Comparing Actual to Estimated Base Rates of “Abnormal” Scores on Neuropsychological Test Batteries: Implications for Interpretation

Brian L. Brooks\textsuperscript{a,b,*}, Grant L. Iverson\textsuperscript{c,d}

\textsuperscript{a}Alberta Children’s Hospital, Calgary, AB, Canada
\textsuperscript{b}University of Calgary, Calgary, AB, Canada
\textsuperscript{c}University of British Columbia, Vancouver, BC, Canada
\textsuperscript{d}British Columbia Mental Health and Addiction Services, Coquitlam, BC, Canada

Accepted 23 November 2009

Abstract

Clinicians can use the prevalence of low scores to help interpret test performance. However, this information is limited for most test batteries. In 2007, Crawford, Garthwaite, and Gault presented Monte Carlo simulation software for estimating the base rates of low scores for any battery of tests. The purpose of this study is to examine the accuracy of a Monte Carlo simulation program for estimating the base rates of low scores. Base rates of low scores were: (a) calculated from large normative samples (actual base rates) for the Neuropsychological Assessment Battery and the Wechsler Adult Intelligence Scale-III/Wechsler Memory Scale-Third Edition and compared to (b) Monte Carlo estimations (estimated base rates). Monte Carlo estimations of the base rates of low scores had good accuracy when compared with the actual base rates of low scores for the two batteries. However, estimated base rates lose considerable accuracy in those with low or high intelligence. Monte Carlo simulation software is a potential option for clinicians to compute the base rates of low scores for any battery with published intercorrelations. However, the Monte Carlo program underestimates the base rates for those with low intelligence and overestimates the base rates for those with high intelligence.

Keywords: Assessment; Base rates; Low scores; Monte Carlo; Estimation; Misdiagnosis

Introduction

Test score variability, large differences between high and low scores, and the presence of some “abnormal” scores in healthy people are psychometric truisms that have long been appreciated by clinicians who administer and interpret neuropsychological tests (see Binder, Iverson, & Brooks, 2009 for a recent review). There has been considerable recent interest in the prevalence of low or “abnormal” scores in healthy people, and the impact these base rates have on test interpretation. Iverson and Brooks (in press) and Brooks, Strauss, Sherman, Iverson, and Slick (2009c) discuss five psychometric principles that should be considered when clinicians are interpreting multiple tests. First, having low scores is common across all batteries and is not a function of any particular battery, age group, or cognitive domain. Second, the number of low scores depends on where a clinician sets the cutoff for identifying low performance. Heaton, Grant, and Matthews (1991) and Heaton, Miller, Taylor, and Grant (2004) considered the prevalence of low scores that are more than 1 \textit{SD} below the mean. However, if a clinician chooses a more stringent criterion for abnormality, such as 2 \textit{SD}s below the mean, then there will be fewer low scores. Third, the number of low scores is related to the number of tests administered and the intercorrelations between the tests. As more tests are administered and interpreted, and as the intercorrelations of test scores get smaller, the chances of healthy people having low scores increases. Fourth, the number of low scores is related to the demographic characteristics of the sample. For example, the
prevalence of low scores can be different on some tests depending on the sex, education level, or ethnic group that is being considered. Finally, the number of low scores is related to the level of intellectual functioning. People with higher intelligence are less likely to obtain low scores than those with lesser intelligence.

Knowing the prevalence of low scores across a battery of tests is a valuable psychometric component of test interpretation. However, clinicians need to know more than the principle that “low scores occur”: They need to have ready-to-use, clinically friendly tables that allow them to consider this information when making clinical decisions. The prevalence of low or “abnormal” test scores is known for a limited number of test batteries and can be used to supplement clinical decision-making. For example, base rates using normative data have been presented for several batteries of cognitive tests, including the Expanded Halstead Reitan Neuropsychological Battery (EHRNB; see Binder et al., 2009; Heaton et al., 1991, 2004; Iverson & Brooks, in press for base rates), the Neuropsychological Assessment Battery (NAB; see Iverson & Brooks, in press; Iverson, Brooks, White, & Stern, 2008b for base rates), and the Wechsler Adult Intelligence Scale – Third Edition/Wechsler Memory Scale – Third Edition battery (WAIS-III/WMS-III; see Iverson & Brooks, in press; Iverson, Brooks, & Holdnack, 2008a for base rates). The base rates of low scores are also known for flexible batteries, including Schretlen and colleagues’ Aging, Brain Imaging, and Cognition (ABC) battery (Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008; Testa & Schretlen, 2006) and Palmer and colleagues’ battery of tests (Palmer, Boone, Lesser, & Wohl, 1998). The prevalence of low scores has also been examined for batteries of memory tests across the lifespan, including the Children’s Memory Scale (CMS; see Brooks, Iverson, Sherman, & Holdnack, 2009b for base rates), the NAB Memory module (see Brooks, Iverson, & White, 2007 for base rates), and the WMS-III (see Brooks, Iverson, Feldman, & Holdnack, 2009a; Brooks, Iverson, Holdnack, & Feldman, 2008 for base rates).

These published studies allow a clinician to incorporate the prevalence of low scores across a battery of tests when interpreting test performance. However, a potential drawback is that the clinician must administer all of the same tests that were used to compute the prevalence of low scores in the studies. In other words, the base rates tables cannot be used if a test, which was included in the base rates analyses, is not included in the clinical assessment. In situations when a battery of tests differs from those presented in the above-mentioned studies, clinicians can calculate the prevalence of low scores using mathematical models. Crawford, Garthwaite, and Gault (2007) critiqued the binomial method (Ingraham & Aiken, 1996) for (a) underestimating the prevalence of low scores when the number of scores examined is low and (b) overestimating the base rates of low scores when the number of scores examined is high. In response, Crawford and colleagues presented a Monte Carlo simulation program, which is designed to estimate the prevalence of low scores for any battery of tests using the intercorrelations between scores. Schretlen and colleagues (2008) compared Monte Carlo estimations to actual base rates using their Aging, Brain Imaging, and Cognition (ABC) battery. On the 10-, 25-, and 43-measure versions of the ABC battery, the differences were <5% for most comparisons and <9% for all comparisons. As the cutoff score became more stringent, the differences between actual and estimated base rates decreased to only a few percentage points. Schretlen and colleagues (2008) concluded that the Monte Carlo estimations “could be clinically useful” (p. 444), particularly when trying to minimize over-interpretation of spurious low scores.

Although the Monte Carlo simulation method allows clinicians to estimate the base rates for low scores for any battery of tests and the use of this method has been suggested when actual base rates are not published (e.g., Binder et al., 2009), little information is known about the consistency or agreement with actual base rates for commonly used batteries. The purpose of this study is to examine whether the Monte Carlo simulation software can accurately estimate the base rates of low scores for test batteries. Based on our clinical experience using the Monte Carlo program and the results presented by Schretlen and colleagues (2008), it is our a priori hypothesis that the estimated base rates will be reasonably close to actual base rates. However, it is also our a priori hypothesis that the Monte Carlo-estimated base rates will be poor at estimating the base rates in those with low or high intelligence. This is because the intercorrelation matrices do not account for intelligence as a potential moderator variable.

Materials and Methods

Participants and Measures

The participants for this study were derived from two different standardization samples from two different test batteries. A brief description of each participant sample and the corresponding test batteries is presented below. The treatment of participants and the collection of data were done in compliance with the Helsinki Declaration.

Neuropsychological Assessment Battery. The NAB (Stern & White, 2003) standardization sample consists of 1,448 healthy, community-dwelling individuals from the USA. Rigorous exclusion criteria were employed in order to prevent the inclusion of
persons who could have a neurological disease, acquired injury, psychiatric illness, treatment/medication, or physical impairment that would negatively impact test performance. Participants for the standardization sample were recruited and tested at five sites across the country, including Rhode Island Hospital, University of Florida Health Sciences Center, Indiana University, University of California at Los Angeles School of Medicine, and the Psychological Assessment Resources offices in Lutz, Florida. These sites were chosen to represent the four geographical regions of the country (Northeast, Midwest, West, and South; see White & Stern, 2003).

The NAB is a comprehensive modular battery of tests normed for adults between the ages of 18–97 years. Demographically adjusted norms are used. The full NAB consists of 36 individual tests across five cognitive domains. From these 36 tests, 30 scores contribute towards five separate index scores and one total battery index score. For this study, the following five index scores were included in the analyses: Attention, Language, Memory, Spatial, and Executive Functions. The Total NAB index was not included in the analyses. A measure of intellectual abilities, the Reynolds Intellectual Screening Test (RIST; Reynolds & Kamphaus, 2003) was co-administered to the NAB standardization sample.

Wechsler Adult Intelligence Scale – III/Wechsler Memory Scale – III. The WAIS-III (Wechsler, 1997a) and WMS-III (Wechsler, 1997b) were developed for use with adults between 16 and 89 years of age. These batteries were co-normed using a stratified, US-representative sample of 1,250 healthy adults. Participants were included in the standardization sample if they were medically and psychiatrically healthy, based on a self-report questionnaire. The exclusion criteria included color blindness, uncorrected hearing loss and/or visual impairment, upper extremity motor problems that might interfere with testing, current treatment for alcohol or drug dependence, consumption of three or more alcoholic beverages on two or more nights per week, having sought attention from a professional for memory or cognitive problems, a history of traumatic brain injury involving loss of consciousness for five or more minutes and/or requiring hospitalization for more than 24 hr, any medical or psychiatric condition that could potentially impact cognitive functioning (e.g., stroke, epilepsy, brain surgery, encephalitis, meningitis, multiple sclerosis, Parkinson’s disease, Huntington’s chorea, Alzheimer’s disease, schizophrenia, or bipolar disorder), or currently receiving any treatment for a medical or psychiatric condition (e.g., electroconvulsive therapy or any antidepressant, anxiolytic, or neuroleptic medication; The Psychological Corporation, 1997).

The WAIS-III is a battery of measures designed to evaluate verbal intelligence, non-verbal intelligence, processing speed, and verbal working memory. The WAIS-III consists of 14 subtests, of which there are 11 subtests that yield four index scores (e.g., Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed). The WMS-III is a battery of memory measures designed to evaluate working memory, learning, immediate and delayed recall, and recognition of information presented in verbal and visual modalities. The WMS-III is comprised of four primary subtests that yield eight index scores (e.g., Auditory Immediate Memory, Visual Immediate Memory, Immediate Memory, Auditory Delayed Memory, Visual Delayed Memory, Auditory Recognition Delayed Memory, General Memory, and Working Memory). In total, there are 12 index scores on the WAIS-III and WMS-III. Intellectual abilities were estimated using the Wechsler Test of Adult Reading (WTAR; The Psychological Corporation, 2001), a measure of single word reading that requires the reading and pronunciation of words with irregular grapheme-to-phoneme translation. The WTAR was co-normed with the WAIS-III (Wechsler, 1997a); therefore, the WTAR reading score can be combined with demographic variables to estimate intellectual abilities (i.e., WTAR-demographically predicted Full Scale IQ, with mean = 100 and SD = 15). The WTAR-demographically predicted FSIQ was chosen as the measure of predicted intellectual abilities in order to avoid potential circularity of having the performance on the WAIS-III/WMS-III stratified by WAIS-III FSIQ.

Analyses

Calculation of the base rates of low index scores was based on considering the performance across all scores simultaneously. Different cutoff scores (i.e., <16th percentile, <10th percentile, ≤5th percentile, and ≤2nd percentile) can be used to determine a “low” or “abnormal” score and there is no current standard for which cutoff to use. In order to keep the presentation of data manageable, only the results for the <10th percentile cutoff will be presented in the figures. Analyses involved comparing the base rates from standardization samples to the Monte Carlo-estimated base rates of low index scores in healthy people for the two test batteries (NAB and WAIS-III/WMS-III). To distinguish between these two analyses, we refer to the base rates from the NAB and WAIS-III/WMS-III standardization samples as “actual” base rates and the Monte Carlo base rates as “estimated”, although we do appreciate that analyses from a standardization sample are still estimates of the population base rates.

Base rate analyses that used standardization data from each test battery (i.e., actual base rates) involved calculating the prevalence of low index scores when all index scores from that battery were considered simultaneously. The actual base rates were calculated for each normative sample in its entirety, as well as by stratifying the samples by levels of intelligence. The NAB sample was stratified by RIST scores, which included: “unusually low” (RIST < 80, n = 28); “low average”
(RIST = 80–89, n = 138); “average” (RIST = 90–109, n = 735); “high average” (RIST = 110–119, n = 316); and “superior” (RIST = 120+, n = 204). There are 29 participants from the NAB standardization sample with missing data points. The WAIS-III/WMS-III sample was stratified by WTAR-demographics predicted FSIQ scores (WTAR-FSIQ), which included: “unusually low” (WTAR-FSIQ < 80, n = 63); “low average” (WTAR-FSIQ = 80–89, n = 155); “average” (WTAR-FSIQ = 90–109, n = 700); “high average” (WTAR-FSIQ = 110–119, n = 251); and “superior” (WTAR-FSIQ = 120+, n = 32).

Calculation of the prevalence of low scores using the Monte Carlo simulation program (i.e., estimated base rates) was completed using the computer program referenced by Crawford and colleagues (2007) (available at: http://www.abdn.ac.uk/~psy086/dept/PercentAbnormKtests.htm). The Monte Carlo program estimates the base rates of low scores based on the test intercorrelations, which are published in the respective test manuals. Test intercorrelations for the NAB index scores are presented in Table C.1 of the NAB Psychometric and Technical Manual (White & Stern, 2003). Test intercorrelations for the WAIS-III/WMS-III index scores are presented in Tables 4.1, 4.3, and 4.17 of the WAIS-III/WMS-III Technical Manual (The Psychological Corporation, 1997).

Results

The base rates of NAB index scores below the 10th percentile are presented in Fig. 1. When considering the actual base rates in the NAB standardization sample, one or more index scores <10th percentile is found in 25.8% of healthy adults. Monte Carlo estimation suggests that one or more index scores <10th percentile is obtained by 29.1% of healthy adults (difference = +3.3%; χ²(1) = 4.00, p = .046, odds ratio [OR] = 1.18 [95% CI = 1.0–1.4]). Estimations of two or more, three or more, four or more, and five low index scores on the NAB were within ±0.3%, ±1.0%, ±1.0%, and ±0.7%, respectively, of the actual base rates. Differences between actual and estimated base rates of low NAB index scores for different cutoff scores were within 3.3% for the 16th percentile cutoff, within 1.6% for the 5th percentile cutoff, and within 1.7% for the 2nd percentile cutoff.

The base rates of WAIS-III/WMS-III index scores below the 10th percentile are presented in Fig. 2. One or more low WAIS-III/WMS-III index scores is found in 36.5% of the standardization sample compared with an estimated 40.0% (difference = +3.5%; χ²(1) = 3.28, p = .07, OR = 1.16 [95% CI = 0.99–1.4]). Differences between actual and estimated base rates are within ±4%, with many estimates being within ±1% of the actual base rates as the number of low scores increases. Differences between actual and estimated base rates of low WAIS-III/WMS-III index scores for different cutoff scores were within 1.5% for the 16th percentile cutoff and within 1.7% for the 2nd percentile cutoff.

Fig. 1. Actual and estimated base rates for the NAB indexes (<10th percentile cutoff). NAB = Neuropsychological Assessment Battery (Stern & White, 2003). Values represent cumulative percentages of the sample with the specified number of low scores, when all scores are considered simultaneously. There are slight variations because of rounding. “Actual” values are derived from base rate analyses with standardization sample (n = 1,448). “Monte Carlo” values are derived from a simulation program that is based on the intercorrelations from the standardization sample. There are five indexes from the NAB (Attention, Language, Memory, Spatial, and Executive Functions) that were considered for these analyses. Index scores <10th percentile are equal to Index < 81. Data for the NAB have been reproduced with special permission from the Publisher, Psychological Assessment Resources, Inc., Lutz, FL, USA from the NAB by Robert A. Stern, PhD, and Travis White, PhD, Copyright 2001, 2003 by PAR, Inc. Further reproduction is prohibited without permission from PAR, Inc.
Chausson, Dray, Faucounau, & Rigaud (2005) illustrated that nearly half of their patients were misidentified as having
across a battery of tests can potentially lead to erroneous clinical decision-making. For example, de Rotrou, Wenisch, Heaton et al., 1991, 2004; Iverson et al., 2008a, 2008b; Palmer et al., 1998; Schretlen et al., 2008; Testa & Schretlen, 2006).

As illustrated in Fig. 3, there are considerable differences in the prevalence of low scores across levels of intelligence. For the NAB, one or more index scores <10th percentile are estimated to occur in 29.1% of adults using the Monte Carlo program (Fig. 1). However, according to analyses with the normative sample, having one or more low index scores is found in 96.2% of adults with unusually low intelligence (difference $= +67.1\%$; $\chi^2(1) = 54.46, p < .001, OR = 61.0$ [95% CI = 10.4–356.0]), 63.8% of adults with low average intelligence (difference $= +34.7\%$; $\chi^2(1) = 69.59, p < .001, OR = 4.3$ [95% CI = 3.0–6.2]), 26.7% with average intelligence (difference $= -2.4\%$; $\chi^2(1) = 1.39, p = .24, OR = 1.1$ [95% CI = 0.9–1.4]), 11.1% with high average intelligence (difference $= -18.0\%$; $\chi^2(1) = 43.84, p < .001, OR = 3.3$ [95% CI = 2.3–4.8]), and in 9.6% with superior intelligence (difference $= -19.5\%$; $\chi^2(1) = 33.93, p < .001, OR = 3.8$ [95% CI = 2.4–6.0]).

On the WAIS-III/WMS-III, the Monte Carlo program estimated that 40.0% of healthy adults would have one or more index scores <10th percentile (Figs. 2 and 3). The actual base rates, stratified by level of WTAR-demographics predicted FSIQ, indicate that one or more index scores <10th percentile is found in 87.3% of adults with unusually low intelligence (difference $= +47.3\%$; $\chi^2(1) = 54.99, p < .001, OR = 10.3$ [95% CI = 4.9–21.5]), 71.6% of adults with low average intelligence (difference $= +31.6\%$; $\chi^2(1) = 56.08, p < .001, OR = 3.8$ [95% CI = 2.6–5.5]), 33.6% with average intelligence (difference $= -6.4\%$; $\chi^2(1) = 7.90, p = .005, OR = 1.3$ [95% CI = 1.1–1.6]), 12.0% with high average intelligence (difference $= -28.0\%$; $\chi^2(1) = 71.99, p < .001, OR = 4.9$ [95% CI = 3.3–7.3]), and in 6.2% with superior intelligence (difference $= -33.8\%$; $\chi^2(1) = 14.92, p < .001, OR = 10.0$ [95% CI = 2.6–38.0]). These differences between the Monte Carlo-estimated base rates and the base rates in those with either low or high intellectual abilities were consistent across various cutoff scores for both test batteries.

Discussion

The presence of some low scores for healthy people across a battery of tests is not a new concept for neuropsychologists. Recently, there have been several descriptive papers that highlight the need for clinicians to consider this information when interpreting test scores. The base rates of low scores, when simultaneously considering all test scores, is now known for several batteries that are used in clinical settings with children, adults, and older adults (Brooks et al., 2007, 2008, 2009b; Heaton et al., 1991, 2004; Iverson et al., 2008a, 2008b; Palmer et al., 1998; Schretlen et al., 2008; Testa & Schretlen, 2006).

Clinicians now have more evidence than ever before which suggests that failing to consider the prevalence of low scores across a battery of tests can potentially lead to erroneous clinical decision-making. For example, de Rotrou, Wenisch, Chausson, Dray, Faucounau, & Rigaud (2005) illustrated that nearly half of their patients were misidentified as having
prolonged dementia when tested at baseline. The misidentification was based on these patients having poor performance on one or two memory measures at baseline but normal memory performance at one-year follow-up. Brooks and colleagues (2008, 2007) reported that one out of four healthy older would meet psychometric criteria for amnestic mild cognitive impairment (i.e., 1 or more memory score \( < 1.5 \times SD \)), but could go as high as one out of two in older adults with lower intelligence. Studies that examine the prevalence of low scores in healthy people and present clinical situations where misdiagnosis can easily occur, strengthen the need to have sophisticated, yet simple-to-use, psychometric analyses for determining whether a low score is common or uncommon in the population.

Unfortunately, there are only a limited number of published tables that can be used and clinicians must administer all tests that were included in the analyses for those tables. For example, using the base rate tables presented for the NAB subtests in Crawford and colleagues' (2008b) study, little is known about the level of accuracy of the Monte Carlo simulation program in comparison with actual base rates. For the present study, comparisons were made between the actual and estimated base rates low scores for two batteries of tests used in adults and older adults—the NAB and the WAIS-III/WMS-III.

The results of this study indicate that the differences between the actual and Monte Carlo-estimated base rates of low scores on a battery of tests are relatively small. When considering the differences on the NAB, the Monte Carlo simulation program was within 4% of the actual base rates using various cutoff scores (it should be acknowledged that the group difference was significant at \( p = 0.046 \), perhaps because of having very few persons with low intelligence relative to those with higher intelligence), and many differences were within 1%–2%. For the WAIS-III/WMS-III, Monte Carlo-estimated base rates were within

![Fig. 3. Prevalence of one or more low index scores (<10th percentile) for Monte Carlo estimates and different levels of intelligence. NAB = Neuropsychological Assessment Battery (Stern & White, 2003); WAIS-III/WMS-III = Wechsler Adult Intelligence Scale – Third Edition/Wechsler Memory Scale – Third Edition (Wechsler, 1997). Values represent cumulative percentages of the sample with the specified number of low scores, when all scores are considered simultaneously. "Monte Carlo" values are derived from a simulation program that is based on the intercorrelations from the standardization samples. “Actual” values, presented for different intelligence subsamples, are derived from base rate analyses with the NAB and WAIS-III/WMS-III standardization samples. Intellectual abilities for the NAB sample are based on RIST scores (mean = 100, SD = 15), where: “unusually low” is RIST < 80 (n = 26); “low average” is RIST = 80–89 (n = 138); “average” is RIST = 90–109 (n = 735); “high average” is RIST = 110–119 (n = 316); and “superior” is RIST = 120+ (n = 204). There are 29 participants from the NAB standardization sample with missing data. There are five indexes from the NAB (Attention, Language, Memory, Spatial, and Executive Functions) that were considered for these analyses. Data for the NAB have been reproduced with special permission from the Publisher, Psychological Assessment Resources, Lutz, FL, USA from the Neuropsychological Assessment Battery by Robert A. Stern, PhD, and Travis White, PhD, Copyright 2001, 2003 by PAR, Inc. Further reproduction is prohibited without permission from PAR, Inc. Intellectual abilities for the WAIS-III/WMS-III sample are based on WTA-R-Demographics Predicted FSIQ scores (WTA-R-Demo; mean = 100, SD = 15), where: “unusually low” is WTA-R-Demo < 80 (n = 63); “low average” is WTA-R-Demo = 80–89 (n = 155); “average” is WTA-R-Demo = 90–109 (n = 700); “high average” is WTA-R-Demo = 110–119 (n = 251); and “superior” is WTA-R-Demo = 120+ (n = 32). There are 12 indexes from the WAIS-III (Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed) and WMS-III (Auditory Immediate Memory, Visual Immediate Memory, Immediate Memory, Auditory Delayed Memory, Visual Delayed Memory, Auditory Recognition Delayed Memory, General Memory, and Working Memory) that were considered for these analyses. Index scores <10th percentile are equal to Index < 81. Data presented for the WAIS-III and the WMS-III are copyright © 1997 by NCS Pearson, Inc. Used with permission. All rights reserved.
5% of the actual base rates, and again many differences were within 1%–2% across the various cutoff scores. These small differences between estimated and actual base rates of low scores suggest that the Monte Carlo simulation program is a viable option for determining the prevalence of low scores.

Crawford’s Monte Carlo simulation program is psychometrically similar to actual base rates for the overall standardization samples, but not for subsamples that are divided by moderator variables. For example, both the actual and estimated base rates indicate that having one or more low index scores is common in healthy people. On the NAB, having one or more index scores <10th percentile is found in 25.8% of people and is Monte Carlo-estimated to occur in 29.1% of people. This should be critically important to clinicians because the NAB Index scores relate to “domains” of functioning—and one of four healthy adults have “impairment” in at least one domain (i.e., attention, language, memory, spatial, or executive functioning). In other words, from a strictly Gaussian perspective, the “false-positive” rate for cognitive impairment on a single domain is 9% when considering a single index score but it is 25.8% when considering the five domain scores simultaneously. On the WAIS-III/WMS-III, having one or more index scores <10th percentile is found in 36.5% of people and is estimated (using Monte Carlo) to occur in 40.0% of people. Both methods (actual and estimated) also identify a similar number of low scores as being “uncommon”. On the NAB, it is “uncommon” to have three or more index scores <10th percentile based on both the actual base rates and on the Monte Carlo-estimated base rates. For the WAIS-III/WMS-III, it is “uncommon” to have five or more index scores <10th percentile based on both calculations of the base rates of low scores.

In the study of Schretlen and colleagues (2008), the Monte Carlo-estimated base rates were relatively similar to the actual base rates for three different versions of the ABC battery (i.e., 10, 25, and 43 subtests) using three different cutoff scores (i.e., T < 40 or <16th percentile, T < 35 or <7th percentile, and T < 30 or <2nd percentile). Rates of accuracy for the unadjusted test scores ranged from 4.0% to 6.1% with the 16th percentile cutoff, from 1.2% to 2.6% with the 7th percentile cutoff, and from 0.0% to 2.3% with the 2nd percentile cutoff. Accuracy between estimated and actual base rates using the adjusted test scores (i.e., adjusted for demographic variables and premorbid IQ) was also fairly good, ranging from 0.1% to 8.5% overall. The results of the present study, when considering the base rates for the entire standardization samples, are relatively consistent with the results presented by Schretlen and colleagues (2008).

Monte Carlo estimation of low scores, based on the examples presented in this study and in Schretlen and colleagues (2008), can be a reasonably accurate method for computing the base rates of low scores on a battery of (up to 20) tests when the intercorrelations are known. In other words, clinicians can have confidence that the estimation of the base rates of low scores using the computer program provided by Crawford and colleagues (2007) and the test intercorrelations will be relatively similar to the actual base rates that would be calculated using a large standardization sample. Clinicians can create shortened versions of these batteries and derive the prevalence of low scores (because the intercorrelations of these shortened batteries can be extracted from the test manuals). Similar to actual base rates, Monte Carlo estimates are restricted to tests which are co-normed (i.e., clinicians cannot create flexible hybrid batteries with known base rates of low scores). However, an advantage of using the Monte Carlo program to compute base rates is that it can be applied to any combination of co-normed tests, whether or not the entire battery administered is irrelevant.

It is critical to appreciate that confidence in the Monte Carlo simulation program should be limited to those studies of people who are demographically similar to the mean of the normative sample. In other words, if certain characteristics are known to moderate test performance (e.g., education, ethnicity, and intelligence), and those factors are not considered in the intercorrelation matrices, then the Monte Carlo-estimated scores will be less accurate in subgroups stratified on those moderator variables. This point was illustrated persuasively in this study. The base rates of low scores vary substantially with level of intelligence (Fig. 3). Having low scores is relatively common in people with lesser intelligence and is relatively uncommon in people with higher intelligence. For example, if a person obtains one index score <10th percentile on the WAIS-III/WMS-III, the Monte Carlo simulation program would estimate that this occurs in 40.0% of healthy adults. However, if the person has a WTAR-demographics predicted FSIQ = 115, then the actual base rates (derived from the standardization sample) would indicate that one or more index scores <10th percentile is found in 12% of healthy people with high average intelligence. Therefore, we strongly caution against using Monte Carlo-estimated base rates for persons who differ substantially from the normative sample’s demographics or with people who have low or high intelligence because these variables are not taken into account for the Monte Carlo calculations. It is hopeful that test publishers will eventually develop scoring software that computes the base rates of low scores for any combination of tests administered and stratifies the results by important moderator variables (e.g., level of intelligence or other demographic variables).

Conflict of Interest

Dr. B. L.B. has no known, perceived, or actual conflict of interest with this research. Dr. G.L.I. has received past research support from Psychological Assessment Resources, Inc. (PAR, Inc.), which is the publisher of the NAB. This study, however, was unfunded.
Acknowledgements

Data for the NAB have been reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, FL, USA from the NAB by Robert A. Stern, PhD, and Travis White, PhD, Copyright 2001, 2003 by PAR, Inc. Further reproduction is prohibited without permission from PAR, Inc. Data presented for the WAIS-III and the WMS-III are copyright © 1997 by NCS Pearson, Inc. Used with permission. All rights reserved.

References


