



HHS Public Access

Author manuscript

Rehabil Psychol. Author manuscript; available in PMC 2018 December 06.

Published in final edited form as:

Rehabil Psychol. 2017 November ; 62(4): 485–495. doi:10.1037/rep0000122.

Uncorrected Versus Demographically-Corrected Scores on the NIH Toolbox Cognition Battery in Persons With Traumatic Brain Injury and Stroke

Kristian P. Nitsch,

Department of Psychology, Illinois Institute of Technology, and Center for Rehabilitation Outcomes Research, Rehabilitation Institute of Chicago, Chicago, Illinois;

Kaitlin B. Casaletto,

Department of Neurology, University of California, San Francisco;

Noelle E. Carlozzi,

Department of Physical Medicine and Rehabilitation, University of Michigan;

David S. Tulskey,

Center on Assessment Research and Translation and Departments of Physical Therapy & Psychological and Brain Sciences, University of Delaware, and Kessler Foundation Research Center, West Orange, New Jersey;

Allen W. Heinemann, and

Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine, and Center for Rehabilitation Outcomes Research, Rehabilitation Institute of Chicago;

Robert K. Heaton

Department of Psychiatry, University of California, San Diego.

Abstract

Objective: The association between demographic characteristics and neurocognitive performance is well established; however, clinicians may have difficulty selecting when to use uncorrected versus demographically corrected scores. We compared these score types in individuals with traumatic brain injury (TBI) and stroke, on the National Institutes of Health Toolbox—Cognition Battery (NIHTB-CB).

Research Method: Adults with TBI and stroke were demographically matched to controls, and completed the NIHTB-CB. Published “corrected scores” are adjusted for age, education, sex, and race/ethnicity; “uncorrected scores” were created using census data to represent the average adult in the U.S. population.

Results: Effect sizes for the TBI and stroke groups versus controls were larger using corrected scores compared with uncorrected scores for the fluid composite (uncorrected to corrected effect sizes: TBI: $d = 0.66$, $p < .001$ to 0.83 , $p < .001$; stroke $d = 0.97$, $p < .001$ to 1.10 , $p < .001$). For the

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Correspondence concerning this article should be addressed to KristianP. Nitsch, MS, Center for Rehabilitation Outcomes Research, Rehabilitation Institute of Chicago, 446 East Ontario Street, Room Number 970, Chicago, IL 60611. knitsch@ric.org

crystallized composite, effect sizes for the TBI and stroke groups versus controls were smaller and nonsignificant using corrected scores (uncorrected to corrected effect sizes: TBI $d = 0.23$, $p = .03$ to 0.20 , $p = .06$; stroke $d = 0.40$, $p < .001$ to 0.17 , $p = .09$). In the injury groups, demographic characteristics accounted for up to 33% of variance in uncorrected scores ($p < .001$), but <5% of variance in corrected scores ($p > .06$).

Conclusions—Corrected scores were more sensitive to neurocognitive impairments in the brain-injured groups. Corrected scores have the advantage of controlling for variance associated with premorbid factors rather than changes in neurological functioning; are more helpful in characterizing *acquired* neurocognitive changes; and can aid in the interpretation of test performance.

Keywords

NIHTB-CB; traumatic brain injury; stroke; cognition; statistical data analysis

Introduction

The relationship between demographic characteristics and neurocognitive test performance is well established. Age, education, gender, race/ethnicity, native language, and other cultural factors each demonstrate a significant association with neuropsychological test results (Diehr et al., 2003; Gasquoine, 2009; Gladsjo et al., 1999; Heaton, Miller, Taylor, & Grant, 2004; Heaton, Ryan, & Grant, 2009; Mungas, Reed, Haan, & Gonzalez, 2005; Norman, Evans, Miller, & Heaton, 2000; Norman et al., 2011; Sherrill-Pattison, Donders, & Thompson, 2000). Importantly, these factors do not operate independently; multifaceted associations among demographic variables have also been observed (Byrd, Touradji, Tang, & Manly, 2004; Casaletto et al., 2015; Heaton et al., 2004; Heaton et al., 2009; Manly, Jacobs, Touradji, Small, & Stern, 2002; Norman et al., 2000; Norman et al., 2011). Moreover, the extent to which demographic characteristics are related to test results often depends on the cognitive abilities being assessed. For example, specific demographic characteristics are related to different domains of cognitive functioning throughout the life span (Casaletto et al., 2015).

Understanding the association between demographic characteristics and cognition is important to interpreting neuropsychological test results for a given individual, and determining whether observed scores reflect acquired brain and cognitive changes versus baseline or premorbid abilities. This is particularly relevant in rehabilitation settings where cognitive impairments can slow recovery of instrumental activities of daily living (Zinn et al., 2004), increase acute rehabilitation length of stay (Rabadi, Rabadi, Edelstein, & Peterson, 2008), increase the likelihood of mortality (Poynter, Kwan, & Vassallo, 2013), and interfere with patient engagement in therapy (Lenze et al., 2004). Thus, it is important for clinicians to consider the role of cognition in the rehabilitation process and to understand how a patient's demographic characteristics may be associated with cognitive performance.

Cognition can be conceptualized in terms of two broad but distinct domains, including “crystallized” and “fluid” abilities (Cattell, 1971). Conceptually, crystallized abilities represent skills that are learned through education and life experiences (e.g., reading). These

abilities develop quickly during childhood and adolescence followed by a period of prolonged stabilization with little or no decline during adulthood (Casaletto et al., 2015). Crystallized cognition is relatively resilient to the deleterious effects of brain injury; tests measuring crystallized abilities are therefore often used as estimates of premorbid neurocognitive functioning, following injury (McFie, 1975; The Psychological Corporation, 2001; Wechsler, 1958). Conversely, fluid cognitive abilities are mutable and reflect ongoing brain processes that develop during childhood, but show age-related declines in adulthood (Wechsler, 1935, 1952). Additionally, fluid cognitive abilities are typically vulnerable (and sensitive) to brain injury or disease. Thus, when examining the cognitive sequelae of injury due to trauma such as traumatic brain injury (TBI) or stroke, greater impairment on tasks of fluid cognition can be anticipated.

Adjusted scores that account for a person's demographic characteristics are thought to improve diagnostic accuracy by eliminating unwanted variance associated with premorbid personal characteristics rather than the influence of an injury on cognitive functioning (Blake, Fichtenberg, & Abeare, 2009), and have therefore become common in clinical practice (Blake et al., 2009; Diehr et al., 2003; Gladsjo et al., 1999; Heaton et al., 2004; Norman et al., 2000; Norman et al., 2011; Taylor & Heaton, 2001). Studies comparing uncorrected and demographically corrected scores among clinical populations and uninjured controls have been reported (Blake et al., 2009; Mungas, Reed, Farias, & Decarli, 2009; O'Connell & Tuokko, 2010; O'Connell, Tuokko, & Kadlec, 2011; Reitan & Wolfson, 1995, 1997; Strong, Donders, & van Dyke, 2005), with some of the earliest examples of demographic correction accounting for the associations between age and cognitive performance (Wechsler, 1935, 1952). Despite this wide-spread use, research findings remain equivocal on whether demographically corrected scores provide more accurate diagnostic classification or are able to better characterize the magnitude of cognitive impairment following neurologic injury when compared to uncorrected scores (Blake et al., 2009; Reitan & Wolfson, 1995, 1997, 2005; Strong et al., 2005). Accordingly, some researchers have argued that demographically corrected scores may lack clinical or neurodiagnostic utility, and in some cases, they suggest that demographically corrected scores may even reduce diagnostic accuracy and compromise sensitivity by removing predictive variance (Blake et al., 2009; Reitan & Wolfson, 2005). For example, correcting for age on cognitive measures used to screen for dementia may be inappropriate, as age is a well-established risk factor for the development of dementia (Sliwinski, Buschke, Stewart, Masur, & Lipton, 1997; Sliwinski, Hofer, Hall, Buschke, & Lipton, 2003). In this situation, using demographically corrected scores that account for age may make dementia screening measures less accurate than unadjusted scores because predictive variance for diagnosing dementia has been eliminated (Sliwinski et al., 1997; Sliwinski et al., 2003). Accordingly, some researchers have called for different normative standards for specific clinical situations (e.g., diagnosis vs. comparison between groups; O'Connell & Tuokko, 2010; Sliwinski et al., 1997; Sliwinski et al., 2003), and have stressed the importance of using demographically corrected scores only when indicated. However, clinicians and researchers may have difficulty deciding when to select and interpret demographically corrected scores versus uncorrected or other score types (e.g., census-weighted, only age-corrected, etc.).

Selecting which score type is most suitable for interpretation relies upon understanding what clinical or research questions are being asked. For instance, uncorrected scores best reflect an individual's "absolute" performance on a measure. Thus, uncorrected scores may be appropriate to use when trying to determine if an individual's basic ability is good enough for everyday tasks that have comparable requirements and can be useful if comparing change across repeated assessments. However, uncorrected scores can be difficult to interpret, as they do not provide information concerning how the test performance compares with normal expectations for the individual. Accordingly, uncorrected scores are not useful in helping determine whether an individual's performance is indicative of changes in performance or "impairment" compared with a relevant reference group. In contrast, scores controlling for demographic characteristics (e.g., age, education, sex, race/ethnicity, etc.) permit judgments regarding whether an individual's performance is atypical compared to their demographically similar peers. Demographically corrected scores, therefore, are those that can be interpreted *relative* to normative expectations for the person being tested. However, because these scores are adjusted for normative demographic factors, they may not convey if an individual has strong or weak ability in an absolute sense, or in relation to requirements for certain everyday tasks (e.g., driving an automobile).

The National Institutes of Health Toolbox–Cognition Battery (NIHTB-CB) normative scores have recently been updated with fully demographically corrected scores (Casaletto et al., 2015). The purpose of this study is to contrast uncorrected and demographically corrected scores on the NIHTB-CB in samples of individuals with traumatic brain injury (TBI) or stroke, with those of uninjured controls to illustrate the differences between these score types and to examine associations between scores, and demographic and clinical variables. As noted, prior research with traditional neurocognitive tests has compared the sensitivity and neurodiagnostic utility of uncorrected and demographically corrected scores among samples with and without neurologic impairments. This study extends the prior work by comparing uncorrected and demographically corrected Composite scores from the recently developed, computerized National Institutes of Health Toolbox–Cognition Battery in two neurologic samples and their controls. Given the Toolbox's novelty and relatively unique features (computerized, nonproprietary, with efficient and broad coverage of abilities, and applicability for ages 3–85), as well as its encouraged use in NIH-funded research, and its increasingly widespread use in both clinical and research settings, it is critical to examine the accuracy of Toolbox scores in detecting and characterizing acquired neurocognitive impairment.

Method

Participants

Our samples included community-dwelling adults who had been diagnosed with TBI or stroke ($n = 395$), and their noninjured, matched-controls ($n = 394$), who were administered the NIHTB-CB in English (see Table 1). Stroke ($n = 184$) and TBI ($n = 211$) participants were recruited at the Rehabilitation Institute of Chicago, University of Michigan, and Washington University in St. Louis; a comparison group (TBI-matched: $n = 184$; stroke-matched: $n = 210$), was drawn from the NIH Toolbox Normative Study and matched for age,

sex, and education (Beaumont et al., 2013). All TBI and stroke participants were at least 1-year postinjury. Injury etiology for the TBI group included motor vehicle crash (54%), fall (28%), gunshot wound/violence (11%), sports-related injury (3%), and other causes (4%). Stroke etiologies included hemorrhagic (28%) and ischemic events (72%); of the stroke group, 38% had right side weakness, 39% had left side weakness, <1% had bilateral weakness, and 16% exhibited no weakness. TBI severity was classified according to the lowest Glasgow Coma Score (GCS) within the first 24 hr after injury; a GCS of 9–12 was classified as moderate injury, and a score ≤ 8 was classified as severe (Teasdale & Jennett, 1974; Traumatic Brain Injury Model Systems National Data Center, 2006). Complicated-mild TBI was classified as a GCS of 13–15 with positive neuroimaging (Williams, Levin, & Eisenberg, 1990). The Modified Rankin Scale was used to classify stroke severity as mild (scores of 1–2), moderate (scores of 3), or severe (scores of 4; van Swieten, Koudstaal, Visser, Schouten, & van Gijn, 1988). We combined complicated-mild TBI and moderate TBI for analyses of injury severity, based on previous research suggesting that the long-term outcomes for individuals with complicated-mild TBI more closely resemble individuals with moderate TBI than they do mild TBI (Kashluba, Hanks, Casey, & Millis, 2008).

The TBI sample's controls were matched on race/ethnicity, as well as age, education and sex. However, because the stroke group contained a disproportionate number of African Americans, the available stroke controls permitted successful matching on only age, education, and sex.

National Institutes of Health Toolbox–Cognition Battery (NIHTB-CB)

The NIHTB-CB is a computerized, 30-min cognitive screen that is a part of the larger National Institutes of Health Toolbox (NIHTB) for the Assessment of Neurological and Behavioral Function (Weintraub et al., 2013a; Weintraub et al., 2013b; Weintraub et al., 2014). The NIHTB-Cognition Battery (CB) consists of seven tests assessing functioning across five cognitive domains including (a) executive function (Zelazo et al., 2014; Zelazo et al., 2013); (b) episodic memory (Bauer et al., 2013; Dikmen et al., 2014); (c) processing speed (Carlozzi, Beaumont, Tulskey, & Gershon, 2015; Carlozzi et al., 2014; Carlozzi, Tulskey, Kail, & Beaumont, 2013); (d) working memory (Tulskey et al., 2013; Tulskey et al., 2014); and (e) language (Gershon et al., 2014; Gershon et al., 2013). These domains can be broadly divided into crystallized and fluid cognitive abilities; composite indices were calculated to represent overall cognitive functioning in both the crystallized and fluid domains (Heaton et al., 2014). The NIHTB-CB, which is suitable for individuals ages 3 through 85 and is available in both English and Spanish, has demonstrated strong convergent and discriminant validity and excellent test–retest reliability (Casaletto et al., 2016; Weintraub et al., 2014).

Crystallized cognitive abilities were evaluated using the Picture Vocabulary Test (Bauer et al., 2013; Dikmen et al., 2014) and Oral Reading Recognition Test (Gershon et al., 2013; Gershon et al., 2014). A crystallized composite score was computed from these tests (Heaton et al., 2014). Fluid cognitive abilities were assessed with (a) Dimensional Change Card Sort (Zelazo et al., 2014; Zelazo et al., 2013); (b) Flanker Inhibitory Control and Attention Test (Zelazo et al., 2014; Zelazo et al., 2013); (c) Picture Sequence Memory Test

(Bauer et al., 2013; Dikmen et al., 2014); (d) List Sorting Working Memory Test (Tulsky et al., 2013; Tulsky et al., 2014); and (e) Pattern Comparison Processing Speed Test (Carlozzi et al., 2015; Carlozzi et al., 2014; Carlozzi et al., 2013). For a detailed description of each individual Toolbox Cognition measure see (Weintraub et al., 2013a; Weintraub et al., 2013b; Weintraub et al., 2014). Heaton et al. (2014) provides a full description of the composite scores.

Developing Normative Standards

For the English version of the NIHTB-CB, demographically corrected normative standards were developed in a group of neurological uninjured adults ($n = 1,038$) in order to determine deviations from expected levels of performances. Details regarding these normative data are outlined in Casaletto et al. (2015). In brief, multiple fractional polynomial models were used to regress normalized, uncorrected NIHTB-CB scores separately for each race/ethnicity (i.e., non-Hispanic White, Black, and Hispanic) on demographic characteristics (e.g., age, education, sex). The residuals from these models were smoothed in order to enhance the homogeneity of the variances across age (Casaletto et al., 2015). The corrected residuals were standardized and rescaled to form T scores. The resulting T score ($M = 50$, $SD = 10$) for each test, therefore, represents an individual's neurocognitive performance compared to age-, education-, sex- and race/ethnicity-matched peers. The uncorrected NIHTB-CB scores represent scores that were weighted to reflect the demographic make-up of the 2010 U.S. Census and then normalized across English-speaking children and adults. These scores are interpreted as an individual's performance compared with the general U.S. population (mean age = 38.2 years; mean education = 13.7 years; female = 51%; White = 68%).

Analyses

We conducted a series of independent t tests to examine differences between the TBI and stroke groups and their respective matched-controls on both the uncorrected and demographically corrected NIHTB-CB scores for the crystallized and fluid composites. We calculated effect sizes (Cohen's d) to describe the magnitude of differences observed between groups overall. Linear regression modeling was conducted to examine the individual contribution of variance accounted for by each of the demographic characteristics (e.g., age, education, race/ethnicity, and sex), and their combinations, on uncorrected and demographically corrected fluid and crystallized composite scores for TBI and stroke groups. Finally, using injury severity classification as the grouping variable, t tests were conducted to examine the sensitivity of the uncorrected versus the demographically corrected fluid scores to injury severity within the TBI and stroke groups.

Results

Crystallized Cognitive Performance

Given that crystallized abilities are not generally considered sensitive to acquired brain injury, we did not expect differences between the neurological groups and their matched controls. However, on the uncorrected scores, there were small, but significant differences between the TBI group and matched controls on crystallized scores ($d = 0.23$, $t = 2.1$, $p = .03$), with controls performing better than the TBI group. Application of the demographically

corrected scores resulted in a reduction in these differences between TBI participants and matched controls, which were no longer significant, consistent with an indicator of premorbid functioning ($d = 0.20$, $t = 1.9$, $p = .06$; see Figure 1). This pattern of findings was more prominent for the stroke group relative to matched-controls on crystallized abilities. Specifically, there was a significant, medium effect for stroke relative to matched controls, with controls performing better than stroke on the uncorrected crystallized composite ($d = 0.40$, $t = 4.1$, $p < .001$); this could be a result of our inability to match these groups on race/ethnicity. Application of the demographically corrected crystallized scores revealed a nonsignificant effect ($d = 0.17$, $t = 1.7$, $p = .09$), as expected of a measure of premorbid functioning (see Figure 2).

Fluid Cognitive Performance

A different pattern of results was revealed on the fluid composite, which is expected to be sensitive to acquired brain dysfunction. Although uncorrected fluid scores demonstrated a medium effect between TBI and matched-controls, with the TBI group performing significantly worse than controls ($d = 0.66$, $t = 6.0$, $p < .001$), the magnitude of difference was larger using the demographically corrected fluid scores ($d = 0.83$, $t = 7.6$, $p < .001$; see Figure 1). The stroke group demonstrated significantly worse uncorrected fluid scores relative to matched controls, with a large effect size ($d = 0.97$, $t = 9.5$, $p < .001$); this effect increased slightly using the demographically corrected scores ($d = 1.10$, $t = 10.8$, $p < .001$; see Figure 2).

Variance Contribution of Demographic Characteristics

Overall, 18% to 29% of the variance in uncorrected crystallized performances was accounted for by education alone in the stroke and TBI groups, respectively (stroke: $F(1, 199) = 45.0$, $p < .001$; TBI: $F(1, 174) = 72.1$, $p < .001$). Race (White vs. other) was significantly related to uncorrected crystallized scores for TBI and stroke, as well (TBI: $R^2 = 0.11$; $F(1, 177) = 21.9$; stroke: $R^2 = 0.18$; $F(1, 202) = 44.7$; $ps < 0.001$). Age was significantly related to uncorrected crystallized scores for the TBI group, $R^2 = 0.06$; $F(1, 174) = 10.3$; $p < .05$; but not the stroke group, $R^2 < 0.01$; $F(1, 199) = 0.13$, $p > .05$. Sex was significantly related to uncorrected crystallized scores for the stroke group, $R^2 = 0.024$; $F(1, 199) = 4.1$; $p < .05$; with males performing significantly better than females, but not the TBI group, $R^2 = 0.01$; $F(1, 174) = 2.0$; $p > .05$. Taken together, demographic characteristics (age, education, sex, race) accounted for 32%–33% of the variance in uncorrected crystallized performances (TBI: $F(4, 171) = 21.4$; stroke: $F(4, 196) = 22.5$; $ps < 0.001$). Conversely, when applying demographic corrections to crystallized scores, demographic characteristics were no longer significantly associated with scores and accounted for less than 4% of the total variance in performances (TBI: $F(4, 173) = 0.67$, $p > .05$; stroke: $F(4, 195) = 1.8$, $p > .05$).

Age demonstrated the strongest associations with uncorrected fluid scores among the demographic factors, accounting for 14%–25% of the variance in fluid performances in the stroke and TBI groups, respectively (TBI: $F(1, 168) = 55.0$; stroke: $F(1, 184) = 30.3$; $ps < 0.001$). Race (White vs. other) demonstrated small, but significant relationships with uncorrected fluid performances for both groups (TBI: $R^2 = 0.07$; $F(1, 171) = 13.3$; stroke: R^2

= 0.09; $F(1, 187) = 18.6, ps < 0.001$). Sex was significantly related to uncorrected fluid scores for the TBI group, $R^2 = 0.024$; $F(1, 168) = 4.2; p < .05$, with females performing better than males, but not the stroke group, $R^2 = 0.006$; $F(1, 184) = 1.05; p > .05$. In contrast, education was significantly related to uncorrected fluid performances for the stroke group, $R^2 = 0.04$; $F(1, 184) = 7.9; p < .05$; but not the TBI group, $R^2 = 0.015$; $F(1, 168) = 2.6; p > .05$. Altogether, demographic variables accounted for 29%–38% of the variance in uncorrected fluid performances across TBI and stroke groups (TBI: $F(4, 165) = 24.9$; stroke: $F(4, 181) = 18.4; ps < 0.001$). Demographically corrected fluid scores were minimally related to demographic characteristics, such that together, these factors accounted for less than 5% of the variance in the demographically corrected fluid composites across the TBI and stroke groups (TBI: $F(4, 167) = 1.6, p > .05$; stroke: $F(4, 184) = 2.3, p > .05$; see Tables 2 and 3).

Sensitivity to Injury Severity

Finally, we examined the sensitivity of the uncorrected versus the demographically corrected fluid scores to injury severity within the TBI and stroke groups. Within the TBI group, uncorrected fluid scores did not differentiate between persons with complicated mild-to-moderate versus severe injuries ($d = 0.12$; $t = 0.79, p > .05$), whereas demographically corrected fluid scores showed that TBI participants with severe injuries evidenced substantially worse impairments ($d = 0.52$; $t = 3.4, p < .001$). Within the stroke group, both types of fluid composite scores demonstrated consistently medium, significant effects between subgroups with mild-to-moderate versus severe injuries ($ds = 0.45$ to $0.54; ps < 0.01$), with individuals with severe stroke injuries performing worse than those with mild/moderate injuries.

Discussion

Crystallized cognitive abilities are typically resilient to neurologic injury; therefore, disparate performances between persons with and without neurological injuries on measures of crystallized cognition would be unexpected. In this study, we did observe significant differences between the neurologically injured individuals and the noninjured controls on uncorrected crystallized composites, despite the samples being matched on most demographics (i.e., age, education, sex). However, these differences reduced and became nonsignificant when applying demographically corrected crystallized composite scores. Conversely, the measures used to calculate fluid composite scores tap domains of cognition sensitive to brain injury. Accordingly, individuals with neurological injuries can be expected to perform more poorly on these measures (i.e., fluid composite) compared with controls without neurologic injuries. The TBI and stroke groups performed worse than noninjured controls on fluid composite scores, regardless of whether uncorrected or demographically corrected scores were used, supporting this conclusion. However, these performance differences were largest when demographically corrected fluid composite scores were applied, indicating increased sensitivity to cognitive dysfunction when controlling for the influences of demographic characteristics. Given that rehabilitation psychologists and neuropsychologists rarely have baseline neurocognitive testing available, demographically corrected scores provide a metric of estimated relative performances compared with

expected (i.e., “baseline”) levels. Therefore, demographically corrected scores are a more sensitive indicator of *changes* in neurocognitive status compared to uncorrected scores as demonstrated here, and are most helpful in identifying and characterizing acquired “impairment” (i.e., changes from baseline levels) compared to uncorrected scores. The improvement of demographically corrected over uncorrected scores is even more impressive when one considers that groups were matched on most demographic characteristics, a process intended to reduce demographic associations at the group level.

Consistent with previous findings (Byrd et al., 2004; Casaletto et al., 2015; Diehr et al., 2003), our results revealed significant associations between demographic characteristics and neurocognitive performance, as reflected in uncorrected scores. Unlike prior research examining the NIHTB-CB, which reported on the relationships between demographic characteristics and neurocognitive performance among neurologically intact adults, the present findings indicate that these associations are also evident among individuals with neurological injuries. For both the TBI and stroke groups, when using uncorrected crystallized composite scores, education had the largest, nonclinical association with performance, such that individuals with more years of schooling performed better; this association would be considered “noise” when looking for acquired deficits. Similarly, race (i.e., “White” vs. “other”) was significantly related to performance among both TBI and stroke participant, with non-White race/ethnicity being associated with poorer performances. Age and sex were less strongly associated with crystallized scores across groups, consistent with findings in non-neurological adults on premorbid metrics, with men outperforming women in the stroke group, but women outperforming men in the TBI group. Older age was associated with better performance on uncorrected crystallized composite scores. Overall, 31%–33% of the variance in uncorrected crystallized composite scores was attributable to demographic characteristics (i.e., “noise”).

Age and race also accounted for a substantial amount of the variance in uncorrected fluid composite scores in both TBI and stroke, with older age and non-White race/ethnicity being associated with poorer performances. These findings are commensurate with the adverse effects of older age and non-White race/ethnicity on fluid cognition demonstrated in neurologically uninjured population studies. However, the impact of education and sex on fluid performance were relatively small across groups (1%–4% of the variance). While educational effects on fluid cognition are demonstrated in uninjured adults, the size of this relationship is commonly smaller than the observed age effects (i.e., small-to-medium educational effects vs. large age effects on fluid cognition; e.g., Casaletto et al., 2015). Similarly, the influence of sex tends to be more variable and modest across fluid domains, most consistently favoring females on processing speed and verbal episodic memory. Taken together, the pattern of demographic associations demonstrated with uncorrected fluid scores within our neurological groups are largely comparable with those observed in uninjured populations, suggesting that they remain robust sources of “noise,” even within the context of brain injury. Notably, in combination, demographic variables accounted for over a third of the variance in uncorrected fluid composite scores across both the TBI and stroke groups.

As previously noted, demographic factors do not operate independently or free of environmental and societal influences that may mediate or moderate their association with

neuropsychological test performance (Byrd et al., 2004; Casaletto et al., 2015; Heaton et al., 2004; Heaton et al., 2009; Manly et al., 2002; Norman et al., 2000; Norman et al., 2011). It is therefore important to recognize that the associations between demographic characteristics and neurocognitive performance are complex and multifaceted, and that viewing the relationship between a single demographic characteristic and testing performance in isolation without considering additional contributing factors may lead to inaccurate conclusions. This is especially true when considering the role of race/ethnicity in neurocognitive performance and how societal and environmental factors may act as mediators. For instance, contributing factors such as limited access to health care, wealth, lower quality of health services, fewer high quality educational opportunities, and disparities in nutrition are potentially modifiable variables that may exacerbate racial/ethnic influences on neuropsychological test performance. Accordingly, when associations between race/ethnicity and neurological functioning are observed, it is important to recognize that underlying and systematic disparities in resources and opportunities for health care, education, and other external variables likely play a large contributory role (Carvalho et al., 2015; Manly, 2006; Manly et al., 2002).

While uncorrected fluid composite scores were unable to differentiate individuals with complicated mild-to-moderate versus severe TBI injuries, demographically corrected fluid composite scores in contrast, demonstrated a large effect; individuals classified as having a severe TBI performed significantly worse than those with complicated mild-to-moderate injuries. Within the stroke sample, both uncorrected fluid composite and demographically corrected fluid composite scores consistently demonstrated medium effects between the subgroups with mild-to-moderate and severe injuries. Taken together, the demographically corrected fluid composite scores showed better overall construct validity with TBI severity, as measured by the Glasgow Coma Scale (GCS), than uncorrected scores. Accordingly, among individuals with TBI, demographically corrected scores more accurately reflect the magnitude of underlying cognitive dysfunction, and provide a more precise indicator of injury severity.

Application of demographically corrected scores substantially reduced or eliminated unwanted, large amounts of variance associated with non-neurological demographic characteristics, on both fluid and crystallized composites. Controlling for the influence of demographic variables on cognitive performance provided a clearer understanding of the consequences of acquired injury, as exemplified by the consistent associations between injury severity and neurocognitive performance using demographically corrected versus uncorrected metrics. These results are relevant to rehabilitation clinicians, as they demonstrate that demographically corrected scores can increase accuracy of injury severity classification. Accurate classification of cognitive abilities and tracking of cognitive changes especially across time is critical in rehabilitation settings wherein cognition is an important outcome measurement. Because injury severity is often associated with typical patterns of recovery and can help predict functional outcomes, this improved classification accuracy can help inform expectations for a patient's recovery from acquired brain injury and drive treatment planning. Accordingly, it is important for clinicians to familiarize themselves with the association between demographic characteristics and cognition, and to know when it is

appropriate to use demographically corrected scores to interpret performance on neurocognitive testing.

Limitations

We acknowledge several limitations to the current study. First, while we were able to match both the TBI and stroke groups to noninjured controls on age, education, and gender, a disproportionate number of African Americans were represented in the stroke group, so an insufficient number of African American uninjured controls were available for stroke group comparisons. This limited our ability to account for the effects of race/ethnicity in uncorrected matched-control comparisons at the group level. Second, we note the use of the Modified Rankin Classification scale for characterizing stroke injury severity as a potential limitation of this study. The Modified Rankin Classification scale has a restricted response set ranging from 0 to 6, with 0 representing asymptomatic individuals, 5 representing individuals with severe disability requiring nursing care, and 6 representing deceased individuals. This classification system is “blunt,” and therefore may inadequately capture subtle functional differences characteristic of differing levels of injury severity, especially since our inclusion requirement that participants be community-dwelling effectively restricted the range of the scale to 0–4. Use of the Modified Rankin Classification scale may have therefore contributed to why the demographically corrected scores did not demonstrate greater sensitivity to injury severity within the stroke group compared to uncorrected scores. Additionally, GCS and Rankin scores were obtained in the acute stage of injury and therefore do not account for differential patterns of injury recovery and long-term outcome.

Importantly, while the findings demonstrate that demographically corrected scores eliminated a large portion of “premorbid” variance in neuropsychological test performance attributable to demographic characteristics (e.g., age, education, race, sex), and resulted in improved injury severity classification among persons in the TBI group, we do not know whether demographically corrected scores are better able to predict functional outcomes. Specifically, given the recent development of the NIHTB-CB and the nascent body of literature available on this measurement system pertaining to criterion validity, as well as limited available information about our participants’ functional status, there is currently insufficient evidence regarding the predictive and concurrent validity of the crystalized and fluid composite scores when compared with other measures of cognition for individuals with TBI or stroke. Accordingly, the available research examining the criterion validity of the NIHTB-CB limits our ability to evaluate the predictive value of these composite scores in regards to activities of daily living, participation and community reintegration, and other functional outcomes that may be of interest to rehabilitation clinicians, such as return to work, independence, and self-care. This limitation is particularly relevant as at least some prior research has demonstrated that uncorrected, absolute scores (i.e., uncorrected scores determined using a general health adult population) may be superior for predicting functional outcomes, especially among TBI populations (Silverberg & Millis, 2009). Accordingly, additional research examining the clinical utility of the NIHTB-CB demographically corrected and uncorrected scores for predicting the functional outcomes relevant to rehabilitation care against commonly used “gold-standard” measures used with TBI and stroke populations is required. Moreover, access to clinician ratings of recovery, as

well as objective indicators of functional recovery (e.g., employment, functional independence, etc.) would be helpful in understanding how injury severity ratings at the time of hospitalization were related to functioning at the time of participation.

An additional limitation of this study is the lack of consistently documented injury-related characteristics (e.g., location of TBI/stroke, collateral neuroimaging, etc.) for participants in the TBI and stroke samples. The absence of this information precludes us from testing several hypotheses that may elucidate some unexpected findings that were observed during this investigation. Specifically, given the expected “hold” characteristics of the NIHTB-CB crystallized composite, we were not anticipating statistically significant differences between the matched controls and the clinical groups on either the uncorrected or demographically corrected scores. Yet, both the TBI and stroke groups performed worse than matched controls when using uncorrected NIHTB-CB crystallized composite scores, although these differences were reduced and became nonsignificant when the demographically corrected scores were used. While this finding could be a result of our inability to match the stroke group and their controls on race/ethnicity, it may also reflect disease-specific characteristics, such as injury location. For instance, because the NIHTB-CB crystallized composite scores are calculated using language-based subtests, and it is possible that some persons experienced injuries that affected language processing areas of the brain, resulting in an overall modest suppression of group mean crystallized scores. While persons exhibiting frank aphasic syndromes were excluded from the study, the screening processes may not have detected subtle impairments in language abilities for either of the clinical groups.

Conclusions

This study represents the first evaluation of uncorrected and demographically corrected NIHTB-CB crystallized and fluid composites scores among a large sample of individuals with TBI and stroke, and uninjured matched controls. Demographically corrected scores substantially reduced premorbid influences in cognitive scores, which increased sensitivity to acquired neurocognitive dysfunction. This increased sensitivity allowed for greater accuracy and precision in distinguishing injury severity, as non-neurologically related variance in cognitive performances was parsed out. Moreover, applying demographically corrected scores resulted in greater separation between neurocognitive performances between the two injury groups and their matched controls without neurological injury. These findings underscore the advantages of using demographically corrected scores when addressing clinical or research questions based on understanding whether observed performances are atypical for a given individual, when compared with demographically comparable peers (i.e., compared with the best available estimates of their premorbid functioning). On the other hand, application of uncorrected neurocognitive scores may be particularly useful when determining absolute levels of functioning or capacity to complete real world tasks that the average adult in society should be able to perform. Selection of the neurocognitive score type should be determined by the research or clinical question being asked (i.e., relative changes vs. total capacity?). Maintaining an understanding of these complex neuro-cognitive relationships is therefore critical for accurate test interpretation, which can subsequently impact what expectations for recovery clinicians have for their patients and how they approach treatment planning.

References

- Bauer PJ, Dikmen SS, Heaton RK, Mungas D, Slotkin J, & Beaumont JL (2013). III. NIH Toolbox Cognition Battery (CB): Measuring episodic memory. *Monographs of the Society for Research in Child Development*, 78, 34–48. <http://dx.doi.org/10.1111/mono.12033> [PubMed: 23952201]
- Beaumont JL, Havlik R, Cook KF, Hays RD, Wallner-Allen K, Korper SP, ... Gershon R (2013). Norming plans for the NIH Toolbox. *Neurology*, 80, S87–S92. <http://dx.doi.org/10.1212/WNL.0b013e3182872e70> [PubMed: 23479550]
- Blake TM, Fichtenberg NL, & Abeare CA (2009). Clinical utility of demographically corrected WAIS-III subtest scores after traumatic brain injury. *The Clinical Neuropsychologist*, 23, 373–384. <http://dx.doi.org/10.1080/13854040802279691> [PubMed: 18671155]
- Byrd DA, Touradji P, Tang MX, & Manly JJ (2004). Cancellation test performance in African American, Hispanic, and White elderly. *Journal of the International Neuropsychological Society*, 10, 401–411. <http://dx.doi.org/10.1017/S1355617704103081> [PubMed: 15147598]
- Carlozzi NE, Beaumont JL, Tulskey DS, & Gershon RC (2015). The NIH Toolbox pattern comparison processing speed test: Normative data. *Archives of Clinical Neuropsychology*, 30, 359–368. <http://dx.doi.org/10.1093/arclin/acv031> [PubMed: 26025230]
- Carlozzi NE, Tulskey DS, Chiaravalloti ND, Beaumont JL, Weintraub S, Conway K, & Gershon RC (2014). NIH Toolbox Cognitive Battery (NIHTB-CB): The NIHTB pattern comparison processing speed test. *Journal of the International Neuropsychological Society*, 20, 630–641. <http://dx.doi.org/10.1017/S1355617714000319> [PubMed: 24960594]
- Carlozzi NE, Tulskey DS, Kail RV, & Beaumont JL (2013). VI. NIH Toolbox Cognition Battery (CB): Measuring processing speed. *Monographs of the Society for Research in Child Development*, 78, 88–102. <http://dx.doi.org/10.1111/mono.12036> [PubMed: 23952204]
- Carvalho JO, Tommet D, Crane PK, Thomas ML, Claxton A, Habeck C, ... Romero HR (2015). Deconstructing racial differences: The effects of quality of education and cerebrovascular risk factors. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 70, 545–556. <http://dx.doi.org/10.1093/geronb/gbu086>
- Casaleto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, & Heaton RK (2015). Demographically-corrected normative standards for the English version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, 21, 378–391. <http://dx.doi.org/10.1017/S1355617715000351> [PubMed: 26030001]
- Casaleto KB, Umlauf A, Marquine M, Beaumont JL, Mungas D, Gershon R, ... Heaton RK (2016). Demographically-corrected normative standards for the Spanish language version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, 22, 364–374. <http://dx.doi.org/10.1017/S135561771500137X> [PubMed: 26817924]
- Cattell R (1971). *Abilities: Their structure, growth, and action*. Cambridge, UK: Cambridge University Press.
- Diehr MC, Cherner M, Wolfson TJ, Miller SW, Grant I, & Heaton RK (2003). The 50- and 100-item short forms of the Paced Auditory Serial Addition Task (PASAT): Demographically corrected norms and comparisons with the full PASAT in normal and clinical samples. *Journal of Clinical and Experimental Neuropsychology*, 25, 571–585. <http://dx.doi.org/10.1076/jcen.25.4.571.13876> [PubMed: 12911108]
- Dikmen SS, Bauer PJ, Weintraub S, Mungas D, Slotkin J, Beaumont JL, ... Heaton RK (2014). Measuring episodic memory across the lifespan: NIH Toolbox Picture Sequence Memory Test. *Journal of the International Neuropsychological Society*, 20, 611–619. <http://dx.doi.org/10.1017/S1355617714000460> [PubMed: 24960230]
- Gasquoine PG (2009). Race-norming of neuropsychological tests. *Neuropsychology Review*, 19, 250–262. <http://dx.doi.org/10.1007/s11065-009-9090-5> [PubMed: 19294515]
- Gershon RC, Cook KF, Mungas D, Manly JJ, Slotkin J, Beaumont JL, & Weintraub S (2014). Language measures of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, 20, 642–651. <http://dx.doi.org/10.1017/S1355617714000411> [PubMed: 24960128]

- Gershon RC, Slotkin J, Manly JJ, Blitz DL, Beaumont JL, Schnipke D, Weintraub S (2013). IV. NIH Toolbox Cognition Battery (CB): Measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development*, 78, 49–69. <http://dx.doi.org/10.1111/mono.12034> [PubMed: 23952202]
- Gladso JA, Schuman CC, Evans JD, Peavy GM, Miller SW, & Heaton RK (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*, 6, 147–178. <http://dx.doi.org/10.1177/107319119900600204> [PubMed: 10335019]
- Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, ...Gershon R (2014). Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *Journal of the International Neuropsychological Society*, 20, 588–598. <http://dx.doi.org/10.1017/S1355617714000241> [PubMed: 24960398]
- Heaton R, Miller S, Taylor J, & Grant I (2004). Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. Lutz, FL: Psychological Assessment, Inc.
- Heaton RK, Ryan L, & Grant I (2009). Demographic influences and use of demographically corrected norms in neuropsychological assessment In Grant I & Adams K (Eds.), *Neuropsychological assessment of neuropsychiatric and neuromedical disorders* (pp. 127–155). New York, NY: Oxford University Press.
- Kashluba S, Hanks RA, Casey JE, & Millis SR (2008). Neuropsychologic and functional outcome after complicated mild traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 89, 904–911. <http://dx.doi.org/10.1016/j.apmr.2007.12.029> [PubMed: 18452740]
- Lenze EJ, Munin MC, Dew MA, Rogers JC, Seligman K, Mulsant BH, & Reynolds CF, III. (2004). Adverse effects of depression and cognitive impairment on rehabilitation participation and recovery from hip fracture. *International Journal of Geriatric Psychiatry*, 19, 472–478. <http://dx.doi.org/10.1002/gps.1116> [PubMed: 15156549]
- Manly JJ (2006). Deconstructing race and ethnicity: Implications for measurement of health outcomes. *Medical Care*, 44, S10–S16. <http://dx.doi.org/10.1097/01.mlr.0000245427.22788.be> [PubMed: 17060816]
- Manly JJ, Jacobs DM, Touradji P, Small SA, & Stern Y (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, 8, 341–348. <http://dx.doi.org/10.1017/S1355617702813157> [PubMed: 11939693]
- McFie J (1975). *Assessment of organic intellectual impairment*. New York, NY: Oxford University Press.
- Mungas D, Reed BR, Farias ST, & Decarli C (2009). Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons. *Psychology and Aging*, 24, 116–128. <http://dx.doi.org/10.1037/a0013421> [PubMed: 19290743]
- Mungas D, Reed BR, Haan MN, & González H (2005). Spanish and English neuropsychological assessment scales: Relationship to demographics, language, cognition, and independent function. *Neuropsychology*, 19, 466–475. <http://dx.doi.org/10.1037/0894-4105.19.4.466> [PubMed: 16060821]
- Norman MA, Evans JD, Miller WS, & Heaton RK (2000). Demographically corrected norms for the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, 22, 80–94. [http://dx.doi.org/10.1076/1380-3395\(200002\)22:1;1-8;FT080](http://dx.doi.org/10.1076/1380-3395(200002)22:1;1-8;FT080) [PubMed: 10649547]
- Norman MA, Moore DJ, Taylor M, Franklin D, Jr., Cysique L, Ake C, ...Heaton RK (2011). Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. *Journal of Clinical and Experimental Neuropsychology*, 33, 793–804. <http://dx.doi.org/10.1080/13803395.2011.559157> [PubMed: 21547817]
- O'Connell ME, & Tuokko H (2010). Age corrections and dementia classification accuracy. *Archives of Clinical Neuropsychology*, 25, 126–138. <http://dx.doi.org/10.1093/arclin/acp111> [PubMed: 20118110]
- O'Connell ME, Tuokko H, & Kadlec H (2011). Demographic corrections appear to compromise classification accuracy for severely skewed cognitive tests. *Journal of Clinical and Experimental*

Neuropsychology, 33, 422–431. <http://dx.doi.org/10.1080/13803395.2010.532114> [PubMed: 21154077]

- Poynter L, Kwan J, & Vassallo M (2013). How does cognitive impairment impact on functional improvement following the rehabilitation of elderly patients? *International Journal of Clinical Practice*, 67, 811–815. <http://dx.doi.org/10.1111/ijcp.12161> [PubMed: 23869683]
- Rabadi MH, Rabadi FM, Edelstein L, & Peterson M (2008). Cognitively impaired stroke patients do benefit from admission to an acute rehabilitation unit. *Archives of Physical Medicine and Rehabilitation*, 89, 441–448. <http://dx.doi.org/10.1016/j.apmr.2007.11.014> [PubMed: 18295621]
- Reitan RM, & Wolfson D (1995). Influence of age and education on neuropsychological test results. *The Clinical Neuropsychologist*, 9, 151–158. <http://dx.doi.org/10.1080/13854049508401597>
- Reitan RM, & Wolfson D (1997). The influence of age and education on neuropsychological performances of persons with mild head injuries. *Applied Neuropsychology*, 4, 16–33. http://dx.doi.org/10.1207/s15324826an0401_3 [PubMed: 16318492]
- Reitan RM, & Wolfson D (2005). The effect of age and education transformations on neuropsychological test scores of persons with diffuse or bilateral brain damage. *Applied Neuropsychology*, 12, 181–189. http://dx.doi.org/10.1207/s15324826an1204_1 [PubMed: 16422659]
- Sherrill-Pattison S, Donders J, & Thompson E (2000). Influence of demographic variables on neuropsychological test performance after traumatic brain injury. *The Clinical Neuropsychologist*, 14, 496–503. <http://dx.doi.org/10.1076/clin.14.4.496.7196> [PubMed: 11262719]
- Silverberg ND, & Millis SR (2009). Impairment versus deficiency in neuropsychological assessment: Implications for ecological validity. *Journal of the International Neuropsychological Society*, 15, 94–102. <http://dx.doi.org/10.1017/S1355617708090139> [PubMed: 19128532]
- Sliwinski M, Buschke H, Stewart WF, Masur D, & Lipton RB (1997). The effect of dementia risk factors on comparative and diagnostic selective reminding norms. *Journal of the International Neuropsychological Society*, 3, 317–326. [PubMed: 9260441]
- Sliwinski M, Hofer SM, Hall C, Buschke H, & Lipton RB (Eds.). (2003). *Optimizing cognitive test norms for detection*. New York, NY: Oxford University Press.
- Strong CA, Donders J, & van Dyke S (2005). Validity of demographically corrected norms for the WAIS-III. *Journal of Clinical and Experimental Neuropsychology*, 27, 746–758. <http://dx.doi.org/10.1081/13803390490919155> [PubMed: 16019650]
- Taylor MJ, & Heaton RK (2001). Sensitivity and specificity of WAIS-III/WMS-III demographically corrected factor scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 7, 867–874. [PubMed: 11771630]
- Teasdale G, & Jennett B (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet*, 2, 81–84. [http://dx.doi.org/10.1016/S0140-6736\(74\)91639-0](http://dx.doi.org/10.1016/S0140-6736(74)91639-0) [PubMed: 4136544]
- The Psychological Corporation. (2001). *Manual for the Wechsler Test of Adult Reading (WTAR)*. San Antonio, TX: Psychological Corporation.
- Traumatic Brain Injury Model Systems National Data Center. (2006). *Traumatic brain injury model systems national data base inclusion criteria*. Retrieved from <https://www.tbinsc.org/StaticFiles/Documents/2015%20TBIMS%20Slide%20Presentation.pdf>
- Tulsky DS, Carlozzi NE, Chevalier N, Espy KA, Beaumont JL, & Mungas D (2013). V. NIH Toolbox Cognition Battery (CB): Measuring working memory. *Monographs of the Society for Research in Child Development*, 78, 70–87. <http://dx.doi.org/10.1111/mono.12035> [PubMed: 23952203]
- Tulsky DS, Carlozzi N, Chiaravalloti ND, Beaumont JL, Kisala PA, Mungas D, ... Gershon R (2014). NIH Toolbox Cognition Battery (NIHTB-CB): List sorting test to measure working memory. *Journal of the International Neuropsychological Society*, 20, 599–610. <http://dx.doi.org/10.1017/S135561771400040X> [PubMed: 24959983]
- van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, & van Gijn J (1988). Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*, 19, 604–607. <http://dx.doi.org/10.1161/01.STR.19.5.604> [PubMed: 3363593]
- Wechsler D (1935). *The range of human capacities* (1st ed). Baltimore, MD: Williams & Wilkins Company <http://dx.doi.org/10.1037/11223-000>

- Wechsler D (1952). *The range of human capacities* (2nd ed). Baltimore, MD: Williams & Wilkins Company.
- Wechsler D (1958). *The measurement and appraisal of adult intelligence* (Vol. 4). New York, NY: Psychological Corporation <http://dx.doi.org/10.1037/11167-000>
- Weintraub S, Bauer PJ, Zelazo PD, Wallner-Allen K, Dikmen SS, Heaton RK, ... Gershon RC (2013a). I. NIH Toolbox Cognition Battery (CB): Introduction and pediatric data. *Monographs of the Society for Research in Child Development*, 78, 1–15. <http://dx.doi.org/10.1111/mono.12031>
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, ... Gershon RC (2013b). Cognition assessment using the NIH Toolbox. *Neurology*, 80, S54–S64. <http://dx.doi.org/10.1212/WNL.0b013e3182872ded> [PubMed: 23479546]
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Slotkin J, ... Gershon R (2014). The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: Validation in an adult sample. *Journal of the International Neuropsychological Society*, 20, 567–578. <http://dx.doi.org/10.1017/S1355617714000320> [PubMed: 24959840]
- Williams DH, Levin HS, & Eisenberg HM (1990). Mild head injury classification. *Neurosurgery*, 27, 422–428. <http://dx.doi.org/10.1227/00006123-199009000-00014> [PubMed: 2234336]
- Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Conway KP, ... Weintraub S (2014). NIH Toolbox Cognition Battery (CB): Validation of executive function measures in adults. *Journal of the International Neuropsychological Society*, 20, 620–629. <http://dx.doi.org/10.1017/S1355617714000472> [PubMed: 24960301]
- Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, & Weintraub S (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, 78, 16–33. <http://dx.doi.org/10.1111/mono.12032> [PubMed: 23952200]
- Zinn S, Dudley TK, Bosworth HB, Hoenig HM, Duncan PW, & Horner RD (2004). The effect of poststroke cognitive impairment on rehabilitation process and functional outcome. *Archives of Physical Medicine and Rehabilitation*, 85, 1084–1090. <http://dx.doi.org/10.1016/j.apmr.2003.10.022> [PubMed: 15241754]

Impact and Implications

This article extends previous research comparing demographically corrected and uncorrected scores from the National Institutes of Health Toolbox–Cognition Battery (NIHT-CB) with traumatic brain injury (TBI) and stroke populations, and demonstrates the increased construct validity of the NIHT-CB for classifying injury severity when using demographically corrected versus uncorrected fluid composite scores. This article highlights the importance of factoring out unwanted variance associated with patient demographic characteristics (e.g., age, education, gender, etc.) when interpreting the impact of acquired brain injury. In clinical practice, knowing when to use demographically corrected scores versus uncorrected scores is important for interpreting a patient’s performance on neuropsychological testing and can assist clinicians in treatment planning by providing a more accurate reflection of current cognitive functioning by removing test variance associated with demographic characteristics. This article demonstrates that NIHTB-CB crystallized composite scores may serve as a measure of premorbid functioning, as it was relatively resistant to the deleterious effects of acquired brain injury.

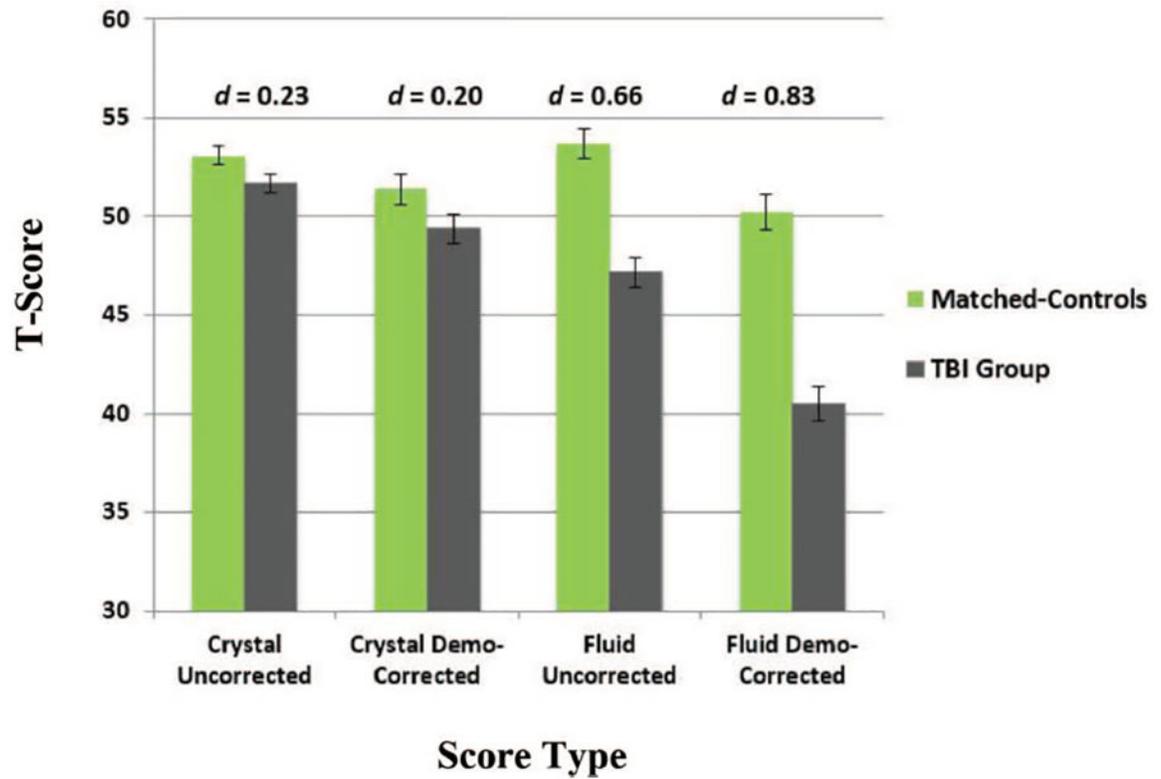


Figure 1.

Effect sizes for score types across fluid and crystallized composites: TBI versus TBI-matched controls. See the online article for the color version of this figure.

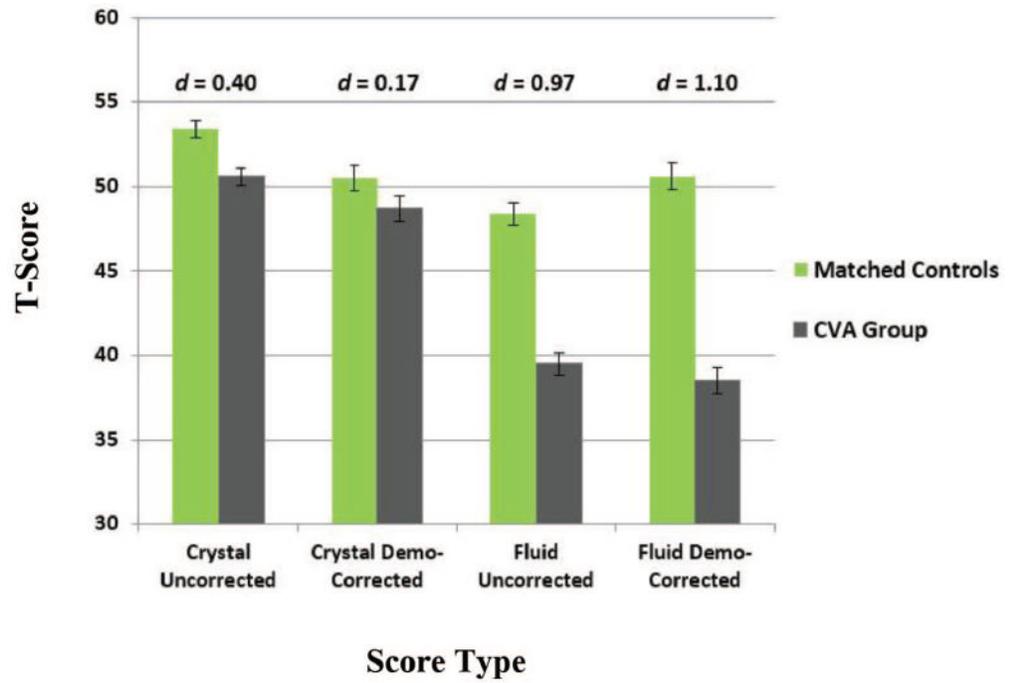


Figure 2. Effect sizes for score types across fluid and crystallized composites: stroke versus stroke-matched controls. See the online article for the color version of this figure.

Demographic and Injury Information for Traumatic Brain Injury (TBI), Stroke, and Matched Control Groups

Table 1

Participant characteristic	TBI (<i>n</i> = 184) Mean (SD)	TBI-Controls (<i>n</i> = 184) Mean (SD)	Stroke (<i>n</i> = 211) Mean (SD)	Stroke-Controls (<i>n</i> = 210) Mean (SD)
Age (years)	39.1 (17.0)	39.5 (17.1)	56.1 (13.0)	56.5 (13.3)
Education (years)	13.7 (2.5)	14.0 (2.7)	13.7 (2.6)	13.9 (2.5)
Sex (% male)	64.7% (<i>n</i> = 119)	62.5% (<i>n</i> = 115)	49.8% (<i>n</i> = 105)	43.8% (<i>n</i> = 92)
Race/ethnicity				
White/Asian	75.0% (<i>n</i> = 138)	71.7% (<i>n</i> = 132)	42.1% (<i>n</i> = 91)	63.8% (<i>n</i> = 134)
African American	20.7% (<i>n</i> = 38)	15.8% (<i>n</i> = 29)	53.1% (<i>n</i> = 112)	22.9% (<i>n</i> = 48)
Hispanic	4.3% (<i>n</i> = 8)	12.5% (<i>n</i> = 23)	3.8% (<i>n</i> = 8)	13.3% (<i>n</i> = 28)
Injury severity				
Complicated/mild	37.6% (<i>n</i> = 65)		28.5% (<i>n</i> = 59)	
Moderate	6.9% (<i>n</i> = 23)		26.6% (<i>n</i> = 55)	
Severe	55.5% (<i>n</i> = 96)		44.9% (<i>n</i> = 93)	

Table 2
 Proportion of Variance Accounted for (R^2) by Demographic Characteristics on Uncorrected vs. Corrected Fluid and Crystallized Composite Scores for TBI Participants

Demographic characteristic	Traumatic brain injury (TBI) group ($n = 184$)			
	Uncorrected fluid	Demo—corrected fluid	Uncorrected crystallized	Demo—corrected crystallized
Age	24.7 ^{**}	<1	5.6 [*]	<1
Education	1.5	3.0 [*]	28.9 ^{**}	<1
Sex	2.4 [*] (F > M)	<1	1.1	<1
Race	7.2 ^{**}	1.4	11.0 ^{**}	<1
Combined	37.7 ^{**}	3.6	33.4 ^{**}	1.5

* $p < .05$.

** $p < .001$.

Proportion of Variance Accounted for (R^2) by Demographic Characteristics on Uncorrected vs. Corrected Fluid and Crystallized Composite Scores for Stroke Participants

Table 3

Demographic characteristic	Stroke group ($n = 211$)			
	Uncorrected fluid	Demo—corrected fluid	Uncorrected crystallized	Demo—corrected crystallized
Age	14.1 ^{**}	<1	<1	<1
Education	4.1 [*]	<1	18.4 ^{**}	<1
Sex	<1	<1	2.0 [*] (M > F)	<1
Race	9.1 ^{**}	3.9 [*]	18.1 ^{**}	2.1 [*]
Combined	28.9 ^{**}	4.8	31.5 ^{**}	3.5

* $p < .05$.

** $p < .001$.