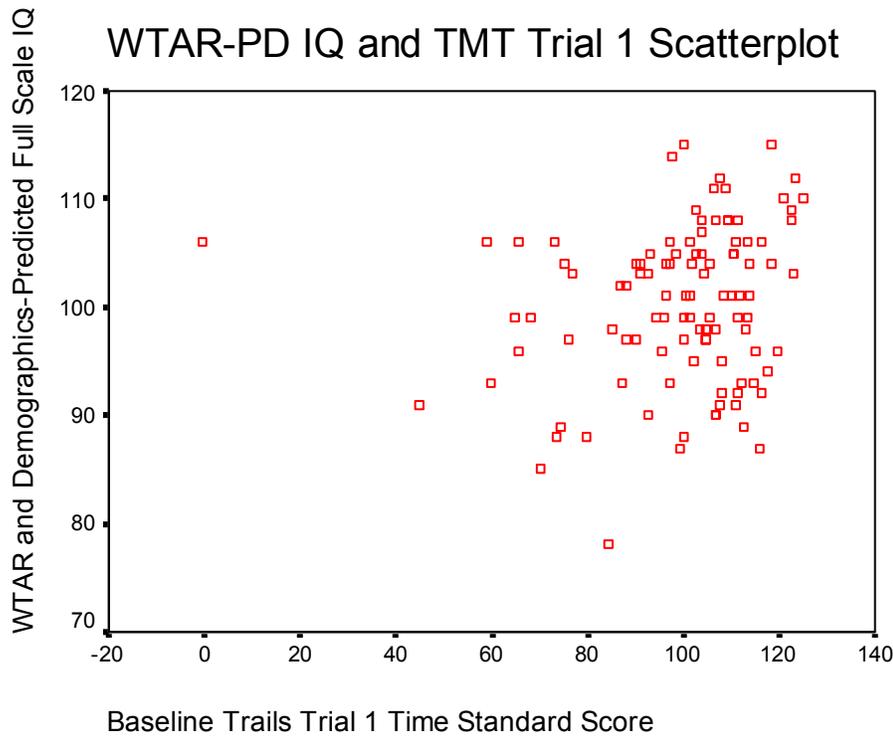
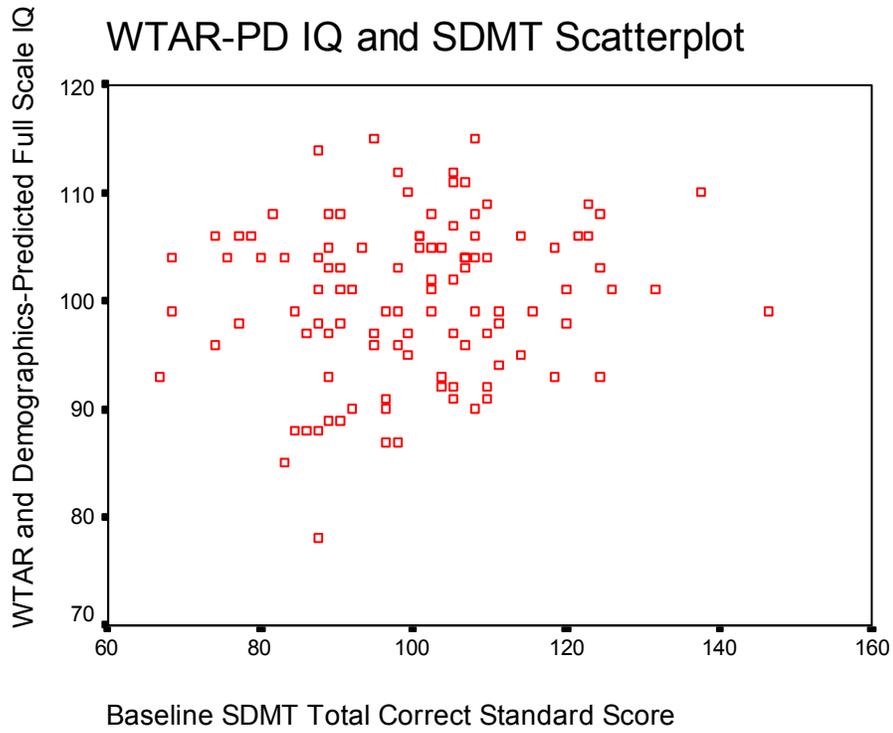


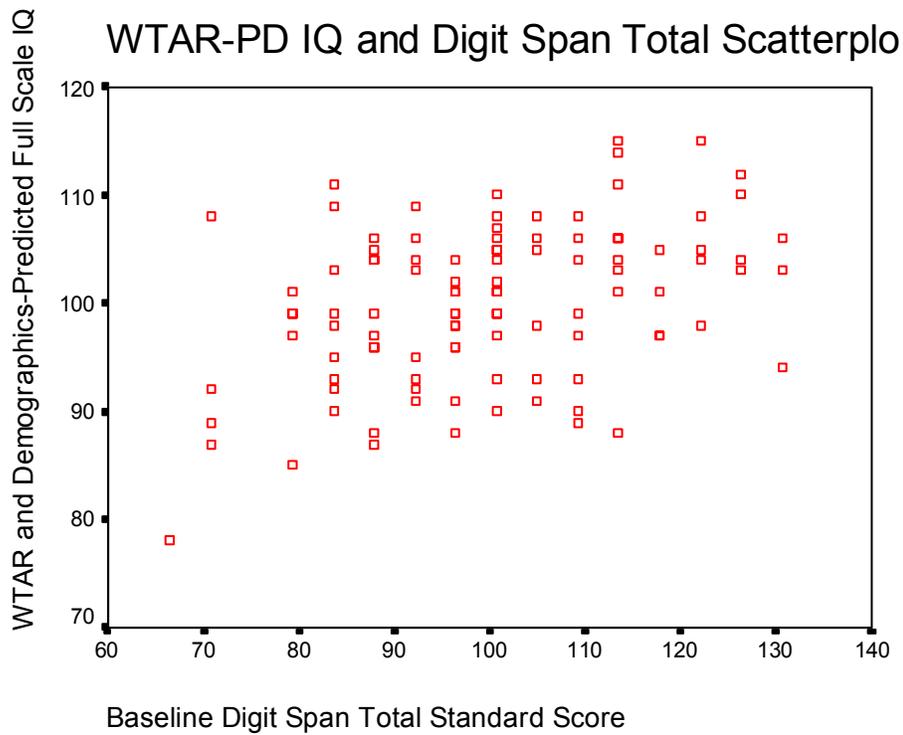
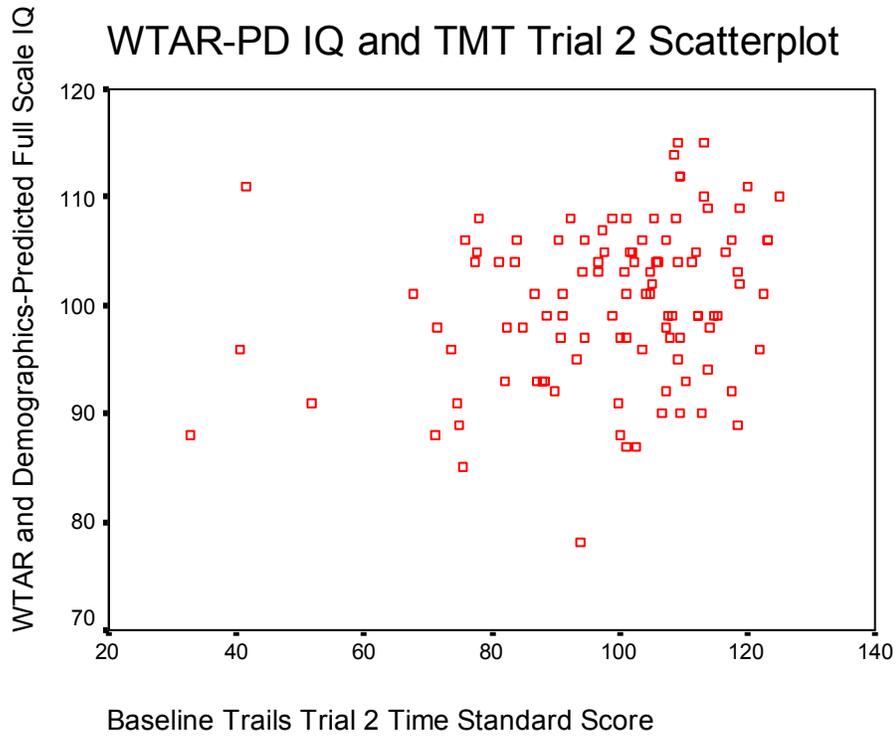
Appendix E.

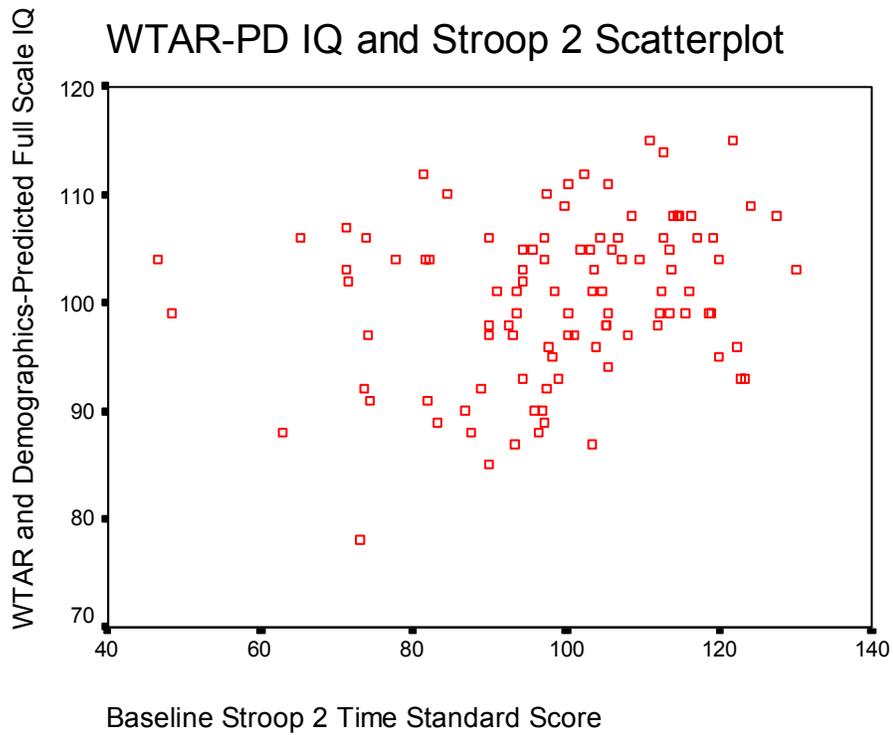
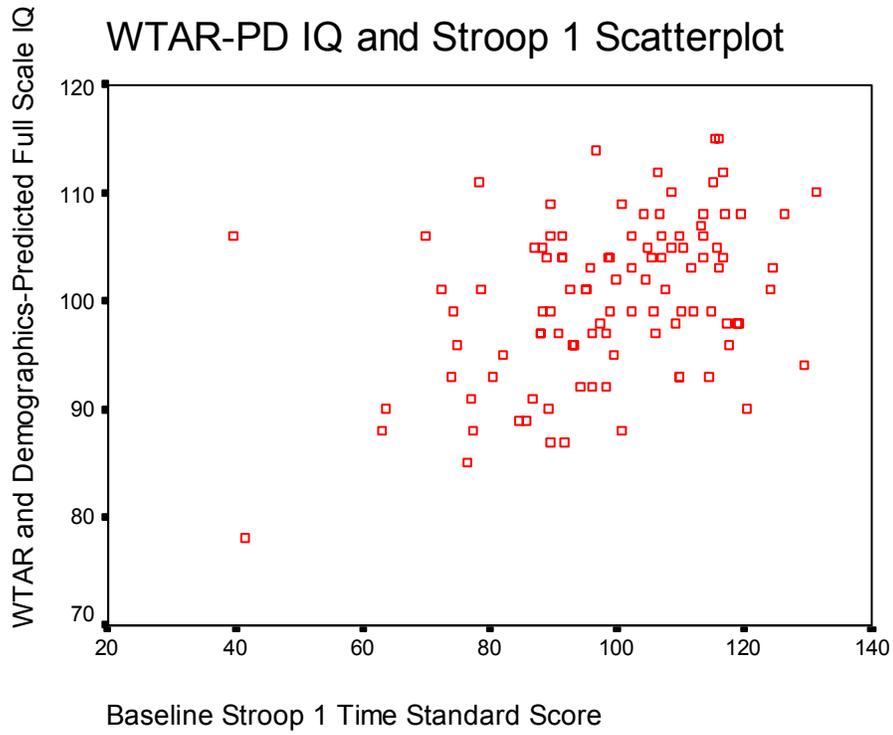
WTAR-Performance & Demographics Baseline Scatterplots











## VITA - Christopher M. Bailey

### **Educational History**

Clinical Psychology Internship – Neuropsychology Track; University of North Carolina; Chapel Hill, NC; August 2007

Ph.D.: Psychology; The Pennsylvania State University; University Park, PA; August 2007; GPA: 3.98.

M.S.: Psychology; The Pennsylvania State University; University Park, PA; August 2003; GPA: 3.97.

B.A.: Psychology; Ohio University; Athens, OH; June 2001; Summa Cum Laude; GPA: 3.93.

### **Honors and Awards**

National Academy of Neuropsychologists (NAN) Student Research Award, 2005.

Teaching Fellowship, 2005.

National Academy of Neuropsychologists (NAN) Student Research Award, 2004.

Gaige Paulsen Scholarship, 2000.

Dean's Scholarship, 1998-2000 (all years eligible).

### **Publications**

Bailey, C. M. & Arnett, P.A. (2006). Motivation and the Assessment of Sports-Related Concussion. In Slobounov, S. & Sebastainelli, W. (eds.) *Foundation of Sports-Related Concussion*. New York: Springer Publishing Co., 171-193.

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, 475-484.

Rosenbaum, A.M., Arnett, P.A., Bailey, C.M., & Echemendia, R.J. (2006). Neuropsychological Assessment of Sports-Related Concussion: Measuring Clinically Significant Change. In Slobounov, S. & Sebastainelli, W. (eds.) *Foundation of Sports-Related Concussion*. New York: Springer Publishing Co., 137-169.