2013-05-14

The State of the Research: Meta-Analysis and Conceptual Critique of Mild Traumatic Brain Injury

Ryan Lance Nelson
Brigham Young University - Provo

The State of the Research: Meta-Analysis and Conceptual Critique of Mild Traumatic Brain Injury

Ryan R. Green

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Erin D. Bigler, Chair
Alan J. Hawkins
Dawson W. Hedges
Michael J. Larson
Mikle D. South

Department of Psychology

Brigham Young University

May 2013

ABSTRACT

The State of the Research: Meta-Analysis and Conceptual Critique of Mild Traumatic Brain Injury

Ryan R. Green
Department of Psychology, BYU
Doctor of Philosophy

Researchers studying the long-term cognitive sequelae of mild traumatic brain injury (mTBI) have produced disparate results. Some studies have shown little to no long-term cognitive effects while others have shown that persistent cognitive sequelae continue to affect a subgroup of patients. Meta-analysis has been used to try to integrate these contrasting results to foster a coherent understanding of the cognitive outcomes following mTBI. However, previous meta-analyses of long-term cognitive sequelae have used studies from a period of mTBI research where methodological rigor has been called into question (Carroll, Cassidy, Holm, Kraus, & Coronado, 2004). Using studies from this period, meta-analysts found little to no effect for long-term cognitive sequelae after mTBI: g = 0.07, d = 0.12 (Binder, Rohling, & Larrabee, 1997), g = 0.11(Frencham, Fox, & Mayberry, 2005), and d = -0.07 (Rohling et al., 2011). The present meta-analysis was conducted to address problems with methodological rigor in the studies used in these previous meta-analyses and address differences in meta-analytic methodology (Pertab, James, & Bigler, 2009). Studies published between January 2003 and August 2010 were rated using the 4-tiered American Academy Neurology (AAN) guidelines for methodological rigor to ensure homogeneity and the methodological rigor of included studies. Seven studies were identified that met criteria for a rating of I or II and five met criteria for the lower ratings of III or IV. When studies of all ratings were combined, a significant effect of g = 0.45 was observed. When only studies rated I and II were combined, a significant effect of g = 0.52 was observed while a significant effect of g = 0.38 was observed when only studies rated III and IV were combined. These effect sizes for long-term cognitive sequelae are much larger than those found in previous meta-analyses. Based on these results, it is likely that methodological rigor and/or heterogeneity amongst included studies can impact meta-analytic effect sizes associated with long-term cognitive sequelae following mTBI. However, analyses did not show that more rigorous studies (i.e., those rated I or II) had significantly higher effect sizes than less rigorous studies (i.e., those rated III or IV), $t(10) = .636$, $p = .845$. This non-significant finding may be a result of the analysis being underpowered given the small k. Significant effects for neuropsychological domain were also observed and are reported. Additionally, a conceptual critique of mTBI is made with recommendations for future development of the rating system that Cappa, Conger, and Conger (2011) have put forth for objectively rating the methodological rigor of neuropsychological studies. Concerns are addressed related to the mTBI literature in the areas of mTBI definition, definition of cognitive impairment, problems with the constructs of post-concussion syndrome (PCS) and persistent post-concussion symptoms (PPCS), heterogeneity of outcome measurement, and unaccounted for variables.

Keywords: mild traumatic brain injury, concussion, meta-analysis, methodological rigor

ACKNOWLEDGMENTS

I wish to express appreciation to my dissertation chair and committee members in training and supervision of this project. I would also like to express my deepest gratitude to Thomas J. Farrer and Mark Lott for their extensive and often exhaustive time and effort they put into making this project a reality.

TABLE OF CONTENTS

The State of the Research: Meta-analysis and Conceptual Critique of Mild Traumatic Brain Injury

Introduction

Traumatic brain injury (TBI) is a serious global health concern and accounts for a large portion of traumatic deaths and disabilities. Conservative epidemiological studies indicate that TBIs occur at an annual rate of about 2 million in the United States alone with associated medical costs upwards of $17 billion (Center for Disease Control, 2003). Some studies place the annual incidence much higher (Ryu, Feinstein, Colantonio, Streiner, & Dawson, 2009). The majority of TBIs (70-90%) are classified as concussions or mild in severity [1] (Cassidy et al., 2004), and mild traumatic brain injury (mTBI) has been termed the "signature injury" of our current military engagements (Hoge et al., 2008; Rona, 2012).

Historically, many definitions of mTBI have been proffered (American Congress of Rehabilitation Medicine Head Injury Interdisciplinary Special Interest Group, 1993; Bigler, 2008; Carroll, Cassidy, Holm, Kraus, & Coronado, 2004; Menon, Schwab, Wright, & Maas, 2010; & Ruff, Iverson, Barth, Bush, & Broshek, 2009). However, to date there is yet to be an universally accepted definition of mTBI. There are many reasons why mTBI has eluded definition. Despite some similarities in the neurologic effects of mTBI, the outcomes may vary depending on the person and the type of injury involved. The biomechanics of mTBI are different depending on the following: the angle of impact, if rotational forces are involved, and which area of the brain was affected (Bigler, 2008). Given that various neurologic pathologies are possible after a concussion, it stands that the cognitive and neurobehavioral sequelae that may follow will also be quite varied thus making definition difficult. Additionally, since the neurological effects of mTBI may not be observed when assessed with conventional

[1] The terms concussion and mild TBI (mTBI) will be used interchangeably.

neuroimaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI), mTBI is often not considered to be severe enough to produce parenchymal abnormalities that can then be used to define the phenomenon.

Interestingly, recent research has shown that the neurologic effects of mTBI can be observed with more sophisticated but less available techniques such as in magnetoencephalography (MEG), single photon emission computed tomography (SPECT), positron emission tomography (PET), diffusion tensor imaging (DTI), and susceptibility-weighted imaging (SWI) (Ashwal et al., 2006, Bigler & Bazarian, 2010, Chen, Kareken, Fastenau, Trexler, & Hutchins, 2003; Fox et al., 2012; Huang et al., 2009; Shenton et al., 2012). Although many studies have been done using these techniques, many more well designed studies are needed before these technologies will be able to more fully elucidate the potential neurological complications following mTBI, which may in turn help disentangle some of the definitional problems. Once this occurs, these technologies are more likely to be made clinically available to diagnose mTBI. Additionally, metabolic processes following mTBI are also beginning to be better understood which may help our understanding of the severity of brain injury and prognosis of recovery. Potential biomarkers of mTBI have been observed in the following neurochemicals: neurofilament light protein, glial fibrillary acidic protein, phosphorylated tau, S100 proteins, and β-amyloid protein, to name a few (de Kruijk et al., 2002; Nygren de Boussard et al., 2004). However, these methods are still being developed and standardized and are of limited clinical availability.

Given the nascent stages of using advanced technologies and biomarkers to define mTBI based on more objective, biological factors, mTBI has historically been defined as the result of external force or rapid acceleration/deceleration forces that disrupt brain function. Mild traumatic

brain injury is diagnosed using behavioral observations, history of the incident, and symptom report. It is generally accepted that mTBI be demonstrated by the following criteria: Glasgow Coma Scale (GCS) from 13 to 15, loss of consciousness (LOC) for less than 30 minutes, and post traumatic amnesia (PTA) of less than 24 hours. An additional component of "alteration in mental status such as being confused or disoriented" is often added to the definition as a descriptive feature of the potential neurological disruption that is associated with mTBI (American Congress of Rehabilitation Medicine Head Injury Interdisciplinary Special Interest Group, 1993; Carroll et al., 2004). Nevertheless, these definitional standards are not universally accepted and researchers often vary in the definitions they use in their studies. The differences in definitions used can potentially cause problems when comparing studies and trying to integrate their findings. Indeed, given the heterogeneity of mTBI definitions used in research, Carroll and coworkers (2004) stated that this definitional "problem has a negative impact on the interpretation and comparison of findings on MTBI" (p. 113).

In light of the considerable concerns that concussions present to global health, and the definitional problems associated with mTBI, several workgroups have been established in order to better understand the prevention, diagnosis, treatment, and prognosis of concussion. These groups include, but are not limited to: the American Congress of Rehabilitation Medicine Head Injury Interdisciplinary Special Interest Group, the CDC Mild Traumatic Brain Injury Work Group, the International Brain Research Foundation (IBRF), the Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine, and the World Health Organization (WHO) Collaborative Centre Task Force on Mild Traumatic Brain Injury.

An example of the work these types of groups have accomplished is important to further establish the background and relevancy for the present study. Acknowledging the high rates of mTBI, high associated costs, and potential for some level of disability in both industrialized and non-industrialized countries, the World Health Organization (WHO) commissioned an extensive literature review on mTBI. The WHO findings, based on the literature from 1980 to 2002, demonstrated the variability of studies based on quality issues of research design and methodological rigor (Carroll et al., 2004; Cassidy et al., 2004). Concurrent with the WHO-sponsored mTBI studies, the International Conferences on Concussion in Sport (McCrory et al., 2009) have addressed diagnostic, assessment, and outcome issues of sports related mTBIs. The indisputable conclusion from both of these major consensus projects is that the vast majority of those who sustain a mTBI recover quickly, successfully returning to baseline level of function within hours, days or within approximately three months. Nevertheless, what remains unresolved, and quite controversial, is whether some individuals with mTBI have cognitive and/or neurobehavioral sequelae, directly related to neurological dysfunction caused by mTBI, which persist beyond three months. Thus, one of the most pressing questions for mTBI researchers centers on whether residual impairments persist beyond an acute stage of recovery, how to predict residual impairment, and how to treat it if it occurs.

Answering these questions has been difficult due to methodological limitations in the literature. For example, after performing the above mentioned comprehensive review of the literature from 1980 to 2002, Carroll et al. (2004) concluded that the mTBI literature "is large and of variable quality" (p. 113). Carroll and colleagues found that only 36% of the studies on treatment, 32% of the studies on diagnosis, and 28% of the studies on prognosis were of acceptable methodological quality. Cassidy (2010) summarized the Carroll et al. (2004) findings

regarding the quality of research in mTBI and stated, "The scientific quality of studies on mTBI was poor up to 2002" (p. e12). Additionally, Carroll and colleagues point out that numerous other concerns plague mTBI research such as inconsistent definitions of mTBI and heterogeneous measurement of mTBI which make a clear consensus of our understanding of the cognitive sequelae of mTBI difficult to elucidate.

Moreover, Dikmen et al. (2009), after completing another systematic review of the literature, concluded that "there is *insufficient evidence* to determine whether mild TBI is associated with cognitive deficits 6 months or longer post-injury" (p. 430, emphasis added). Similarly, Konrad and colleagues (2010) point out that, "Previous research on long-lasting consequences of mTBI yielded ambiguous results. Some investigators reported no or only subtle differences between control groups and patients several years after mTBI…By contrast, other researchers found significant cognitive impairments even many years after mTBI. Thus, the debate about long-term deficits of mTBI *remains unresolved*" (p. 9, emphasis added). Although the majority of individuals who experience mTBI recover fully within hours to days to a couple of months following the injury (Ponsford et al., 2000; Reitan & Wolfson, 1999; Voller et al., 1999), it also appears that there may exist a subgroup of individuals, often referred to as the "miserable minority," who experience residual long-term deficits (Iverson, 2010; Miles et al., 2008). It can be further seen that our understanding of the long-term sequelae of mTBI is unclear given the various estimates of individuals who appear to suffer long-term sequelae which ranges from zero to thirty-three percent (Binder, 1986; Binder, Rohling, & Larrabee, 1997; Frencham, Fox, & Maybery, 2005; McCrea et al., 2003; Miles et al., 2008; Pertab et al., 2009).

Review of Previous Meta-Analyses

Notwithstanding the difficulties and discrepancies in mTBI research, several meta-analyses have been conducted in order to attempt to synthesize our understanding of the potential residual cognitive deficits associated with mTBI (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Binder et al., 1997; Frencham et al., 2005; Pertab, James, & Bigler, 2009; Rohling et al., 2011; Schretlen &Shapiro, 2003). Several of these meta-analyses will be reviewed here.[2] In their frequently cited study, Binder et al. (1997) aggregated data from 8 studies with 11 samples and found an overall small effect of mTBI on post-acute cognitive impairment ($d = 0.12$, $p < .03$). This effect was reduced and nonsignificant when using the more conservative g statistic which corrects for sample size ($g = 0.07$). When domains of neuropsychological functioning were evaluated, they found a small but significant effect for attention and concentration ($d = 0.20$, $p < .006$; $g = 0.17$, $p < .02$). From this effect, they reasoned that measures of attention may be the most susceptible to chronic dysfunction after mTBI. Although Binder and colleagues (1997) concluded that neuropsychological performance was reduced by less than 5% compared to controls, that this reduction in performance may be accounted for by measurement error, and that the average effect of mTBI on neuropsychological performance is *undetectable*, they nevertheless indicated that their results may support one of two hypotheses: 1) there is an association between mTBI and greater cognitive impairment in a small percentage of mTBI patients (i.e., the "miserable minority") or 2) there are small reductions in cognitive functioning in a large percentage of mTBI patients. The fact that these authors indicated two possible hypotheses regarding observed residual deficits in mTBI is important because their meta-analysis is often cited as evidence that suggests there are no residual deficits following mTBI.

[2] Meta-analysts will use different methods of coding whether they want the effect demonstrated as a negative or positive number. This difference in methodology accounts for why some authors report results with positive or negative numbers. In this paper, I use the original authors' method of reporting.

Frencham et al. (2005) conducted a follow-up to the Binder et al. (1997) meta-analysis. Frencham and colleagues included studies published since the Binder et al. study up to 2003 and extended the previous meta-analysis by including not only studies assessing mTBI in the post-acute stage (i.e., post 3 months), but also studies assessing mTBI within 3 months of injury. Seventeen studies met their inclusion criteria. These 17 studies yielded a significant overall effect, $g = 0.32$, $p < .001$ when all stages post-injury were aggregated. When neuropsychological domains were broken down and these data aggregated across all stages post-injury, speed of processing measures yielded the largest significant effect, $g = 0.47$, $p < .001$ with working memory/attention, memory, and executive functioning also yielding significant effects, $g = 0.25$, $g = 0.30$, and $g = 0.30$, respectively. When studies reporting post-acute data were analyzed ($k = 5$), a nonsignificant effect size of $g = 0.28$ was reported. However, no power analyses were reported in the study. When these five studies were pooled with Binder and colleagues' post-acute data, the effect size was considerably reduced, $g = 0.11$. Frencham et al. found a significant moderating effect for time post-injury on neuropsychological performance across all stages post-injury that accounted for 22% of the variance in the effect of mTBI. They indicated that "the effect size tended toward zero with increased time since injury" (p. 344). In the post-acute stage, the relationship between time since injury and effect size failed to reach significance. Frencham and colleagues concluded that this nonsignificant result likely supports the view that the majority of recovery after mTBI occurs within the first 3 months and that subsequent improvement in neuropsychological performance is of limited statistical and clinical significance. Like Binder et al. (1995), Frencham et al. (2005) indicated that it is "possible that a sub-sample in the studies summarized did have more severe cognitive deficits, and that the effect of their results has been lost (in terms of statistical significance) by the pooling of data…[and] that a subgroup of

approximately 15% may experience protracted cognitive recoveries" (p. 347). Again, this statement is important to note because there is quite a strong movement in neuropsychology to suggest that there is no long-term impairment in mTBI and if there is impairment, it is accounted for by factors other than the brain injury (e.g., psychological factors, compensation seeking, chronic pain, poor effort, premorbid conditions etc.).

Pertab and colleagues (2009) sought to clarify opposing views in the mTBI literature regarding cognitive sequelae post three months. These authors re-analyzed the Binder et al. (1997) and Frencham et al. (2005) data and argued that several constructs affect long-term mTBI outcome including: 1) the mechanism of injury, 2) which diagnostic criteria are employed, 3) which assessment tools are utilized, and 4) whether symptomatic groups are considered separately. They concluded that, in the studies used in the Binder et al. and Frencham et al. meta-analyses, there was "significant statistical heterogeneity in the following areas: (a) the effect sizes of neuropsychological measures employed in the post-acute phase (>3 months) and marked qualitative heterogeneity, (b) in the criteria used to define mTBI and mTBI severity, and (c) in the populations and mechanisms of injury from which the mTBI samples were selected" (p. 504). Pertab et al. (2009) argued that because of the heterogeneity observed in the studies used in the meta-analysis by Binder et al. (1997) and Frencham et al. (2005), it is likely that these meta-analyses have overlooked important information regarding the long-term outcomes following mTBI and have, therefore, given an inaccurate assessment of these long-term outcomes. Furthermore, these authors found that Verbal Paired Memory (g = -0.52), Coding Tasks (g = -0.33), and Digit Span (g = -0.33) remained statistically significant post three months when compared to control groups. Pertab and colleagues pointed out that group statistics likely conceal

a nested minority of patients who continue to suffer long-term cognitive sequelae following mTBI.

Rohling and colleagues (2011) sought to clarify their own initial meta-analytic findings, and conducted re-analyses of the Binder et al. (1997), Frencham et al. (2005), and Pertab et al. (2009) meta-analyses with the purpose of analyzing epochs of time post injury and using a random effects model; a model that had up to that point not been used in mTBI meta-analyses. Rohling and colleagues found differences across epochs: 1) < 7 days post injury (d = -0.39), 2) 8-30 days post injury (d = -0.32), 3) 31-92 days post injury (d = -0.14), and 4) > 93 days post injury (d = -0.07). These authors concluded that it is unlikely that a "highly impaired, but undetected, subgroup of mTBI patients of any appreciable size" (p. 619) exists but that these results may not generalize to individuals with complicated mTBI, history of multiple concussions, and/or individuals with unequivocal neurological abnormalities such as hemiplegia.

The Continued Debate

The debate regarding mTBI meta-analyses has continued in the literature (Larrabee, Binder, Rohling, & Ploetz, 2013; Rohling, Larrabee, & Millis, 2012). Ruff and Jamora (2009), for example, point out many of the concerns with current meta-analyses and affirm many of the conclusions from Pertab et al. (2009) by pointing out the following concerns: 1) different criteria for diagnosing mTBI were used in meta-analysis source studies, 2) time intervals between injury onset and assessment, and attrition rates varied considerably in source studies, 3) test batteries used to assess cognitive abilities varied substantially in source studies, 4) control groups were often poorly matched in source studies, 5) flawed sampling procedures in source studies, 6) although critical to our understanding of mTBI, no attempt has been made to meta-analyze

emotional and physical outcomes following mTBI, and 7) a separation between those in litigation was not consistently reported in source studies.

Subsequent to this critique by Ruff and Jamora (2009) and in response to Pertab et al. (2009), Rohling and colleagues (2012) conducted an empirical study based on the hypothetical distributions presented by Pertab and coworkers to test whether a nested *miserable minority* of mTBI survivors would actually be lost with group statistics. Rohling and colleagues point out two important findings in their study. First, they state that their effect size (d = -1.60) from their hypothetic sample of mTBI patients which did not consist of an impaired subgroup is not equivalent to the effect size (d = -1.02) from a hypothetical sample that does comprise a subgroup of impaired mTBI patients as asserted by Pertab et al. (2009). Second, they state that both of these effect sizes estimates are much too large to genuinely represent the data found in meta-analyses (i.e., previous meta-analyses have found effect sizes in the 0.07 to 0.12 range).

Bigler and colleagues (2013) then responded to Rohling et al. (2011) and reaffirmed their initial findings regarding the limitations of previous meta-analyses published in Pertab et al. (2009). Bigler and coworkers assert the following: 1) it is still possible that a nested minority exist in previous meta-analyses and given that no power analyses were conducted in previous meta-analyses there is no way to know whether these studies were well enough powered to detect an effect if it were present, 2) the source studies used in previous meta-analyses have methodological limitations (e.g., American Academy of Neurology [AAN] criteria differences), 3) problems with data transparency exist in previous meta-analyses (e.g., adding zeros in meta-analytic datasets when data was not given in source studies), 4) limitations in statistical and methodological assumptions exist in previous meta-analyses, 5) limitations in research design in the source studies of previous meta-analyses, 6) the possibility of a lack of sensitivity in

neuropsychological assessment in source studies, and 7) the likelihood that previous meta-analyses perpetuate type II error by indicating that all mTBI patients get better.

Most recently, Larrabee, Binder, Rohling, and Ploetz (2013) published a rejoinder to Bigler et al. (2013) again reaffirming their initial critique of Pertab et al. (2009). Among these authors' more substantial assertions are the following: 1) their statistical analyses (including *Q*, tau^2, and I^2) preclude the possibility of a nested minority of mTBI patients who suffer long-term cognitive sequelae, 2) that including patients with different etiologies of mTBI does not lead to problems of heterogeneity, 3) that it is not problematic to code zeros for unreported statistical data, and 4) that re-analysis of previous meta-analytic data did not show important differences when AAN clinical practice guidelines were used to rank the methodological rigor of source studies.

Despite these debates, previous meta-analyses (Belanger et al., 2005; Binder et al., 1997; Frencham et al., 2005; Pertab et al., 2009; Rohling et al., 2011; Schretlen &Shapiro, 2003) have nevertheless used source studies that were published during the era of mTBI research that the WHO investigators (Carroll et al., 2004) originally pointed out is methodologically problematic. These prior meta-analyses involved studies up to and inclusive of 2002, but for the most part have not included studies published after 2002. Thus, debates about the merits and demerits of previous meta-analyses are unlikely to be resolved using data from this time period; particularly given that large scale, prospective, longitudinal studies that use an orthopedic-injury control group have yet to be undertaken (see Bigler et al., 2013; Iverson, 2010; Larrabee et al., 2013; Pertab, James, & Bigler, 2009; Rohling et al., 2011).

The Current Meta-Analysis

The goal, therefore, of the current meta-analysis is to improve upon some of the limitations recognized in previous meta-analyses. I intend to update the previous comprehensive reviews regarding the state of the mTBI literature and meta-analyses by using studies from August 2003 to January 2010, but apply greater front-end selection criteria in an attempt to achieve greater uniformity and rigor in defining acceptable studies to include in the analyses according to the WHO guidelines (Carroll et al., 2004; Cassidy, 2010). Cappa, Conger, and Conger (2011) have shown the importance of establishing conclusions from meta-analyses involving TBI based on the front-end selection process to better determine which studies to include in the meta-analysis. Therefore, the guidelines set forth by the AAN in their clinical practice guidelines (Edlund, Gronseth, So, & Franklin, 2004) which provide a four-point scale to assess methodological rigor based on a priori conditions of the experimental design, and on independence of data collection, analysis, and investigators, will be used in the current meta-analysis to determine methodological rigor and reduce heterogeneity among included studies from 2003 to 2010 (see Tables 1, 2, 7, and 8 below). Edlund and coworkers state that the AAN developed clinical practice guidelines "to assist its members in clinical decision making-particularly in situations of controversy or variation in practice " (2004, p. 6). Therefore, given that long-term outcome following mTBI has been hotly debated and continues to be a controversial topic, the process of rating methodological rigor of scientific studies set forth by the AAN would be inherently helpful to improve homogeneity among articles and reduce the amount of potential bias in articles selected for meta-analysis, and would therefore likely further improve our understanding of long-term cognitive outcomes in mTBI.

Regarding improving our understanding of potential bias in scientific studies, Edlund and colleagues (2004) point out that bias and systematic error is the tendency to inaccurately measure the variables of interest. They further indicate that it is not possible to directly study the amount of "bias" in a study but that by using "well-established principles of good study design we can estimate the *risk* of bias of a study" (p. 18). Regarding the 4-tiered AAN classification system, these authors stated, "studies graded class I are judged to have a low risk of bias; studies graded class II are judged to have moderate risk of bias; studies graded class III are judged to have a moderate to high risk of bias; studies graded class IV are judged to have a very high risk of bias" (p. 18). Indeed, as Borenstein, Hedges, Higgins, and Rothstein (2009) point out, "While a meta-analysis will yield a mathematically accurate synthesis of the studies included in the analysis, if these studies are a biased sample of all relevant studies, then the mean effect computed by the meta-analysis will reflect this bias" (p. 277). Thus, given the concerns related to potential bias in scientific studies of mTBI, the AAN rating system will be used in the current meta-analysis to determine and rank the methodological rigor of selected studies so an understanding of the potential bias within each article can be estimated and the merits of the meta-analysis can be better evaluated.

By nature, mTBI represents the mildest form of TBI. Therefore, the supposition with regards to residual effects from a mild injury would be that any residual effects, if present, would also likewise be mild. Accordingly, to detect subtle cognitive or neurobehavioral effects, studies included in a meta-analysis of mTBI should only be those meeting a standard of methodological rigor (e.g., the above mentioned AAN criteria) in order to reduce, as much as possible, any experimental error or bias that may have entered into the data. Intuitively, this would imply those with the greatest rigor in experimental design. Thus, by separating studies in the current meta-

analysis based on methodological rigor, it may be possible to reduce the amount of error associated with data collection/analysis (Cook & Campbell, 1979) and thereby minimize the possibility that the mild effects of those who do experience residual deficits will be lost when pooled with those who experience a good recovery. Moreover, when combining the effect sizes from the most methodologically rigorous studies and combining the less methodologically rigorous studies, comparisons between the levels of rigor can be performed to evaluate whether there is a significant difference between the effect sizes produced by the studies based on their ranking of rigor.

A further concern with regards to previous meta-analytic methodology (e.g., Binder et al., 1997; Frencham et al., 2005) will be addressed in the current meta-analysis. This concern is with the methodological handing of instances where non-significant data for neuropsychological tests or domains were not provided by authors. Specifically, some authors of previous meta-analyses have used a method which potentially minimizes the effect sizes associated with long-term cognitive effects of mTBI by adding zeros to their data bases when authors of source studies did not provide statistical information regarding non-significant results. More specifically, when the authors of source studies found nonsignificant results, but did not give explicit statistical data, those nonsignificant results were replaced by previous meta-analysts with an unspecified number of zeros in their meta-analytic data sets. When this method has been used, authors did not make explicit how many zeros were entered into the data sets and in which neuropsychological domains. It is therefore difficult to objectively discern whether this methodology is appropriate. In fact, it is arguably not a viable way to handle these instances as it will inevitably bias resultant effect sizes toward zero. Thus, for the current and updated meta-analysis only studies with actual data, which were presented in the articles, were used. In instances in which authors did not

provide statistical information when nonsignificant results were found, authors were contacted by email and asked to provide the additional statistical information. Several authors were contacted and only two responded with data. However, these studies whose authors provided data did not meet final inclusion criteria and therefore the provided data were not used in the current data set.

In summary, three primary concerns are being addressed by the current meta-analysis: 1) conducting an updated review of studies published from 2003 to 2010 to ensure the most recent and rigorous studies are being incorporated, 2) using AAN criteria for rating methodological rigor of studies to limit potential bias and increase homogeneity, and 3) using a methodology that does not add zeros to the data sets when statistical information is not available. Addressing these concerns, I hypothesize that the current meta-analysis will demonstrate an overall effect size significantly greater than zero indicating poorer cognitive performance being associated with individuals who have experienced mTBI compared to controls in studies from 2003 to 2010. Additionally, given that systematic bias based on potential methodological problems is a concern for researchers and meta-analysts, this concern will be directly addressed by comparing the effect sizes of studies that received higher methodological rigor ratings (e.g., I and II) to the effect sizes of studies that received a lower methodological rigor rating (e.g., III and IV). It is hypothesized, therefore, that the cognitive impairment effect size in the mTBI groups in studies receiving the higher methodological rigor rating of I or II will be larger than the cognitive impairment effect size in the mTBI groups in studies receiving a lower ranking of III or IV. By testing this second hypothesis, I will be better able to determine whether the AAN clinical practice guidelines rankings are good indicators of the potential systematic error associated with studies that vary in methodological rigor. Accordingly, the hypotheses of the current study are:

1) Cognitive performance of mTBI group < Cognitive performance of control group
2) Effect size in studies receiving an AAN rating of I and II > Effect size in studies receiving an AAN rating of III or IV.

Methods

Search Procedures

An online literature search was conducted through the PubMed database using the search term "mild traumatic brain injury." Search dates were limited to those articles published between January 2003 and August 2010. This resulted in 1,631 articles being identified. As some of the articles in the original search were not relevant, the authors narrowed selection by reviewing titles and abstracts for the terms: *mild traumatic brain injury, mild head injury, minor head injury, concussion, postconcussive syndrome, mild closed head injury, Glasgow Coma Scale/Score, brain trauma, and post traumatic amnesia* (see Figure 1).

A similar search was conducted using the online database PsychINFO in order to ensure a more comprehensive search. The same procedures as outlined above were used. This search yielded 477 results. No studies were found in addition to the search results of the PubMed search.

Again, a similar search was conducted using the online database PsychEXTRA in order to search for unpublished studies (e.g., theses and dissertations) and directly assess publication bias. It has been pointed out that studies which are not published may be systematically different than published studies (Vevea & Woods, 2005). No additional studies were found.

To further assure no articles were missed, seven major journals (Archives of Clinical Neuropsychology, Applied Neuropsychology, Brain Injury, Journal of Clinical and Experimental Neuropsychology, Journal of the International Neuropsychological Society, Neuropsychology,

and The Clinical Neuropsychologist) which publish in the area of neuropsychology and mTBI outcome were then reviewed using the methods described above to determine if any articles had been overlooked in the original online database searches. No new articles were identified.

Inclusion Criteria

To be included in the present study, articles had to meet several criteria in order to establish further homogeneity and permit calculation of effect sizes for cognitive domains of functioning. 1) Studies published or unpublished from January 2003 to August 2010 had to include mTBI patients and a non-brain injured control group(s) in their comparisons. If authors did not separate severity of TBI in their analyses even if mTBI subjects were included, the study was excluded. 2) Patients had to be evaluated using validated clinical measures in the post-acute stage of recovery (i.e., post 3 months) and have experienced only one mTBI. 3) Participants had to be 16 years old or over (given that this the age at which the Wechsler Adult Intelligence Scale –IV begins). 4) Participants were included in studies based on history of TBI by presenting to a hospital where mTBI was diagnosed or medical records were available to substantiate the mTBI history. 5) Studies needed enough information to calculate effect sizes. In instances where studies did not have information to calculate effect sizes, but met all other criteria, the authors were contacted in order to procure the necessary data to calculate effect sizes. If authors responded with the needed statistical information, the data was included. 6) Studies must have been written in English or an English translation must have been available. 7) Articles not relevant to the current study (e.g., review articles and animal models) were excluded. 8) Studies which included participants with previous psychiatric conditions were also excluded given that history of psychiatric condition can account for a large proportion of the variance of individuals with persistent post concussion symptoms (Luis, Vanderploeg, & Curtiss, 2003). 9) For

longitudinal studies (e.g., Heitger et al., 2006), data was used from the three-month time point unless the range of time post mTBI indicated the possibility of individuals not being three months post TBI in which case we used the six-month follow-up data. Figure 1 below details the inclusion/exclusion process.

Rating of Methodological Rigor

As previously stated, the AAN ratings were used to objectively determine methodological rigor. Each article from the 2003 to 2010 timeframe was ranked according to the 4-tier AAN clinical practice guideline criteria (See Edlund et al., 2004 and Tables 2, 7, and 8 below for details). The attempt was made to follow the guidelines as explicitly as possible.

Three independent doctoral students rated studies based on the AAN clinical practice guideline criteria (Edlund et al., 2004). Each rater was responsible for rating two-thirds of the articles. Table 1 illustrates the overlap among the raters and the articles which they were individually responsible for rating. This method was used to ensure that each article was rated independently by at least two different raters. Raters used separate spread sheets to rate the rigor of the articles based on the six domains assessed by the AAN ratings (e.g., comparison group, study design, patient spectrum, reference standard, completeness, and masking). Table 2 below briefly outlines the criteria used to determine the overall rating of an article (see Edlund et al., 2004 for further details). Once the six domains were assessed, a final rating was given for the article based on the lowest rating for each of the six domains. When all articles had been rated by all three raters, the raters then compared their ratings of the studies assigned to them. If discrepancies were observed between ratings, the original article was reviewed by both raters together to determine why the discrepancy occurred. (The original methodology for the study indicated that when discrepancies between raters were observed and difficult to rectify, the original article would be reviewed by

three independent neuropsychologists [two of which are board certified] to determine the most accurate rating. However, this was not necessary as discrepancies were easily rectified.) Interrater reliability was not calculated because it was determined that an accurate rating was more important than interrater reliability. All ranked studies were examined for potential moderator variables and data that could be used to calculate effect sizes. It was necessary to contact two authors whose studies were given a rating of II (Miles et al., 2008; Ge et al., 2009) as

Figure 1

Flow Chart of Inclusion/Exclusion Process

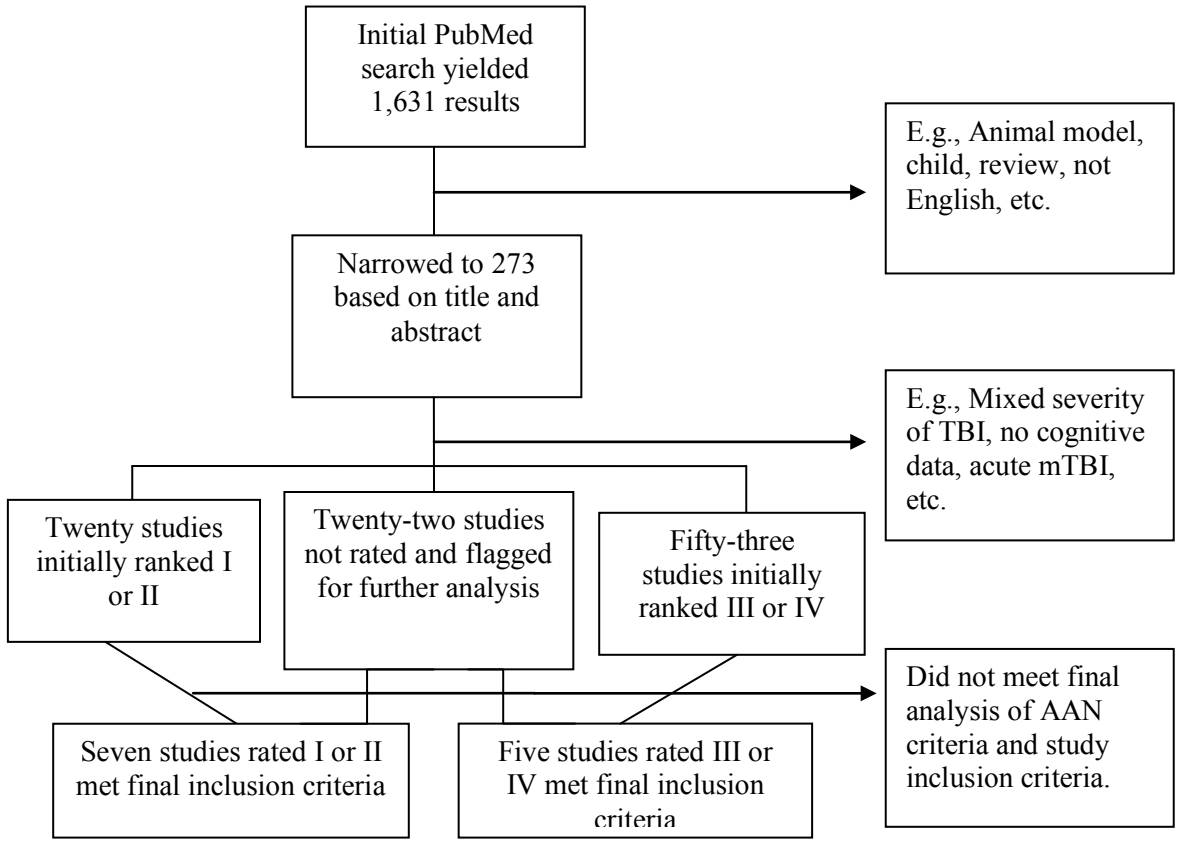


Table 1

Overlap Among Raters

AAN Rater	First Third of 1,631 Articles	Second Third of 1,631 Articles	Third Third of 1,631 Articles
1	X		X
2	X	X	
3		X	X

Table 2

AAN Guidelines

Rating of Diagnostic Article	Rating of Prognostic Article
Class I: Evidence provided by a prospective study in a broad spectrum of persons with the suspected condition, using a reference (gold) standard for case definition, where test is applied in a blinded evaluation, and enabling the assessment of appropriate tests of diagnostic accuracy. All patients undergoing the diagnostic test have the presence or absence of the disease determined.	**Class I:** Evidence provided by a prospective study of a broad spectrum of persons who may be at risk for developing the outcome (e.g. target disease, work status). The study measures the predictive ability using an independent gold standard for case definition. The predictor is measured in an evaluation that is masked to clinical presentation and, the outcome is measured in an evaluation that is masked to the presence of the predictor. All patients have the predictor and outcome variables measured.
Class II: Evidence provided by a prospective study of a narrow spectrum of persons with the suspected condition, or a well designed retrospective study of a broad spectrum of persons with an established condition (by "gold standard") compared to a broad spectrum of controls, where test is applied in a blinded evaluation, and enabling the assessment of appropriate tests of diagnostic accuracy.	**Class II:** Evidence provided by a prospective study of a narrow spectrum of persons at risk for having the condition, or by a retrospective study of a broad spectrum of persons with the condition compared to a broad spectrum of controls. The study measures the prognostic accuracy of the risk factor using an acceptable independent gold standard for case definition. The risk factor is measured in an evaluation that is masked to the outcome.
Class III: Evidence provided by a retrospective study where either persons with the established condition or controls are of a narrow spectrum, and where the reference standard, if not objective, is applied by someone other than the person that performed the test	**Class III:** Evidence provided by a retrospective study where either the persons with the condition or the controls are of a narrow spectrum. The study measures the predictive ability using an acceptable independent gold standard for case definition. The outcome, if not objective, is determined by someone other than the person who measured the predictor.
Class IV: Any design where test is not applied in an independent evaluation OR evidence provided by expert opinion alone or in descriptive case series without controls.	**Class IV:** Any design where the predictor is not applied in an independent evaluation OR evidence provided by expert opinion or case series without controls.

it was unclear whether the samples used in these two studies overlapped. We were informed by the lead contact in those studies that the samples did not overlap and both articles were included. Additionally, when data to calculate effect sizes was not presented, attempts were made to contact authors for additional data and this data, if received, was added to the analyses.

Data Extraction and Preparation

All reported neuropsychological data were not reviewed for coding until the final inclusion/exclusion process had been completed. Once the inclusion/exclusion process had been completed, neuropsychological data were extracted. Neuropsychological domains were identified to include attention/concentration, executive functioning, expressive language, IQ, perceptual reasoning, premorbid IQ estimates, processing speed, psychomotor speed, receptive language, response speed, sensory/perceptual processing, verbal abstract reasoning, verbal learning and memory (both contextual and rote), verbal working memory, visual memory, and visual and spatial working memory. In order to test the first hypothesis as described above, means, standard deviations, and number of patient and control participants were entered into Biostat's Comprehensive Meta Analysis II to perform calculations. Positive direction of an effect size indicated poorer performance by the mTBI group. Random effects models were employed to be consistent with previous meta-analysis (Rohling et al., 2011). Both Cohen's *d* and Hedges' *g* have been calculated in previous meta-analyses. However, given that Cohen's *d* has been reported to inflate effect sizes when smaller samples are used, Hedges' *g* (Hedges & Olkin, 1985) with a bias correction factor was used in the current meta-analysis in order to present a more conservative estimate. Hedges' *g* is a standardized mean difference formula with a correction factor where $f(m)$ = bias correction factor, n_c = number of control subjects, n_t =

number of mTBI subjects, s_c = standard deviation of control subjects, and s_t = standard deviation of mTBI subjects.

Equation 1

Hedges' g

$$g_c = \mathcal{J}(n_c + n_t - 2)\frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{(n_t-1)s_t^2+(n_c-1)s_c^2}{n_t+n_c-2}}}$$

Sample demographic and descriptive data were extracted including the study's country of origin, setting of the study (e.g., university, hospital), proportion of male and female, age, handedness, education, litigation status, time post-injury, MTBI definition, and etiology of injury (see Appendix 1 for an example of the code book). Studies did not consistently report all sample descriptions (see Tables 3, 4, 5, and 6 below).

Results

Search and rating procedures from January 2003 to August 2010 produced a total of 12 studies (k = 12). Tables 3 through 6 below present demographic and descriptive data for the 12 studies. Seven of those studies were given an AAN ranking of I or II and five were given an AAN ranking of III or IV (see Tables 7 and 8). In studies ranked I-II, a total of 141 mTBI participants and 140 control participants were assessed. In studies ranked III-IV, a total of 191 mTBI participants and 121 control participants were assessed. A total of 188 means and standard deviations were entered for analysis; 102 of which were from the articles with an AAN rating of I or II and 86 of which were from the articles with an AAN rating of III or IV. When all rated articles were combined (see Table 9), an overall effect of g = 0.45 was observed. The Classic Fail-safe analysis indicated that 71 studies reporting null results would be necessary to reduce this finding to nonsignificance. The Q statistic was not significant indicating a reasonable

Table 3

Demographic Data for Studies Ranked I and II

Author(s)	Year	Group	N	Sex	Mean Age (SD/Range)	Handedness	Differences in Demographics	% Attrition
Blanchet et al.	2009	MTBI	13	M/5F	26.31 (5.23)	10R/3L	Matched according to age, gender, handedness, and education	
		Control	12	6M/6F	26 (5.34)	10R/2L		
Ge et al.	2009	MTBI	21	6F/15M	34.1 (8.6/22-54)		Matched by education, gender, and age	0.0%
		Control	18*	4F/14M	36.1 (10.6)			
Konrad et al.	2010	MTBI	33	16F/17M	36.7 (12.4)		Matched by age, gender, and education	25.0%
		Control	33	16F/17M	37 (12)			
Kraus et al.	2007	MTBI	22	9M/13F	35.85 (9.39)		No significant difference on estimated premorbid IQ	9.1%
		Control	18	7M/11F	32.83 (10.65)			0.0%
Lee et al.	2008	MTBI	28	5F/23M	30.3(8.6)	27R/0L/1A	Matched by gender, education, handedness, and age	22.2%
		Control	18	3F/15M	34.3 (8.9)	16R/2L		
Little et al.	2010	MTBI	12		31.2 (SEM = 2.71)		Matched by age, education, and premorbid IQ	
		Control	12		30.8 (SEM: 3.04)			
Miles et al.	2008	MTBI	12	6F/11M**	33.44 (18-58)		Matched by age and gender	29.00%
		Control	29	14F/15M	35 (18-61)			

*Reported both 18 and 20. **Does not account for attrition of 5 participants whose sexes were not detailed. ***Possible significant differences for gender (Comparison of 5 different groups; not just mTBI & Control)

Table 4

Demographic Data for Studies Ranked III and IV

Author(s)	Year	Group	N	Sex	Mean Age (SD/Range)	Handedness	Non-significant Differences in Demographics	% Attrition
Hattori et al.	2009	MTBI	15	12M/3F	45 (11/27-60)		No significant difference on age & education	3.33%
		Control	15	3M/12F	43 (9/28-58)			
Heitger et al.	2006	MTBI	37	24M/13F	29.1 (12.7/15-56)		Matched by age, education, gender, and IQ	
		Control	37	24M/13F	29.2 (12.6/15-57)			
Kwok et al.	2008	MTBI	15	9M/6F	39.13 (11.46)		No significant difference on age, education, gender, IQ, or BDI-II	51.6%
		Control	19	11M/8F	44.47 (7.49)			38.71%
Meyers & Rohling	2004	MTBI	57**	43M/14F	36.93 (15.1)	51R/6L/0A	No significant difference on all demographic data	
		Control	30**	15M/15F	38.6 (18.89)	29R/1L/0A		
Ord et al.	2009	MTBI	67	42M/25F	38.9 (11.4)		No significant difference for age, education, & ethnicity	
		Control	20	16M/4F	33.2 (10.6)			

** Number of participants varied between tests administered

Table 5

Study Details for Studies Ranked I and II

Author(s) (year)	Country of Origin	Site of Study	Time Post mTBI	Definition of mTBI	Definition of RCI	Symptom Validity	Litigation and Disability Claim
Blanchet et al., (2009)	Canada	Uni/Hosp	Range = 4-99 months	LOC < 30 min, PTA < 24 hrs, GCS = 13-15; transitory neurologic abnormalities; confusion or disorientation			1 in litigation
Ge et al., (2009)	USA	Hosp	Range = 6 months to 7 years	LOC < 30 min, PTA ≤ few hours, GCS ≥ 13			
Konrad et al., (2010)	Germany	Uni/Hosp	Range = 4.75 – 7.25 years	LOC < 3 0min, PTA < 24 hrs, GCS = 13-15		Excluded if failed effort testing	
Lee et al., (2008)	USA	Hosp	1 year	LOC < 30 min, PTA present, GCS = 13-15			
Little et al., (2010)	USA	Uni/Hosp	>1 year	LOC < 30 min, PTA < 24 hrs		Excluded if failed effort testing	Excluded if history of litigation
Miles et al., (2008)	USA	Hosp	6 months	LOC < 20 min, PTA < 24 hrs, GCS = 13-15			

RCI = Residual Cognitive Impairment. Hosp = Hospital. Uni = University.

Table 6

Study Details for Studies Ranked III and IV

Authors(s) (year)	Country of Origin	Site of Study	Time Post mTBI	Definition of mTBI	Definition of RCI	Symptom Validity Measures	Litigation and Disability Status
Hattori et al., (2009)	USA	Uni/Hosp	>6 months	LOC < 30 min, PTA < 24 hrs, GCS = 13-15		PASAT scores not different from controls	6 in litigation
Kwok et al., (2008)	China	Uni/Hosp	3 months	LOC < 30 min, PTA < 24 hrs, GCS = 13-15			
Heitger et al., (2006)	New Zealand	Uni/Hosp	6 months*	LOC ≤ 15 min, PTA < 24 hrs, GCS 13-15, hrs,			0 in litigation
Meyers & Rohling (2004)	USA	Uni/Hosp	>6 months	LOC ≤ 20 min		9 effort measures; could fail only one	0 in litigation
Ord et al., (2009)	USA	Private Practice (records available)	>1 year	LOC < 30 min, PTA < 24 hrs, GCS = 13-15; no neurological signs; no abnormalities on neuroimaging associated with head injury		Used at least 2 effort measures	"most" were involved in litigation

RCI = Residual Cognitive Impairment. Hosp = Hospital. Uni = University. *The six-month data was used in this longitudinal study.

homogeneity amongst studies. When studies ranked I-II were combined, an overall effect of g = 0.52 was observed. The Classic Fail-safe analysis indicated that 24 studies reporting null results would be necessary to reduce this finding to nonsignificance. The Q statistic was not significant indicating a reasonable homogeneity amongst studies. When studies ranked III-IV were combined, an overall effect of g = 0.38 was observed. The Classic Fail-safe analysis indicated that 8 studies reporting null results would be necessary to reduce this finding to nonsignificance.

The Q statistics were not significant in all analyses indicating a reasonable homogeneity amongst studies. Given that the Q statistics were nonsignificant and given the small k in the study, further moderator analyses were not deemed necessary. Although Q statistics were calculated and were not significant in all analyses, this does not, however, necessarily indicate that there is not important variability among effect sizes that may be due to error stemming from sources other than subject-level sampling error. That is, a nonsignificant Q-value may be due to the small k and small sample sizes within the included studies (Lipsey & Wilson, 2001). Additionally, Figure 2 is a funnel plot representing the relative effect size spread for the overall combined effect for all studies (g = 0.45). (See Appendix 2 for the funnel plots displaying the effect of studies ranked I or II [g = 0.52] and for the effect of studies ranked III or IV [g = 0.38], respectively.) Funnel plots are often used as a visual assessment of publication bias (Borenstein et al., 2009). As Borenstein and colleagues point out, using the standard error on the Y axis is advantageous for detecting potential publication bias because "smaller studies" (that are less likely to be published) will tend to have more standard error and will be spread out along the bottom half of the plot making it easier to visually detect asymmetry in the plot. They further point out that "In the *absence* of publication bias, the studies will be distributed symmetrically about the mean effect size, since the sampling error is random. In the *presence* of publication

bias the studies are expected to follow the model, with symmetry at the top, a few studies missing in the middle, and more studies missing near the bottom" (emphases in original, p. 283). As can be seen in the Figure 2, overall there are gaps in the top of the funnel which represents "large" studies with small standard errors and in the bottom of the funnel which represents "small" studies with large standard errors indicating potential missing studies in these two areas. Analyses of effects for neuropsychological domains, which had at least five data points available for analysis, were also calculated; all of which were significant (see Table 10). In studies rated I-II, verbal working memory demonstrated the largest effect ($g = 0.57$), followed by processing speed ($g = 0.56$), verbal memory for a list ($g = 0.54$), attention/concentration ($g = 0.48$), executive functioning ($g = 0.36$), and visual memory ($g = 0.33$). In studies rated III-IV, significant effects for neuropsychological domain were observed in expressive language ($g = 1.02$), followed by attention/concentration ($g = 0.64$), verbal memory ($g = 0.55$), perceptual reasoning ($g = 0.52$), processing speed ($g = 0.33$), executive functioning ($g = 0.33$), visual memory ($g = 0.32$), and verbal working memory ($g = 0.31$).

To test the second hypothesis, an independent *t*-test was conducted using each study's mean effect size (*g)* and grouped by whether the study received an AAN rating of I-II or III-IV. In this way, I was able to compare whether significant differences in effect size can be detected in studies with variable methodological rigor as a function of AAN rating. The *t*-test was nonsignificant, $t(10) = .636$, $p = .845$. Cohen (1992) points out that, for a significant difference between means to be observed with a medium effect size (i.e., $d = 0.50$), power of .80, and $\alpha = 0.05$, 64 observances in each group would be necessary. It is possible, therefore, that the analysis was underpowered; the I-II group had only seven mean effect sizes and the III-IV group had only

Table 7

AAN Criteria for Articles rated I or II

Author (date)	Comparison Group	Study Design	Patient Spectrum	Reference Standard	Completeness	Masking	Final Rating
Blanchet et al. (2009)	I	II	I	I	I	II	II
Ge et al. (2009)	I	II	I	I	I	II	II
Konrad et al. (2010)	I	II	I	I	I	II	II
Kraus et al. (2007)	I	II	I	I	I	II	II
Lee et al. (2008)	I	I	I	I	I	II	II
Little et al. (2010)	I	II	I	I	I	II	II
Miles et al. (2008)	I	II	I	I	I	II	II

Table 8

AAN Criteria for Articles rated III or IV

Author (date)	Comparison Group	Study Design	Patient Spectrum	Reference Standard	Completeness	Masking	Final Rating
Hattori et al. (2009)	I	II	I	I	I	III	III
Heitger et al. (2009)	I	I	I	I	I	III	III
Kwok et al. (2008)	I	I	I	I	I	III	III
Meyers & Rohling (2004)	I	II	I	II	I	III	III
Ord et al. (2008)	I	II	I	I	I	III	III

five mean effect sizes accounting for the total *k* of 12. It should be noted, however, that the I-II group did display a larger effect size ($g = 0.52$) compared to the III-IV group ($g = 0.38$) and it is therefore possible that with more observations in each group a significant difference may be detected.

Table 9

Meta-analytic Data

Study Author(s)	*g*	Standard Error	Variance	Lower Limit	Upper Limit	*z* value	*p* value
Studies Rated I and II							
Blanchet et al.	0.33	0.39	0.15	-0.44	1.09	0.83	0.41
Ge et al.	0.97	0.33	0.11	0.31	1.62	2.9	0.001
Konrad et al.	0.67	0.25	0.06	0.18	1.16	2.66	0.01
Kraus et al.	0.13	0.31	0.1	-0.49	0.75	0.42	0.67
Visual Memory	0.33	0.12	0.02	0.09	0.58	2.68	0.007
Lee et al.	0.65	0.3	0.09	0.05	1.25	2.13	0.03
Little et al.	0.35	0.4	0.16	-0.43	1.14	0.88	0.37
Miles et al.	0.35	0.34	0.12	-0.32	1.02	1.03	0.3
Total	**0.52**	**0.12**	**0.015**	**0.28**	**0.76**	**4.27**	**0.0001**
Studies Rated III and IV							
Hattori et al.	0.56	0.36	0.13	-0.15	1.27	1.54	0.12
Heitger et al.	0.31	0.23	0.05	-0.15	0.76	1.33	0.18
Kwok et al.	0.29	0.34	0.12	-0.38	0.96	0.85	0.39
Meyers & Rohling	0.76	0.24	0.06	0.29	1.22	3.18	0.001
Ord et al.	0.03	0.25	0.06	-0.47	0.52	0.1	0.92
Total	**0.38**	**0.14**	**0.02**	**0.12**	**0.65**	**2.85**	**0.004**
Overall Total	**0.45**	**0.09**	**0.01**	**0.28**	**0.62**	**5.26**	**0.0001**

Table 10

Meta-analytic Data for Neuropsychological Domain

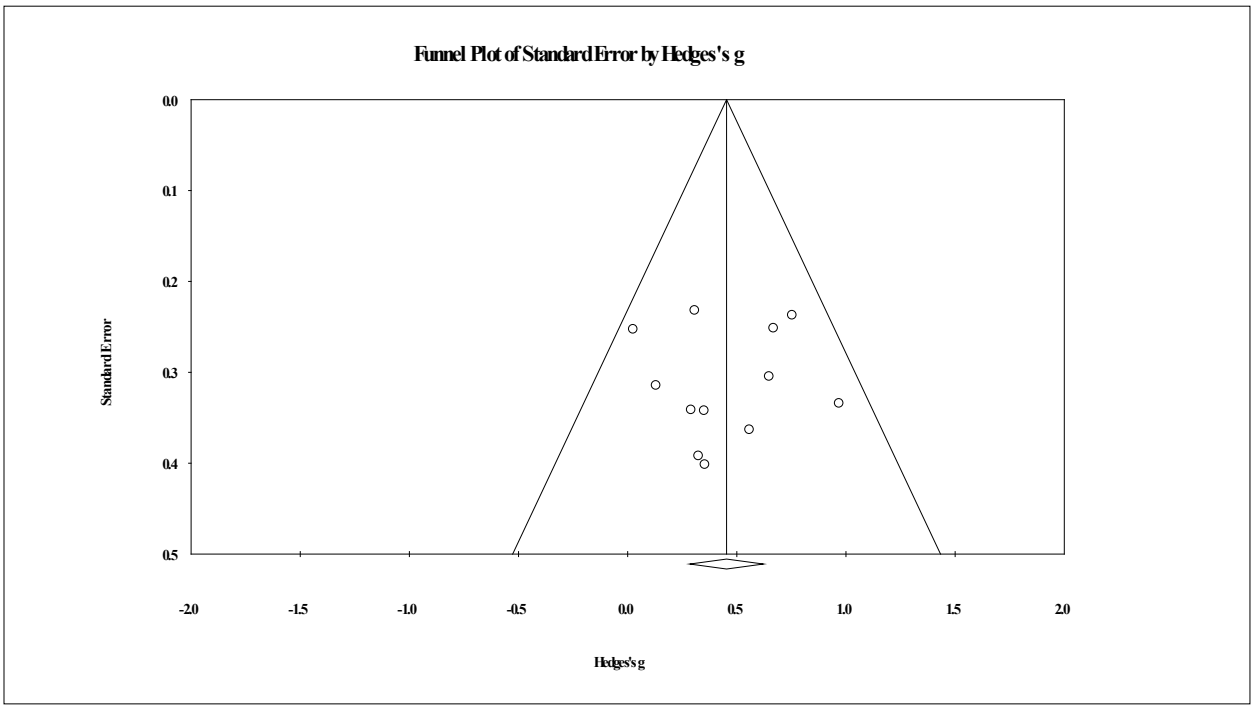
Domain	*g*	Standard Error	Variance	Lower Limit	Upper Limit	*z* value	*p* value
Studies Rated I and II							
Verbal Working Memory	0.57	0.14	0.02	0.29	0.84	4.01	0.001
Processing Speed	0.56	0.13	0.02	0.3	0.82	4.21	0.001
Verbal Memory (list)	0.54	0.05	0.003	0.43	0.65	9.84	0.001
Attention/Concentration	0.48	0.11	0.01	0.25	0.72	4.09	0.001
Executive Functioning	0.36	0.11	0.01	0.14	0.59	3.21	0.001
Visual Memory	0.33	0.12	0.02	0.09	0.58	2.68	0.007
Studies Rated III and IV							
Expressive Language	1.02	0.23	0.05	0.57	1.47	4.46	0.001
Attention Concentration	0.64	0.09	0.01	0.46	0.83	6.8	0.001
Verbal Memory (list)	0.55	0.09	0.01	0.38	0.72	6.41	0.001
Perceptual Reasoning	0.52	0.15	0.02	0.22	0.81	3.44	0.001
Processing Speed	0.33	0.11	0.01	0.11	0.56	2.89	0.004
Executive Functioning	0.33	0.11	0.01	0.12	0.54	3.09	0.002
Visual Memory	0.32	0.14	0.02	0.04	0.59	2.26	0.024
Verbal Working Memory	0.31	0.1	0.01	0.12	0.5	3.16	0.002

Discussion

Taking the recommendations from Carroll et al. (2004) this updated meta-analysis on the long-term cognitive outcomes following mTBI included studies from January 2003 to August 2010 and sought to increase homogeneity in study inclusion. An established criteria to rate the methodological rigor of studies was employed (Edlund et al., 2004) to increase homogeneity within our sample and to limit potential sources of error that can enter into studies with less rigorous methodologies (Cook & Campbell, 1979). Additionally, to get the most objective sense of the data, I did not enter zeros into the data set when authors indicated null results for either

Figure 2

Funnel Plot for all Studies around the Combined Effect



neuropsychological tests or neuropsychological domain and instead attempted to procure data from authors. Several authors were asked to provide additional data and only two responded. However, the studies whose authors provided these additional data did not meet final inclusion criteria and therefore the data was not entered into the current data set.

Search and rating procedures for the time period from 2003 to 2010 produced a total of 12 studies (k = 12) that met inclusion criteria. Seven of those studies were given an AAN ranking of I or II and five were given an AAN ranking of III or IV. In studies ranked I-II, a total of 141 mTBI participants and 140 control participants were assessed. In studies ranked III-IV, a total of 191 mTBI participants and 121 control participants were assessed. A total of 188 means and standard deviations were analyzed; 102 of which were from articles with an AAN rating of I or II and 86 of which were from articles with an AAN rating of III or IV.

Results from this meta-analysis indicate statistically significant long-term cognitive effects following mTBI can be observed in studies from January 2003 to August 2010. When all studies were combined, a significant effect size of $g = .45$ was shown which represents a small to moderate effect size based on Cohen's (1992) guidelines. When studies met criteria for an AAN rating of I or II, a significant effect for long-term cognitive sequelae following mTBI was observed ($g = 0.52$) which represents a moderate effect size based on Cohen's guidelines. When studies met criteria for an AAN rating of III or IV, a small but significant effect for long-term cognitive sequelae following mTBI was observed ($g = 0.38$). These differences in effect sizes between studies rated I-II and III-IV, however, were not significantly different when a *t*-test was performed. Thus, it is possible that methodological rigor as defined by AAN criteria does not statistically predict larger effect sizes for studies rated I and II. It is possible, however, that the *t*-test was underpowered and given the difference in effect sizes between studies rated I and II ($g = 0.52$) and those rated III and IV ($g = 0.38$), if more studies meeting our inclusion criteria and AAN criteria were available, a significant effect may be detected. Cohen (1992) points out that, for a significant difference between means to be observed with a medium effect size (i.e., $d = .50$), power of .80, and $\alpha = .05$, 64 observances in each group would be necessary. This study only had seven in one group and five in the other.

In studies rated I-II, significant effects for neuropsychological domain (with at least five data points available for analysis) were from $g = 0.57$ for verbal working memory to $g = 0.33$ for visual memory (see Table 10 above). In studies rated III-IV, significant effects for neuropsychological domain (with at least five data points available for analysis) were from $g = 1.02$ for expressive language to $g = 0.31$ for verbal working memory (see Table 10 above).

These results are in contrast to the findings of previous meta-analyses (see Table 11 below). Differences in meta-analytic methodology may account for these differences. First, we used data from January 2003 to August 2010 to determine if studies in mTBI have demonstrably improved methodological quality since Carroll et al. (2004) first identified the poor quality of the mTBI literature. Second, we used AAN's established criteria for clinical practice guidelines (Edlund et al., 2004) to determine the methodological quality of a study in order to give us a reasonable assurance of limited bias and increased heterogeneity within the source studies that met inclusion criteria. Despite this quality assurance process, or perhaps because of it, we were able to identify only seven studies that met AAN criteria for a rating of I or II and which also met our inclusion criteria and only five that met AAN criteria for a rating of III or IV and which also met our inclusion criteria. Therefore, the paucity of quality research in mTBI up to August 2010 continues particularly as it relates to experimenters being blind to group when administering neuropsychological assessments. Previous meta-analyses did not use this type of scrupulous filtering process in the selection of articles that was based on an accepted rating of methodological rigor and therefore the front-end assurance of homogeneity among included studies cannot be as easily verified. Lastly, we did not enter zeros into our data set when authors reported null effects for neuropsychological test or neuropsychological domain as previous meta-analysts have done. We did not enter zeros into our data set because it is difficult to determine how many zeros should be entered into the data set if information is missing from published studies. The process of adding zeros in place of missing data then becomes completely subjective and, if not reported in the meta-analysis, not open to peer review.

Table 11

Comparison of Effect Sizes Reported by Meta-analysts

Meta-analysis	Effect Size
Binder et al., 1997	$g = 0.07$
Frencham et al., 2005	$g = 0.11$
Rohling et al., 2011	$g = -0.07$
Present Meta-analysis	$g = 0.52$

Although this study demonstrated some improvements from previous meta-analyses there are nevertheless limitations to the study. For instance, criteria were used to rate the methodological rigor of neuropsychological studies that were not specifically developed for neuropsychology. It is possible that the criteria employed by AAN do not meet the exact needs of neuropsychology. It is because of this that the criteria in the Cappa et al. (2011) meta-analysis need to be further developed and honed to the specific needs of neuropsychology. Once this rating system is in place, authors will have an a priori outline of what is expected in terms of outstanding methodological rigor and thus additional methodologically sound meta-analyses can be conducted to further our understanding of the potential long-term sequelae of mTBI.

A second limitation of the current study was related to the AAN criterion assessing "blinding" or "masking." According to this criterion, a study component is considered blind if the assessor is not aware of the participants' group affiliation. There were no studies that met inclusion criteria in which an assessor of a neuropsychological test was blind to group affiliation. Therefore, if any one component of a study in which the authors were blind to group affiliation (e.g., neuroradiologist was blind to group when rating brain scans) the study was determined to meet the criteria for being blind. These studies were downgraded to a rating of II instead of I because the assessor of the primary variable of interest was not blind to group. Additionally, studies which used computerized testing were determined to meet "masking" criteria for those

measures given the experimenters less direct influence on the participant and were given a rating of II. That is, computerized measures have been empirically shown to foster less participant test anxiety and negative evaluations of the testing condition compared to paper- and-pencil testing which may influence test performance (Collerton et al., 2007; Fritts & Marszalek, 2010).

It is of note that an extraordinary increase in the amount of mTBI publications has continued since the cutoff point (i.e., August 2010) of the current study. Recall that my initial search from January 2003 to August 2010 resulted in 1,631 articles. A cursory PubMed search dated 04/25/2013, restricted to articles published since August 2010, and with the search terms "mild traumatic brain injury neuropsychological" yielded 200 articles. Thus, in 32 months' time, this amounts to approximately six published studies per month. It is true, however, that not all of these studies will be experimental (e.g., reviews, case studies, etc.) and not all will be related to mTBI (e.g., some studies involving stroke or more severe brain injury were also captured in the search). However, even if 1/3 of these studies were related to the neuropsychological outcomes following mTBI, this would still yield approximately two published studies per month on the topic. Although studies confirming (e.g., Dean & Sterr, 2013; Zhou et al., 2013) and disconfirming (Lange, Iverson, Brubacher, Madler, & Heran, 2012) long-term sequelae of mTBI continue to be published, I am still unaware of any large scale studies that are prospective, longitudinal, and utilize an orthopedically injured control group to better understand the potential long-term effects of mTBI. These types of studies are critical to our understanding of the potential long-term sequelae of mTBI and are needed for a better meta-analytic understanding of the long-term sequelae of mTBI.

Conceptual Critique

Given that the type of studies that are most beneficial for meta-analysis have simply not been conducted and given that multiple concerns remain in the field of mTBI research, the following critique is offered in an attempt to further our knowledge regarding mTBI.

Determining Methodological Rigor for Meta-Analyses

As mentioned above, Cappa et al. (2011) showed the importance of establishing conclusions from meta-analyses involving TBI based on a front-end selection process. The importance of using a front-end selection process to select studies to include in a meta-analysis is two-fold: first, to have reasonable assurance of the methodological quality of studies and second, to establish a reasonable heterogeneity amongst included studies. These authors developed an excellent system of rating methodological rigor based on a rating scale from "A" to "I" with an overall rating from zero to nine points. They used this system to determine front-end, methodologically-based inclusion criteria that increase the likelihood that included studies will be less influenced by potential bias and increase homogeneity within the studies meeting inclusion criteria. Although this system has not been established as a gold standard and continues to be developed, it shows excellent promise for becoming a system by which methodological rigor can be assessed in studies specific to neuropsychology.

In order to further the development of Cappa and colleagues (2011) rating criteria, I offer the following suggestions for improvement. For instance, based on criterion A and E, credit would not be given to cross-sectional studies in which description of participants lost to attrition would not be reported as it should be in prospective/longitudinal studies. Thus, well designed cross-sectional studies would be downgraded both for criterion A (not prospective /longitudinal) and Criterion E (no report of those lost to attrition because they are measured only once).

Further, these authors indicated that a study would not meet criterion B if the study did not use multivariate modeling. However, studies using univariate statistical models give useful information and no explanation is given by Cappa and colleagues why multivariate modeling is methodologically preferred over univariate. Moreover, the definition of "blinding of assessor" in criterion I is important, but does not fully capture the idea of an assessor being blind to group. A more specific definition such as "assessor of neuropsychological functioning was blind to group affiliation" should be added to criterion I in order to make explicit the need to limit experimenter influence on test performance. Computerized testing also needs to be addressed in the definition of blinding and specific recommendations made regarding computerized testing should be proffered. In reference to criterion H, it appears that Cappa and colleagues had specific reasons for stating that "The *majority* of the variables [must be] assessed using standardized measures for which normative data exists and/or reliability and validity analyses have been conducted" (p. 544, emphasis added). In rating the methodological rigor of an article, it is foreseeable that by stating *all the variables* (rather than the "majority") must be assessed using standardized measures may unnecessarily lower the rating of some well-designed studies that use experimental means to measure cognitive performance. However, it should also be noted that, when meta-analyzing cognitive performance data, experimental test procedures should not be included quite simply because we do not know the reliability and validity of these measures (Pertab et al., 2009). Additionally, in order to objectively assess the specific reliability and validity of a measure, the lower limit acceptable for reliability and validity coefficients (e.g., is it 0.80, 0.90, etc?) should be specified and met in order to receive credit for criterion H. Criterion J should be added which stipulates that symptom validity measures should be given in order to test for effort. Although symptom validity testing is still under development and is itself a hotly

debated topic (Bigler, 2012), given that effort has been shown to significantly impact neuropsychological test performance, as symptom validity testing continues to improve it will be important to assess and control for it (An, Zakzanis, & Joordens, 2012; Fox, 2011). Penultimately, a criterion K can be added delineating definitions for diagnoses used for mTBI and criteria for residual impairment following an mTBI. When definitional analyses are available, the variance associated with different definitions can be assessed and the differences between definitions examined. Lastly, once their criteria for assessing methodological rigor have been refined, it will be important to conduct studies to determine which criteria account for the most variance in long-term sequelae after mTBI. In this way, amount of variance accounted for by criteria can help weight the importance of each criterion and more credence can then be given to those criteria that account for more variance in outcome.

As this system of rating methodological rigor continues to develop, it will be a much needed addition to neuropsychological research, to our ability to determine methodological rigor, and to improve our meta-analytic understanding of the long-term cognitive sequelae of mTBI and other neuropsychological constructs of interest. Once a rating system has been put in place, I propose that all neuropsychological studies be rated and ranked according to that standard. This ranking should become as essential as an abstract. Just as an abstract gives a brief detailing of a study, a standardized ranking would also give a brief detailing of the methodological rigor of a study. Thus, consumers of the research can be aware of possible concerns regarding the validity and potential bias (Cook & Campbell, 1979) associated with each ranking and further critique studies based on this information. Edlund et al. (2004) state:

> An important step in developing a guideline is to measure the risk of bias in each included study. Bias, or systematic error, is the study's tendency to

> inaccurately measure the intervention's effect on the outcome. It is not possible to directly measure the bias of a study…However, using well-established principles of good study design, we can estimate the *risk* of bias of a study. (p. 18)

Using a system of rating methodological rigor can inform consumers of potential levels of bias associated with a given study, thus allowing informed evaluation and decision making regarding the merits of the inferences and conclusions of the study to improve clinical practice.

The Problem of mTBI Definition

As indicated above, definitional problems of mTBI have long been disputed (Bigler, 2008; Carroll et al., 2004; Ruff et al., 2009). These problems stem in large part due to the means of assessing mTBI. Currently, it is largely accepted that to diagnose mTBI it must be as a result of external force or rapid acceleration/deceleration forces that disrupt brain function and that three criteria, based on three means of behavioral assessment, are needed in order to diagnose mTBI following the trauma. These criteria include: Glasgow Coma Scale (GCS) from 13 to 15, loss of consciousness (LOC) for less than 30 minutes, and post traumatic amnesia (PTA) of less than 24 hours. However, qualifiers such as "focal neurological signs," "altered consciousness," or "period of confusion" are often added and definitionally sufficient for diagnosing a concussion (American Congress of Rehabilitation Medicine Head Injury Interdisciplinary Special Interest Group, 1993; Carroll et al., 2004).

These criteria are problematic because they are often associated with measurement error. For example, GCS is often not assessed within a reasonable timeframe to determine severity of injury (e.g., not assessed at time of injury or shortly thereafter) and sometimes it is not assessed at all. Additionally, although some have suggested that GCS should be measured approximately

30 minutes following a concussion (American Congress of Rehabilitation Medicine, 1993; Carroll et al., 2004), there is no agreed upon, empirically based timeframe during which GCS should be assessed in order to understand the severity of the head injury. Although Carroll and colleagues are well intended in their attempt to address this concern, there is still left an extremely large margin of potential error when GCS is measured. These authors state:

> We agree with the American Congress of Rehabilitation Medicine definition, which specifies that the GCS score of 13–15 be assessed after 30 minutes post-injury. However, we recognize the practical concern that individuals with MTBI will rarely be assessed at an emergency department within this time frame. Therefore, although an assessment of GCS score just after 30 minutes post-injury remains the ideal, our proposed definition permits diagnostic use of a GCS score assessed by a qualified healthcare provider at the first opportunity. (p. 115)

Although their definition allows for diagnostic use of GCS score "at the first opportunity," there is no telling when this first opportunity may occur, if at all. Thus, the question then becomes does an individual who has a GCS of 13 at 30 minutes post injury have the same level of injury as an individual with a GCS of 13 at 120 minutes post injury? Or 240 minutes post injury?

LOC and PTA have similar problems with being assessed and measured. LOC, for instance, is usually assessed by asking the patient if he/she lost consciousness after the accident. However, patients who have been acutely concussed (or those trying to recall the details of their accident days or weeks after the injury occurred) are typically unable to accurately self-report (Ruff et al., 2009). Additionally, if an individual loses consciousness, it is difficult for them to

accurately gauge how long the period of unconsciousness actually lasted. It is similarly difficult for them to tease apart LOC from PTA. The question is often, "Did I lose consciousness or do I simply not recall what happened?" It is clearly better to have collateral data from an outside observer reporting on whether an individual was unconscious and, if so, for how long. However, this is also problematic as the outside observers rarely actually time the duration of the LOC and therefore rely on memory and approximations to gauge the length of LOC if present. Moreover, is an LOC of 30 seconds equivalent to the level of neurologic injury as an LOC of 30 minutes? Should these times of LOC be equated in a definition of mTBI? Although attempts to alleviate this problem have been made by using "grades" of mTBI based on more specific criteria (American Academy of Neurology, 1997; Leclerc, Lassonde, Delaney, Lacroix, & Johnston, 2001), grading of mTBI still suffers from the same problems of assessment and measurement that are associated with assessing GCS, LOC, and PTA as outlined here.

Assessing PTA is similarly problematic in that many individuals have been told about what happened at the time of their accident several times which can cause a memory trace, albeit an often false (or at least biased) memory trace, that is often difficult to distinguish from what the individual actually recalls from the accident. As Ruff and colleagues (2009) point out, "it is essential to determine what the patient remembers versus what he or she has been told or has surmised" (p. 6). These authors further point out that assessing PTA can be problematic due to intoxication, psychiatric concerns (e.g., acute stress disorder, post-traumatic stress disorder), or psychogenic amnesia caused by significant emotional trauma. Thus, individuals are often left piecing together and making sense of the experience based on their own recollections, what others have told them, and those details that are simply lost in the complexity of the process. This confusion between actual memory and a memory trace that was created by hearing the story

of the accident (with relevant details possibly missing) makes assessing PTA quite problematic. Moreover, there is evidence that PTA and LOC are not necessarily part of the sine qua non of mTBI (Smits et al., 2007). Smits and colleagues (2007) showed that those with mTBI may or may not demonstrate LOC and PTA. These authors point out, for example, that of their 2,462 study participants who consecutively presented to the emergency department those individuals in their study who had mTBI and who required neurosurgical intervention did not significantly depend on whether PTA and LOC were present (six patients with PTA and LOC versus four patients with no PTA and LOC). Interestingly, although all participants in the study met criteria for mTBI on behavioral observation (e.g., GCS, PTA, LOC) these 10 patients nevertheless displayed indications for neurosurgery which included "isolated depressed skull fracture (n=1), epidural haematoma (n=4), subdural haematoma (n=4) and a combination of epidural and subdural haematoma (n=1)" (p. 1,362).

Given these problems, researchers are trying to determine more objective, biological markers to define and diagnose mTBI. Brain imaging has been a focus of identifying biological markers for psychiatric concerns for many years (Farah & Gilihan, 2012). These methods are also being used to identify biological markers of brain injury that potentially will aid with the definitional problems discussed above and lead to better prediction of outcome and treatment options (Bigler & Bazarian, 2010; Mayer et al., 2010).

Although this literature is too large to fully review here, a few instances will illustrate the point. For example, Huang and colleagues (2009) point out that mTBI can often be difficult to objectively diagnose given lack of external injuries and because neurological damage if present is often not detected on conventional acute MRI or CT. Given the capability of MEG to detect the low-frequency neuronal magnetic signal generated following TBI and the capability of DTI

to identify abnormalities of white matter tracts, Huang and colleagues (2009) integrated the two imaging modalities to determine whether these technologies would be more sensitive in detecting subtle neuronal injury in mTBI than conventional neuroimaging. These researchers' data showed four relevant findings: 1) MEG and DTI combined are more sensitive than conventional MRI and CT to detect neuronal damage in mTBI, 2) MEG slow waves tended to originate from cortical gray matter areas that experienced de-afferentation from axonal damage to white matter tracts, 3) post-concussive symptoms are consistent with MEG and DTI findings, and 4) in some cases where DTI did not show abnormalities in white matter MEG was still able to detect abnormal neurological signals.

Chen and colleagues (2003) offer another example of an advanced imaging technology which detected neurological changes and subsequent neuropsychological deficits associated with mTBI. These researchers used PET to compare a group of mTBI patients to a group of controls during resting state and during a spatial working memory task. Findings indicated that no differences were found between groups in resting state fluro-2-deoxy-D-glucose (FDG) uptake in frontal and temporal regions of interest. However, differences were identified during the spatial working memory task with patients displaying smaller increases in regional cerebral blood flow (rCBF) in right prefrontal cortex compared to controls. Chen and colleagues indicate that their findings suggest that a cognitive challenge may be necessary to identify neurological changes associated with mTBI that are not detectable with conventional neuroimaging or on advanced imaging techniques during resting states.

In summary, as brain imaging modalities such as MEG, SPECT, PET, DTI, SWI continue to improve, they will become more useful in determining structural and functional neurological changes and will likely help alleviate not only definitional problems associated with mTBI but

also help determine neurologically based prognostic indicators and treatment foci (Huang et al., 2009; Mayer et al., 2010; Niogi & Mukherjee, 2010; Niogi et al., 2008; Shenton et al., 2012; Toledo et al., 2012; Zhang et al., 2006). In this way, mTBI diagnosis can be based on radiographic substantiation rather than by symptoms alone that are often self-report and often unreliable in nature.

Along with neuroimaging techniques, neuro-metabolic changes have also been studied as biomarkers that have potential to further our understanding of brain pathology following mTBI (McCrea, 2008). Indeed, lumbar puncture and peripheral blood based measurements of metabolic processes following brain injury such as neurofilament light protein, glial fibrillary acidic protein, phosphorylated tau, S100 proteins, neuron specific enolase, and β-amyloid protein may continue to show reliable and valid neurologic changes that can be used to define head injury severity and which may correlate with neuropsychological performance (de Kruijk, et al., 2002; Nygren de Boussard, et al., 2004).

Again, this literature is becoming quite large and I'm not able to fully review it here. Nevertheless, an example will illustrate the potential of these methods in helping understand definitional, prognostic, and treatment factors associated with mTBI. de Kruijk and colleagues (2002) conducted a prospective study of 79 patients seen in the emergency department (ED) less than six hours after injury and seen again six months later for follow up. Findings indicated that of the 79 patients 22 (28%) reported one or more post-traumatic complaints at the six-month follow up. In these patients, a twofold increase in severity of cognitive and vegetative complaints at six-month follow up was associated with increased blood concentrations of the biochemical serum markers S-100B and neuron specific enolase at first presentation to the ED. These researchers found that headache, dizziness, and nausea at initial ED visit were strongly

associated with the severity of post-traumatic complaints at six-month follow up. Results further indicated that those with normal serum markers and no neurobehavioral symptoms upon arrival at the ED recovered fully (n = 10) indicating that those who recover fully and those who don't can be better predicted based on these serum markers.

As measurement of neuro-metabolic change continues to develop, we will likely be in a better position to help us understand the nuances of the potential neurobehavioral impairments associated with mTBI and help re-define our understanding of mTBI. Indeed, lumbar puncture to identify potential metabolic processes following TBI is still recommended in the emergency department alongside CT scans when ruling out hemorrhagic injury following stroke or head injury when headache, neck pain, or LOC are present (Mark et al., in press) simply because CT scans do not reveal approximately 20% of hemorrhagic injuries.

In conclusion, the concern about definitional issues is so pressing that Cassidy and colleagues (2004) state:

> There is an urgent need for workable clinical and surveillance definitions of MTBI and subsequent studies to validate various methods of capturing cases. Until there is some consistency of definitions and appropriate validation of them, studies of the incidence of MTBI will remain so heterogeneous that we will be unable to compare the incidence rates. (p. 2004)

However, since 2004 when Cassidy and colleagues made this statement, little progress has been made in demonstrating a reliable and valid definition of mTBI. Thus, we are no closer to being able to compare incidence rates or capture cases. Advanced technologies such as imaging and serum biomarkers may help elucidate the elusive definition of mTBI. Once biomarkers are identified, correlations to neurocognitive and

neurobehavioral sequelae can then be better understood and teased apart from potential non-neurological causes of decreased cognitive performance (e.g., effort, psychological concerns, sleep deficiencies, etc.).

The Problem of Residual Cognitive Impairment Definition

In their recent meta-analysis, Rohling et al. (2011) stated that it remains possible but unlikely that a small number of individuals may continue to experience long-term cognitive sequelae following mTBI. Theses authors correctly point out that other factors such as poor effort, premorbid conditions (such as attention deficit/hyperactivity disorder or learning disorders), pain, or fatigue need to be ruled out before determining whether cognitive deficits are associated with the mTBI. However, in order to determine whether long-term cognitive deficits are associated with mTBI, there needs to be an accepted definition of residual cognitive impairment following mTBI.

Few efforts have been made to determine reliable and valid criteria against which residual impairment can be weighed. For example, in addition to Iverson and Brooks' (2011) work on defining residual cognitive impairment using the Neuropsychological Assessment Battery (NAB), there have been only two other primary attempts to define residual cognitive impairment following acquired brain injury (e.g., Reitan and Wolfson's work on the Halstead Reitan Neuropsychological Battery [1985, 1993], and Golden and colleagues' work on the Luria-Nebraska Neuropsychological Battery [1985, 2000]. Unfortunately, these three definitions of cognitive impairment have relied on "fixed" batteries of neuropsychological assessment which have fallen out of favor in more recent years for "flexible" batteries making the definitions by the above mentioned researchers less user friendly to today's neuropsychologist. Iverson and Brooks (2011) correctly point out a couple of points that bear mentioning here. First, they correctly state

that there is no agreed upon definition of cognitive impairment that is based on sound statistical parameters stating that even their own proposed criteria need additional research to refine the criteria. Second, they point out that there is no accepted and empirically-validated psychometric criteria for identifying cognitive disorders most likely due to the fact that there is no agreement on the definition of a "low score" (i.e., 1SD, 2SD) that would help determine whether an individual experiences a cognitive disorder. These two points (among others) are then left entirely up to the individual neuropsychologist to resolve on a patient to patient basis.

Given the heterogeneity in causes of mTBI (car accident vs. fall), heterogeneity in biomechanical forces involved (rotational force vs. translational force), and the heterogeneity in individual premorbid variability/differences (gender, age, intelligence, education, ethnicity), how can we determine what residual impairment is for a group of individuals in a research study (when these variables for the most part are not being considered) let alone for the individual who is in our office being assessed? What about clinical significance as opposed to statistical significance (Jacobson & Truax, 1991)? Is a standard score of 95 on a measure of intelligence indicative of impairment relative to a normative group? What about relative to the patient's premorbid level of functioning, let's say, if his premorbid intellectual functioning was measured at 120? Can an "average" score be an impaired score for an individual who was once able to fly jet aircraft in combat and who previously scored in the superior range? It is true, that perhaps an individual like this may be the most highly functioning individual in his new office job, but he/she was once a fighter pilot.

Moreover, *inter*individual variability is ostensibly accounted for by normative data when appropriate norms are used. But *intra*individual variability (Binder & Binder, 2011; Iverson & Brooks, 2011) remains unaccounted for and needs to be addressed in a definition of residual

cognitive impairment. That is, because variability exists in any given individual's neuropsychological tests results (Binder & Binder, 2011; Iverson & Brooks, 2011), a definition of what is considered normal and aberrant variability within that individual's neuropsychological test results needs to be considered when defining residual cognitive impairment (Iverson & Brooks, 2011). Because there is no accepted definition of residual cognitive impairment following mTBI, one cannot objectively state whether the individual continues to experience cognitive impairment relative to the patient his/herself, and, in a clinical setting the decision is left to the assessing provider to make that subjective determination without a definition and empirically-based guideline.

Interestingly, researchers (as opposed to clinicians) implicitly use statistically significant differences between groups as an ostensible analogue for a definition of residual cognitive impairment. That is, when a statistically significant difference between an mTBI patient group and a control group is observed, that statistical difference is implied to mean that there is "residual impairment" in the mTBI group in that domain of cognitive functioning. But how do we know there is residual impairment? It is possible that other factors may account for the statistical difference observed especially when we don't even know what we mean when we say "residual impairment" and when most studies don't take into consideration variables such as gender, intellectual abilities, ethnicity, education, cognitive reserve, biomechanical forces, etc. (Bigler, 2008; Farias et al., 2012; Iverson & Brooks, 2011; Pertab et al., 2009). What exactly do researchers want to imply when statistically significant differences are observed in the patient group? Do they actually mean to imply that the mTBI group displayed residual cognitive impairment(s)?

Additionally, traditional pencil-and-paper neuropsychological tests have been criticized as possibly not being sensitive enough to detect the changes associated with mTBI in order to detect residual impairment (Bigler, 2008; Bigler &Bazarian, 2010; Collie et al., 2006; Mayer et al., 2010). Thus, more sophisticated means of testing should continue to be developed. Computerized assessment is becoming more automated and more available and has been used quite extensively now in the assessment of sports-related brain injuries (Cernich, Reeves, Sun, Bleiberg, 2007; Iverson, Lovell, & Collins, 2005; Leclerc et al., 2001). Computerized testing using cognitive neuroscience measures has been and continues to be configured to be compatible with brain imaging techniques thereby increasing the sensitivity to neurologic dysfunction (Chen, Johnston, Collie, McCrory, & Ptito, 2007; Mayer, Mannell, Ling, Gasparovic, & Yeo, 2011; Mayer et al., 2012; Scheibel et al., 2007). As cognitive neuroscience measures of brain function become standardized and norms created, we will be better able to detect neurologic changes that occur on the scale of milliseconds thereby making these measures more sensitive than traditional pencil-and-paper neuropsychological tests which typically measure on the scales of minutes and seconds. As these measures become more accurate in describing potential neurocognitive dysfunction following mTBI, they can also be a tool used to help us refine our definition of residual impairment.

The Problem of PCS and PPCS

The term post-concussion syndrome (PCS) has been used to describe the complex, controversial cluster of physical, cognitive, behavioral, and emotional symptoms that may occur in association with mTBI, and the term persistent post concussion symptoms (PPCS) has been used when these symptoms continue post three months (Arciniegas, Anderson, Topkoff, & McAllister, 2005; Bigler, 2008; Hall, Hall, & Chapman, 2005; Satz et al., 1999). PCS and PPCS

symptoms often include headache, fatigue, insomnia, dizziness, concentration/attention difficulties, memory loss, irritability, sensitivity to sensory stimuli, and emotional instability/dysregulation (Bazarian et al., 1999; Bigler, 2008; Ryan & Warden, 2003; Satz et al., 1999).

PCS symptoms have been criticized as being nonspecific to mTBI patients and have thus been viewed as having little to no value when ruling in/out mTBI as the cause of these symptoms (Lees-Haley, Fox, & Courtney, 2001; Silva, Donnell, Kim, & Vanderploeg, 2012). Although it is true that many PCS symptoms overlap with and could be accounted for by other concerns (e.g., psychological factors, medication, medical problems, premorbid psychiatric history, etc), that PCS and PPCS still have definitional problems, and are in fact considered works in progress (Hall, Hall, & Chapman, 2006; Smith, 2006), it does not necessarily stand that these symptoms could not be accounted for neurologically in some individuals and therefore related to mTBI (Shenton et al., 2012).

Related to the above discussion regarding definitional problems, the potential for advanced neuroimaging techniques to further our understanding of the neurological involvement in PCS and PPCS is also quite possible. Shenton and coworkers (2012) conducted a thorough review of the MRI and DTI literature involving mTBI and found overwhelming evidence in the 43 DTI/mTBI studies they reviewed in favor of DTI being able to detect white matter abnormalities in mTBI above conventional neuroimaging. They state:

> Given the different magnet strengths, with some [studies] conducted on a 1.5 T magnet, and others conducted on a 3 T magnet…, as well as differences in the analysis methods employed, and the dependent measures used, as well as differences in the selection of brain regions to investigate, in addition to

> differences in the post-injury time of the study, and differences in whether subjects had positive or negative findings on conventional CT or MRI, it is surprising that there is as much convergence and consistency with respect to the detection of brain abnormalities in mTBI using DTI. (p. 180)

Shenton and colleagues (2012) go on to point out that, despite the various regions of interest studied using DTI, the corpus callosum was consistently determined to be particularly susceptible to injury in mTBI as identified on DTI. However, in patients who recover fully from the concussion injury, there is evidence that a healing process which reverses corpus callosum damage may occur and that more severe corpus callosum damage without this healing process may be associated with poorer long-term outcome (Rutgers et al., 2008). It is therefore possible that these more severe neurological changes and lack of healing process are a part of the causal factors associated with PCS and which may lead to PPCS.

Despite the difficulties conventional imaging modalities have in detecting neurologic abnormalities, Arciniegas and colleagues (2005), after conducting a review of the literature stated that "biomechanical and cytotoxic consequenses of mild TBI [such as calcium and magnesium regulation, free radical formation, neurotransmitter excitotoxicity, inflammatory responses, disruption of vascular homeostasis]may be substantial despite an ostensibly 'mild' mechanism of injury" (p. 313). It is precisely these biomechanical and cytotoxic consequences that advanced imaging technologies are becoming more sensative to detect and it is possibly these very processes that may lead to long-term complications for some individuals following mTBI. As advanced technologies continue to be developed, our understanding of the neurological sequealae associated with mTBI will be further elucidated and therefore the etiology of symptoms whether they be psychogenic, neurogenic, or otherwise will be more

distinguishable. Being able to differentiate between the various etiologies of PCS-like symptoms is important for clinicians to correctly diagnose in order to prescribe appropriate treatment and lead to better outcomes (Potter & Brown, 2012; Ruff, 2005).

The Problem of Heterogeneous Outcome Measurement

Researchers have long lamented the difficulty of fully integrating and understanding the heterogeneity of outcome data in TBI research (Nightingale, Soo, & Tate, 2007; Willemse-van Son, Ribbers, Verhagen, & Stam, 2007).The need for a homogeneous system of outcome measurement can also be seen in recent meta-analytic techniques. Cappa and colleagues (2011) very nicely demonstrated the problem with the heterogeneity of outcome measurement in neuropsychology. Although these researchers did not focus solely on mTBI, they nevertheless found a significant *Q* statistic for neuropsychological outcome measures and found that that outcome measure was a significant moderator of the TBI severity and outcome relationship. That is, the link between injury severity and outcome varied based on the neuropsychological outcome measure that was used. Cappa and coworkers point out that this level of heterogeneity in outcome measure makes an integrated understanding of the long-term sequelae of TBI difficult.

In response to these laments, Wilde and colleagues (2010), who are a part of the interdisciplinary Common Data Elements (CDE) workgroup for common outcome measures in TBI research, developed a 3-tier system to give guidance to researchers, agencies, and other populations to facilitate the adaption of a common set of outcome measures to better understand the effects of all severity ranges of TBI (Nightingale, Soo, & Tate, 2007; Thurmond et al., 2010; Willemse-van Son, Ribbers, Verhagen, & Stam, 2007). The 3-tier system Wilde and coworkers (2010) developed consisted of core (tier-one), supplemental (tier-two), and emerging (tier-three) measures. Firstly, these researchers identified 12 outcome domains which are relevant to

understanding brain injury. They then identified six criteria by which they would evaluate an outcome measure's utility of assessing these domains. Finally, they identified nine tier-one core measures, twenty tier-two supplemental measures, and nine tier-three emerging measures. Wilde and colleagues describe the rational for the 3-tier system and state:

> In the first tier, core measures included valid, robust, and widely applicable outcome measures with proven utility in TBI from each identified domain, including global level of function, neuropsychological impairment, psychological status, TBI-related symptoms, executive functions, cognitive and physical activity limitations, social role participation, and perceived health-related quality of life. In the second tier, supplemental measures were recommended for consideration in TBI research focusing on specific topics or populations. In the third tier, emerging measures included important instruments currently under development, in the process of validation, or nearing the point of published findings that have significant potential to be superior to some older ("legacy") measures in the core and supplemental lists and may eventually replace them as evidence for their utility emerges. (p. 1650)

Once this type of system is adopted, a better homogeneity amongst outcome measures used in research will be achieved and less confusion will be had regarding the integration of heterogeneous measurement. This is important because having a system that guides research based on accepted outcome measurement will influence researchers to begin using the system and will then allow for a better, more fully integrated understanding of the long-term cognitive sequelae of TBI in general and mTBI specifically.

The Problem of Unaccounted for Variables

In order to better understand the nuances in mTBI, it is important that careful analysis of all potentially influential variables be conducted. For example, mechanism of injury has been viewed as an important variable because it may be associated with different neuropsychological profiles and outcomes. Bigler (2008) points out that a fall injury with translational forces may be associated with different levels of neurological impairment when compared with an injury sustained in a motor vehicle accident (MVA) with rotational forces. Moreover, Bigler (2008) further indicated that various types of MVA may also be associated with various levels of neurological impairment depending on whether the vehicle rolled, spun, and whether it was hit from behind, from the side, or from the rear. Each of these aspects associated with the wide spectrum of mechanisms of injury needs to be carefully addressed in the literature. Additionally, demographic information such as sex, age, premorbid intellectual functioning, handedness, education, and family history of neurological and/or psychiatric concerns etc. also need to be carefully studied (Iverson & Brooks, 2011).

In the current meta-analysis, 11 of 12 source studies reported the sex of participants, 11 of 12 source studies reported information on the age of participants, 3 of 12 source studies reported on handedness, 10 of 12 source studies reported on education, 2 of 12 source studies reported on ethnicity, and 7 of 12 source studies reported on psychiatric comorbidity of the samples used. Variables of particular interest that need to be accounted for in research studies because of their known influence on neuropsychological performance at a bare minimum include sex, ethnicity, education, and intellectual functioning (Iverson & Brooks, 2011). Interestingly, the nuances associated with many other aspects of assessing cognitive functioning can be further seen in a recent study by Karremans and colleagues (2009). These researchers tested men and

women who interacted with both female and male research assistants (RA). They found that when men interacted with female RAs their cognitive functioning was "impaired" when compared with when they interacted with male RAs. Thus, it is clear that there are many variables that potentially effect neuropsychological outcome and they must be accounted for in order for us to fully appreciate, understand, and diagnose mTBI.

Conclusion

Given that the literature up to 2002 has been viewed as lacking in methodological rigor (Carrol et al., 2004), we sought to further assess the state of the literature from January 2003 to August 2010 by using the AAN clinical practice guidelines to rank studies published during this time frame. Only seven studies were identified that met criteria to be given an AAN rating of I or II and five studies were found that met criteria for an AAN rating of III or IV. This small *k* is consistent with the view that methodological rigor continues to be a problem in research on mTBI, and is further evidence that there continues to be a dearth of research from January 2003 to August 2010 that utilized a high standard of methodological rigor to understand the long-term sequelae of mTBI. Improved methodological rigor is important as we continue to build our understanding of mTBI and seek to reduce error associated with less rigorous research designs.

Additionally, this study further emphasizes the potential influence that heterogeneity amongst studies included in meta-analyses may have on meta-analytic findings. This study represents the usage of the most rigorous and homogeneous inclusion criteria when compared to previous meta-analyses on mTBI to date. The larger effect sizes found in the current meta-analysis (i.e., $g = 0.52$ for studies rated I or II, $g = 0.38$ for studies rated III or IV, and an overall combined effect size of $g = 0.45$) when compared to the smaller effect sizes found in previous meta-analyses indicate that, as Konrad and colleagues (2010) stated, the debate about long-term

cognitive sequelae following mTBI remains unresolved. The problem of heterogeneity amongst included studies and the within study heterogeneity as described above are likely important variables that need to be considered in future studies.

Moreover, problems associated with using meta-analysis to justify the claim that no individuals who experience mTBI were addressed (e.g., definitional problems, measurement problems, etc.). Iverson (2010) has stated that, "…clinicians should not use the results of meta-analysis to state, *unequivocally,* that MTBI *cannot* cause residual cognitive difficulties in individual patients. This is simply an over-generalization and is invariably inaccurate at some point or other" (emphasis in original, p. 1252). Clearly concerns still exist in various aspects of research on mTBI. Definitional standards of mTBI and residual impairment of mTBI should be addressed in order to facilitate better understanding of the phenomenon and epidemiology of mTBI. Better, more consistently used instruments in evaluating cognitive functioning after mTBI should also continue to be pursued. Of particular interest are neuroimaging techniques, other potential biomarkers, and computerized assessment in future studies of mTBI.

As this work continues, it is likely that using more advanced technologies including more sophisticated means of measuring neuropsychological ability will help us define and correctly identify neurologic involvement, residual impairments, and inform treatment of symptoms associated with mTBI. Indeed, President Barak Obama recognizes the need for advancements in advanced brain-based technologies to better understand brain-behavior relationships (The White House: Office of the Press Secretary, 2013). President Barak Obama recently unveiled the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative that will allocate 100 million dollars in research support to,

> "…accelerate the development and application of new technologies that will enable researchers to produce dynamic pictures of the brain that show how individual brain cells and complex neural circuits interact at the speed of thought. These technologies will open new doors to explore how the brain records, processes, uses, stores, and retrieves vast quantities of information, and shed light on the complex links between brain function and behavior."

After all, our ultimate concern as mental health professionals when assessing mTBI should be to use clinical history, appropriate assessment measures, and informed clinical judgment to diagnose and inspire empirically validated treatments which improve the well being and quality of life of our patients. This is the explicit goal of the BRAIN initiative as it "…aims to help researchers find new ways to treat, cure, and even prevent brain disorders, such as Alzheimer's disease, epilepsy, and traumatic brain injury."

References

References marked with an asterisk indicate studies included in the meta-analysis.

American Academy of Neurology. (1997). Practice parameter: The management of concussion in sports (summary statement). Report of the quality standards subcommittee. *Neurology*, *48*, 581–585.

American Congress of Rehabilitation Medicine. (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation, 8*, 86-7.

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology, 27*(8), 849-857. doi: 10.1093/arclin/acs085

Arciniegas, D. B., Anderson, C. A., Topkoff, J., & McAllister, T. W. (2005). Mild traumatic brain injury: A neuropsychiatric approach to diagnosis, evaluation, and treatment. *Neuropsychiatric Disease and Treatment, 1*(4), 311-327.

Ashwal, S., Babikian, T., Gardner-Nichols, J., Freier, M.C., Tong, K.A., & Holshouser, B.A. (2006). Susceptibility-weighted imaging and proton magnetic resonance spectroscopy in assessment of outcome after pediatric traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *87*, 50–58.

Bazarian, J. J., Wong, T., Harris, M., Leahey, N., Mookerjee, S., & Dombovy, M. (1999). Epidemiology and predictors of post-concussive syndrome after minor head injury in an emergency population. *Brain Injury, 13*(3), 173-189. doi: 10.1080/026990599121692

Belanger, 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*(3), 215-227. doi: 10.1017/s1355617705050277

Bigler, E. D. (2008). Neuropsychology and clinical neuroscience of persistent post-concussive syndrome. *Journal of the International Neuropsychological Society, 14*(1), 1-22. doi: 10.1017/s135561770808017x

Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 632-640. doi: 10.1017/s1355617712000252

Bigler, E. D., & Bazarian, J. J. (2010). Diffusion tensor imaging: A biomarker for mild traumatic brain injury? *Neurology, 74*(8), 626-627. doi: 10.1212/WNL.0b013e3181d3e43a

Bigler, E. D., Farrer, T. J., Pertab, J. L., James, K., Petrie, J. A., & Hedges, D. W. (2013). Reaffirmed limitations of meta-analytic methods in the study of mild traumatic brain injury: A response to Rohling et al. *The Clinical Neuropsychologist*. doi:10.1080/13854046.2012.693950.

Binder, L. M., & Binder, A. L. (2011). Relative subtest scatter in the WAIS-IV standardization sample. *The Clinical Neuropsychologist, 25*(1), 62-71. doi: 10.1080/13854046.2010.533195

Binder, L. M., Rohling, M. L., & Larrabee, G. J. (1997). A review of mild head trauma: I. Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology, 19*(3), 421-431. doi: 10.1080/01688639708403870

*Blanchet, S., Paradis-Giroux, A.-A., Pépin, M., & McKerral, M. (2009). Impact of divided attention during verbal learning in young adults following mild traumatic brain injury. *Brain Injury, 23*(2), 111-122. doi: 10.1080/02699050802649688

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, UK: John Wiley& Sons, Ltd.

Cappa, K. A., Conger, J. C., & Conger, A. J. (2011). Injury severity and outcome: A meta-analysis of prospective studies on TBI outcome. *Health Psychology, 30*(5), 542-560. doi: 10.1037/a0025220

Carroll, L. J., Cassidy, J. D., Holm, L., Kraus, J., & Coronado, V. G. (2004). Methodological issues and research recommendations for mild traumatic brain injury: The WHO collaborating centre task force on mild traumatic brain injury. *Journal of Rehabilitation Medicine : Official Journal of the UEMS European Board of Physical and Rehabilitation Medicine*, *43*, 113-125.

Cassidy, J.D. (2010). Epidemiology of mild brain injury: Quality of the scientific evidence. *Journal of Neurology, Neurosurgery & Psychiatry*, *81*, e12.

Cassidy, D. J., Carroll, L. J., Peloso, P. M., Borg, J., von Holst, H., Holm, L., et al. (2004). Incidence, risk factors and prevention of mild traumatic brain injury: Results of the WHO collaborating centre task force on mild traumatic brain injury. *Journal of Rehabilitation Medicine*, *43*, 28-60.

Center for Disease Control (2003). Report to congress on mild traumatic brain injury in the united states: Steps to prevent a serious public health problem. Retrieved October 23, 2009, from http://www.cdc.gov/NCIPC/tbi/mTBI/mTBIreport.pdf

Cernich, A., Reeves, D., Sun, W., & Bleiberg, J. (2007). Automated Neuropsychological Assessment Metrics sports medicine battery. *Archives of Clinical Neuropsychology*, *22* (Suppl. 1), S101–114.

Chen, J.K., Johnston, K.M., Collie, A., McCrory, P., & Ptito, A. (2007). A validation of the post concussion symptom scale in the assessment of complex concussion using cognitive testing and functional MRI. *Journal of Neurology Neurosurgery and Psychiatry*, *78*, 1231–1238.

Chen, S.H., Kareken, D.A., Fastenau, P.S., Trexler, L.E., & Hutchins, G.D. (2003). A study of persistent post-concussion symptoms in mild head trauma using positron emission tomography. J*ournal of Neurology Neurosurgery and Psychiatry*, *74*, 326–332.

Cohen, B.A., Inglese, M., Rusinek, H., Babb, J.S., Grossman, R.I.,& Gonen, O. (2007). Proton MR spectroscopy and MRI volumetry in mild traumatic brain injury. *American Journal of Neuroradiology*, *28*, 907–913.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 115–159.

Collerton, J., Collerton, D., Arai, Y., Barrass, K., Eccles, M., Jagger, C., . . . Kirkwood, T. (2007). A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ pilot study. *Journal of the American Geriatrics Society, 55*(10), 1630-1635. doi: 10.1111/j.1532-5415.2007.01379.x

Collie, A., Makdissi, M., Maruff, P., Bennell, K., & McCrory, P. (2006). Cognition in the days following concussion: Comparison of symptomatic versus asymptomatic athletes. *J Neurology Neurosurgery and Psychiatry*, 77, 241–245.

Cook, T. D., & Cambpell D. T. (1979). *Quasi-experimentation: Design & analysis issues for filed settings*. Boston, MA: Houghton Mifflin Company.

Dean, P. J. A., & Sterr, A. (2013). Long-term effects of mild traumatic brain injury on cognitive performance. *Frontiers in Human Neuroscience*. doi: 10.3389/fnhum.2013.00030

de Kruijk, J. R., Leffers, P., Menheere, P. P. C. A., Meerhoff, S., Rutten, J., & Twijnstra, A. (2002). Prediction of post-traumatic complaints after mild traumatic brain injury: Early symptoms and biochemical markers. *Journal of Neurology, Neurosurgery & Psychiatry, 73*(6), 727-732. doi: 10.1136/jnnp.73.6.727

Dikmen, S. S., Corrigan, J. D., Levin, H. S., Machamer, J., Stiers, W., & Weisskopf, M. G. (2009). Cognitive outcome following traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 24*(6), 430-438. doi: 10.1097/HTR.0b013e3181c133e9

Edlund,W., Gronseth, G., So, Y., & Franklin, G. (2004). *Clinical practice guideline process manual: For the quality standards subcommittee (QSS) and the therapeutics and technology assessment subcommittee (TTA).* St. Paul: American Academy of Neurology.

Farah, M. J., & Gillihan, S. J. (2012). The puzzle of neuroimaging and psychiatric diagnosis: Technology and nosology in an evolving discipline. *American Journal of Bioethics Neuroscience*, 1, 31-42.

Farias, S. T., Chand, V., Bonnici, L., Baynes, K., Harvey, D., Mungas, D., . . . Reed, B. (2012). Idea density measured in late life predicts subsequent cognitive trajectories: Implications for the measurement of cognitive reserve. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 67*(6), 677-686. doi: 10.1093/geronb/gbr162

Fox, D. D. (2011). Symptom validity test failure indicates invalidity of neuropsychological tests. *The Clinical Neuropsychologist, 25*(3), 488-495. doi: 10.1080/13854046.2011.554443

Fox, W. C., Park, M. S., Belverud, S., Klugh, A., Rivet, D., & Tomlin, J. M. (2013). Contemporary imaging of mild TBI: The journey toward diffusion tensor imaging to assess neuronal damage. *Neurological Research*, *35*, 223-232.

Frencham, K. A. R., Fox, A. M., & Maybery, M. T. (2005). Neuropsychological studies of mild traumatic brain injury: A meta-analytic review of research since 1995. *Journal of Clinical and Experimental Neuropsychology, 27*(3), 334-351. doi: 10.1080/13803390490520328

Fritts, B. E., & Marszalek, J. M. (2010). Computerized adaptive testing, anxiety levels, and gender differences. *Social Psychology of Education, 13*(3), 441-458. doi: 10.1007/s11218-010-9113-3

*Ge, Y., Patel, M. B., Chen, Q., Grossman, E. J., Zhang, K., Miles, L., . . . Grossman, R. I. (2009). Assessment of thalamic perfusion in patients with mild traumatic brain injury by true FISP arterial spin labelling MR imaging at 3T. *Brain Injury, 23*(7-8), 666-674. doi: 10.1080/02699050903014899

Golden, C., J., Freshwater, S. M., & Vayalakkara, J. (2000). The Luria-Nebraska neuropsychological battery. In G. Groth-Marnat (Ed.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration* (pp. 263-289). New York: Wiley.

Golden, C., Purish, A., & Hammeke, T. (1985). *Manual for the Luria-Nebraska neuropsychological battery*. Los Angeles: Western Psychological Services.

Hall, R. C. W., Hall, R. C. W., & Chapman, M. J. (2005). Definition, diagnosis, and forensic implications of postconcussional syndrome. *Psychosomatics: Journal of Consultation Liaison Psychiatry, 46*(3), 195-202. doi: 10.1176/appi.psy.46.3.195

Hall, R. C. W., Hall, R. C. W., & Chapman, M. J. (2006). Postconcussional syndrome: A work in progress (response to Dr. Smith). *Psychosomatics, 47*(3), 272. doi: 10.1176/appi.psy.47.3.272

*Hattori, N., Swam, M., Stobbe, G. A., Uomoto, J. M., Minoshima, S., Djang, D., Krishnananthan, R., & Lewis, D. H. (2009). Differential SPECT activation patterns associated with PASAT performance may indicate frontocerebellar functional dissociation in chronic mild traumatic brain injury. *Journal of Nuclear Medicine*, *50*, 1054-61.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

*Heitger, M. H., Jones, R. D., Dalrymple-Alford, J. C., Frampton, C. M., Ardagh, M. W., & Anderson, T. J. (2006). Motor deficits and recovery during the first year following mild closed head injury. *Brain Injury, 20*(8), 807-824. doi: 10.1080/02699050600676354

Hoge, C. W., McGurk, D., Thomas, J., Cox, A. L., Engel, C. C., & Castro, C. A. (2008). Mild traumatic brain injury in U.S. soldiers returning from Iraq. *The New England Journal of Medicine, 358*(5), 453-463. doi: 10.1056/NEJMoa072972

Huang, M.-X., Theilmann, R. J., Robb, A., Angeles, A., Nichols, S., Drake, A., . . . Lee, R. R. (2009). Integrated imaging approach with MEG and DTI to detect mild traumatic brain injury in military and civilian patients. *Journal of Neurotrauma, 26*(8), 1213-1226. doi: 10.1089/neu.2008.0672

Independent Review Group. *Rebuilding the trust: report on rehabilitative care and administrative processes at Walter Reed Army Medical Center and National Naval Medical Center*. Alexandria, VA: April 2007. (Accessed December 13, 2010, at http://www.ha.osd.mil/dhb/recommendations/2007/IRG-Report-Final.pdf.)

Iverson, G. L. (2010). Mild traumatic brain injury meta-analyses can obscure individual differences. *Brain Injury, 24*(10), 1246-1255. doi: 10.3109/02699052.2010.490513

Iverson, G. L., & Brooks, B. L. (2011). Improving accuracy for identifying cognitive impairment. In M. R. Schoenberg & J. G. Scott (Eds.), *The little black book of neuropsychology: A syndrome-based approach* (pp. 923-950). New York, NY US: Springer Science + Business Media.

Iverson, G.L., Lovell, M.R., & Collins, M.W. (2005). Validity of ImPACT for measuring processing speed following sports-related concussion. *Journal of Clinical and Experimental Neuropsychology*, *27*, 683– 689.

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*(1), 12-19. doi: 10.1037/0022-006x.59.1.12

*Konrad, C., Geburek, A. J., Rist, F., Blumenroth, H., Fischer, B., Husstedt, I., . . . Lohmann, H. (2010). Long-term cognitive and emotional consequences of mild traumatic brain injury. *Psychological Medicine, FirstView*, 1-15. doi: doi:10.1017/S0033291710001728

*Kraus, M. F., Susmaras, T., Caughlin, B. P., Walker, C. J., Sweeney, J. A., & Little, D. M. (2007). White matter integrity and cognition in chronic traumatic brain injury: A diffusion tensor imaging study. *Brain: A Journal of Neurology, 130*(10), 2508-2519. doi: 10.1093/brain/awm216

*Kwok, F. Y., Lee, T. M. C., Leung, C. H. S., & Poon, W. S. (2008). Changes of cognitive functioning following mild traumatic brain injury over a 3-month period. *Brain Injury, 22*(10), 740-751. doi: 10.1080/02699050802336989

Lange, R. T. P., Iverson, G. L. P., Brubacher, J. R. M. D., Madler, B. P., & Heran, M. K. M. D. (2012). Diffusion tensor imaging findings are not strongly associated with

postconcussional disorder 2 months following mild traumatic brain injury. *Journal of Head Trauma Rehabilitation May/June, 27*(3), 188-198.

Leclerc, S., Lassonde, M., Delaney, J.S., Lacroix, V.J., & Johnston, K.M. (2001). Recommendations for grading of concussion in athletes. *Sports Medicine*, *31*, 629– 636.

Larrabee, G. J., Binder, L. M., Rohling, M. L., & Ploetz, D. M. (2013). Meta-analytic methods and the importance of non-TBI factors related to outcome in mild traumatic brain injury: Response to Bigler et al. (2013). *The Clinical Neuropsychologist, 27*(2), 215-237. doi: 10.1080/13854046.2013.769634

*Lee, H., Wintermark, M., Gean, A. D., Ghajar, J., Manley, G. T., & Mukherjee, P. (2008). Focal lesions in acute mild traumatic brain injury and neurocognitive outcome: CT versus 3T MRI. *Journal of Neurotrauma, 25*(9), 1049-1056. doi: 10.1089/neu.2008.0566

Lees-Haley, P. R., Fox, D. D., & Courtney, J. C. (2001). A comparison of complaints by mild brain injury claimants and other claimants describing subjective experiences immediately following their injury. *Archives of Clinical Neuropsychology, 16*(7), 689-695. doi: 10.1016/s0887-6177(00)00092-5

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*Little, D. M., Kraus, M. F., Joseph, J., Geary, E. K., Susmaras, T., Zhou, X. J., . . . Gorelick, P. B. (2010). Thalamic integrity underlies executive dysfunction in traumatic brain injury. *Neurology, 74*(7), 558-564. doi: 10.1212/WNL.0b013e3181cff5d5

Luis, C. A., Vanderploeg, R. D., & Curtiss, G. (2003). Predictors of postconcussion symptom complex in community dwelling male veterans. *Journal of the International Neuropsychological Society*, 9, 1001-1015.

Mark, D. G., Hung, Y.-Y., Offerman, S. R., Rauchwerger, A. S., Reed, M. E., Chettipally, U., … Ballard, D. W. (in press). Nontraumatic subarachnoid hemorrhage in the setting of negative cranial computed tomography results: External validation of a clinical and imaging prediction rule. *Annals of Emergency Medicine*. Retrieved from http://ac.els-cdn.com/S0196064412015089/1-s2.0-S0196064412015089-main.pdf?_tid=dc11fb3a-c348-11e2-acad-00000aacb360&acdnat=1369273027_c0b436e76d540d6ef2fd2a1a8e6888da

Mayer, A. R., Ling, J., Mannell, M. V., Gasparovic, C., Phillips, J. P., Doezema, D., … Yeo, R. A. (2010). A prospective diffusion tensor imaging study in mild traumatic brain injury. *Neurology, 74*(8), 643-650. doi: 10.1212/WNL.0b013e3181d0ccdd

Mayer, A. R., Mannell, M. V., Ling, J., Gasparovic, C., & Yeo, R. A. (2011). Functional connectivity in mild traumatic brain injury. *Human Brain Mapping, 32*(11), 1825-1835. doi: 10.1002/hbm.21151

Mayer, A. R., Yang, Z., Yeo, R. A., Pena, A., Ling, J. M., Mannell, M. V., ... Mojtahed, K. (2012). A functional MRI study of multimodal selective attention following mild traumatic brain injury. *Brain Imaging and Behavior, 6*(2), 343-354. doi: 10.1007/s11682-012-9178-z

McCrea, M. A. (2008). *Mild traumatic brain injury and post concussion syndrome: The new evidence base for diagnosis and treatment*. New York, New York: Oxford University Press.

McCrea, M, Guskiewicz, K. M., Marshall, S. W., Barr, W, Randolph, C, Cantu, R. C., Onate, J. A., Yang, J, Kelly, J. P. (2003). Acute effects and recovery time following concussion in

collegiate football players: The NCAA concussion study. *Journal of the American Medical Association*, 290, 2556–2563.

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M., & Cantu, R. (2009). Consensus statement on concussion in sport: The 3rd international conference on concussion in sport, held in Zurich, November 2008. *Journal of Clinical Neuroscience*. 16, 755-763.

Menon, D. K., Schwab, K., Wright, D. W., & Maas, A. I. (2010). Position Statement: Definition of traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 91*(11), 1637-1640. doi: http://dx.doi.org/10.1016/j.apmr.2010.05.017

*Meyers, J. E., & Rohling, M. L. (2004). Validation of the Meyers Short Battery on mild TBI patients. *Archives of Clinical Neuropsychology, 19*(5), 637-651. doi: 10.1016/j.acn.2003.08.007

*Miles, L., Grossman, R. I., Johnson, G., Babb, J. S., Diller, L., & Inglese, M. (2008). Short-term DTI predictors of cognitive dysfunction in mild traumatic brain injury. *Brain Injury, 22*(2), 115-122. doi: 10.1080/02699050801888816

Nightingale, E. J., Soo, C. A., & Tate, R. L. (2007). A systematic review of early prognostic factors for return to work after traumatic brain injury. *Brain Impairment, 8*(2), 101-142. doi: 10.1375/brim.8.2.101

Niogi, S. N., & Mukherjee, P. (2010). Diffusion tensor imaging of mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 25*(4), 241-255. doi: 10.1097/HTR.0b013e3181e52c2a

Niogi, S. N., Mukherjee, P., Ghajar, J., Kolster, R. A., Sarkar, R., Lee, H., …McCandliss, B. D. (2008). Extent of microstructural white matter injury in postconcussive syndrome

correlates with impaired cognitive reaction time: A 3T diffusion tensor imaging study of mild traumatic brain injury. *American Journal of Neuroradiology,* 29, 967-973.

Nygren de Boussard, C., Fredman, P., Lundin, A., Andersson, K., Edman, G., & Borg, J. (2004). S100 in mild traumatic brain injury. *Brain Injury, 18*(7), 671-683. doi: 10.1080/02699050310001646215

*Ord, J. S., Greve, K. W., Bianchini, K. J., & Aguerrevere, L. E. (2010). Executive dysfunction in traumatic brain injury: The effects of injury severity and effort on the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology, 32*(2), 132-140. doi: 10.1080/13803390902858874

Pertab, J. L., James, K. M., & Bigler, E. D. (2009). Limitations of mild traumatic brain injury meta-analyses. *Brain Injury, 23*(6), 498-508. doi: 10.1080/02699050902927984

Ponsford, J., Willmont, C., Rothwell, A., Cameron, P., Kelly, A.-M., Nelms, R., . . . Ng, K. (2000). Factors influencing outcome following mild traumatic brain injury in adults. *Journal of the International Neuropsychological Society, 6*(5), 568-579. doi: 10.1017/s1355617700655066

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. Tucson: Neuropsychology Press.

Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). Tucson: Neuropsychology Press.

Reitan, R. M., & Wolfson, D. (1999). The two faces of mild head injury. *Archives of Clinical Neuropsychology, 14*(2), 191-202. doi: 10.1016/s0887-6177(98)00012-2

Rohling, M. L., Larrabee, G. J., & Millis, S. R. (2012). The "miserable minority" following mild traumatic brain injury: Who are they and do meta-analyses hide them? *The Clinical Neuropsychologist*, *26*, 197-213.

Rohling, M. L., Binder, L. M., Demakis, G. J., Larrabee, G. J., Ploetz, Danielle M. & Langhinrichsen-Rohling, J. (2011). A meta-analysis of neuropsychological outcome after mild traumatic brain injury: Re-analyses and reconsiderations of Binder et al. (1997), Frencham et al. (2005), and Pertab et al. (2009). *The Clinical Neuropsychologist*, *25*, 608-623.

Rona, R. J. (2012). Long-term consequences of mild traumatic brain injury. *British Journal of Psychiatry*, *201*, 172-174.

Ruff, R. (2005). Two decades of advances in understanding of mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 20*(1), 5-18. doi: 10.1097/00001199-200501000-00003

Ruff, R. M., Iverson, G. L., Barth, J. T., Bush, S. S., Broshek, D. K., and the NAN Policy and Planning Committee. (2009). Recommendations for diagnosing a mild traumatic brain injury: A national academy of neuropsychology education paper. *Archives of Clinical Neuropsychology, 24*, 3-10.

Ruff, R. M., & Weyer Jamora, C. (2009). Myths and mild traumatic brain injury. *Psychological Injury and Law, 2*(1), 34-42. doi: 10.1007/s12207-009-9029-4

Rutgers, D. R., Toulgoat, F., Cazejust, J., Fillard, P., Lasjaunias, P., & Ducreux, D. (2008). White matter abnormalities in mild traumatic brain injury: A diffusion tensor imaging study. *American Journal of Neuroradiology, 29*(3), 514-519. doi: 10.3174/ajnr.A0856

Ryan, L. M., & Warden, D. L. (2003). Post concussion syndrome. *International Review of Psychiatry*, *15*, 310-316.

Ryu, W. H. A., Feinstein, A., Colantonio, A., Streiner, D. L., & Dawson, D. R. (2009). Early identification and incidence of mild TBI in Ontario. *The Canadian Journal of Neurological Sciences/ Le Journal Canadien Des Sciences Neurologiques, 36*(4), 429-435.

Satz, P., Alfano, M. S., Light, R., Morgenstern, H., Zaucha, K., Asarnow, R. F., & Newton, S. (1999). Persistent post-concussive syndrome: A proposed methodology and literature review to determine the effects, if any, of mild head and other bodily injury. *Journal of Clinical and Experimental Neuropsychology, 21*(5), 620-628. doi: 10.1076/jcen.21.5.620.870

Scheibel, R.S., Newsome, M.R., Steinberg, J.L., Pearson, D.A., Rauch, R.A.,Mao, H., Troyanskaya, M., Sharma, R.G.,&Levin, H.S. (2007). Altered brain activation during cognitive control in patients with moderate to severe traumatic brain injury. *Neurorehabilitation and Neural Repair*, *21*, 36– 45.

Schretlen, D. J., & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry, 15*(4), 341-349. doi: 10.1080/09540260310001606728

Shenton, M. E., Hamoda, H. M., Schneiderman, J. S., Bouix, S., Pasternak, O., Rathi, Y., . . . Zafonte, R. (2012). A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. *Brain Imaging and Behavior, 6*(2), 137-192. doi: 10.1007/s11682-012-9156-5

Silva, M. A., Donnell, A. J., Kim, M. S., & Vanderploeg, R. D. (2012). Abnormal neurological exam findings in individuals with mild traumatic brain injury (mTBI) versus psychiatric and healthy controls. *The Clinical Neuropsychologist, 26*(7), 1102-1116. doi: 10.1080/13854046.2012.723753

Smith, D. H. (2006). Postconcussional symptoms not a syndrome. *Psychosomatics, 47*(3), 271-272. doi: 10.1176/appi.psy.47.3.271

Smits, M., Hunink, M. G. M., Nederkoorn, P. J., Dekker, H. M., Vos, P. E., Kool, D. R., . . . Dippel, D. W. J. (2007). A history of loss of consciousness or post-traumatic amnesia in minor head injury: 'Conditio sine qua non' or one of the risk factors? *Journal of Neurology, Neurosurgery & Psychiatry, 78*(12), 1359-1364. doi: 10.1136/jnnp.2007.117143

The White House: Office of the Press Secretary. (2013, April 02). Fact sheet: BRAIN initiative. Retrieved from http://www.whitehouse.gov/the-press-office/2013/04/02/fact-sheet-brain-initiative

Thurmond, V. A., Hicks, R., Gleason, T., Miller, A. C., Szuflita, N., Orman, J., & Schwab, K. (2010). Advancing integrated research in psychological health and traumatic brain injury: Common data elements. *Archives of Physical Medicine and Rehabilitation, 91*(11), 1633-1636. doi: http://dx.doi.org/10.1016/j.apmr.2010.06.034

Toledo, E., Lebel, A., Becerra, L., Minster, A., Linnman, C., Maleki, N., . . . Borsook, D. (2012). The young brain and concussion: Imaging as a biomarker for diagnosis and prognosis. *Neuroscience and Biobehavioral Reviews, 36*(6), 1510-1531. doi: 10.1016/j.neubiorev.2012.03.007

Vevea, J. L., & Woods, C. M. (2005). Publication Bias in Research Synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*(4), 428-443. doi: 10.1037/1082-989x.10.4.428

Voller, B., Benke, T., Benedetto, K., Schnider, P., Auff, E., & Aichner, F. (1999). Neuropsychological, MRI and EEG findings after very mild traumatic brain injury. *Brain Injury, 13*(10), 821-827. doi: 10.1080/026990599121214

Wilde, E. A., Whiteneck, G. G., Bogner, J., Bushnik, T., Cifu, D. X., Dikmen, S., . . . von Steinbuechel, N. (2010). Recommendations for the use of common outcome measures in traumatic brain injury research. *Archives of Physical Medicine and Rehabilitation, 91*(11), 1650-1660.e1617. doi: http://dx.doi.org/10.1016/j.apmr.2010.06.033

Willemse-van Son, A. H. P., Ribbers, G. M., Verhagen, A. P., & Stam, H. J. (2007). Prognostic factors of long-term functioning and productivity after traumatic brain injury: A systematic review of prospective cohort studies. *Clinical Rehabilitation, 21*(11), 1024-1037. doi: 10.1177/0269215507077603

Zetterberg H, H. M. J. M., & et al. (2006). Neurochemical aftermath of amateur boxing. *Archives of Neurology, 63*(9), 1277-1280. doi: 10.1001/archneur.63.9.1277

Zhang, L., Heier, L.A., Zimmerman, R.D., Jordan, B., & Ulug, A.M. (2006). Diffusion anisotropy changes in the brains of professional boxers. *AJNR American Journal of Neuroradiology*, *27*, 2000–2004.

Zhou, Y., Kierans, A., Kenul, D., Ge, Y., Rath, J., Reaume, J., . . . Lui, Y. W. (2013). Mild traumatic brain injury: longitudinal regional brain volume changes. *Radiology*. doi: 10.1148/radiol.13122542

Appendix 1

mTBI Meta-Analysis Code Book

General Instructions
1. Note that each code below is numbered. When coding, please highlight empirical evidence in support of the code in the study report, and mark it with the coding reference number.
2. You should indicate a code when plausible. However, if insufficient evidence exists in the study report to make a plausible coding, then fill in the coding blanks with "-.99" to indicate unknown.

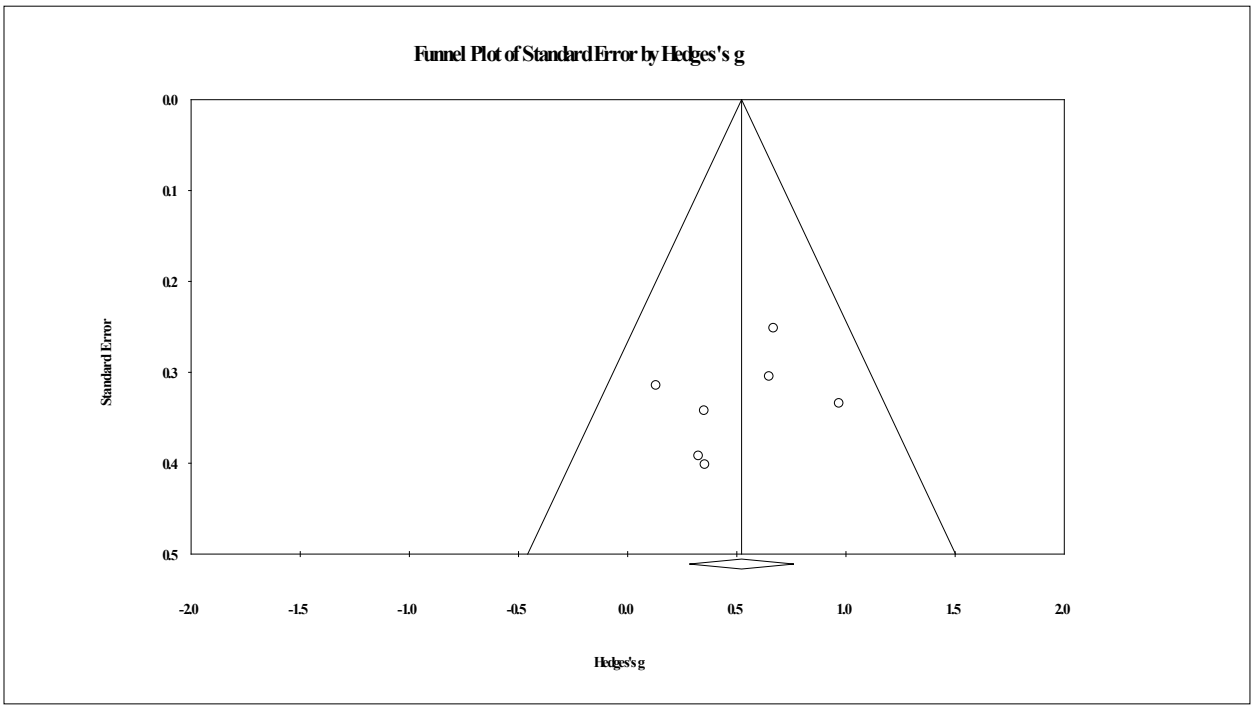
Code #	Code Description	Code
1	**Study Identification Number:**	
2	**Title:**	
3	**Authors:**	
4	**Country:**	
	Site: 1. University 2. Hospital 3. Private Clinic 4. Other	
5	**Year of Study:**	
6	**Study Design:** 1. Quasi experimental a. Prospective b. Retrospective c. Crossectional	
7	**Patient N:**	
8	**Patient sex ratio:**	
9	**Patient age (M, SD, range):**	
10	**Patient Handedness:** 1. R 2. L 3. A	
11	**Patient Marital status:** 1. Single 2. Married	

	3. Divorced	
12	**Patient SES:**	
13	**Patient Education (M, SD, Range):**	
14	**% patient attrition:**	
15	**Control N:**	
16	**Control sex ratio:**	
17	**Control age (M, SD, range):**	
18	**Control Handedness:** 4. R 5. L 6. A	
19	**Control Marital status:** 4. Single 5. Married 6. Divorced	
20	**Control SES:**	
21	**Control Education (M, SD, Range):**	
22	**% control attrition:**	
23	**Domain Assessed:** 1. Attention/concentration/working memory 2. Processing speed 3. Expressive language 4. Receptive language 5. Verbal memory 6. Visuospatial ability 7. Perceptual reasoning 8. Visuospatial memory 9. Executive functioning	
24	**Time post TBI (M, SD, range):**	
25	1. **Definition of MTBI:** a. LOC b. PTA c. GCS	

	d. Other	
26	**Etiology:** 1. Sports 2. MVA 3. Fall/blow to head 4. Assault	
27	**Minority Proportion in Sample (For this study minority is defined as someone belonging to a non-Caucasian group)** *Note:* This number should be a decimal (two places) of total sample	
28	**Illegal substance use:**	
29	**Psychiatric comorbidity**	
30	**Prior head trauma:**	
31	**Litigation:** 1. Yes 2. No	
32	**Comparison Type:** 1. Healthy control 2. Orthopedic injury 3. Other	
33	**Measures used:**	
34	**Patient mean, SD, Stat test value (e.g., F-test value), p value**	
35	**Control mean, SD, Stat test value (e.g., F-test value), p value**	

Appendix 2

Funnel Plot for Studies Rated I and II



Funnel Plot for Studies Rated III and IV