# A COMPENDIUM OF NEUROPSYCHOLOGICAL TESTS

## Administration, Norms, and Commentary

2006

Third Edition

ESTHER STRAUSS
ELISABETH M. S. SHERMAN
OTFRIED SPREEN

## TYPES OF NORMS p 46

It was not so long ago that neuropsychologists relied exclusively on raw scores and cut-offs for determining level of performance, and that normative samples consisted solely of "normals" and "brain-damaged" individuals. Over time, the field has shifted toward the notion of general population norms comprised of a large sample of individuals characterized by specific demographic characteristics. Use of these norms then allows the user to better determine whether the individual differs from the general population. While IQ and achievement tests have for many years set the standard and perfected techniques for constructing large, representative normative datasets, neuropsychological datasets have only recently begun to include a variety of individuals reflecting the composition of the general population.

At the same time, demographically corrected normative datasets, which represent only a subset of the general population most similar to the patient under scrutiny, are increasingly used and demanded by practitioners. Demographic corrections have come to be routinely applied to most normative data in neuropsychology, first in the form of age-corrected norms, and in some cases, gender-corrected norms, and later, as education-corrected norms. As more and more neuropsychological tests become commercially produced, some norms specifically include a certain proportion of individuals who represent groups defined by race/ethnicity, in order to make norms as representative of the general population as possible. This has also allowed additional levels of within-group norming. This gradual shift away from raw scores as the basis of interpretation to scores adjusted for multiple demographic

factors occurred because of an acknowledgment by clinicians and researchers that demographics are significantly correlated with performance on neuropsychological tests, as they are for most cognitive tests in common usage (see test reviews in this volume for specific references, as well as sources below). The reasons for this shift, and the ways in which general population and demographically corrected normative datasets can be used effectively, will be reviewed in this chapter.

## NORMS: POPULATION-BASED VERSUS DEMOGRAPHICALLY ADJUSTED

p. 47

There are two schools of thought regarding how closely matched the norms must be to the demographic characteristics of the individual being assessed, and these views are diametrically opposed. These are: (1) that norms should be as representative of the general population as possible, and (2) that norms should approximate, as closely as possible, the unique subgroup to which the individual belongs. The latter view is the central tenet of Mitrushina et al.'s text (Mitrushina et al., 2005), which is essentially a guidebook for choosing norms that most closely fit the demographic characteristics of the individual patient being assessed. Although most neuropsychologists assume that the latter is always preferable, this is not necessarily the best choice at all times. Because a test's sensitivity, specificity and impairment cutoffs (see Chapter 1) depend to some extent on the norms selected, choosing norms necessitates a trade-off between the risk of making false negative errors and the risk of making false positive errors. Thus, the use of broadly representative versus demographically specific norms will depend on the purpose of the testing.

At ti[illegible] the individual to

At times, it will be paramount to compare the individual to all other persons of the same age in the general population. Determining a diagnosis of mental retardation or learning disability would be one example. At other times, the goal will be to compare the individual to the best-matched demographic subgroup. Here, the goal might involve mapping out an individual's relative strengths and weaknesses in order to inform diagnostic considerations, plan classroom accommodations or return to work, or to obtain a best estimate of premorbid level in a dementia work-up. In many clinical situations, the assessment will involve both approaches because the results serve to address questions at many levels, including diagnosis, individual strengths and weaknesses, and areas in need of treatment/accommodation. For a more complete discussion of these and related issues, see *Ethnicity, Race and Culture.*

## STRATIFIED GENERAL POPULATION NORMS

The rationale for stratifying norms according to demographic characteristics is to maximize the likelihood that norms are representative of the general population. Major tests are usually stratified based on age, gender, education, ethnicity/race,

socioeconomic status (SES), the latter defined as either actual SES, occupation, or education/parental education. Additional variables include geographic region and urban versus rural residence. Stratification occurs according to representative proportions such as those provided by U.S. Census data to best approximate the composition of the general population.

Although each of these demographic factors tend to be treated as separate stratification variables, it is important to note there is considerable overlap between supposedly separate demographic characteristics. In most cases, some demographic characteristics (e.g., ethnicity, geographic location) are actually surrogate variables for other more fundamental variables that influence test scores (e.g., SES, education).

When nationally representative norms are considered, the sample should have an even distribution of demographic variables across ages, unless certain demographics are more heavily weighted in some age groups versus others. For example, because of factors related to parity and longevity, there is a larger ethnic minority representation among young children in the United States than in adults, and a higher representation of women in the oldest age bands. Because the demographic composition of nations changes over time, this suggests additional caution in using normative data that are outdated, as the norms may no longer be representative of the population being evaluated.

p-47

## Types of Norms for Minority Groups

Norms that take into consideration ethnicity, race, and/or culture typically consist of separate normative datasets that include only specific members such as African Americans or Spanish speakers (e.g., Manly et al., 1998; Ardila et al., 1994; Heaton et al., 2004; Pontón, 2001). There are other methods to adjust scores, including: bonus points (i.e., adding a constant to the score of all members of a subgroup so that the group mean is equivalent to the majority group's), separate cutoffs for each subgroup, and banding (i.e., treating all individuals within a specific score range as having equivalent scores to avoid interpretation of small score differences) (Sacket & Wilk, 1994). With the exception of the use of subgroup-specific cutoffs, these techniques are not commonly used in neuropsychology.

Of historical interest is the fact that neuropsychology's recent shift toward demographically corrected scores based on race/ethnicity and other variables has occurred with surprisingly little fanfare or controversy despite ongoing debate in other domains of psychology. For example, when race-norming was applied to pre-employment screening in the United States to increase the number of minorities chosen as job applicants, the result was the Civil Rights Act of 1991, which outlawed race-norming for applicant selection or referral (see Sackett & Wilk, 1994; Gottfredson, 1994; and Greenlaw & Jensen, 1996, for an interesting historical review of the ill-fated attempt at race-norming the GATB).

Wilk, 1994; Gottfredson, 1994; and Greenlaw & Jensen, 1996, for an interesting historical review of the ill-fated attempt at race-norming the GATB).

## Limitations of Demographically Adjusted Norms

There are some cases where adjustment for demographic influences might be questioned. For example, age and education are risk factors for dementia; removing the effects of these demographic variables might therefore remove some of the predictive power of measures of cognitive impairment (Sliwinski et al., 1997, 2003). O'Connell et al. (2004) have recently reported that use of age- and education-corrected normative data failed to improve the diagnostic accuracy of the 3MS, a version of the MMSE, when screening for dementia or cognitive

impairment. Instead, they recommended that the unadjusted normative data be used for screening purposes. In a similar vein, Reitan and Wolfson (1995, 1996) have argued that age and education corrections are appropriate for normal, healthy individuals, but that they are not needed for brain-damaged subjects. This view has been disputed, however (e.g., Lezak et al., 2004; also see *Test Selection* in this volume). There is also evidence that when "corrective norms" are applied, some demographic influences remain, and overcorrection may occur, resulting in score distortion for some subgroups and a risk of increased false negatives (e.g., Fastenau, 1998).

Some of the limitations of ethnically adjusted norms are summarized by Sattler (2001), including the fact that they may provide (1) a basis for negative comparisons between groups, (2) lower expectations for children from groups that differ culturally and linguistically from the majority, and (3) have little relevance outside of the specific geographic area in which they were collected. Gasquoine (1999) also argues that within-group norming in neuropsychology has limited merit due to the complexities inherent in measuring ethnicity and culture; subgroups can be divided exponentially, limited only by the number of demographic variables available. Figure 2–2 shows the various subgroups that can be derived from subdividing within the U.S. Hispanic population alone.

One alternative to within-group norming is to directly take into account the influence of specific variables on test scores, since these factors presumably operate across ethnicity and culture. These include factors such as English language fluency, acculturation, length of residence in the United States, education, quality of education, and SES, including quality of the home environment, health/nutritional status, income, and degree and persistence of poverty (Gasquoine, 1999). Therefore, norms that correct for ethnicity may be correcting for the "wrong" variable, particularly if other variables appear to better account for observed differences between groups. For instance, differences between minority groups and the majority culture on cognitive tests may remain even after important variables such as number of years of education are accounted

for (e.g., Farias et al., 2004; Shuttleworth-Edwards et al., 2004), but procedures that take into account quality of education by coding for quality directly (e.g., Shuttleworth-Edwards et al., 2004) or using literacy corrections yield few group differences (e.g., Byrd et al., 2004; Manly et al., 1998; 2002).

It is important to note, as well, that variables such as the persistence of poverty and education level/quality within and across minority groups cannot be fully dissociated. For instance, in some Western countries, the prevalence of continuous periods of poverty versus temporary instances of poverty may be much higher in some minority subgroups than in children from the majority culture with similar SES. It is difficult to conceptualize how norms could be made to correct for these multiple factors, but multivariate corrections on large, varied samples, including cross-validation, might be feasible.

Lastly, there are individuals who, despite apparent similarities such as race, differ substantially from the overall characteristics of the target normative subgroup, and who actually may be more closely matched to another group that does not share obvious racial, cultural, or ethnic characteristics. For example, in the United States, the terms "Hispanic" and "African American" are associated with a unique sociopolitical environment related to SES, education, and health status. However, these terms have limited applicability to individuals who may share only superficial characteristics such as race or language (e.g., Shuttleworth-Edwards et al., 2004), but who are recent immigrants to the United States, or who are citizens of other countries. Thus, demographically corrected norms developed in the United States may have limited utility for (1) individuals who differ from the target subgroup in important aspects such as quality of education, or (2) in other countries where ethnic subgroups may have different educational/SES correlates. However, given the state of the field, the provision of normative data for minority groups is a step in the right direction, as is increased awareness on the part of practitioners and researchers that diversity is an inherent characteristic of human populations that needs to be reflected in our neuropsychological tools and practice.

neuropsychological tools and practice.

p.51

## Caveats in the Clinical Use of Demographically Based Norms for Minority Groups

As we have already alluded to, within-group norms significantly affect sensitivity and specificity with regard to tests that employ cutoffs for assigning individuals to specific groups, such as those based on diagnosis (e.g., diagnostic criteria for dementia or language disorder), or for assigning individuals to treatment (e.g., Aricept trial, language intervention). The costs of a false positive may be high in one subgroup (e.g., African American elders undergoing a dementia workup), but false negatives might be of larger consequence in another subgroup (i.e., Hispanic preschoolers in an early screening program). In the former, a false positive is associated with clearly adverse consequences (e.g., false dementia diagnosis). However, in the latter, a false negative might be worse if it means losing access to a service that might have far-reaching benefits

(e.g., a screening program that provides free language intervention, literacy materials, and parent support for at-risk inner-city preschoolers). Ideally, the decision to use adjusted cutoff and within-group norms should be based on a full understanding of the context in which they are destined to be used, including the base rate of the particular disorder within the subgroup, the sensitivity and specificity of the measure, and the costs of false positive and false negative errors.

Importantly, with regard to the WAIS-III/WMS-III, the Psychological Corporation explicitly states that demographically adjusted scores are not intended for use in psychoeducational assessment, determination of intellectual deficiency, vocational assessment, or any other context where the goal is to determine absolute functional level (IQ or memory) in comparison to the general population. Rather, demographically adjusted scores are best used for neurodiagnostic assessment in order to minimize the impact of confounding variables on the diagnosis of cognitive impairment. That is, they should be used to infer strengths and weakness relative to a presumed pre-morbid standard (The Psychological Corporation, 2002). Therefore, neuropsychologists need to balance the risks and benefits of using within-group norms, and use them with a full understanding of their implications and the situations in which they are most appropriate. p.51

Finally, eliminating score differences across demographic groups with demographic corrections or within-group norms will not adjust the current disparities in life circumstances and outcomes that are at times reflected in test scores, nor will it adjust for the relative lack of neuropsychological models that include sociocultural factors (e.g., Campbell et al., 2002; Pérez-Arce, 1999). Above all else, as noted by Ardila et al. (1994), "the clinical neuropsychologist must entertain the notion that human diversity does not translate into human deficiency" (p. 5).

## PRACTICAL CONSIDERATIONS IN THE USE OF DEMOGRAPHICALLY CORRECTED NORMS

p. 57

By necessity, neuropsychologists often adopt a compromise between using population-wide norms and within-group norms. Almost all tests provide age-based scores, but the availability of norms based on education and minority status (ethnicity/race/culture) varies greatly across tests. As a result, practicing clinicians are therefore only able to demographically correct some scores but not others, unless a uniform battery such as the WAIS-III/WMS-III or NAB is employed. The problem is that if only some scores are adjusted for moderator variables (e.g., education), and then compared with non-corrected scores, false attributions of impairment may occur (Kalechstein et al., 1998).

As a practical solution, given equivalent sampling quality, the a priori selection of the normative set should be primarily guided by the particular moderating variable that is most likely to affect the classification of test performance (Kalechstein et al., 1998). For example, on tests of psychomotor

speed, it would be preferable to match on the basis of age rather than education, if a choice was required. In contrast, on tests of verbal achievement, education would likely be the primary selection criterion.

In each test review in this volume, we have included information on demographic influences on test performance. With this information, users can determine for themselves whether demographic corrections are necessary, and whether test scores need to be interpreted within a demographically relevant context.