The Potential Clinical Utility of Methods for Estimating Prior Standing in Specific Cognitive Domains: A Feasibility and Illustration Study

Stephen Correia
University of Rhode Island

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation
Correia, Stephen, "The Potential Clinical Utility of Methods for Estimating Prior Standing in Specific Cognitive Domains: A Feasibility and Illustration Study" (2001). Open Access Dissertations. Paper 974. https://digitalcommons.uri.edu/oa_diss/974
THE POTENTIAL CLINICAL UTILITY OF METHODS FOR ESTIMATING PRIOR
STANDING IN SPECIFIC COGNITIVE DOMAINS: A FEASIBILITY AND
ILLUSTRATION STUDY

BY

STEPHEN CORREIA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
CLINICAL PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2001
ABSTRACT

This study evaluated the viability of a potential strategy for improving the detection of cognitive decline over currently available IQ-based methods. The proposed strategy makes use of differential cognitive effects across different neurocognitive disorders. It involves examining estimated/obtained difference scores (D-scores) for the specific cognitive domain(s) (SCDs) most affected by a disorder. The current study undertook a broad feasibility test of the strategy as a preliminary step in the development of specific cognitive domain estimation methods (SCDEMs). Clinical and control group score distributions were reconstructed from IQ and SCD test (SCDT) means and standard deviations reported in previously published studies of mild Alzheimer’s disease (mild AD), chronic alcohol abuse (CAA), and mild traumatic brain injury (mTBI). For each test, the percentage of area shared by the two reconstructed clinical and control distribution curves was calculated. Percent overlap values for tests measuring the same SCD were then pooled across studies of the same disorder and averaged, thereby forming indexes that served to estimate SCD sensitivity. Comparable IQ indexes were also formed. The average SCD and IQ overlap values were then compared. The main result suggests that diagnostic accuracy could be improved considerably for mild AD, and, to a lesser extent, for CAA and mTBI, by using SCD versus IQ D-scores. The development of SCDEMs appears clinically worthwhile, especially given their potential application to a disorder of such importance as AD, although their utility may be lower or considerably lower, for other disorders.
ACKNOWLEDGEMENT

This project is dedicated to my wife, Hyunjung. Without her loving and constant support, this would not have been possible. This project would also not have been possible without the guidance and encouragement of David Faust, Ph.D.
# Table of Contents

List of Abbreviations vi

Introduction 1

Informal Methods of Estimating Prior IQ 3

Formal Methods of Estimating Prior IQ 5

Demographically Based Actuarial Methods for Estimating Prior IQ 6

Current Ability Methods for Estimating Prior IQ 7

  Vocabulary and Hold-Don’t-Hold Strategy 7

  Lezak’s Best Performance Method 9

Methods Combining Demographic and Current Ability Measures 10

Reading-Based Methods of Estimating Prior IQ: The NART 12

  Revisions of the NART 13

  Cognitive Decline and NART Performance 13

  Use of the NART in disorders other than dementia 14

  Retrospective accuracy of the word-reading methods 14

  Modifications and extensions of the word-reading method 15

  Combined demographic and word-reading methods 16

Other Approaches to the Estimation of Prior IQ 17

Methods for Predicting Prior Intelligence Level in Children 18

Summary 18

A New Approach for Improving Methods of Estimating Prior Functioning 20

  Limitations of IQ-Based Methods 20

  Preliminary Specific Cognitive Domain Estimation Methods 23

  The New Approach 25

Purpose of the Study 26

Method 29

  Study Selection 30

  Selection of IQ and SCDT Distributions for Reconstruction 32

  Correlations Among SCDTs and IQ 33

  Distribution Reconstruction and Analysis Procedures 37
List of Abbreviations

| Abbreviation | Description |
|--------------|-------------|
| AC           | Attention/Concentration Index |
| AD           | Alzheimer’s Disease |
| AI           | Attention Index |
| AMNART       | American version of the Nelson Adult Reading Test |
| AVLT         | Auditory Verbal Learning Test |
| BD           | Block Design |
| BFRT         | Benton Facial Recognition Test |
| BNT          | Boston Naming Test |
| BPM          | Best Performance Method |
| BVRT         | Benton Visual Retention Test |
| CAA          | Chronic Alcohol Abuse |
| CCRT         | Cambridge Contextual Reading Test |
| CT           | Category Test, also Computerized Tomography |
| DH           | Dominant Hand |
| DMI          | Delayed Memory Index |
| DRS          | Dementia Rating Scale |
| DS           | Digit Span subtest |
| DSy          | Digit Symbol subtest |
| FCSRT        | Free and Cued Selective Reminding Task |
| GMI          | General Memory Index |
| HII          | Halstead Impairment Index (also modified-HII or mHII) |
| IMI          | Immediate Memory Index |
| IQ           | Intelligence Quotient |
| LM-I         | Logical Memory subtest (immediate recall) |
| LM-II        | Logical Memory subtest (delayed recall) |
| MCI          | Mild Cognitive Impairment |
| MMSE         | Mini-Mental State Examination |
| MQ           | Memory Quotient |
| NART         | National Adult Reading Test |
| Abbreviation | Description |
|--------------|-------------|
| NAART        | North American National Adult Reading Test |
| NDH          | Non-Dominant Hand |
| OA           | Object Assembly subtest |
| PA           | Paired Associates subtest |
| Pegs         | Grooved Pegboard Test |
| PID          | Progressive Idiopathic Dementia |
| PSI          | Processing Speed Index |
| ROCFT        | Rey-Osterreith Complex Figure Test |
| SCD          | Specific Cognitive Domain |
| SCDT         | Specific Cognitive Domain Test |
| SCDEM        | Specific Cognitive Domain Estimation Method |
| SD           | Standard Deviation |
| SEE          | Standard Error(s) of Estimate |
| SEM          | Standard Error(s) of Measurement |
| SILS         | Shipley Institute for Living Scale |
| SRT          | Seashore Rhythm Test |
| SSPT         | Speech-Sounds Perception Test |
| Stroop int.  | Stroop Interference score |
| STW          | Spot The Word test |
| TBI/mTBI     | Traumatic Brain Injury/mild Traumatic Brain Injury |
| TMT-A        | Part A of the Trail Making Test |
| TMT-B        | Part B of the Trail Making Test |
| TPT          | Tactual Performance Test |
| VF           | Verbal Fluency |
| VR-I         | Visual Reproduction subtest (immediate recall) |
| VR-II        | Visual Reproduction subtest (delayed recall) |
| VsDI         | Visual Delay Index |
| VsII         | Visual Immediate Index |
| WAIS         | Wechsler Adult Intelligence Scale |
| WCST         | Wisconsin Card Sorting Test |
| WIS          | Wechsler Intelligence Scales |
| Acronym | Description                          |
|---------|--------------------------------------|
| WISC    | Wechsler Intelligence Scale for Children |
| WMI     | Working Memory Index                  |
| WMS     | Wechsler Memory Scale                 |
INTRODUCTION

Most neuropsychological evaluations are conducted to identify the presence and extent of cognitive decline in the context of an established or suspected disorder, such as traumatic brain damage or dementia. The results of these evaluations can have important diagnostic, prognostic, legal, and treatment implications, and thus minimizing errors in detecting cognitive decline is an important practice goal. Norms provide a means of ranking an individual's test score relative to a representative criterion group. Such rankings help to determine the presence and relative extent of impairment. However, normative comparison alone may be insufficient for evaluating cognitive decline in any given patient. This is because individuals have different prior baselines across cognitive domains. Thus, the assumption underlying to the use of norms to evaluate change (i.e., that an individual originally ranked somewhere around the middle of the distribution) may not hold. For example, an individual may have fallen within either extreme of the normal distribution prior to the onset of decline. Accordingly, the same score might indicate significant decline for one individual and not for another. Therefore, the evaluation of cognitive decline can be aided by comparison to an individual's own prior baseline.

Often, the best baseline data are prior cognitive test results. Unfortunately, such data are frequently either unavailable, or cannot be obtained in a timely manner. Even when they are available, clinicians must carefully consider whether the data provide a representative baseline for comparison. The representativeness of prior test data depends on a number of factors, such as the reliability and temporal stability of the previously administered tests. Also, a person's pattern of strengths and weaknesses may change with
age and education (Kaufman, 1990; Reynolds, 1997; Woodruff-Pak, 1997). Additionally, events intervening between the time of the previous evaluation and the onset of an event or condition prompting a referral may limit the value of prior test results. Consider, for example, a 40-year-old patient with 12 years of education who is referred for assessment of cognitive functioning following recent mild to moderate traumatic brain injury (TBI). High school IQ scores might provide a potentially helpful baseline against which to compare the patient’s current performance. However, if the patient had been a heavy drinker since the late teens, the high school IQ scores might no longer be representative of the more recent pre-TBI baseline. That is, they might not aid in disentangling the relative contribution of alcohol abuse from TBI in accounting for cognitive decline, especially if current performance falls below that of the high school IQ scores.

Given the potential limitations of accessing or using prior cognitive test results, clinicians frequently estimate a patient’s cognitive baseline. Typically, this is done using indirect, inferential methods -- both formal and informal -- to estimate a patient’s prior IQ score. IQ scores are popular for this purpose because they provide an index of global cognitive ability, and correlate more or less strongly with virtually every cognitive domain (Crawford, 1992; Kaufman, 1990; Schinka & Vanderploeg, 2000). Thus, one score can assist in setting expectations for performance across many cognitive domains. Accordingly, clinicians typically use the IQ estimate as a benchmark for forming expectations about a patient’s performance in other cognitive domains, and take deviations from these expectations as potential indicators of disorder or loss. Thus, the diagnostic process often involves two inferential links: the estimation of prior IQ, and,
then, the use of the estimated prior IQ to formulate expectancies for performance in specific cognitive domains.

Estimated (or actual) baseline IQ (i.e., before the onset of a condition affecting cognitive functioning) is typically referred to in the literature as “premorbid” IQ. According to Graves et al. (1999), a better term might be “no-morbid” IQ. In most cases, I will refer to it as “prior” IQ. This avoids the implicit suggestion raised by the term “premorbid” that disorder is now present. It also avoids the suggestion raised by the term “no-morbid” that the estimate reflects a person’s pristine intellectual ability in the absence of any previous cognitive disorder or risk factors.

Informal Methods of Estimating Prior IQ

Because prior IQ estimation is such a common practice, a review of the literature on this topic is warranted, especially because alternative procedures will proposed. In formulating prior IQ estimates, practitioners often rely on clinical judgment (Smith-Seemiller, Franzen, Burgess, & Prieto, 1997). However, voluminous literature suggests that cognitive data integration of this sort is limited (e.g. Arkes, 1981; Dawes, Faust, & Meehl, 1989; Faust et al., 1988; Kareken & Williams, 1994; Meehl, 1954). First, it may be difficult to subjectively evaluate the validity of the data used in forming the estimates. This may be particularly true of data bearing an uncertain relationship with IQ (e.g., occupational history, military history, or hobbies). Second, information bearing on the strength of association between extra-test data and IQ often is not readily available or is not broadly recognized.¹ In the absence of this knowledge, clinicians might rely too

¹The WAIS-III Technical Manual (Psychological Corporation, 1997) provides some improvement in this regard as it contains correlations between WAIS-III IQ scores and several frequently used neuropsychological tests.
heavily on information that is weakly correlated with IQ, and too little on more valuable information. Third, even when clinicians are aware of the correlations among variables, they tend not to adjust their estimates accordingly or optimally (Kareken & Williams, 1994), something that can be very difficult to do subjectively. For example, optimal combination might involve weighting Variable X 2.7 times more than Variable Y and 1.63 times less than Variable Z. Fourth, clinicians tend to be overconfident in their judgments (Kareken & Williams, 1994), and the more data they have, the more overconfident they tend to become. However, judgment accuracy does not necessarily depend on the quantity of data per se, but rather on the degree to which the data are valid and non-redundant (Dawes et al., 1989). Overconfidence tends to lead to overly extreme predictions, rather than properly regressed prediction.

Actuarial judgment involves the use of predetermined or a prespecified means of data combination, and is based on empirically established relations (Meehl, 1954). Well over 100 studies (see Dawes et al., 1989) show that actuarial methods almost always equal, and frequently exceed, the accuracy of clinical judgment, thereby making it a better overall method. Accordingly, actuarial methods for estimating prior IQ ought to be at least as accurate, and quite possibly more accurate than estimates based on clinical judgment alone. Certainly, on a patient-by-patient basis, there will be instances in which clinical judgment is more accurate than actuarial judgment. For example, a clinician might reject a prior IQ estimate derived from a demographic regression formula when the patient’s history clearly suggests high intellectual ability in the presence of low standing on the variables that enter the formulae (Schinka & Vanderploeg, 2000). However, determining when to countervail an actuarial conclusion appears to be much more
difficult than it might seem, and in many instances such countervailing overturns an
otherwise correct decision (Faust et al., 1988). Studies addressing this issue show that
clinicians tend to call exceptions too frequently and that, for each incorrect actuarial
conclusion that is corrected, there will be one, or more than one, correct actuarial
conclusion spoiled, leading to no overall net gain or, worse, a loss in overall judgmental
accuracy (Dawes et al., 1989). Thus, there are strong grounds to believe that formal or
actuarial methods will not decrease, and may well increase, the accuracy of prior ability
estimates. As will be discussed later, the methods described here incorporate actuarial or
statistical judgment methods.

Formal Methods of Estimating Prior IQ

There are four general formal, or quasi-formal, approaches for estimating prior IQ
(Spreen & Strauss, 1998). The first employs actuarial regression formulae based on a
patient's standing on background demographic variables. The second is based on current
performance on WIS subtests (e.g., Vocabulary) thought to be resistant to brain damage.
The third takes the highest current WIS subtest score or highest level of achievement in
everyday tasks as a marker of prior IQ and as the standard against which performance on
other tests is compared. The fourth is also an actuarial regression model but instead uses
performance on word-reading tests. Other methods include combinations or hybrids of
the four methods just mentioned. All of these methods attempt to capitalize on markers of
cognitive ability that are, or are thought to be, highly correlated with prior IQ, yet
minimally or relatively unaffected by brain impairment.
Demographically Based Actuarial Methods for Estimating Prior IQ

Using data from the 1955 WAIS standardization sample, Wilson, Rosenbaum, Brown, Rourke, and Whitman (1978) developed regression formulae for estimating patients’ Wechsler Adult Intelligence Scale (WAIS) VIQ, PIQ, and FSIQ scores based on the age, race, sex, education, and occupation. Among these variables, education was the strongest predictor of IQ, accounting for about 44%, 31%, and 43% of the variance in VIQ, PIQ, and FSIQ scores, respectively. The remaining variables accounted for about another 10% of the variance. The standard errors of estimate (SEE) were 10.2, 11.4, and 10.2 for the VIQ, PIQ, and FSIQ formulae, respectively.

Following Wilson et al.’s (1978) lead, Barona, Reynolds, and Chastain (1984) developed demographic regression formulae for estimating WAIS-R IQ scores. The formulae were based on the data contained in the WAIS-R standardization sample. Barona et al. added the variables of geographical region of residence, and urban versus rural residence to those that Wilson et al. (1978) used. The squared multiple correlations between all variables in Barona formulae were .38, .24, and .36 for VIQ, PIQ, and FSIQ, respectively, slightly lower than the Wilson formulae. The formulae SEE were 11.79, 13.23, and 12.14 for VIQ, PIQ, and FSIQ, respectively, slightly larger than the Wilson formulae. Due to the effects of regression toward the mean, Barona et al. warned that serious over- or under-estimation could occur for individuals with IQs below 69 or above 120, respectively. A subsequent revision to these formulae (Barona & Chastain, 1986), based on a subset of the WAIS-R sample, had similar SEE and generally failed to improve predictive accuracy over the original 1984 formulae (Paolo & Ryan, 1992).
Studies on the Barona and Wilson formulae have generally shown that they perform only slightly better than chance in estimating patients' prior IQ (e.g., Bolter, Gouvier, Veneklasen, & Long, 1982; Goldstein, Gary, & Levin, 1986; Hawkins, 1995; Schinka, 2000 #241; Karzmark, Heaton, Grant, & Matthews, 1985; Klesges, Fisher, Vasey, & Pheley, 1985; Silverstein, 1987; Sweet, Moberg, & Tovain, 1990). That is, they do not provide much improvement over assuming that a patient's prior IQ was in the Average range. This assumption will be correct about 50% of the time (Hawkins, 1995; Schinka & Vanderploeg, 2000) assuming a normal distribution of scores. That is, the Average range for the Wechsler scales falls between scores of 90 and 110; by definition this score range corresponds to the 25th to 75th percentile. This limited predictive power is reflected by the relatively large SEE of these formulae -- about 10 points for the Wilson formulae (Karzmark et al., 1985; Wilson et al., 1978) and about 12 points for the Barona formulae (Barona et al., 1984; Eppinger, Craig, Adams, & Parsons, 1987). These values, especially the latter, approach the 15-point standard deviation of the Wechsler scales, and therefore produce a distribution of predicted scores that is nearly the same as that of scores on the test.

**Current Ability Methods for Estimating Prior IQ**

**Vocabulary and the Hold–Don’t-Hold Strategy**

Results from early studies suggested that performance on the WIS Comprehension, Information, Picture Completion, and Object Assembly subtests was relatively preserved (i.e., they “hold”) in various pathological brain conditions. In contrast, performance on the Digit Span, Digit Symbol, Arithmetic, and Block Design subtests tended to deteriorate (i.e., “don’t hold”) (Kaufman, 1990; Lezak, 1995; Vogt &
Based on this information, Wechsler (1944) developed a Deterioration Index to aid in the detection of cognitive loss. The index was a ratio based on the “hold” and “don’t hold” subtests of the Wechsler-Bellevue Intelligence Test. Wechsler (1958) revised the index for the WAIS and replaced Comprehension with Vocabulary as a “hold” subtest, and Arithmetic with Similarities as a “don’t hold” subtest. McFie (1975) subsequently proposed using only the Vocabulary and/or Picture Completion subtests as markers of prior IQ.

The Hold–Don’t-Hold strategy has been generally discredited for a number of reasons (Kaufman, 1990; Larrabee, Haley, Largen, & Levin, 1985; Lynch & McCaffery, 1997; Nelson & McKenna, 1975; Spreen & Strauss, 1998). For one, performance on the Vocabulary and other “hold” subtests may well be attenuated in a number of disorders or conditions, including verbal learning disability, chronic epilepsy, dementia, traumatic brain injury, left hemisphere stroke, inadequate educational opportunity, or impoverished environments (Axelrod, Vanderploeg, & Rawlings, 1999; Christensen & MacKinnon, 1992; Kaufman, 1990; Larrabee et al., 1985; Putnam, Ricker, Ross, & Kurtz, 1999; Russell, 1972; Vogt & Heaton, 1977; Yates, 1956). Accordingly, the Hold–Don’t-Hold strategy is likely to underestimate prior IQ in these, and possibly other conditions. Additionally, the method also increases the likelihood of over- or underestimation of prior IQ by ignoring chance fluctuations in subtest performance, by failing to account for the effects of regression toward the mean (Reynolds, 1997), and by overlooking age-related changes in the pattern of WAIS subtest performance (Kaufman, 1990).
Lezak’s Best Performance Method

Lezak’s (1995) best performance method (BPM) takes the patient’s highest level of performance as the “standard against which all other aspects of the patient’s current performance are compared” (p. 106). According to Lezak, a patient’s highest level of prior functioning might be reflected in such variables as their highest current WIS subtest score, their highest occupation level, or other special accomplishments or proficiencies. The BPM (Lezak, 1995) assumes that among neurologically normal individuals, test scores tend to cluster around a mean level of performance and that for cognitively impaired individuals, their highest obtained test score provides the best marker of their previous level of performance.

This latter assumption ignores error variance and normal variability in test scores. Normal individuals often show wide discrepancies between their best performance and their more typical or average performances. As a result BPM will virtually always overestimate prior IQ. This can be demonstrated by considering a normal individual with a FSIQ of 100 and mean subtest score of 10. Given the average 7-point range of subtest scatter (Matarazzo, Daniel, Prifitera, & Herman, 1988), this individual’s score on any one subtest is likely to be 10 ± 3.5. Because an obtained score reflects both “true” score and measurement error, any individual’s highest subtest score will contain a positive error component. Thus, the person’s highest subtest score is likely to be 13.5, which is 1.16 SD above his or her mean subtest score of 10. According to the BPM, this individual’s prior FSIQ would correspondingly be estimated to be 1.16 SD above the mean, or about 117 (High Average range) -- nearly 40 percentile points above the person’s actual FSIQ of 100 (Average range).
Mortensen, Gade, and Reinisch (1991) illustrated this problem when they used the BPM to estimate IQ scores of subjects in the Dutch WAIS-R standardization sample. The samples' mean estimated FSIQ using the BPM was 117, whereas their actual mean FSIQ was 100. This result demonstrated clearly that the BPM overestimates prior IQ, often by a wide margin, and creates a considerable risk for false-positive errors.

**Methods that Combine Demographic Variables and Current Ability Measures**

Two recently developed methods estimate prior IQ using demographic variables in combination with WAIS-R subtest scores. The Oklahoma Premorbid Intelligence Estimate (OPIE) (Krull, Scott, & Sherer, 1995) consists of three formulae derived from the WAIS-R standardization sample. Each formula combines age, education, occupation, and race, with a patient’s raw score on either the Vocabulary or Picture Completion subtest, or both subtests. In a subsequent revision (OPIE-R), these authors recommended using the Vocabulary formula when the patient’s Vocabulary raw score is 4 or more points higher than their Picture Completion raw score, and using the Picture Completion formula when the reverse is true. If Vocabulary and Picture Completion do not differ by at least 4 points, the formula incorporates both subtests (Williamson, Krull, & Scott, 1996).

Vanderploeg and Schinka (1994; 1995) developed 33 regression equations for predicting prior IQ -- one for each of the 11 WAIS-R subtests -- based on data from the WAIS-R standardization sample. Each equation combined age, gender, race, education, and occupation with a current subtest score. Vanderploeg, Schinka, and Axelrod (1996) identified the three most robust equations for estimating VIQ, PIQ, and FSIQ (3 equations each, 9 total). These equations used the Vocabulary, Information, or Picture
Completion subtests. For each IQ index, the highest estimate produced by the three
equations is taken as the patient’s estimated prior IQ.

Relative to the Barona method, the OPIE and BEST-3 methods have smaller
SEEs (6.29 to 11.86, across methods). Consequently, they have higher squared multiple
correlations (.41 to .78, across methods), and higher correlations between estimated and
obtained IQ (.74 to .87, across methods) (Ritchie, Lam, & Rankin, 1996; Vanderploeg et
al., 1996). Regarding accuracy, the BEST-3 formulae produced overestimates of about 5
points (Vanderploeg et al., 1994; 1995). The OPIE formulae correctly classified 63% of a
cross-validation sample into the same qualitative IQ category as their obtained FSIQ
scores. None of the OPIE IQ estimates were off by more than two categories (Krull et al.,
1995). One study, however, found no significant difference in the accuracy of the three
methods (Axelrod, Vanderploeg, & Schinka, 1999). Axelrod et al. attributed the
inconsistent research outcomes to unidentified sample differences across the studies.

One major methodological limitation to these studies is that they compare
formula-estimated IQ to the actual IQ obtained concurrently, rather than previously.
Further, the prediction formulae use subtests that are incorporated in the obtained IQ
score, that is, the independent and dependent variables, in part, share identical variance
(i.e., certain subtest scores). Consequently, it is almost inevitable that the OPIE and
BEST-3 formulae outperform the Barona method. However, a variety of conditions could
affect standing on the subtests used by these methods. Thus, examining the utility of
current scores to predict other current scores may not accurately reflect the capacity of
these formulae to postdict prior status. Clearly, retrospective designs are needed. Further
heightening these concerns, the OPIE and BEST-3 methods incorporate potentially
problematic features of the BPM: The BEST-3 takes the highest of three VIQ, PIQ, and FSIQ estimates, whereas the OPIE selects the higher of two subtest scores.

**Reading-Based Methods of Estimating Prior IQ: The National Adult Reading Test**

Several researchers have developed methods for estimating prior IQ that use current reading ability. These methods were initially based on observations that reading ability appeared to be relatively preserved in dementia (Nelson & McKenna, 1975). These methods attempt to capitalize on a “hold” skill, and thus are variants of the Hold–Don’t-Hold method.

Nelson and O'Connell (1978) developed the New Adult Reading Test (NART) (subsequently renamed the National Adult Reading Test). The NART requires patients to read aloud a graded list of 50 phonemically irregular words. Performance is scored for pronunciation errors; thus, lower scores indicate better performance. Nelson and O'Connell (1978) theorized that the utility of single word-reading for estimating prior IQ depends less on application of grapheme-phoneme conversion rules (reading skills) and more on vocabulary (i.e., previous familiarity with the words); hence the inclusion of only phonemically irregular words on the NART.

Regression equations for predicting WAIS IQ from NART error score yielded SEE's of 7.6, 9.4, and 7.6 for VIQ, PIQ, and FSIQ, respectively (Nelson & O'Connell, 1978); results that were essentially replicated in a cross-validation study (Crawford, Parker, Stewart, Besson, & De Lacey, 1989). Education level, social class and, to a lesser extent, age, were significantly correlated with NART performance and together accounted for 49% of the variance in NART score (Crawford, Allan, Cochrane, & Parker, 1990; Crawford, Stewart, Garthwaite, Parker, & Besson, 1988).
Revisions of the NART. Several revisions and modifications of the NART have been undertaken. Crawford (1992) created a revision known as the NART-R UK, or NART-2, for estimating WAIS-R IQ scores. Ryan and Paolo (1992) adapted the NART for use with older Americans. Blair and Spreen (1989) developed a version known as the New Adult Reading Test – Revised (NART-R), or the North American Adult Reading Test (NAART) for use with North American populations. Schwartz and Saffran developed an American version called the AMNART, sometimes called the ANART (unpublished report cited in Grober & Sliwinski, 1991). Grober, and Sliwinski (1991) subsequently developed a regression model based on AMNART errors and years of education to predict WAIS-R VIQ. These various revisions of the NART accounted for 59% to 69% of the variance in predicted IQ. The SEE of the various regression formulae ranged from 7 points (for VIQ) to 12 points (for PIQ).

Cognitive decline and NART performance. The NART and its revisions were developed for differentiating patients with dementia from normals. Therefore, its clinical utility hinges on the resilience of single word reading to dementia. Numerous studies have addressed this issue. Among non-demented elderly, NART performance appears to be stable over periods of up to 6 years (Schmand, Geerlings, Cees, & Lindeboom, 1998). Patients with mild, and even very mild dementia successfully read approximately 3 to 5 fewer NART items than normal controls (Paolo, Tröster, Ryan, & Koller, 1997; Storandt, Stone, & LaBarge, 1995). The NART performance of patients with mild dementia generally declines slowly (i.e., less than 3 raw score points) over periods of 1 to 3 years (Berry et al., 1994; Fromm, Holland, Nebes, & Oakley, 1991; O'Carroll, Baikie, & Whittick, 1987; Paque & Warrington, 1995; Schmand et al., 1998; Stebbins, Wilson,
Gilley, Bernard, & Fox, 1990). However, NART performance declines rapidly among patients with moderate-to-severe dementia (Brayne & Beardsall, 1990; Grober & Sliwinski, 1991; Maddrey, Cullum, Weiner, & Filley, 1996; Patterson, Graham, & Hodges, 1994; Schmand et al., 1998; Stebbins, Wilson, Gilley, & Fox, 1988; Taylor et al., 1996). This decline in NART performance among patients with moderate to severe dementia is not necessarily a limitation, because such patients are likely to be easily identified by other means, thereby obviating the need for the NART.

**Use of the NART in disorders other than dementia.** Although the NART was developed to identify dementia, this measure and its revisions are often used to estimate prior functioning in other disorders (O'Carroll, 1995). Studies have provided mixed support for this practice. For example, NART performance appears to be relatively stable among the patients with closed head injury (Crawford, Parker, & Besson, 1988; Watt & O'Carroll, 1999), early idiopathic Parkinson’s disease (Lees & Smith, 1983), and depression (O'Carroll, 1995). In contrast, NART performance was attenuated among patients with Huntington’s disease and Korsakoff’s syndrome (Crawford, Parker et al., 1988). Moreover, NART performance may well be reduced in the presence of dominant hemisphere damage.

**Retrospective accuracy of the word-reading methods.** The results of the few studies examining the retrospective (rather than concurrent) accuracy of NART-estimated IQ show that they tend to underestimate IQ scores obtained 3 to 5 years earlier by about 2 to 4 points. Further, the D-score standard deviations in these studies are rather large, ranging from about 8 to 10 points (Berry et al., 1994; Carswell, 1992), with the IQ of
about one in five individuals being underestimated by at least 10 points (Carswell, Graves, Snow, & Tierney, 1997).

**Modifications and extensions of the word-reading method.** There have been a number of extensions and modifications of the NART. Beardsall and Brayne (1990) developed an equation for predicting full NART scores from performance on its first 25 words (Short NART). This test showed a high correlation with the full NART (.83 to .93) (Beardsall & Brayne, 1990; Crawford, Parker, Allan, Jack, & Morrison, 1991), and produced VIQ estimates that were only minimally less accurate than those based on the full NART (Crawford et al., 1991). Although these results suggested that the Short-NART could be used with reasonable confidence, Crawford et al. (1991) argued that it requires administrative and scoring adjustments, and thereby introduces needless potential clerical error into the calculation of NART-estimated IQs. Moreover, the full NART does not take long to administer and is generally well tolerated by patients.

The Spot-the-Word Test (STW) (Baddely, Emslie, & Nimmo-Smith, 1993) requires patients to identify the real word in each of a series of word/non-word pairs presented aurally or visually. The test was intended, in part, to reduce the frequency and extent of IQ underestimates among patients who may be familiar with a NART item aurally but not orthographically, or vice-versa. Research on the STW, however, has yielded generally disappointing results (Beardsall & Huppert, 1997; Law & O'Carroll, 1998; Watt & O'Carroll, 1999).

Beardsall and Huppert developed The Cambridge Contextual Reading Test (CCRT) (Beardsall, 1998; Beardsall & Huppert, 1997; Beardsall & Huppert, 1994). The CCRT places the NART-2 within meaningful sentences, which patients read aloud. The
intention was to reduce the likelihood of underestimating prior IQ among patients who mispronounce NART-2 words they likely know. The final regression formula combined CCRT score with sex and years of education. It accounted for 68% of the variance in VIQ estimation and had a SEE of 7.80 -- values similar to those reported for other word-reading methods. Studies on the CCRT show that placing NART-2 items in sentences significantly reduces overall pronunciation errors, and that this effect is strongest for patients with mild to moderate AD or who have poor word reading ability (Beardsall, 1998; Beardsall & Huppert, 1994; Conway & O'Carroll, 1997; Law & O'Carroll, 1998). Although this result suggests that the CCRT might be superior to the NART for such patients, firm conclusions await retrospective confirmation.

The Reading subtest of the Wide Range Achievement Test–Revised (WRAT-R) is commonly used to screen reading ability and shows moderate to strong correlations with IQ (Cooper & Fabroni, 1988; Kareken, Gur, & Saykin, 1995; Spruill & Beck, 1986). Kareken, Gur, and Saykin (1995) developed regression formulae for estimating WAIS-R IQ scores that combined WRAT-R reading score with demographic variables. The squared multiple correlations were .67, .62, and .72 for VIQ, PIQ, and FSIQ, respectively, with respective SEEs of 10.42, 11.82, and 10.24, slightly larger than the SEEs of other word reading measures.

Combined demographic and word-reading methods. For the NART, CCRT, WRAT-R, AMNART, and STW (but not the NAART), the addition of demographic variables to regression equations yielded slightly smaller SEEs (Beardsall, 1998; Blair & Spreen, 1989; Crawford, Nelson, Blackmore, Cochrane, & Allan, 1990; Grober & Sliwinski, 1991; Raguet, Campbell, Berry, Schmitt, & Smith, 1996; Watt & O'Carroll,
This suggests that demographic and word-reading methods are both valid, yet somewhat independent (i.e., non-redundant) estimators of IQ (Crawford, Allan et al., 1990; Raguet et al., 1996).

**Other Approaches to Estimating Prior IQ**

Schlottman and Johnsen (1991) developed the Intellectual Correlates Scale (ICS) for estimating prior IQ in brain-damaged individuals on the basis of changes in interests and attitudes following brain insult. The ICS, however, has demonstrated poor reliability over time (Raguet et al., 1996).

Wrobel and Wrobel (1996) attempted to improve upon the accuracy of the Barona IQ estimates among psychiatric patients by including results from the Minnesota Multiphasic Personality Inventory. Unfortunately, the SEEs in Wrobel and Wrobel’s formulae were as large as 14 points, thereby seriously limiting their utility.

Reynolds (1997) proposed an IQ estimation method that makes use of IQ correlations among close family members. Although for monozygotic twins, the method would capture about 64% of the predictive variance, it would capture only 25% of the variance for parent-child pairs, and less for more distant relations.

Zachary, Crumpton, and Speigel (1985) developed two regression formulae for predicting prior FSIQ using the Vocabulary subtest of The Shipley Institute for Living Scale (SILS) (Shipley, 1930). SILS-estimated/obtained WAIS-R FSIQ correlations were high (i.e., .85 and .87) in a sample of male psychiatric patients. The formulae SEEs were 6.21 and 6.26 (Zachary, Crumpton, & Spiegel, 1985). Even though the multiple-choice format of the SILS may make it a somewhat easier test than the WIS Vocabulary test, which requires generation of a response (Yuspeh, Vanderploeg, &
Kershaw, 1998), it remains susceptible to the same limitations of other Hold–Don’t-Hold methods.

**Methods for Predicting Prior Intellectual Level in Children**

Attempts at developing methods for estimating prior IQ in children (e.g., Reynolds & Gutkin, 1979; Sellers, Burns, & Guyrke, 1996; Vanderploeg, Schinka, Baum, Tremont, & Mittenberg, 1998) have generally yielded disappointing results. This is not unexpected given that children appear to have greater variability in IQ over time than adults and that IQ in children can be affected by numerous factors (Franzen, Burgess, & Smith-Seemiller, 1997; Franzen, Robbins, & Sawicki, 1989; Sattler, 1992).

**Summary**

The Average IQ range of 90 to 110 encompasses 50% of the normal distribution. Accordingly, a clinician should be correct about 50% of the time, by always assuming that a patient’s prior IQ fell within this range. The challenge is to develop methods that permit greater accuracy (Schinka & Vanderploeg, 2000). A number of methods have been developed to accomplish this aim, but none have been highly successful and some have serious flaws. The BPM systematically overestimates IQ, often by a gross margin. The Hold–Don’t-Hold strategy has a number of weaknesses that have lead to it’s being discredited. Demographic regression approaches, such as the Barona method, do not appear to perform much better than chance. Methods such as the OPIE and BEST-3, which combine demographic variables with current WAIS-R subtest performance, are potentially superior to the Barona method, but it remains unclear if this is a statistical artifact stemming from contamination of independent and dependent variables, or if it reflects a true gain in accuracy. In any case, current studies of these methods have serious
limitations that almost completely undermine their interpretability. Methods that combine
demographic variables with performance on single-word reading tests, such as the NART
and AMNART, appear to perform better than methods relying on demographic variables
or single-word reading performance alone. However, even the best of these methods have
error ranges of ±7 points. Across the demographic and word-reading methods, PIQ is
consistently estimated with much less accuracy than VIQ or FSIQ. Consistent with a very
large body of independent research on clinical versus actuarial judgment, clinical
judgment methods for estimating prior IQ apparently do no better, and may do worse,
than the best of the actuarial methods,

The main impetus for developing prior IQ estimation methods is to improve the
detection of cognitive decline. It is quite possible that present approaches, as constituted,
have been pushed as far as they can go, and that at least two basic factors constrain
further increments in predictive utility. The first involves using estimated prior IQ as a
benchmark for forming expectations about patients’ standing in specific cognitive
domains (SCDs). This practice introduces a extra inferential link into the overall
diagnostic decision process and thereby may produce more error, or far more error, than
would be present if prior standing on the SCD most affected by the suspected disorder(s)
could be estimated directly. The second factor is related to the first and arises from
potential differences in the extent to which IQ versus SCDs are affected by various
neurocognitive disorders. For various disorders, functioning in certain cognitive domains
is likely affected earlier, and to a greater, or far greater extent, than IQ. Therefore, a new
approach that capitalizes on the positive aspects of the current methods (e.g., the use of
statistical prediction), and that: a) attempts to remove the extra inferential link, and b)
focuses on the cognitive domain(s) most likely to be affected by the disorder might lead to greater success.

A NEW APPROACH FOR IMPROVING METHODS OF ESTIMATING PRIOR FUNCTIONING

Limitations of IQ-Based Methods

Intelligence scales were designed to assess global intellectual capacity and not specific cognitive abilities related to brain function (Kaufman, 1990; Putnam et al., 1999; Sattler, 1992; Woodruff-Pak, 1997). Yet, IQ scores are related more or less strongly to virtually every specific cognitive domain (SCD). Correlations between scores on intelligence and SCD tests (SCDTs) generally range from about .20 to .70 (Crawford, 1992; Psychological Corporation, 1997). Given correlations of this magnitude, one would expect considerable variability in SCD standing across individuals with the same intellectual level.

This line of reasoning suggests that IQ D-scores will vary in their capacity to detect clinically relevant decline in SCD functioning. This has potentially important consequences because neurocognitive disorders differ with respect to the pattern, magnitude, and base rate of SCD impairment; and these differences can have important diagnostic, treatment, and prognostic implications. The ability of IQ D-scores to detect cognitive decline is further reduced in neurological conditions that are associated with SCD impairment in the face of relatively preserved, or only modestly attenuated, IQ levels. Possible examples of such disorders include ventromedial frontal cortical damage, mild-to-moderate traumatic brain injury (mTBI), chronic alcohol abuse (CAA), and Alzheimer's disease (AD), particularly in its earliest stages. PIQ D-scores can be more
sensitive to certain types of brain dysfunction than either VIQ or FSIQ D-scores (Gouvier, Bolter, Veneklasen, & Long, 1983) and therefore, might be better able to detect cognitive decline. However, methods for estimating prior PIQ are less, or considerably less accurate than for VIQ or FSIQ, suggesting that PIQ D-scores should be used cautiously.

Dissociation between SCD functioning and IQ performance probably reflects several factors, in particular the relative insensitivity of intelligence scales to some changes in brain state (Damasio, 1994; Schlosser & Ivison, 1989). Dissociation between IQ and SCDT performance was demonstrated in a study involving 35 mild AD patients, who were administered the WAIS-III and the Wechsler Memory Scale-III (WMS-III) (Psychological Corporation, 1997). As expected, the mild AD patients had lower WAIS-III IQ and WMS-III scores than the standardization sample, but the discrepancy was much greater for the WMS-III scores. Additionally, the mild AD group’s WMS-III scores were less variable than their IQ scores, suggesting that the group’s memory abilities were more consistently depressed than their IQ scores. For example, respective standard deviations for the VIQ and the WMS-III General Memory Index were 13.1 versus 8.6. The study suggests that, all other things being equal, a method for estimating prior WMS-III versus WAIS-III standing would provide a more accurate and sensitive means of detecting AD. Improved detection of AD (and possibly other disorders) could permit earlier and potentially more effective interventions (McLendon & Doraiswamy, 1999; Woodruff-Pak, 1997). Moreover, early diagnosis affords AD patients and their families more time to plan for long-term care and to develop coping strategies.
The foregoing discussion suggests that the clinical utility of IQ D-scores may be constrained by the magnitude and base rate of IQ impairment in certain disorders. If a disorder minimally effects intellectual level, then IQ D-scores scores will be of limited value in diagnosing it. Alternatively, if the disorder has a greater effect on a SCD than on IQ, SCD D-scores might provide a more powerful diagnostic aid than IQ D-scores. This implies that for certain disorders, it might be possible to improve on the diagnostic accuracy of IQ D-scores by estimating prior functioning on the SCD (or SCDs) most affected by the disorder.

Unfortunately, formal methods for estimating prior SCDT standing (specific cognitive domain estimation methods, SCDEMs) are generally, although not entirely lacking (Williams, 1997). As a result, clinicians commonly link expectations about how a patient ought to perform on a variety of neuropsychological tests to estimated prior IQ. This process involves two far from perfect inferential links: (1) estimating prior IQ, and (2) then using the IQ estimate to formulate expectations for performance on other tests. The use of SCDEMs, if available, might improve clinical accuracy simply by removing one of these inferential links. Suppose, for example, that A (IQ estimation) predicts B (prior IQ) at .50 accuracy, and that B subsequently predicts C (SCD test performance) at .50 accuracy. The potential result is an overall validity coefficient of .25 (.50 X .50). If, instead, one attempts to predict C by finding the strongest direct correlates (i.e., by dropping the middle step in the chain), matters may improve greatly, even if the direct predictors show a lower correlation than that which exists between A and B.
Preliminary Specific Cognitive Domain Estimation Methods

Schlosser and Ivison (1989) conducted one of the few attempts at developing an SCDEM. They developed a potential method for improving diagnostic accuracy for AD by estimating prior WMS Memory Quotient (MQ) score. They derived regression formulae based on age, and errors on the NART and/or the Schonell Graded Word Reading Test in a sample of 65 healthy elderly Australian subjects aged 65 to 89 years. The combined formulae had a multiple correlation of .74 and a SEE of 7.21. The maximum MQ D-score among the 65 normal subjects was 21 points, a result that occurred in only 2% of the subjects. In contrast, 14 of 16 patients (87.5%) with presumed AD had MQ D-scores of at least that magnitude. Schlosser and Ivison asserted that MQ D-scores might provide a quick and economical screen for dementia. Moreover, their results imply that it might be possible to estimate prior memory ability with sufficient accuracy to aid in clinical decision-making.

There are a number of limitations to Schlosser and Ivison’s (1989) study. First, the regression formulae were not cross-validated. Second, the two groups, especially the AD sample, were quite small. Third, the severity of dementia among the 16 AD patients was not clearly specified. However, the subjects were recruited from hospitals and nursing homes, and it seems likely that they had at least mild, and quite possibly moderate, dementia. This raises the question of whether the method would be sensitive enough to differentiate between individuals with mild or early dementia and non-demented patients who present with memory complaints -- the situation where it would be most useful. Finally, a partial replication of the study using the WMS-R raised doubt
that instruments such as the NART are effective predictors of memory functioning (O'Carroll et al., 1994).

Other investigators have attempted to develop SCDEMs on the basis of performance on single-word reading measures. For example, Crawford, Moore, and Cameron (1992) investigated the utility of the NART for predicting prior verbal fluency in a sample of 142 British subjects. NART errors correlated .67 with verbal fluency. The NART-based regression equation had a SEE of 9.09 -- about the size of the SD for the verbal fluency test thereby adding almost nothing to the determination of discrepancies from expectation.

Crawford, Obonsawin, and Allan (1998) examined the NART as a predictor of Paced Auditory Serial Addition Test (PASAT) raw scores in 152 healthy British subjects. The PASAT is a widely used test for assessing information processing and sustained and divided attention. A NART-based regression formula for predicting PASAT score yielded a multiple R of .52 and a SEE of 34.87. Crawford et al. (1998) suggested that NART-predicted PASAT scores could provide an estimate of prior ability on this task, which in turn could aid in detecting decline in information processing and in sustained and divided attention. In a sample of normal subjects and patients with either schizophrenia or bipolar disorder, Hawkins et al. (1993) found a strong correlation (.83) between raw scores on a measure of single word reading ability (Gates-MacGinitie Reading Vocabulary Test) and on a measure of naming ability (the Boston Naming Test).

There is so little additional work in this area, that, beyond some assurances that prediction of prior abilities in SCDs seems potentially feasible, determining which variables serve as the best predictors is, presently mostly a matter of educated conjecture.
Further, preliminary to this, one needs to examine whether, or the extent to which, accuracy might be increased if prior functioning in SCDs could be predicted.

The New Approach

Recently, two groups of authors (Franzen et al., 1997; Schinka & Vanderploeg, 2000) independently called for the development of SCDEMs as a means of improving diagnostic accuracy. The authors did not, however, offer guidelines for developing such methods. The present project attempts to address this omission. It proposes a potential new approach or strategy for improving diagnostic accuracy through the development of SCDEMs, and then undertakes an initial feasibility analysis of the approach's potential clinical utility. The approach capitalizes on divergent patterns of SCD impairment in different neurocognitive disorders. It involves identifying the SCDs that might optimally differentiate between those with and without a particular neurocognitive disorder. This information could then be used to prioritize and guide the development of SCDEMs, which, in turn, would permit calculation of SCD D-scores that could aid in improving detection of cognitive decline and diagnostic accuracy. If the disorder in question causes more decline in a SCD than in IQ, then, all other things being equal, SCD D-scores ought to detect the decline more effectively than IQ D-scores.

The validity of the proposed strategy rests upon two basic assumptions: (a) that in certain neurocognitive disorders, the base rate and/or magnitude of impairment in some SCDs (as measured by their associated SCDDTs) are greater than the base rate and magnitude of impairment in global intellectual functioning; and (b) that it is possible to develop SCDEMs for estimating prior SCD standing. That is, prior SCD standing is, to some extent, predictable given an optimal set of variables sufficiently correlated with
the SCD and sufficiently resistant to disruption by the neurocognitive condition. To the extent that either of these two assumptions fails to hold, the clinical utility of the method would be reduced or completely undermined.

PURPOSE OF THE STUDY

The present study attempted to provide an initial test of the proposed strategy’s potential to improve the detection of cognitive decline over IQ in particular neurocognitive disorders. It addressed three related, main questions. First, for specific disorders, are certain SCDs particularly effective in differentiating patients from controls? Second, would diagnostic accuracy for those disorders be improved over current, IQ-based methods should it be possible to estimate prior standing on the identified SCDs? Third, would the potential increase in diagnostic accuracy be sufficient to warrant development of SCDEMs? The results of this study were expected to help guide and prioritize the future development of SCDEMs.

The feasibility of the strategy was examined for mild AD, CAA, and mTBI. The method involved use of summary data reported in previously published studies of these disorders. The decision to study mild AD, CAA, and mTBI as opposed to other disorders was based on several factors: (a) these disorders are common reasons for referrals to neuropsychologists, (b) it may be difficult to detect the cognitive decline associated with them, (c) it may be hard to differentiate them from other competing etiologies, (d) the SCDT performance of patients with these disorders may be disproportionately impaired relative to their performance on IQ tests, and (e) these disorders can be associated with very high costs to individuals and to society.
For example, the annual incidence of TBI in the US is approximately 2 million, and the vast majority of cases are mild. Across all categories of TBI, the annual cost for acute care alone is estimated at $25 billion (Williamson, Scott, & Adams, 1996). Alcohol problems affect about 10% of the US adult population, or approximately 30 million individuals, and cost US society an estimated $100 billion annually in treatment and in lost productivity (Drug Abuse USA, 1996; Parsons, 1996; US Census Bureau, 1998).

Alzheimer's disease is a progressive degenerative brain disorder. Estimates of the incidence and prevalence of AD vary widely due to methodological differences across studies (Bondi, Salmon, & Kaszniak, 1996). However, it is estimated that approximately 10% of individuals over the age of 65 years have dementia, and that AD accounts for as many as 50% of those cases (Nixon, 1996). About 4 million Americans currently have AD, with an annual cost to US society of about $100 billion annually (Alzheimer’s Association, 1999; Gilliard & Rabins, 1999). Given current prevalence and incidence rates, approximately 10 million people in the US may suffer from AD by the year 2030 (Woodruff-Pak, 1997). Diagnosis of AD is usually made by exclusion and can be confirmed only at autopsy (McKhann et al., 1984). Differential diagnosis of AD may be difficult at times. For example, research indicates that of patients with dementia, less than 75% are correctly classified by dementia subtype (e.g., AD vs. vascular dementia) (Barrett, Haley, Harrell, & Powers, 1997; Chui & Zhang, 1997; Ryan, 1994). Accurate and early detection of AD is critical to initiating appropriate treatment early in the disease process, when it might be most effective (McLendon & Doraiswamy, 1999). This, in turn, may help to reduce emotional and financial burden to families and to society.
The results of the study were expected to support the strategy. That is, it was anticipated that for each disorder being studied, analysis would show that one or more SCDs (as measured by their corresponding SCDTs) exceeded the accuracy of IQ measures. Further, it was hypothesized that this difference would be of sufficient magnitude (i.e., 10% or greater) to suggest that the development of SCDEM is worthwhile, that is, could improve detection of cognitive decline and reduce false-positive errors, thereby increasing diagnostic accuracy. The 10% benchmark is somewhat arbitrary, but it is intended only as a rough guide.
METHOD

For the sake of clarity, a brief overview of the method will be presented first, followed by a more detailed description. The method involved reconstructing SCDT and IQ distributions based on means and SDs reported in previously published mild AD, CAA, and mTBI studies, and then calculating the percentage of overlap in the reconstructed distributions. The percent overlap values comprised the study’s primary data units and provided a measure of the extent to which a neurocognitive disorder affected SCDT and IQ standing, respectively. Thus, these values were viewed as indirect markers of SCDT and IQ sensitivity to the disorder. Estimated and obtained distributions should overlap to a much greater extent among normal versus abnormal individuals.

The SCDT percent overlap values were then pooled (summed) by cognitive domain and averaged to form SCD Indexes. Comparable IQ Indexes were calculated. The SCD Indexes were then subtracted from the IQ Indexes. The obtained values were viewed as a marker of the potential sensitivity of the SCD versus IQ to the disorder, and provided an indication of the relative ability of SCD D-scores versus IQ D-scores to detect cognitive decline associated with the disorder. The magnitude of any discrepancy was viewed as a measure of the potential improvement in diagnostic accuracy that might be obtained by developing SCDEM.

Arguably, the strategy would have been tested more rigorously had actual SCDEM been developed and examined for clinical utility across several disorders. However, the effort involved in developing even a single SCDEM and examining its clinical utility in a single clinical sample with a single disorder is substantial. Moreover, it might well be premature to undertake such a laborious and potentially expensive
project without first examining the viability of the strategy. It appears reasonable, then to have accepted somewhat less control and rigor in order to permit a broader test of the proposed strategy.

**Study Selection**

A total of 5 previously published studies each for mild AD, CAA, and TBI seemed sufficient to test the proposed strategy. Study selection guidelines were as follows:

(a) The studies needed to use adequate diagnostic methods for forming clinical groups.

(b) The studies needed to contain clearly reported clinical- and control-group means and SDs or standard errors of the mean (SEM) for commonly used SCDTs and for Wechsler IQ scores. (When SEM were reported, they were algebraically converted to SDs.) Full- or short-form Wechsler Intelligence Scales scores were deemed acceptable. Some of the identified studies used a short-form IQ test and reported only subtest means and SDs, not prorated IQ. A few such studies (2 of 8 for mild AD, 1 of 9 for CAA, and 0 of 5 for mTBI) were included, provided that results for at least 4 subtests were available and could be used to prorate IQ means and SDs.

(c) The studies needed to include properly formed or matched control groups. It was assumed that control groups provided an acceptable estimate of clinical groups’ prior SCDT and IQ standing. Unanticipated difficulties were encountered in identifying studies that simultaneously met this as well as the other guidelines. As a result, some studies were selected that lacked control
groups but that met the other criteria. In the absence of control groups, SCD and IQ normative samples were used as the basis for comparison, as they likely provided a reasonable, although less exact, estimate of the clinical group's prior standing. Different samples can vary along a number of demographic and performance characteristics. Thus, using normative rather than control samples as the standard of comparison introduces potential error into the analysis. Most studies for mild AD (5 of 8) and for CAA (8 of 9) did have control groups, but not for TBI (only 1 of 5 studies). When norms were used as the standard of comparison, an effort was made to select normative samples that were demographically similar to the clinical group. However, for the Wechsler Intelligence Scale and Memory Scale comparisons, clinical performances were always compared to the standardization samples in the absence of a control group.

Based on these guidelines, 22 studies (8 mild AD, 9 CAA, and 5 TBI) were consecutively selected, starting from an extensive search of approximately 800 PsychInfo and PubMed abstracts, and proceeding, where indicated, to a full review of the studies. Many studies that initially appeared promising turned out to be unsuitable once the full report was examined. More studies were used than originally planned (22 vs. 15) in an effort to increase the trustworthiness of the findings, especially in light of some of the allowances that were made during study selection. A description of each selected study is provided in Appendix A.

The most common reasons for study exclusion were the absence of reported, or clearly reported, means and/or SDs; use of only one or two WIS subtests; use of
experimental rather than standardized neuropsychological tests; use of heterogeneous clinical samples (e.g., a “brain damaged” or a “dementia” group comprised of subjects with multiple etiologies) -- a particularly troublesome problem in identifying mild AD and mTBI studies; and use of summary scores, (i.e., scores on several SCDTs collapsed into a single summary score). Other reasons for study exclusion included lack of clearly reported means and SDs (e.g., difficult to read bar graphs with error bands), omission of means and SDs for key tests, and very small sample sizes (i.e., < 5).

Selection of IQ and SCDT Distributions for Reconstruction

The IQ measures included VIQ, PIQ and FSIQ from various versions of the Wechsler Intelligence Scales or, in one case, the Wechsler-Bellevue Intelligence Scale. The Processing Speed Index was the only factor score selected for distribution reconstruction because of its reported high sensitivity to brain damage in general (Hawkins, 1998).

Most of the studies reported means and SDs for several tests. Typically, distributions were reconstructed for most of the SCDTs included in a study’s cognitive test battery. This decision was intended to reduce experimenter bias and permit relative consistency in selecting SCDTs. However, certain exceptions were made according to four guidelines. First, SCDTs that were highly correlated with IQ (i.e., colinearity between the variables was present) were omitted because the strategy will not work in such cases. This is because, colinearity indicates that the SCDT and IQ measures are tapping the same, or highly similar abilities. Thus, including a high number of SCDTs having colinearity with IQ might well lead to a spurious rejection of the strategy. Second, SCDTs were selected that appeared to hold promise for the development of SCDEMs
using the demographic variables of age, education, and gender. Third, emphasis was
given to SCDTs that are commonly used in clinical practice. Fourth, to streamline the
analysis for the WMS-III, which provides several indexes, distributions were
reconstructed only for the Working Memory, Immediate Memory, and General Memory\(^2\)
Indexes (IMI, WMI, and GMI, respectively), thereby covering all of the theorized stages
of memory processing. Distributions were reconstructed for other WMS-III indexes (e.g.,
Visual Immediate Index) when the reported means and SDs appeared as sensitive to the
disorder as IMI, WMI, and GMI.

Correlations among the Selected SCDTs and IQ

Memory loss is a cardinal feature of Alzheimer’s disease (Bondi et al., 1996;
Nixon, 1996) and may be affected to a relatively greater degree than IQ in the early
stages of the disorder (Bornstein & Chelune, 1988). Measures evaluating retention of new
information over brief delays have consistently been shown to be effectively differentiate
AD patients from controls and other patient groups (Albert & Moss, 1992; Butters,
Salmon, & Butters, 1997).

Accordingly, the memory SCDTs selected for distribution reconstruction included
the Wechsler Memory Scales, the Auditory Verbal Learning Test, the Free and Selective
Reminding Task, and the Benton Visual Retention Task. Research shows generally low
to moderate correlations between these memory measures and IQ (e.g., Bishop, Dickson,
& Allen, 1990; Mitrushina, Boone, & D’Elia, 1999; Psychological Corporation, 1997;
Query & Megran, 1983; Sivan, 1992; Spreen & Strauss, 1998).

\(^2\) A global delayed memory index (Psychological Corporation, 1997).
Patients with AD can exhibit language impairment, particularly in naming and verbal fluency (Bondi et al., 1996; Nixon, 1996). Accordingly, distributions were reconstructed for the Boston Naming Test and for verbal fluency tests. The Boston Naming Test generally shows modest correlations with WAIS-III IQ (.38, .42, and .44 for VIQ, PIQ, and FSIQ respectively) (Mitrushina et al., 1999; Psychological Corporation, 1997). The verbal fluency measures used here involve generating words starting with a specific letter or belonging to a designated category as rapidly as possible within a short time period. Fluency to letter cue is often considered to be an executive functioning task (i.e., initiation), whereas semantic fluency is more often considered to be a language task. It was sometimes difficult to determine from the reported studies whether the letter cue or semantic cue task had been administered, and thus verbal fluency was grouped with other language measures. Verbal fluency is moderately correlated with WAIS-III VIQ, PIQ, and FSIQ, (.61, .48, and .59, respectively) (Psychological Corporation, 1997).

Attentional abilities can also be impaired in AD (Lezak, 1995; Nixon, 1996). Part A of the Trail Making Test is often described as being sensitive to attention, visual scanning, eye-hand coordination speed, and information processing. In general, the correlations between this test and IQ scores ranges from modest to moderate levels (−.27 to −.46) (Psychological Corporation, 1997; Tremont, Hoffman, Scott, & Adams, 1998), although lower correlations have been reported (Yeudall, Reddon, Gill, & Stefanyk, 1987).

Alzheimer’s disease is also associated with impairment in executive functioning and abstract reasoning (Bondi et al., 1996) Part B of the Trail Making Test is often viewed as tapping executive functioning; its correlation with IQ scores tends to be
moderate (i.e., -.42 to -.66) (Psychological Corporation, 1997; Tremont et al., 1998), although lower correlations have been reported (Yeudall et al., 1987). The Porteus Mazes are considered to be sensitive to executive impairment, particularly planning ability (Lezak, 1995). Studies have suggested modest correlations between performance on the Porteus Mazes and IQ (e.g., Porteus, 1965; Watson & Klett, 1974). The Stroop Interference Task is considered to be a measure of complex attention and executive functioning. Successful performance requires inhibiting a prepotent response in favor of an unusual one. Correlations between the Stroop and IQ vary across studies but appear to be generally modest (Mitrushina et al., 1999; Spreen & Strauss, 1998).

Visuospatial impairment is also commonly seen in AD patients (Nixon, 1996). The Benton Facial Recognition Test (BFRT) was designed to assess recognition of unfamiliar faces, or a sub-domain of visuospatial ability. Trahan’s (1997) study suggested, in most cases, non-significant relations between BFRT performance and IQ in a group of patients with cerebrovascular disorder.

Some of the mild AD studies reported results for the Dementia Rating Scale (DRS) and for the Mini Mental State Examination (MMSE). These tests screen cognitive functioning across a number of domains and, therefore, are sometimes referred to as omnibus tests. As such, they were not grouped with SCDTs in the present study. However, given their wide use in clinical practice, it was decided to reconstruct their distributions for separate evaluation. The DRS correlates moderately with WAIS-III VIQ, PIQ, and FSIQ (.59, .58, and .61, respectively) (Psychological Corporation, 1997). Individuals with higher IQ and higher levels of education achieve higher MMSE scores
(Spreen & Strauss, 1998). One study showed strong correlations between the MMSE and WAIS VIQ and PIQ (.78 and .66, respectively) (Folstein, Folstein, & McHugh, 1975).

As was the case with mild AD, correlations between IQ and the SCDTs selected for distribution reconstruction in the CAA and mTBI studies were generally modest (with some exceptions to be covered later), and it seems unnecessary to go into full details. However, a brief description -- similar to that provided above for mild AD -- of the cognitive dysfunction typically associated with CAA and mTBI helps provide a context for subsequent discussion.

Early in the course of abstinence, chronic alcohol abusers tend to exhibit mild-to-moderate impairment in executive functioning, particularly in abstract reasoning and problem solving. These patients also tend to show impairment in perceptual-motor and perceptual-spatial abilities, as well as in learning and memory (somewhat greater for non-verbal as opposed to verbal information) (Allen & Landis, 1997; Lezak, 1995; Parsons, 1996). Most studies show that cognitive functioning improves with sustained abstinence, but the recovery process may take up to 5 years (Parsons, 1996).

In the case of mTBI, many individuals perform in the impaired range on tasks of complex attention and cognitive set shifting. Deficits may also occur in learning and memory, and in abstraction and problem solving (Dikmen, Temkin, & Armsden, 1989; Kay, 1986; Levin, Benton, & Grossman, 1982; Levin et al., 1987; D. J. G. Williamson et al., 1996). Decreased information processing efficiency is common (Gronwall & Wrightson, 1974, 1981; Lezak, 1995) in mTBI. It may contribute to attenuated performance on other cognitive tasks, and might also underlie the memory and concentration complaints of mTBI patients (D. J. G. Williamson et al., 1996). Moreover,
improvement in speed of information processing tends to parallel reduction in post-concussion symptoms (Dikmen et al., 1989; Gronwall, 1976). Post-concussion symptoms may include blurred vision, headaches, dizziness, anxiety, depression, and sleep disturbance (Levin et al., 1987; Rutherford, 1989). Results of well-controlled studies indicate that by 1 to 3 months post-injury, most mTBI patients do not differ to a statistically significant degree from matched controls on tests of neuropsychological functioning (e.g., Dikmen et al., 1989; Levin et al., 1987). Given that most mTBI and CAA patients improve or recover over time, the current strategy might be most useful in identifying cognitive decline during the period before recovery occurs, or in the minority of cases with persistent deficits.

Distribution Reconstruction and Analysis Procedures

**Distribution Reconstruction Procedures**

SCDT and IQ distributions were reconstructed based on reported IQ and SCDT means and SDs. All reconstructed distributions were arbitrarily set to a sample size of 100 hypothetical subjects, thereby permitting consistency and uniformity in the reconstructed distributions. Because the shape of a distribution is dependent on its mean and standard deviation, increasing or decreasing sample size has little or no effect on the general shape of distribution curves. Few studies reported score ranges. Consequently, reconstructed distributions were set either to a test’s entire range of scores, or, for tests with large ranges (e.g., WAIS IQ), were set to scores that corresponded to approximately ±3.5 to 4.0 SD of the mean. It is understood that this work involves approximations and that here and elsewhere, adjustments to the original data may be necessary, but might also alter, to varying degrees, the true nature of underlying data. However, such
approximations and adjustments are not at all uncommon in initial stages of development or in initial feasibility examinations. Further, in the present case, these assumptions generally do not work in favor of the hypotheses, and they permit a much broader analysis than would otherwise be possible.

The distributions were reconstructed using the STANDARDIZE and NORMDIST commands in Excel 98™ (Microsoft, 1997-98). The STANDARDIZE command returns the z-score for all test scores in a normal distribution with a given mean and SD. The NORMDIST command returns the cumulative area under the curve for each score in a normal distribution. For example, scores that fall at 1 SD below the mean and 1 SD above the mean receive a NORMDIST value of .34 and .84, respectively. Therefore, the area under the curve associated with a particular score (i.e., score band – the interval between the next lowest and next highest scores) had to be calculated by subtraction as follows:

\[ \text{Area of score}_n = \text{NORMDIST}_{n+1} - \text{NORMDIST}_{n-1}. \]

However, summing across all of the score bands results in twice the area under the curve. This is because the score bands overlap such that 50% of the area of a score band also belongs to the next highest score band and the other 50% belongs to the next lowest score band. Accordingly, the area of each score band was divided by 2 before being multiplied by 100 (the number of hypothetical subjects in each sample). The resulting values provided a reconstructed normal frequency distribution with a mean and SD identical to that reported in the original study. The distributions were then graphed using Excel’s graphing functions. For each single SCDT, the graphs of the reconstructed clinical and control distributions were then plotted and juxtaposed to aid visualization of the degree to which they overlapped. Selected graphs are presented in Appendix B for purposes of illustration.
Calculating Percent Overlap

For each distribution, the test score closest to the point at which the two distributions intersected was identified (Intersect scores). The z-scores for these two Intersect scores were obtained using Excel’s STANDARDIZE function. The area under the curve between the mean and the two z-scores was obtained from a published table (Berkowicz, Ewen, & Cohen, 1976, Table B, p. 300). This value was added or subtracted (as appropriate) to or from 50% (which is the area of each curve falling above or below the mean) to determine the percentage of area for each distribution falling either above or below the Intersect score. The region of interest for the distribution with the higher mean was the percentage of area lying below the Intersect score; for the distribution with the lower mean, the region of interest was the percentage of area lying above the Intersect score. These percentages were summed to provide the shared amount of the total available area under the two curves. This value was termed the “percent overlap.” This value corresponds to the percentage of subjects in the reconstructed distributions that have overlapping scores, when both performance level and frequency of scores at the varying performance levels are considered.

To clarify, consider the example of the juxtaposed, reconstructed clinical and control distributions for Part A of the Trail Making Test (TMT-A) in Botwinick et al.’s (1986) mild AD study (the authors inverted the TMT-A scores so that lower scores indicated poorer performance, see Appendix B). The mean TMT-A score was 14.0 ± 7.7 for the mild AD group and 27.6 ± 7.8 for the control group. Accordingly, the graphs of the reconstructed distributions overlapped such that the mild AD distribution fell to the left of (i.e., was shifted lower than) the control distribution. The reconstructed
distributions intersected at a TMT-A score of 21 seconds (Intersect score). For the mild AD distribution, the z-score for the Intersect score was .91. Reference to the table of the area under the normal curve revealed that 31.86% of that area falls between the mean and \( z = .91 \). As the mild AD group’s Intersect score fell above the mild AD mean, the area of interest was that extending beyond (i.e., higher than) the Intersect score. Accordingly, 18.14% (i.e., 50% – 31.86%) of the mild AD distribution fell above the Intersect score.  

The z-score for the control distribution at this Intersect score was \(-.85\). Reference to the table revealed that 30.23% of the area under the normal curve falls between the mean and \( z = -.85 \). As the control group’s Intersect score falls below the control group mean, the area of interest was that extending below (i.e., lower than) the Intersect score. Accordingly, 19.77% (i.e., 50% – 30.23%) of the control distribution fell below a score of 21. Summing the two areas of interest revealed that the two distributions overlapped by 38% (18.14% + 19.77% = 37.91%). This general procedure was followed for all reconstructed clinical/control IQ and SCDT distribution pairs, with modifications when necessary.

In some cases the clinical and control distributions had similar means but very different SDs (see Appendix B for an example). It was not uncommon in such cases for the reconstructed distribution with the smaller SD to be entirely contained within only a portion of the distribution with the larger SD. The graphic appearance was of two curves -- one broad and one narrow -- superimposed on one another. The percent overlap values in these cases may be quite large, suggesting that D-scores based on this SCD are likely to have little clinical utility. However, in some of these situations, the large percent overlap value might have masked certain effects. For example, only a small percentage of
the narrower distribution might fall above or below an Intercept score, whereas a large percentage of the broader distribution may fall above or below that same score. In such situations, the SCDT could increase accuracy beyond IQ measures, with the Intersect score actually providing a potential cut point. However, this begs the question of the need to develop SCDEMs for measures that reliably detect disorder on the basis of such cut scores. Instances in which the overlap of the distributions suggested potential cut points are discussed on a study-by-study basis in Appendix A.

Determining the Relative Sensitivity of SCDTs versus IQs

For each disorder, the percent overlap values for each SCDT were pooled according to the cognitive domain they measure, and then summed and averaged. The decision of which SCDT would be grouped in which domain was made in accordance with common clinical and research practices and with reference to previously published SCDT descriptions (e.g., Lezak, 1995; Mitrushina et al., 1999; Spreen & Strauss, 1998). Each SCDT was grouped in only one SCD. These average percent overlap values for the SCDs were referred to as SCD indexes. Analogous IQ indexes were calculated. The SCD and IQ indexes provided a measure of the average extent to which clinical and control groups overlapped on IQ tests and on tests measuring the same SCD. As such, the indexes were also considered to provide a marker of the degree of SCD versus IQ sensitivity to the disorder. The smaller the value of the index, the greater the sensitivity.

SCDT and IQ indexes were as follows:

(a) The ALL IQ index was the average percent overlap for all reconstructed IQ distributions (i.e., VIQ, PIQ, and FSIQ).
(b) The VIQ/FSIQ index was the average percent overlap for all reconstructed VIQ and FSIQ distributions. PIQ was omitted because of the relatively low accuracy with which it can be estimated.

(c) The MEMORY index was comprised of the subtests and indexes of the WMS and its revisions (but not the Working Memory and Attention/Concentration indexes), the Auditory Verbal Learning Test, the Free and Cued Selective Reminding Test, the recall conditions of the Benton Visual Retention Test, the recall conditions of the Rey-Osterreith Complex Figure Test, and the memory score from the Tactual Performance Test.

(d) The ATTENTION index included measures of attention, working memory, and processing speed. SCDTs comprising this index were selected subtests and indexes from the WMS and its revisions (Working Memory and Attention/Concentration indexes, Digit Span, and Mental Control), Part A of the Trail Making Test, the WAIS-III Processing Speed Index, and the WAIS Digit Symbol subtest.

(e) The LANGUAGE index was comprised of verbal fluency measures (as previously discussed) and the Boston Naming Test.

(f) The CONSTRUCTION index was used only for the mTBI studies and was comprised of the WIS Block Design and Object Assembly subtests.

(g) The Perceptual Organizational index (PCPT-ORG) included measures of visuospatial ability, and of auditory and tactile perceptual organization. Tests comprising this index included the Block Design subtest (for the CAA studies only), the Benton Facial Recognition Test, the copy condition of the Benton
Visual Retention Test, the copy condition for the Rey-Osterreith Complex Figure Test, the total time and location scores for the Tactual Performance Test, the Seashore Rhythm Test, and the Speech-Sounds Perception Test.

(h) The EXECUTIVE index included measures of abstract reasoning, concept formation, response inhibition, and the ability to alternate between cognitive sets. Tests comprising this index included the various Stroop scores, Part B of the Trail Making Test, the Category Test, and the various scores from the Wisconsin Card Sorting Test.

(i) The OTHER index included SCOTs that were too few (i.e., less than 4 percent overlap values) to form a reliable index. The SCOTs comprising this index differed considerably across conditions.

(j) The OMNIBUS index included the Dementia Rating Scale and Mini-Mental State Examination, which were used in some of the mild AD studies.

For each disorder, each SCOT index was subtracted from each of the two IQ indexes (ALL IQ and VIQ/FSIQ). The resultant values were considered to be markers of a SCD’s sensitivity to the disorder relative to IQ. Positive values suggested that the disorder attenuated SCD to a greater extent than IQ; negative values suggested that IQ was more sensitive to the disorder. As a secondary analysis, Bonferroni-corrected pairwise t-tests were used to determine the degree to which the difference scores exceeded chance levels. This was intended to enhance confidence in the results but was not meant to replace the previously set 10% criterion: that a difference between the SCOT and IQ indexes of 10 percentage points would be required to provide support for the rationale and for the development of SCDEMs.
RESULTS

The percent overlap values for each pair of reconstructed clinical and control group distributions are presented by disorder in Tables 1, 4, and 7. All individual percent overlap values reported in text and tables are rounded off to the nearest whole number, and all percent overlap values are based on normal distributions.

As followed from the procedures used, all of the reconstructed distributions were normal. This occurred because the reconstructed distributions were based on reported means and SDs rather than on subject-by-subject data. Thus, it was not possible to portray skewness in the reconstructed distributions. Because all of the reconstructed distributions were normal, skewness in the original data appeared as truncated graphs.

Skewness in test score distributions reflects ceiling and floor effects. In extreme cases, floor effects might reflect the complete absence of the ability being measured, whereas ceiling effects might occur on tests where performing normally means achieving a perfect, or near perfect score. In most cases, however, ceiling and floor effects probably indicate that the test lacks sufficiently easy or difficult items to adequately differentiate among those with low or high standing on the abilities being measured. Their presence suggest that even greater separation between clinical and control distributions would likely occur if the test’s upper or lower ability ranges were extended to more fully capture the range of variability in the sample. Consequently, in the presence of ceiling and floor effects (i.e., truncated reconstructed distribution graphs), the percent overlap values most likely underestimate the actual separation between clinical and control group ability levels. As such, their presence in the reconstructed distributions probably does not
systematically favor the research hypotheses and, in some cases may well work against them, although clarifying such matters is problematic in the absence of raw data.

Mild Alzheimer's Disease

Table 1 lists each of the selected mild AD studies, summarizes the sample sizes, provides the means and SDs for each test selected for distribution reconstruction, and lists the percentage of area shared by each reconstructed clinical and control distribution pair. Table 2 lists the mean ± SD percent overlap for each mild AD index, and provides the number and range of the individual SCDT percent overlap values comprising them. The SCDTs that comprised the mild AD OTHER index were The Bender-Gestalt Test, the copy condition of the Benton Visual Retention Test, the Benton Facial Recognition Test, the Stroop interference score, Part B of the Trail Making Test, and the Porteus Maze Test.

Table 3 summarizes the pairwise differences between the various SCDT and IQ indexes; the table entries are the difference between the IQ and SCDT index values (i.e., IQ index minus SCDT index). Inspection of Table 3 shows that the greatest difference occurred between MEMORY and the two IQ indexes (ALL IQ and VIQ/FSIQ). That is, across the mild AD studies, the amount of overlap between the reconstructed clinical and control distributions was considerably less for memory SCDTs than it was for IQ. The differences were statistically significant. This suggests that memory is more sensitive to mild AD than IQ is, and thus, that memory SCD D-scores might be more efficacious than IQ D-scores at detecting cognitive decline in mild AD, and improving diagnostic accuracy. Moreover, the magnitude of difference between the MEMORY index and the ALL IQ and VIQ/FSIQ indexes easily exceeds the 10% criteria for supporting the development of memory SCDEMs.
Table 1. Summary of mild AD comparisons by study: sample sizes, reported means and SDs, and percent overlap values for each reconstructed distribution pair.

| Study                  | Test   | Comparison Groups                | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|------------------------|--------|----------------------------------|---------------|-------------|--------------|------------|---------------|
| Bigler et al., 1985    | WAIS   | mild AD (n = 42) vs. strd. sample | 95.50         | 15.40       | 100.00       | 15.00      | 88            |
|                        |        | VIQ                              | 86.30         | 16.30       | 100.00       | 15.00      | 66            |
|                        |        | PIQ                              | 91.70         | 14.90       | 100.00       | 15.00      | 66            |
|                        |        | FSIQ                             |               |             |              |            |               |
|                        |        | SCDTs                            |               |             |              |            |               |
|                        |        | WMS MQ                           | 91.70         | 14.90       | 100.00       | 15.00      | 55            |
| Botwinick et al., 1986 | WAIS   | mild AD (n = 30) vs. controls (n = 18) | 97.00         | 11.25       | 128.00       | 19.95      | 30            |
|                        |        | Prorated FSIQ                     |               |             |              |            |               |
|                        |        | SCDTs                            |               |             |              |            |               |
|                        |        | Bender-Gestalt                    | 8.70          | 2.20        | 10.80        | 0.90       | 47            |
|                        |        | BNT                              | 42.70         | 22.40       | 71.50        | 9.50       | 33            |
|                        |        | BVRT copy                        | 7.90          | 2.90        | 9.40         | 0.90       | 55            |
|                        |        | BVRT recall                       | 2.30          | 1.80        | 5.80         | 1.40       | 27            |
|                        |        | TMT-A¹                           | 14.00         | 7.70        | 27.60        | 7.80       | 38            |
|                        |        | VF                               | 16.10         | 10.30       | 28.80        | 6.60       | 44            |
|                        |        | WMS DS(f)                        | 6.30          | 1.20        | 6.80         | 1.10       | 83            |
|                        |        | DS (b)                           | 3.50          | 1.40        | 5.30         | 1.40       | 46            |
|                        |        | LM I                             | 2.10          | 1.70        | 9.00         | 2.20       | 08            |
|                        |        | MC                               | 4.90          | 2.30        | 7.30         | 1.80       | 55            |
|                        |        | PA                               | 6.00          | 2.50        | 13.50        | 3.40       | 21            |
| Study                  | Test    | Comparison Groups                  | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|------------------------|---------|------------------------------------|---------------|-------------|--------------|------------|---------------|
| Haxby et al., 1990     | WAIS    | mild AD (n = 11) vs. controls (n = 29) |               |             |              |            |               |
|                        |         | FSIQ                               | 117.00        | 8.00        | 126.00       | 10.00      | 61            |
|                        |         | SCDTs                              |               |             |              |            |               |
|                        |         | BFRT                               | 42.00         | 5.00        | 44.0         | 4.00       | 81            |
|                        |         | BNT                                | 35.00         | 7.00        | 37.00        | 4.00       | 71            |
|                        |         | Porteus Mazes                      | 12.80         | 3.90        | 15.40        | 1.60       | 55            |
|                        |         | Stroop int.                        | 24.00         | 8.00        | 37.00        | 8.00       | 42            |
|                        |         | TMT-B                              | 192.00        | 155.00      | 82.00        | 40.00      | 44            |
|                        |         | VF                                 | 34.00         | 13.00       | 42.00        | 15.00      | 77            |
|                        |         | WMS LM I                           | 11.00         | 5.00        | 22.00        | 5.00       | 27            |
|                        |         | LM II                              | 2.00          | 4.00        | 17.00        | 5.00       | 9             |
|                        |         | VR I                               | 7.00          | 4.00        | 10.00        | 3.00       | 65            |
|                        |         | VR II                              | 1.00          | 1.00        | 7.00         | 3.00       | 11            |
|                        |         | DRS                                | 132.00        | 6.00        | 142.00       | 2.00       | 18            |
| Kirk & Kertesz, 1991   | WAIS-R  | AD (n = 38) vs. strd. sample        |               |             |              |            |               |
|                        |         | FSIQ                               | 81.00         | 12.70       | 100.00       | 15.00      | 49            |
|                        |         | VIQ                                | 79.70         | 13.40       | 100.00       | 15.00      | 55            |
|                        |         | PIQ                                | 83.50         | 12.60       | 100.00       | 15.00      | 47            |
|                        |         | SCDTs                              |               |             |              |            |               |
|                        |         | WMS MQ                             | 75.50         | 14.60       | 100.00       | 15.00      | 41            |
|                        |         | DRS                                | 99.60         | 20.00       | 141.30       | 3.00       | 5             |
| Study                  | Test    | Comparison Groups | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|------------------------|---------|-------------------|---------------|-------------|--------------|------------|---------------|
| Petersen et al., 1999  | WAIS-R  | v. mild AD (n = 234) vs. control (n = 48) |               |             |              |            |               |
|                        | FSIQ    |                   | 83.90         | 9.01*       | 101.80       | 10.71*     | 36            |
|                        | VIQ     |                   | 86.60         | 10.39       | 102.50       | 9.18       | 42            |
|                        | PIQ     |                   | 82.10         | 11.78       | 100.40       | 12.41      | 44            |
| SCDTs                  | AVLT learning |                   | 21.50         | 7.70        | 35.50        | 9.20       | 40            |
|                        | % ret.  |                   | 8.30          | 19.80       | 62.10        | 24.58      | 27            |
|                        | BNT     |                   | 34.70         | 13.60       | 50.30        | 7.65       | 43            |
|                        | FCSRT learning |                  | 25.60         | 16.26       | 58.40        | 12.24      | 25            |
|                        | FCSRT % ret. |                 | 39.50         | 43.84       | 86.70        | 16.83      | 38            |
|                        | VF      |                   | 24.40         | 12.02       | 35.10        | 10.71      | 64            |
|                        | WMS-R LM I |                 | 8.60          | 5.54        | 21.30        | 6.12       | 27            |
|                        | LM II   |                   | 2.80          | 4.16        | 15.3         | 7.65       | 27            |
|                        | VR I    |                   | 14.40         | 6.24        | 25.70        | 6.12       | 36            |
|                        | VR II   |                   | 4.10          | 4.85        | 17.60        | 7.65       | 27            |
|                        | DRS     |                   | 112.70        | 13.44       | 134.30       | 6.12       | 24            |
|                        | MMSE    |                   | 22.60         | 3.54        | 28.30        | 1.53       | 24            |
### Mild AD Comparisons (cont.)

| Study                          | Test            | Comparison Groups                  | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|--------------------------------|-----------------|------------------------------------|---------------|-------------|--------------|------------|---------------|
| Psych. Corp., 1997             | WAIS-III        | mild AD (n = 35) vs. strd. sample  | 86.60         | 13.10       | 100.00       | 15.00      | 64            |
|                                |                 | FSIQ                               | 92.20         | 13.10       | 100.00       | 15.00      | 77            |
|                                |                 | VIQ                                | 81.70         | 13.20       | 100.00       | 15.00      | 51            |
|                                |                 | SCDTs                              |               |             |              |            |               |
|                                |                 | WMS-III GMI                         | 60.40         | 8.90        | 100.00       | 15.00      | 9             |
|                                |                 | IMI                                | 62.90         | 11.40       | 100.00       | 15.00      | 16            |
|                                |                 | WMI                                | 80.40         | 16.90       | 100.00       | 15.00      | 55            |
|                                |                 | PSI                                | 79.60         | 14.40       | 100.00       | 15.00      | 49            |
| Storandt et al., 1984          | WAIS            | mild AD (n = 42) vs. controls (n = 42) | 92.00         | 17.55       | 124.00       | 12.45      | 28            |
|                                |                 | Prorated FSIQ                       |               |             |              |            |               |
|                                |                 | SCDTs                              |               |             |              |            |               |
|                                |                 | Bender-Gestalt                      | 4.38          | 2.05        | 2.26         | 1.21       | 50            |
|                                |                 | BNT                                | 36.10         | 22.34       | 70.83        | 8.25       | 25            |
|                                |                 | BVRT form C                        | 4.71          | 1.47        | 6.93         | 0.34       | 37            |
|                                |                 | BVRT errors                        | 2.14          | 1.83        | 5.43         | 1.76       | 30            |
|                                |                 | TMT-A                              | 4.50          | 2.64        | 7.21         | 1.68       | 33            |
|                                |                 | VF                                 | 16.24         | 7.38        | 5.17         | 3.53       | 44            |
|                                |                 | WMS DS                             | 9.14          | 2.18        | 11.64        | 2.21       | 59            |
|                                |                 | LM                                 | 2.02          | 1.94        | 9.65         | 2.07       | 10            |
|                                |                 | MC                                 | 4.50          | 2.64        | 7.21         | 1.68       | 52            |
|                                |                 | PA rcl.                            | 5.61          | 2.59        | 12.89        | 3.16       | 20            |
|                                |                 | PA rec.                            | 5.61          | 2.59        | 12.89        | 3.16       | 41            |
### Mild AD Comparisons (cont.)

| Study                  | Test | Comparison Groups                  | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|------------------------|------|------------------------------------|---------------|-------------|--------------|------------|---------------|
| Weingartner et al, 1981| WAIS | mild PID\(^3\) (n = 14) vs. controls (n = 14) |               |             |              |            |               |
|                        |      | FSIQ                               | 100.40        | 6.10        | 111.40       | 10.00      | 47            |
|                        |      | PIQ                                | 85.00         | 10.00       | 112.00       | 6.00       | 9             |
|                        |      | SCDTs                              | 82.00         | 11.10       | 113.00       | 6.30       | 7             |
|                        |      | WMS MQ                             |               |             |              |            |               |

\(^1\)TMT-A values are the reciprocal of time to complete X 1000.

\(^2\)SD algebraically converted from SEM.

\(^3\)PID = progressive idiopathic dementia.

SCDT abbreviations: AVLT = Rey Auditory-Verbal Learning Test (% ret. = % retention); BFRT = Benton Facial Recognition Test; BNT = Boston Naming Test; BVRT = Benton Visual Retention Test; DRS = Dementia Rating Scale; DS = Digit Span (f = forward, b = backward); FCSRT = Free and Cued Selective Reminding Task (% ret. = % retention); GMI = General Memory Index; IMI = Immediate Memory Index; LM = Logical Memory subtest (I-immediate recall, II-delayed recall); MC = Mental Control subtest; MMSE = Mini-Mental State Examination; MQ = memory quotient; PA = Paired Associates subtest (rc. = recall, rec. = recognition); PSI = Processing Speed Index; TMT-A, TMT-B = Parts A and B of the Trailmaking Test; Stroop int. = Stroop Interference score; VF = Verbal Fluency; VR = Visual Reproduction subtest (I-immediate recall, II-delayed recall); WMI = Working Memory Index; WMS = Wechsler Memory Scale
Table 2. Average amount of shared area under the normal curve: mild AD clinical vs. control (or standardization) samples.

| Index       | n | % overlap (mean ± SD) | Range |
|-------------|---|-----------------------|-------|
| ALL IQ      | 17| 50.59 ± 19.21         | 9 - 88|
| VIQ/FSIQ    | 12| 53.58 ± 18.52         | 42 - 88|
| MEMORY      | 25| 27.24 ± 14.81         | 7 - 65|
| ATTENTION   | 9 | 52.22 ± 14.27         | 33 - 83|
| LANGUAGE    | 8 | 50.13 ± 18.52         | 44 - 47|
| OTHER       | 7 | 53.43 ± 13.15         | 42 - 81|

1Table entries are the number of juxtaposed reconstructed clinical versus control distribution pairs.

Table 3. Differences in average percentage of overlap: mild AD clinical vs. control (or standardization) samples1.

| Index       | ALL IQ | VIQ/FSIQ |
|-------------|--------|----------|
| MEMORY      | 23^2   | 26^3     |
| ATTENTION   | -2     | 1        |
| LANGUAGE    | 1      | 3        |
| OTHER       | -3     | 0        |

1Table entries are results of IQ index minus SCDT index (see Table 2).

^2t(1,41) = 4.44, p < .05

^3t(1,36) = -4.71, p < .05

Three studies (i.e., Haxby et al., 1990; Kirk & Kertesz, 1991; Petersen et al., 1999) reported results for the Dementia Rating Scale (DRS) and/or the Mini-Mental State Examination (MMSE). These are omnibus tests of cognitive functioning commonly used in the assessment and staging of dementia. Haxby et al. and Petersen et al. used control groups; Kirk and Kertesz did not. DRS norms from Schmidt et al., 1994 (reported in Spreen & Strauss, 1998) provided the comparison group for Kirk and Kertesz’s sample. The mean percent overlap value for the reconstructed omnibus distributions was 18%. This was 33% less than ALL IQ and 36% less than VIQ/FSIQ (both significant at p < .05). This result suggests that omnibus measures may be more efficacious than memory measures in differentiating patients with mild AD from normals; however, the number of
comparisons in the Omnibus index was quite small, thereby raising questions about the trustworthiness of the result.

The present results raise the question of whether memory SCDTs might also be efficacious in identifying patients with mild cognitive impairment (MCI). These patients are at increased risk for developing AD but do not yet meet diagnostic criteria for it (Jones & Ferris, 1999). This issue was addressed by reconstructing and analyzing memory and IQ distributions for a group of MCI patients and controls (Petersen et al., 1999). For the MCI group, the mean percent overlap was 86% for ALL IQ and 48% for MEMORY. Thus, there was 38% less clinical versus control overlap, on average, for the memory SCDTs than for the IQ measures. This result is not necessarily unexpected since MCI is typically defined by memory impairment in the context of relatively preserved functioning in other SCDs and in daily functioning. Nonetheless, the result supports the clinical utility of memory SCDEMs in MCI and suggests that memory SCD D-scores would be more effective, or much more effective, than IQ D-scores at detecting MCI-related cognitive decline.

Chronic Alcohol Abuse

Table 4 lists each of the selected CAA studies, summarizes the sample sizes, provides the means ± SD for each test selected for distribution reconstruction, and lists the percentage of shared area for each reconstructed distribution pair. Table 5 lists the mean ± SD percent overlap for each CAA index, and provides the number and range of the individual SCDT percent overlap values comprising them. The CAA OTHER index was comprised of the Grooved Pegboard Test and a verbal fluency test.
Table 6 summarizes the pairwise differences between the various SCDT and IQ indexes; the table entries are the differences between the IQ and SCDT index values (i.e., IQ index minus SCDT index). Inspection of Table 6 reveals negative differences between the IQ indexes and the MEMORY and ATTENTION indexes. This result suggests that memory and attention measures are no more sensitive, and possibly less sensitive to cognitive decline in CAA. Accordingly, the result raises the possibility that IQ D-scores are more efficacious than memory or attention SCD D-scores in detecting cognitive decline associated with CAA.

Table 6 also reveals that the OTHER index was significantly more sensitive to CAA than ALL IQ or VIQ/FSIQ. This result is difficult to interpret because of the discrepant abilities tapped by the tests comprising the index (the Grooved Pegboard Test and a verbal fluency measure), although mental or motor slowing and executive dysfunction could conceivably contribute to poor performance on both measures. Moreover, the index is comprised of a small number of percent overlap values, thereby calling into question the trustworthiness of the result.

Also, the mean percent overlap for the EXECUTIVE index was 9% and 12% lower than for ALL IQ and VIQ/FSIQ, respectively (see Table 6). These differences approached or met the 10% criteria, but were not statistically significant. This result was, however, in the expected direction given reports of impaired executive functioning in CAA (e.g., Parsons, 1996). It suggests quite tentatively that executive SCD D-scores might be more efficacious than IQ D-scores at detecting cognitive decline in CAA.
Table 4. Summary of CAA comparisons by study: sample sizes, reported means and SDs, and percent overlap values for each reconstructed distribution pair.

| Study                  | Test   | Comparison Groups                                | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|------------------------|--------|--------------------------------------------------|---------------|-------------|--------------|------------|---------------|
| Barron & Russell, 1992 | WAIS-R | CAA (n = 40) vs. control (n = 40)                |               |             |              |            |               |
|                        |        | FSIQ                                             | 88.65         | 11.48       | 100.75       | 13.14      | 62            |
|                        |        | VIQ                                              | 91.62         | 12.57       | 101.97       | 15.64      | 70            |
|                        |        | PIQ                                              | 85.02         | 11.24       | 98.47        | 10.71      | 54            |
|                        |        | SCDTs                                            |               |             |              |            |               |
|                        |        | TPT DH                                           | 13.75         | 5.74        | 7.79         | 3.91       | 52            |
|                        |        | TPT NDH                                          | 12.64         | 5.49        | 7.17         | 3.86       | 54            |
| Dao-Castellana et al., 1998 | WAIS    | CAA (n = 17) vs. control (n = 9)                |               |             |              |            |               |
|                        |        | FSIQ                                             | 95.00         | 11.00       | 107.00       | 10.00      | 57            |
|                        |        | SCDTs                                            |               |             |              |            |               |
|                        |        | Stroop errors                                    | 4.10          | 3.80        | 0.07         | 1.00       | 40            |
|                        |        | Stroop int time                                  | 131.00        | 29.00       | 93.00        | 18.00      | 20            |
|                        |        | VF                                               | 56.00         | 15.00       | 74.00        | 16.00      | 56            |
|                        |        | WMS MQ                                           | 98.00         | 10.00       | 109.00       | 7.00       | 65            |
| Jones & Parsons, 1971  | WAIS   | CAA (n = 40) vs. controls(n = 40)                |               |             |              |            |               |
|                        |        | FSIQ                                             | 102.90        | 8.72        | 99.70        | 13.32      | 78            |
|                        |        | SCDTs                                            |               |             |              |            |               |
|                        |        | CT errors                                         | 74.95         | 28.83       | 59.93        | 26.07      | 78            |
| Study/McLachlan, 1974 | Test | CAA (n = 22) vs. controls (n = 22) | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|----------------------|------|---------------------------------|---------------|-------------|-------------|-----------|--------------|
| Wechsler-Bellevue (WB) | FSIQ | 125.09 | 7.52 | 129.77 | 7.75 | 75 |
| | VIQ | 125.86 | 6.40 | 125.68 | 7.03 | 78 |
| | PIQ | 120.27 | 9.16 | 129.50 | 8.91 | 61 |
| SCDTs | CT errors | 57.32 | 25.08 | 41.91 | 24.69 | 76 |
| | SRT | 3.95 | 2.65 | 3.59 | 2.41 | 97 |
| | SSPT | 4.91 | 2.92 | 3.59 | 2.41 | 76 |
| | TPT location | 4.18 | 2.28 | 4.00 | 2.31 | 94 |
| | TPT memory | 7.55 | 1.47 | 7.55 | 1.26 | 94 |
| | TPT total time | 14.99 | 6.65 | 11.42 | 3.75 | 65 |
| | TMT-A | 9.00 | 1.51 | 9.77 | 0.69 | 61 |
| | TMT-B | 4.91 | 2.41 | 7.86 | 1.78 | 48 |
| | WB BD | 11.23 | 2.29 | 13.77 | 2.15 | 58 |
| | DSy | 9.95 | 1.92 | 12.59 | 1.85 | 49 |
| Omnibus Test | mod. HI II | 2.82 | 1.37 | 1.64 | 1.43 | 67 |
CCA comparisons (cont.)

| Study                  | Test | Comparison Groups | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|------------------------|------|-------------------|---------------|-------------|--------------|------------|---------------|
| Oscar-Berman et al., 1993 | WAIS-R | Young CAA (n = 8) vs. Young controls (n = 14) |               |             |              |            |               |
|                        |      | FSIQ              | 113.70        | 13.60       | 115.5        | 15.90      | 92            |
|                        |      | VIQ               | 115.40        | 16.40       | 115.20       | 18.30      | 95            |
|                        |      | PIQ               | 110.20        | 9.20        | 111.90       | 15.10      | 76            |
|                        |      | SCDTs             |               |             |              |            |               |
|                        |      | WMS-R AC          | 110.40        | 14.90       | 110.60       | 14.40      | 98            |
|                        |      | DMI               | 118.50        | 16.60       | 121.60       | 14.60      | 88            |
|                        |      | GMI               | 118.20        | 15.50       | 121.30       | 15.30      | 96            |
|                        | WAIS-R | Old CAA (n = 15) vs. Old controls (n = 15) |               |             |              |            |               |
|                        |      | FSIQ              | 109.50        | 15.40       | 114.80       | 14.70      | 86            |
|                        |      | VIQ               | 110.90        | 14.50       | 112.50       | 13.70      | 95            |
|                        |      | PIQ               | 105.50        | 14.90       | 115.30       | 16.50      | 75            |
|                        |      | SCDTs             |               |             |              |            |               |
|                        |      | WMS-R AC          | 105.30        | 15.60       | 10.950       | 15.60      | 89            |
|                        |      | DMI               | 106.60        | 16.60       | 112.50       | 18.40      | 79            |
|                        |      | GMI               | 128.40        | 17.00       | 124.70       | 19.90      | 90            |
| Study                        | Test          | Comparison Groups | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|-----------------------------|---------------|-------------------|---------------|-------------|--------------|------------|---------------|
| Psych. Corp., 1997          | WAIS-III      | CAA (n = 28) vs. std. |               |             |              |            |               |
|                             | FSIQ          |                   | 106.10        | 13.50       | 100.00       | 15.00      | 86            |
|                             | VIQ           |                   | 92.20         | 13.10       | 100.00       | 15.00      | 64            |
|                             | PIQ           |                   | 101.20        | 14.50       | 100.00       | 15.00      | 83            |
|                             | SCDTs         |                   |               |             |              |            |               |
|                             | WAIS-III PSI  |                   | 97.70         | 12.50       | 100.00       | 15.00      | 90            |
|                             | WMS-III IMI   |                   | 102.50        | 16.50       | 100.00       | 15.00      | 93            |
|                             | GMI           |                   | 105.20        | 14.20       | 100.00       | 15.00      | 86            |
|                             | VsII          |                   | 96.00         | 14.50       | 100.00       | 15.00      | 89            |
|                             | VsDI          |                   | 97.90         | 13.50       | 100.00       | 15.00      | 93            |
|                             | WMI           |                   | 98.00         | 9.30        | 100.00       | 15.00      | 77            |
| Smith & Burt, 1973          | WAIS          | CAA (n = 26) vs. controls (n = 26) |               |             |              |            |               |
|                             | FSIQ          |                   | 115.60        | 9.50        | 124.20       | 8.47       | 63            |
|                             | VIQ           |                   | 116.90        | 10.00       | 123.70       | 11.56      | 75            |
|                             | PIQ           |                   | 112.00        | 9.50        | 128.60       | 8.10       | 35            |
| Study                  | Test       | Comparison Groups                  | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|-----------------------|------------|-----------------------------------|---------------|-------------|--------------|------------|---------------|
| Smith & Burt (cont.)  | SCDTs      | CAA (n = unspec.) vs. controls (n = unspec.) | 56.80         | 24.90       | 40.70       | 15.60      | 67            |
|                       | CT errors  |                                   | 4.40          | 2.60        | 3.32        | 2.30       | 81            |
|                       | SRT        |                                   | 6.50          | 2.50        | 4.60        | 2.80       | 72            |
|                       | SSPT       |                                   | 29.00         | 10.20       | 30.40       | 7.60       | 85            |
|                       | TMT-A      |                                   | 66.50         | 18.40       | 66.10       | 21.00      | 96            |
|                       | TMT-B      |                                   | 3.30          | 1.90        | 3.70        | 1.80       | 91            |
|                       | TPT location|                                 | 6.80          | 0.90        | 7.00        | 1.20       | 94            |
|                       | TPT memory |                                 | 16.40         | 8.10        | 62.10       | 24.58      | 57            |
|                       | TPT total time |                             |               |             |             |             |               |
| Omnibus Tests         | HII        |                                   | 0.42          | 0.26        | 0.28        | 0.17       | 64            |
| Wilson et al., 1987   | WAIS       | CAA (n = 18) vs. controls (n = 20) | 114.00        | 10.70       | 117.00      | 11.30      | 89            |
|                       | FSIQ       |                                   | 113.80        | 11.60       | 114.80      | 11.60      | 96            |
|                       | VIQ        |                                   | 114.00        | 10.70       | 117.50      | 11.60      | 87            |
|                       | PIQ        |                                   |               |             |             |             |               |
|                       | SCDTs      |                                   | 32.80         | 3.60        | 35.20       | 1.30       | 52            |
|                       | ROCFT Copy |                                   | 14.80         | 5.60        | 21.50       | 5.20       | 54            |
|                       | ROCFT DR   |                                   | 9.80          | 2.50        | 11.90       | 2.50       | 68            |
|                       | WAIS BD    |                                   | 10.40         | 2.60        | 10.00       | 2.50       | 87            |
|                       | DS         |                                   | 5.20          | 1.10        | 5.60        | 1.10       | 72            |
|                       | WCST cat.  |                                   |               |             |             |             |               |
### CCA comparisons (cont.)

| Study                      | Test            | Comparison Groups | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|----------------------------|-----------------|-------------------|---------------|-------------|--------------|------------|---------------|
| Wilson et al. (cont.)      | WMS LM I        |                   | 8.90          | 2.30        | 9.80         | 2.40       | 85            |
|                            | LM II           |                   | 14.40         | 4.60        | 15.30        | 4.70       | 92            |
|                            | MQ              |                   | 111.10        | 12.60       | 116.20       | 15.40      | 83            |
|                            | PA I            |                   | 16.20         | 3.30        | 16.70        | 2.90       | 92            |
|                            | PA II           |                   | 9.20          | 0.90        | 9.40         | 1.00       | 91            |
|                            | VR I            |                   | 9.60          | 2.90        | 10.90        | 2.70       | 82            |
|                            | VR II           |                   | 7.30          | 3.50        | 9.90         | 3.30       | 70            |
| Yohman et al., 1985        | WAIS            | CAA (n = 37) vs.  | 14.40         | 4.60        | 11.10        | 12.60      | 83            |
|                            |                 | controls (n = 20) |               |             | 16.20        | 3.30       | 92            |
|                            | Prorated FSIQ   |                   | 104.00        | 11.70       | 112.00       | 11.70      | 73            |
|                            | SCDTs           |                   |               |             |              |            |               |
|                            | Pegs DH         |                   | 85.00         | 15.50       | 65.50        | 11.40      | 46            |
|                            | Pegs NDH        |                   | 93.00         | 76.10       | 29.50        | 10.70      | 53            |
|                            | TMT-B           |                   | 95.00         | 43.40       | 43.40        | 19.90      | 66            |
|                            | WCST cat.       |                   | 3.70          | 2.00        | 4.40         | 1.90       | 86            |
|                            | persever.       |                   | 12.20         | 8.60        | 8.50         | 8.10       | 83            |

SCDT abbreviations: AC = Attention/Concentration index; BD = Block Design; CT errors = Category Test error score; DMI = Delayed Memory Index; DS = Digit Span; GMI = General Memory index; IMI = Immediate Memory Index; mod. HII = modified Halstead Impairment Index; Pegs = Grooved Pegboard Test (DH = dominant hand; NDH = non-dominant hand); LM = Logical Memory (I = immediate recall; II = delayed recall); PA = Paired Associates (I = immediate recall; II = delayed recall); PSI = Processing Speed Index; ROCFT = Rey-Osterreith Complex Figure Test (DR = delayed recall); SRT = Seashore Rhythm Test; SSPT = Speech Sounds Perception Test; Stroop int. = Stroop interference; TMT-A, TMT-B = Parts A and B of the Trail Making Test; TPT = Tactile Performance Test (DH = dominant hand, NDH = non-dominant hand); VF = verbal fluency; VR = Visual Reproduction (I = immediate recall; II = delayed recall); VsDI = Visual Delay Index; VsII = Visual Immediate Index; WCST = Wisconsin Card Sorting Test (cat. = number of categories; persever. = perseverative errors).
Table 5. Average amount of shared area under the normal curve: CAA clinical vs. control (or standardization samples).

| Index         | n  | % overlap (mean ± SD) | Range |
|---------------|----|-----------------------|-------|
| ALL IQ        | 24 | 75.21 ± 15.06         | 35 – 96|
| VIQ/FSIQ      | 17 | 78.47 ± 12.67         | 57 – 96|
| MEMORY        | 19 | 85.05 ± 11.13         | 54 – 96|
| ATTENTION     | 8  | 79.50 ± 16.51         | 49 – 98|
| PCPT-ORG      | 13 | 70.54 ± 16.20         | 52 – 97|
| EXECUTIVE     | 12 | 66.55 ± 22.32         | 20 – 96|
| OTHER         | 3  | 51.67 ± 5.13          | 46 – 56|

1Table entries are the number of juxtaposed reconstructed clinical versus control distribution.

Table 6. Differences in average percentage of overlap: CAA clinical vs. control (or standardization) samples.

| SCDT Index   | ALL IQ | VIQ/FSIQ |
|--------------|--------|----------|
| MEMORY       | -10    | -7       |
| ATTENTION    | -4     | -1       |
| PCPT-ORG     | 5      | 8        |
| EXECUTIVE    | 9      | 12       |
| OTHER        | 24     | 27       |

1Table entries are rounded differences in the average percentage of clinical versus control overlap: IQ – SCDT (see Table 5).

Two of the CAA studies reported scores for the Halstead Impairment Index and for a modified Halstead Impairment Index. Being omnibus indexes of cognitive functioning, they were treated separately from SCDTs in the present analysis. On average, these indexes had 10% less overlap than ALL IQ and 13% less than VIQ/FSIQ, suggesting that they are more sensitive to CAA than IQ. However, the small number of comparisons renders the results rather tenuous.

3 Modified by Long and McLachlan (1974) to include the Trail Making Test, use of only 7 of the 10 tests comprising the HII, and a modified scoring procedure for the Finger Tapping Test.
Mild Traumatic Brain Injury

Table 7 lists each of the selected mTBI studies, summarizes the sample sizes, provides the means and SDs for each test selected for distribution reconstruction, and lists the percentage of area shared by each reconstructed distribution pair. Table 8 lists the mean ± SD percent overlap for each mTBI index, and provides the number and range of the individual SCDT percent overlap values comprising them. The mTBI OTHER index consisted of a Verbal Fluency measure and the Picture Completion subtest.

Table 9 summarizes the pairwise differences between the various SCDT and IQ indexes; the table entries are the differences between IQ and SCDT index values (i.e., IQ index minus SCDT index). Inspection of the Table 9 reveals that CONSTRUCTION and OTHER SCDT indexes were less sensitive to mTBI than either the ALL IQ or VIQ/FSIQ indexes. This suggests that construction SCD D-scores might be less effective than IQ D-scores in identifying cognitive decline in mTBI, although the small number of comparisons calls for a rather tentative conclusion. Interpretation of the OTHER SCDT index is difficult due to the divergent cognitive abilities it encompasses.

Attentional SCDTs were somewhat more sensitive to mTBI than IQ measures. The difference approached the 10% benchmark when PIQ was removed from the IQ comparison. This result allows the tentative conjecture that attention SCD D-scores might be more efficacious than IQ D-scores in detecting cognitive decline associated with mTBI. The results also suggest that attentional SCDTs share more variance with PIQ than with VIQ/FSIQ. That SCD D-scores tapping attention might improve detection of cognitive decline in mTBI is consistent with reports in the literature describing impairments in such functions, and in other functions with a co-dependency on
attentional capacities, such as working memory and processing speed (e.g., D. J. G. Williamson et al., 1996). None of the remaining SCD versus IQ comparisons approached the 10% criteria. Of interest, the relatively small differences that were obtained are not consistent with the marked effects sometimes assumed to occur within these cognitive domains in mTBI in comparison to changes in IQ.
Table 7. Summary of mTBI comparisons by study: sample sizes, reported means and SDs, and percent overlap values for each reconstructed distribution pair.

| Study                  | Test            | Comparison Groups                      | Clinical | Clinical | Control | Control | % Shared |
|------------------------|-----------------|----------------------------------------|----------|----------|---------|---------|----------|
|                        |                 |                                        | Mean     | SD       | Mean    | SD      | Area     |
| Bassett & Slater, 1990 | WAIS-R/WISC-R   | mCHI (n = 19) vs. controls (n = 29)    |          |          |         |         |          |
|                        |                 | FSIQ                                   | 93.10    | 15.70    | 110.20  | 17.30   | 60       |
|                        |                 | VIQ                                    | 95.00    | 14.90    | 108.30  | 18.50   | 68       |
|                        |                 | PIQ                                    | 92.80    | 14.90    | 109.70  | 15.20   | 58       |
|                        |                 | SCDTs                                  |          |          |         |         |          |
|                        |                 | TMT-A                                  | 32.80    | 13.90    | 27.10   | 11.90   | 82       |
|                        |                 | TMT-B                                  | 69.10    | 23.90    | 57.60   | 25.90   | 86       |
|                        |                 | VF                                     | 10.70    | 3.30     | 13.90   | 3.70    | 65       |
|                        |                 | WCST persev.                           | 14.40    | 11.60    | 8.40    | 6.60    | 68       |
|                        |                 | non-persev.                            | 8.50     | 8.40     | 4.80    | 3.70    | 65       |
|                        |                 | Cat.                                   | 5.60     | 1.00     | 5.80    | 0.70    | 85       |
|                        |                 | WMS LM I                               | 6.80     | 2.60     | 7.90    | 3.20    | 83       |
|                        |                 | LM II                                  | 5.10     | 2.60     | 6.20    | 3.00    | 84       |
|                        |                 | VR I                                   | 10.80    | 3.20     | 11.20   | 2.90    | 94       |
|                        |                 | VR II                                  | 9.80     | 3.50     | 10.90   | 2.90    | 85       |
| Corrigan et al., 1987  | WAIS-R          | CHII (n = 102) vs. std.                |          |          |         |         |          |
|                        |                 | VIQ                                    | 84.07    | 9.60     | 100.00  | 15.00   | 50       |
|                        |                 | PIQ                                    | 76.34    | 14.60    | 100.00  | 15.00   | 42       |
|                        |                 | SCDTs                                  |          |          |         |         |          |
|                        |                 | CT errors                              | 71.23    | 29.75    | 35.74   | 22.76   | 49       |
| Study                        | Test     | Comparison Groups                       | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|-----------------------------|----------|-----------------------------------------|---------------|-------------|--------------|------------|---------------|
| Johnstone et al., 1995      | WAIS-R   | CHI\(^1\) (n = 97) vs. strd.            |               |             |              |            |               |
|                             | FSIQ     | 93.42                                   | 10.18         | 100.00      | 15.00        | 75         |
|                             | VIQ      | 92.79                                   | 10.90         | 100.00      | 15.00        | 74         |
|                             | SCDTs    | WMS-R AI                                | 92.19         | 16.42       | 100.00       | 15.00      | 80            |
|                             |          | DMI                                     | 88.32         | 18.56       | 100.00       | 15.00      | 72            |
|                             |          | GMI                                     | 89.21         | 17.1        | 100.00       | 15.00      | 73            |
|                             |          | VbMI                                    | 91.33         | 14.87       | 100.00       | 15.00      | 77            |
|                             |          | TMT-A                                   | 40.47         | 22.81       | 23.80        | 6.80       | 38            |
|                             |          | TMT-B                                   | 99.48         | 55.16       | 57.70        | 16.60      | 47            |
| Psych. Corp., 1997          | WAIS-III | mod.-svr. CHI (n = 22) vs. strd.         |               |             |              |            |               |
|                             | FSIQ     | 86.50                                   | 10.90         | 100.00      | 15.00        | 59         |
|                             | VIQ      | 89.60                                   | 12.40         | 100.00      | 15.00        | 66         |
|                             | PIQ      | 84.50                                   | 13.80         | 100.00      | 15.00        | 59         |
|                             | SCDTs    | WAIS-III WMI                            | 89.80         | 13.10       | 100.00       | 15.00      | 71            |
|                             |          | PSI                                     | 73.40         | 10.70       | 100.00       | 15.00      | 30            |
|                             |          | WMS-III GMI                             | 81.90         | 16.50       | 100.00       | 15.00      | 50            |
|                             |          | IMI                                     | 78.90         | 17.70       | 100.00       | 15.00      | 52            |
|                             |          | VsDI                                    | 74.30         | 13.90       | 57.70        | 16.60      | 40            |
|                             |          | VsIMI                                   | 74.90         | 13.90       | 100.00       | 15.00      | 39            |
|                             |          | WMI                                     | 91.90         | 11.90       | 100.00       | 15.00      | 75            |
| Study            | Test | Comparison Groups                  | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|------------------|------|------------------------------------|---------------|-------------|--------------|------------|---------------|
| Uzzell et al., 1988 | WAIS | Mild-mod. CHI w/VFD² (n = 54) vs. strd. |               |             |              |            |               |
|                  |      | FSIQ                               | 97.10         | 15.43³      | 100.00       | 15.00      | 92            |
|                  |      | VIQ                                | 99.70         | 18.31       | 100.00       | 15.00      | 98            |
|                  |      | PIQ                                | 94.10         | 16.90       | 100.00       | 15.00      | 75            |
|                  |      | SCDTs                               |               |             |              |            |               |
|                  |      | WAIS BD                            | 9.90          | 3.67        | 10.00        | 3.00       | 90            |
|                  |      | DSy                                | 8.50          | 3.67        | 10.00        | 3.00       | 82            |
|                  |      | OA                                 | 8.70          | 2.94        | 10.00        | 3.00       | 83            |
|                  |      | PC                                 | 9.00          | 2.94        | 10.00        | 3.00       | 87            |
|                  |      | WMS DS(b)                          | 4.20          | 1.47        | 5.26         | 1.13       | 69            |
|                  |      | TMT-A                              | 49.60         | 30.13       | 25.37        | 9.17       | 37            |
|                  |      | TMT-B                              | 118.90        | 72.75       | 66.02        | 34.17      | 67            |
### Study Test

|                     | Comparison Groups | Clinical Mean | Clinical SD | Control Mean | Control SD | % Shared Area |
|---------------------|-------------------|---------------|-------------|--------------|------------|---------------|
| **WAIS**            | Mild-mod. CHI no VFD\(^2\) (n = 35) vs. strd. |               |             |              |            |               |
| FSIQ                |                   | 103.10        | 12.42\(^3\) | 100.00       | 15.00      | 88            |
| VIQ                 |                   | 101.70        | 12.42       | 100.00       | 15.00      | 90            |
| PIQ                 |                   | 104.30        | 13.02       | 100.00       | 15.00      | 87            |
| **SCDTs**           |                   |               |             |              |            |               |
| WAIS BD             |                   | 11.60         | 2.96        | 10.00        | 3.00       | 79            |
| DSy                 |                   | 10.10         | 2.96        | 10.00        | 3.00       | 99            |
| OA                  |                   | 11.40         | 2.96        | 10.00        | 3.00       | 82            |
| PC                  |                   | 10.50         | 2.96        | 10.00        | 3.00       | 92            |
| WMS DS(b)           |                   | 4.70          | 0.59        | 5.26         | 1.13       | 69            |
| TMT-A               |                   | 36.30         | 14.20       | 25.37        | 9.17       | 61            |
| TMT-B               |                   | 77.10         | 33.72       | 66.02        | 34.17      | 87            |

\(^{1}\)CHI group is mixed injury severity.  
\(^{2}\)VFD = visual field defect.  
\(^{3}\)SDs calculated algebraically from reported SEM

SCDT abbreviations: AI = Attention Index; CT errors = Category Test error score; DMI = Delayed Memory Index; DS(b) = Digit Span (b = backwards); DSy = Digit Symbol; GMI = General Memory Index; IMI = Immediate Memory Index; LM = Logical Memory (I = immediate, II = delayed); OA = Object Assembly; TMT-A, TMT-B = Parts A and B of the Trail Making Test; VF = verbal fluency; VsDI = Visual Delayed Index; VsIMI = Visual Immediate Memory Index; VR = Visual Reproduction (I = immediate, II = delayed); WCST = Wisconsin Card Sorting Test (cat. = categories; persever. = perseverative errors; non-persever. = non-perseverative errors); WMS = Wechsler Memory Scale
Table 8. Average amount of shared area under the normal curve: mTBI clinical vs. control (or standardization samples).

| Index         | \( n^1 \) | % overlap (mean ± SD) | Range |
|---------------|---------|------------------------|-------|
| ALL IQ        | 16      | 71.31 ± 16.35          | 42 - 98 |
| VIQ/FSIQ      | 11      | 74.55 ± 15.64          | 50 - 98 |
| MEMORY        | 11      | 68.36 ± 19.51          | 39 - 94 |
| ATTENTION     | 11      | 65.27 ± 21.86          | 30 - 99 |
| CONSTRUCTION  | 4       | 83.50 ± 4.65           | 79 - 90 |
| EXECUTIVE     | 8       | 69.25 ± 15.92          | 47 - 87 |
| OTHER         | 3       | 81.33 ± 13.36          | 65 - 92 |

\(^1\)Table entries are the number of juxtaposed reconstructed clinical versus control distribution.

Table 9. Differences in average percentage of overlap: mTBI clinical vs. control (or standardization) samples\(^1\).

| Index         | ALL IQ | VIQ/FSIQ |
|---------------|--------|----------|
| MEMORY        | 3      | 6        |
| ATTENTION     | 6      | 9        |
| CONSTRUCTION  | -12    | -9       |
| EXECUTIVE     | 2      | 5        |
| OTHER         | -10    | -7       |

\(^1\)Table entries are rounded differences in the average percentage of clinical versus control overlap: IQ – SCDT (see Table 2).
DISCUSSION

The present study was an initial feasibility test of a new approach to estimating prior cognitive abilities. The analysis examined the extent to which SCD D-scores might be more efficacious than IQ D-scores at detecting cognitive decline associated with particular disorders. The results could help to guide and prioritize the development of SCDEM for improving diagnostic accuracy in specific disorders.

Across the mild AD studies, the mean percent overlap was 23% to 26% less for memory SCDTs than for IQ, a result that achieved statistical significance. That is, there was 23% to 26% less overlap, on average, between the reconstructed clinical and control SCDT distributions than there was between the reconstructed clinical and control IQ distributions. Larger differences between the mean SCDT and IQ percent overlap values indicate greater SCDT sensitivity to the disorder. Accordingly, the result suggests that memory D-scores might be more efficacious than IQ D-scores for detecting cognitive decline in mild AD and at identifying individuals with the disorder. This finding is consistent with Schlosser and Ivison’s (1989) report that WMS MQ D-scores were superior to IQ D-scores for distinguishing AD patients from normal controls. The current results expand on Schlosser and Ivison’s findings by providing the first systematic demonstration across different patient samples of the potential superiority of memory D-scores versus IQ D-scores for identifying cognitive decline in mild AD. Moreover, the result exceeded the 10% benchmark set as a rough guideline for determining whether further development of SCDEM might be justified.

Some of the percent overlap values for memory SCDTs were quite small, thereby raising the question of whether dementia severity in some samples exceeded mild levels.
If so, the present result might provide an inflated estimate of the potential efficacy of memory D-scores. However, the finding that memory SCDTs had 38% less overlap than IQ measures, even among MCI patients, argues against a more negative interpretation.

Across the mild AD studies, the mean percent overlap was 33% to 36% percent less for omnibus measures than for IQ, although the limited number of individual comparisons in this analysis raises questions about the trustworthiness of the result. This caveat notwithstanding, the result indicates that omnibus measures were more sensitive than IQ to mild AD. The result probably follows from the wide net these measures cast in screening cognitive functioning. Because these measures tap cognitive functioning broadly, they are likely to capture impairment not only in memory but in other SCDs as well. Poor overall scores on these measures could be achieved via impairment in any SCD. Thus, although D-scores based on MMSE and DRS total scores may be useful for distinguishing patients with AD from normals, they may have less utility for distinguishing between AD and other neurocognitive disorders, especially when the pattern of cognitive impairment is a key differentiating feature.

The potential value of the strategy for identifying mild AD patients more efficiently may be considerable. For example, the present results showed that up to 26% more mild AD cases could be identified through the use of memory D-scores and appropriate cutoffs than might be identified using IQ D-scores. Given reported US incidence and prevalence rates for AD (Hebert et al., 1995), this result translates to annual potential identification of over 20,000 new mild AD cases that might otherwise remain undetected. Although this figure assumes certain optimal conditions, an increase in accuracy by 5% to 10% may well be realistic. Further, the MCI reanalysis
demonstrated the possibility of identifying patients in a prodromal phase of AD, thereby providing the opportunity to introduce treatments that might slow or prevent conversion to AD.

Across the CAA studies, executive measures showed 12% less overlap than VIQ and FSIQ combined, tentatively suggesting that executive D-scores might be more efficacious than IQ D-scores at identifying cognitive decline in CAA. Strong correlations (i.e., colinearity) between IQ and some of the selected executive SCDTs, especially the Category Test, might well have constrained this examination of differential effectiveness. Nonetheless, this result is consistent with research indicating diminished executive and abstract reasoning abilities in CAA patients (e.g. Parsons, 1996). In contrast, the finding that memory SCDTs were less sensitive than IQ to CAA was unexpected, especially in light of research indicating disproportionately decreased memory functioning in CAA (e.g., Parsons, 1996). The reason for this outcome is uncertain, but it might reflect some idiosyncrasies of the selected studies, or baseline IQ differences between the control and clinical subjects in these studies. Results suggesting that the combination of the Grooved Pegboard Test and a verbal fluency test were more sensitive than IQ to CAA are difficult to interpret because the two measures seem to tap different abilities. However, this result could reflect executive dysfunction (e.g., decreased initiation and impaired motor planning and sequencing) and/or generally reduced performance speed. Alternatively, the sensitivity of the combination of Grooved Pegboard and verbal fluency tests could be a reflection of both tests being valid yet non-redundant indicators of cognitive dysfunction in CAA.
The reconstructed clinical and control Halstead Impairment Index (HII) distributions showed 10% to 13% less overlap than the reconstructed IQ distributions. However, because HII is a composite index, it is unlikely to aid in differential diagnosis when the pattern of impairment is a distinguishing feature.

Although not always clearly specified, the severity of TBI across the selected studies appeared mild to moderate. Even so, the mean percent overlap values were generally similar for SCDs versus IQ, with discrepancies of less than 10%. In fact, constructional tests appeared less sensitive than IQ to mild to moderate TBI. For the TBI studies, attentional measures showed 6% to 9% less overlap than IQ. This result is congruent with previous reports of impaired attentional functions in mTBI (e.g., D. J. G. Williamson et al., 1996). However, the size of the difference is relatively small and suggests that attention D-scores would not provide much of an advantage over IQ D-scores at detecting cognitive decline in mTBI, at least for the attention SCDs included in the present study. The failure to find a more robust discrepancy between SCDs and IQ across the mTBI studies could partly reflect patient status, that is, many patients eventually recover following mild to moderate head injury and there simply may have been relatively small overall differences to detect.

Finding that memory D-scores might exceed IQ D-scores at detecting mild AD was expected: memory impairment is the cardinal feature of mild AD, and memory measures are the most sensitive psychometric indicators of the disorder, even in its earliest stages (e.g. Bondi et al., 1996). Still, almost all of the research on estimating prior ability -- even for detecting dementia -- has focused on estimating prior IQ, thereby suggesting that any such expectancy has not guided research. The current findings
suggest that research efforts aimed at improving AD detection may be more fruitful if
directed at developing memory SCDEMs, rather than at prior IQ estimation. Indeed, it
seems likely that more exact and refined methods than were used here might well
improve the sensitivity of memory D-scores. For example, predictive variables, or more
exactly, postdictive variables could be identified and combined through much more exact
means, such as discriminative function analysis and multiple regression.

In contrast to the mild AD results, it was unexpected that memory SCDTs would
be found to be less sensitive than IQ to CAA, and that attention SCDTs would turn out to
be only very slightly more sensitive than IQ to mTBI. These counterintuitive results
indicate that it may be difficult to identify accurately or optimally the differential
cognitive effects and magnitude of impairment that best characterizes a disorder and
differentiates it from others. The method used in the present study provides a viable
method for characterizing these patterns. That is, by refining and extending the method
used here, (i.e., calculating and comparing percent overlap values across multiple studies
of a disorder) it might be possible to identify the dimensions and levels of functioning
that best characterize disorders. Knowing these dimensions and levels and applying them
to the formation of properly developed decision procedures could help considerably in
discriminating between those with and without a particular disorder. When combined
with prior test results or with estimates of prior SCD standing, such information could
provide an important key to accurate diagnosis.
Limitations

The limited support for the strategy obtained in the CAA and mTBI analyses could reflect a number of methodological limitations. First, colinearity between IQ and some SCDT measures indicates that they may be, in large part, measuring the same function(s). Such colinearity likely limits the possibility of achieving or detecting a difference in sensitivity to disorder. Second, the tests used in the selected studies might not have been the most sensitive to CAA and mTBI. For example, the present results provided some very limited evidence that the Grooved Pegboard Test might be particularly sensitive to CAA, but only one study used this test. Also, a more difficult test of complex attention and information processing, such as the PASAT, might have been better able to separate mTBI and control groups. Third, the overall level of overlap in the SCDT and IQ distribution pairs across the mTBI and CAA studies was fairly high (see tables 5 and 8). Although this may be a unique feature of the selected studies, it may also reflect a lack of strong SCDT and IQ performance decrements in these disorders. Fourth, although studies were screened for the presence of adequately diagnosed clinical samples, the results are, nevertheless, constrained by the appropriateness of the subject selection procedures used in the original studies. For example, the inclusion of non-demented subjects or subjects with other types of dementia could have affected the outcome of the mild AD analysis. Similarly, the inclusion of mTBI subjects who were nearly recovered from their injuries could have decreased effects.

The lack of more robust support for the strategy in the CAA and mTBI analyses does not discredit the strategy for improving detection of cognitive decline in particular disorders. Indeed, formal selection and combination of the variables that based on
comparison of predicted and obtained performance, optimally discriminate between those with and without a disorder may improve diagnosis considerably. Accordingly, the strategy might have some utility in CAA and in mTBI if variables with more discriminative power, either singly or in combination, were identified. Moreover, the strategy may be useful in diagnosing other disorders that disrupt cognitive functioning. The CAA and mTBI results do suggest, however, that efforts to improve IQ estimation should not be broadly or completely abandoned in favor of developing SCDEM.

The ranges of percent overlap values comprising the SCD and IQ indexes appear to have been quite large, suggesting that some of the values may have been outliers (see tables 2, 5, and 8). This may reflect considerable variability across studies of the same disorder in terms of patient characteristics. However, given the relatively small number of individual percent overlap values comprising these indexes, and the exploratory nature of this project, all percent overlap values were retained in forming the indexes. This decision could have masked, or attenuated certain effects. However, wide ranges of percent overlap values occurred for both IQ and SCD indexes, suggesting that including all of the values probably did not systematically work for or against the hypothesis.

Normative samples were sometimes used as a substitution for control groups. Given sufficient demographic similarity, normative samples should provide a reasonable representation of the clinical group’s prior ability level. However, this procedure is suboptimal because differences may exist between any given selected sample and a normative sample. For example, contamination could have occurred between the demographic variables used to form the normative groups and the outcome variables in
the present study. Such effects may be difficult to detect, however, and it seems unlikely
they would have consistently altered results in a particular direction, even if present.

All of the reconstructed distributions were normal, and it was not possible to
account for skewness, or ceiling and floor effects, in the original data. When ceiling and
floor effects were present in the original data, the reconstructed distributions were
truncated, accordingly, at either end of score range. This truncated effect was seen in
minority of cases suggesting that it probably did not affect the results greatly. Further,
ceiling and floor effects generally indicate that the clinical and control groups would have
performed even more disparately if the test contained sufficiently easy or difficult items.
Consequently, ceiling and floor effects seem unlikely to have caused systematic
underestimation of percent overlap, in turn spuriously supporting the strategy, although
firm conclusions are difficult without the original raw test data.

Although the methods followed in this study were imprecise and often involved
estimations and approximations, they seem to have been sufficient to achieve the
investigative aims. It is not unusual, in exploratory work, both across the soft and hard
sciences, to set up approximations, make educated guesses, and proceed in the face of
considerable ambiguity. Certainly, the proposed strategy could be tested more rigorously
by developing and using actual SCDEMgs. However, such an endeavor was well beyond
the scope of this particular project, especially given the general lack of SCDEMgs and the
need for a broader initial analysis.

One curious and remarkably consistent result was that SCOT indexes were 3%
more sensitive than VIQ/FSIQ as opposed to ALL IQ in all but two comparisons. This
implies that SCOTs consistently share slightly greater variance with PIQ than with either

75
VIQ or FSIQ. The result is generally consistent with previous reports that PIQ may be more sensitive to brain damage than VIQ (Gouvier et al., 1983). No particular source of methodological artifact could be identified that could have produced this result.

Summary

The results of this initial feasibility study provide relatively strong support for the proposed strategy for one of the three disorders, specifically, for the development of memory SCDEMs for detecting mild AD. The main result suggests that memory D-scores might be more efficacious than IQ D-scores at detecting cognitive decline in mild AD, and at improving detection of the disorder. The results of the CAA and mTBI analyses provide less, or considerably less, support for the strategy. For these two conditions, SCDEMs, at least in the cognitive domains, or via the composites examined here, might not meaningfully improve diagnostic accuracy beyond properly derived IQ D-scores. The lack of robust support for the strategy in CAA and mTBI might reflect several factors, including, but not limited to colinearity between IQ and some of the SCDTs selected for distribution reconstruction. The CAA and mTBI results do not discredit the strategy, as it may well have utility under different circumstances, or with different disorders. Moreover, the strategy and the method used to test its feasibility may help prioritize and guide the development of SCDEMs in other neurocognitive disorders. Even if turns out that SCDEMs cannot be viably developed, the current method could help to identify which cognitive variables, in which combination, and with which D-score cutoffs best distinguish among neurocognitive disorders. This could lead to more accurate characterization of disorders and could be key to improving diagnostic accuracy in certain areas of neuropsychology.
Implications and Possible Future Directions

Diagnostic accuracy is often essential to treatment and care planning and may serve to improve patient outcomes, lessen caregiver stress, decrease financial strain on patients and their families, and reduce public health care burden. The strategy developed and tested here provides a potential means of improving diagnostic accuracy, at least for mild AD. This is important because early detection of AD may permit implementation of newly available treatments at a stage of illness where they are likely to be most effective at slowing cognitive decline.

Determining the potential clinical utility of SCD D-scores more accurately requires development and testing of actual SCDEMs. The current results suggest that priority might be placed, at least initially, on the development of memory SCDEMs. Results of previous studies offer a potential guide for developing viable SCDEMs (e.g., Crawford et al., 1992; Crawford et al., 1998; Hawkins et al., 1993; Schlosser & Ivison, 1989). The first step would involve a search for a set of variables that are strongly related to SCD performance, yet are relatively unaffected by the disorder in question. Once identified, these variables could be combined via multiple regression. The validity of SCDEMs could be initially and tentatively assessed using cross-sectional designs; however, retrospective designs would be necessary to firmly test their validity.
APPENDIX A

Characteristics of the Selected Studies

Mild Alzheimer’s Disease Studies

Bigler, Hubler, Cullum, and Turkheimer (1985) obtained brain CT scans from patients with early AD. Based on these scans, they estimated ventricular volume and calculated an index of cerebral atrophy. They examined the relationship between these measurements and WAIS and WMS performance. There were 42 subjects, 23 men and 19 women, with a mean education level of $13.1 \pm 3.5$ years and a mean age of $67.9 \pm 9.9$ years. There was no control group. Although not specified in the report, it appears that subjects were drawn from university medical centers in western United States. Bigler et al. (1985) reported WAIS IQ indexes and WMS MQ only. In the present study, reconstructed distributions for Bigler et al.’s sample were compared with reconstructed distributions based on the WAIS old age standardization sample and with the standardization sample for the WMS. The patients in Bigler et al.’s sample were reasonably similar to WAIS old age standardization sample in terms of age and education, although they were older than the WMS standardization sample. This is not expected to have impacted the results significantly since MQ is age corrected and because Bigler et al. reported MQ scores, not raw scores.

Botwinick, Storandt, and Berg (1986) followed 18 subjects with mild AD for four years. Subjects’ annual performance on 16 cognitive tests was compared with that of 30 control subjects matched on age, sex, and socioeconomic status. In the present study, distributions were reconstructed based on the means and SDs reported from the first of the annual evaluations. WAIS FSIQ scores were prorated from reported means and SDs.
for the WAIS Information, Comprehension, Digit Symbol, and Block Design subtests. Regarding potentially misleading overlap values and possible cut points in the reconstructed (normal) distributions: 78% of the mild AD but only 15% of the control groups fell below a VF score of 22 words; 75% of the mild AD but only 8% the control groups fell below a score of 58 on the 85-item BNT; and 49% of the mild AD and only 8% of the control groups fell below a Benton Visual Retention Test copy score of 8.

Haxby et al., (1990) conducted a longitudinal study of regional cerebral metabolic rates and neuropsychological functioning in 11 patients with mild AD and 29 controls matched on age, sex, and education. Mean follow-up duration was 26 months. In the present study, distributions were reconstructed based on the reported means and SDs from the initial evaluation. Regarding potentially misleading overlap values and possible cut points in the reconstructed (normal) distributions, 61% of the mild AD group but only 5% of the control group scored above 149 seconds on the TMT-B.

Kirk and Kertesz (1991) compared the spontaneous drawings of 38 patients with probable AD with those of 39 controls. The groups did not differ significantly in terms of age or education. The severity of dementia in the AD group was not specified but given their capacity to complete the research test battery, were probably mild-to-moderately impaired, not severely impaired. The control and AD patients' drawings were compared, however, the control group's IQs and MQ were not reported. Therefore, for the present study, the reconstructed mild AD IQ and MQ distributions were compared to reconstructed distributions based on the WAIS-R and WMS standardization samples. The AD group appeared to be reasonably similar to the standardization samples demographically. Distributions for the drawing scores were not reconstructed because the
variables on which they were scored seemed unique to this study. The authors also administered the DRS to patients and this was treated separately.

The Psychological Corporation (1997) conducted a number of small studies aimed at characterizing the WAIS-III/WMS-III performance of patients with various diagnoses including mild AD, CAA, and TBI. The WAIS-III/WMS-III was administered to 35 mild AD subjects with a mean age of 72.2 years (SD = 7.8). This group was better educated than the same-aged subgroup of the standardization sample: Most (48.6%) of the mild AD subjects had at least 16 years of education, whereas only 14% of the same-aged standardization sample fell in this education range. The higher education level of the mild AD group likely did not unfairly favor the research hypothesis. Lower versus higher education levels are generally associated with higher rates of dementia and with relatively greater impairments at comparable levels of pathological process than (Bondi et al., 1996). Consequently, patients with comparable levels of mild AD who have more education would be expected, on average, to score better on cognitive tests than those with average or below average education levels. Given their higher education level, one might expect that the prior IQ level of this particular mild AD sample was somewhat higher than that of the standardization sample mean of 100 and SD of 15. Accordingly, it seems unlikely that comparing the reconstructed mild AD distributions to the WAIS-III/WMS-III standardization sample's mean and SD would have biased the results toward a spurious conclusion of exaggerated decline in the mild AD group.

For the Psychological Corporation (1997) study, distributions were reconstructed for the WAIS-III VIQ, PIQ and FSIQ indexes. The WAIS-III Working Memory Index (WMI) was omitted in favor of the WMS-III WMI, because it included a measure of non-
verbal attention span that is not included in the WAIS-III WMI. Moreover, examination of the mild AD group’s reported WAIS-III WMI and WMS-III WMI means and SDs revealed that the latter was a more sensitive measure. The VCI and POI indexes were omitted because examination of the reported mild AD means and SDs revealed that these indexes were not as sensitive to mild AD as was the PSI index. Inspection of the reported means and SDs for the mild AD group clearly indicated performance below or far below the standardization group on most WMS-III indexes.

Petersen et al. (1999) characterized patients with mild cognitive impairment (MCI) using a combined cross-sectional and longitudinal design. They compared performance on neuropsychological tests for 234 controls, 76 patients with MCI, and 106 patients with either very mild AD or mild AD. Scores on the Clinical Dementia Rating scale for these groups were 0.5, 0.5, and 1.0 for MCI, very mild, and mild AD, respectively. Patients were assigned to groups using diagnostic consensus. However, the basis for differentiating the very mild AD patients from the MCI patients was unclear. For the present study, distributions were reconstructed for the 48 subjects with very mild AD, a decision that reflected the concern in the present study with improving methods for early detection of AD. The very mild AD group was younger and less educated than the control group. Although these differences were statistically significant, their magnitude was relatively small (very mild AD group: mean age = 75.6, mean education = 12.5, control group: mean age = 79.8, mean education = 13.3). Rather than SDs, Petersen et al. reported standard errors of the mean, which were converted to SDs for the purposes of the present study. The percent overlap values for the reconstructed distributions are listed in table 1.
Using discriminant function analysis, Storandt, Botwinick, Danziger, Berg, and Hughes (1984) developed a brief, 10-minute, battery of four cognitive tests to aid differentiation of patients with mild AD from healthy controls. The sample included 32 mild AD and 32 control subjects matched on the variables of age, sex, and social position. The percent overlap values for the reconstructed distributions are listed in table 1.

Weingartner et al. (1981) compared the cognitive test performances of 14 patients diagnosed with progressive idiopathic dementia (PID) thought to be in the earliest stages of AD, to that of 14 normal controls matched on age, education, and socioeconomic status. Weingartner et al. reported means and SDs for WAIS FSIQ and PIQ and for WMS MQ. The percent overlap values for the reconstructed distributions are listed in table 1.

**Chronic Alcohol Abuse Studies**

Barron and Russell (1992) sought to determine whether a common alcoholic WAIS pattern resulted from right hemisphere damage and whether it could be characterized as loss of fluid intelligence. They compared patients with either right hemisphere damage, left hemisphere damage, or alcoholism, to normal controls matched on age and education. There were 40 subjects in each group. The alcoholic subjects were inpatients at least 35 years old who had been drinking heavily for at least 20 years. At the time of their participation in the study, all subjects had been abstinent for over 2 weeks following detoxification.

Using neuropsychological tests and PET imaging, Dao-Castellana et al. (1998) investigated the possible presence of frontal dysfunction in neurologically normal, chronic alcohol abusers. Subjects were 17 chronic alcohol abusers and nine same-aged
controls. The CAA group had a 13 ± 9-year history of drinking, with an average recent daily alcohol consumption of 243 ± 126 grams. The control group had a significantly higher education level than the CAA group, which authors thought might reflect the reportedly poor social and professional adaptation of the CAA patients. The lower education level of the CAA group may have biased results in favor of the present study because education is positively correlated (more or less strongly) with several cognitive tests. However, examination of the percent overlap values (table 4) suggests that this potential biasing effect had minimal impact on the overall results. The only IQ index reported by the authors was WAIS FSIQ and the only memory measure was WMS MQ. Also reported were VF and Stroop interference and error scores. Of note, 14% of the CAA subjects and 24% of the control subjects fell below a Stroop error score of zero. Also, about 61% of the CAA subjects and only 1% of the control subjects had Stroop error scores greater than three.

Jones and Parsons (1971) investigated abstraction ability in matched groups of CAA, brain damaged, and control subjects (n = 40, each group). The alcoholic subjects averaged 7.23 years of heavy drinking. They were tested after withdrawal symptoms had subsided, with an average interval between admission and testing of 42.43 days. WAIS FSIQ mean and SD is reported for the control group. For the CAA group, WAIS FSIQ was estimated using the Shipley-Hartford test. The Category Test was used to measure abstraction ability.

Long and McLachlan (1974) compared matched controls and alcoholic patients on multiple cognitive tests and on a general index of cerebral dysfunction (Modified Halstead Impairment Index). The alcoholic subjects had been drinking heavily for about
9 years. All had been detoxicated at the time of testing. The average duration from admission to time of testing was 11.41 days.

Oscar-Berman, Clancy, and Weber (1993) evaluated discrepancies between IQ and memory performance in alcoholic men. Subjects were 59 men divided into four groups: young and old alcoholics, and young and old normal controls. All alcoholic subjects had been drinking for at least 5 years, and had been abstinent for at least 4 weeks prior to testing. All participants were from similar socioeconomic backgrounds and did not differ on level of education. The authors compared the groups on the WAIS and WAIS-R IQ indexes and on the WMS and WMS-R. For the purposes of the present study, only the more contemporary WAIS-R and WMS-R indexes were compared.

The Psychological Corporation (1997) reported on the WAIS-III/WMS-III performances of 28 CAA patients with a mean age of 53.3 ± 10.2 years. The duration and severity of alcohol abuse in the CAA subjects was not specified; all were detoxified prior to testing, although the time between detoxification and testing was not specified. Overall, the CAA group was better educated than the same-aged subgroup of the standardization sample; however, most subjects in both groups had 12 years of education. Distributions were reconstructed for VIQ, PIQ, FSIQ, and the PSI factor score. Distributions were also reconstructed for the WMS-III IMI, GMI, WMI, and for the Visual Immediate and Visual Delay Indexes.

Smith, Burt, and Chapman (1973) sought to determine if CAA patients from middle and upper socioeconomic backgrounds showed the same pattern of neuropsychological impairment found in two previous studies of CAA patients from lower socioeconomic backgrounds. Subjects were males aged 35-55 years admitted to a
private hospital for treatment of alcoholism. The sample size could not be determined. The duration and severity of alcohol abuse was not specified. All patients had been detoxified prior to testing, but the duration of abstinence was not specified. Subjects were reported to have above average IQ and education levels. A control group of similar age and education, but with even higher IQs, formed the comparison group.

Wilson, Kolb, Odland, and Wishaw (1987) compared patients with CAA (n = 49) to a control group (n = 60) and to groups of patients discrete unilateral frontal, temporal, or parietal lesions on a series of neuropsychological tests. The CAA and control groups were comparable in terms of average age (41.4 and 41.9 years for CAA and controls, respectively). The control group had a higher education level (11.8 years vs. 10.2 years for the CAA group) – a small but statistically significant difference (based on t-tests conducted by the present author). This difference could have favored the research hypothesis in the present study. The CAA subjects had a 5-to-20-year history of alcohol abuse; all were abstinent for more than 2 months at the time of testing.

Yohman, Parsons, and Leber (1985) compared 37 middle-aged alcoholics to 20 non-alcoholic controls matched on age and education. Neuropsychological tests were given 7 weeks and 13 months after detoxification. For the present study, test results from the first assessment were selected for distribution reconstruction because some subjects reportedly resumed drinking at some point prior to the 13-month follow-up testing. The mean alcohol consumption of the CAA group was 9.0 ounces for 13 years. For this study, IQ scores were prorated from reported WAIS Comprehension, Digit Symbol, Digit Span, and Block Design subtests.
Mild Traumatic Brain Injury Studies

Bassett and Slater (1990) evaluated 19 adults with mTBIs and 10 with severe TBIs. Distributions were reconstructed only for the mild injury and for matched control groups. The average duration between the time of the injury and testing was 3 weeks. IQ was measured either with the WAIS-R or WISC-R. The percent overlap values for the reconstructed distributions are presented in Table 7. Of note, 43% of the mild CHI group and 8% of the control group produced a non-perseverative error score of 10 or greater on the WCST.

Corrigan, Agresti, and Hinkeldy (1987) sought to re-examine and extend previous research on the psychometric properties of the Halstead Category Test (CT). Subjects were drawn from a rehabilitation setting that included patients with closed head injury, likely rather severe in nature. The mean age of the CHI group was 27.70 ± 10.03. The authors did not use a control group. Therefore, for the present study, the CT and IQ performances of the CHI group were compared to that of normative samples. Dodrill’s 1987 norms (reported in Mitrushina et al., 1999) were used for comparison on the CT because this sample was reasonably similar to Corrigan et al.’s in terms of age and education. Corrigan et al. reported results only for the WAIS-R PIQ and VIQ indexes, not for FSIQ.

Johnstone, Hexum, and Ashkanazi (1995) administered the WRAT-R Reading subtest (WRAT-R) as an estimate of prior overall cognitive ability. They then subtracted WRAT-R z-scores from cognitive test z-scores, thereby obtaining a difference z-score that were intended to estimate decline in intelligence and SCDs following TBI. Subjects were 97 TBI patients referred as outpatients for neuropsychological evaluation. Subjects’
mean age was 33.24 ± 1.28 years; their mean education level was 12.59 ± 2.39 years. According to the authors, lack of information precluded a grading of injury severity. There was no control group, therefore, reconstructed distributions for the WAIS-R (only VIQ and FSIQ were reported) and WMS-R were compared with the tests’ standardization samples. Means and SDs were also reported for the TMT-A and TMT-B. Johnstone et al. used norms published by Fromm-Auch and Yeudall (1983) in their conversion of TMT-A and TMT-B raw scores to z-scores. Accordingly, these norms were used for comparison with the reconstructed TMT distributions of the TBI group. Distributions were not reconstructed for the WRAT-R since it was used to estimate prior IQ. As such, the authors must have expected performance on it to be relatively preserved following TBI. Of note, 59% of the TBI group and 13% of the standardization group fell above a score of 35 seconds on the TMT-A; 82% of the clinical group and 6% of the standardization group fell above a score of 82 seconds on the TMT-B.

The Psychological Corporation (1997) published results on WAIS-III and WMS-III performances of 22 TBI patients with moderately-to-severely injured TBI (closed head injury). Subjects had a mean age of 26.9 ± 11.5 years. The duration between time of injury and testing was 6-18 months. Reconstructed TBI distributions were compared to the WAIS-III/WMS-III standardization sample. This was an appropriate comparison given reasonable demographic congruity between the clinical and standardization samples. Distributions were reconstructed for the three main IQ indexes and for the PSI. For the WMS-III, distributions were reconstructed for WMI, IMI, and GMI. Additionally, distributions were reconstructed for the Visual Immediate and Visual Delay indices given
results suggesting that these indices might be very slightly more sensitive to TBI than some of the other WMS-III indices.

Uzzell, Dolinskas, and Langfitt (1988) studied the impact of visual field defects (VFDs) as sequelae of head injury on neuropsychological functioning. Subjects were 159 head-injured patients classified into four groups on the basis of the presence or absence of VFDs and according to injury severity (minor-to-moderate or severe). Uzzell et al. did not use a control group. Distributions were reconstructed for the minor-to-moderate head injured groups only. The percent overlap values for the reconstructed distributions are reported separately for the VFD and non-VFD groups in Table 7. Reconstructed distributions were compared to the WAIS and WMS standardization samples, given reasonable demographic congruity between these samples and Uzzell et al.’s. Also, given reasonable demographic congruity between samples, Dodrill’s 1987 norms (reported in Mitrushina et al., 1999) were used as the comparison group for the TMT-A and TMT-B. Of note, 63% of the VFD group and only 5% of the standardization sample fell above a TMT-A score of 40 seconds; 55% of the VFD group, whereas only 11% of the standardization distribution fell above a TMT-B score of 108 seconds.
APPENDIX B

Selected Reconstructed Distribution Graphs

(for illustration)
(Botwinick et al., 1986) Trails A (inverted scores): mAD vs. controls

Y axis = (π% area of each score band × 100)/2

Trails A inverted scores (37.91% shared area)

mAD

ctrl
(Psych. Corp., 1997) WAIS-III VIQ: mAD vs. strd

WAIS-III VIQ score
(77.02% shared area)
(Psych. Corp., 1997) WMS-III General Memory Index:
mAD vs. strd
(Storandt et al., 1994) Associates Learning-recognition: mAD vs. control

Graph showing the comparison of associates learning-recognition scores between the mAD and control groups. The y-axis represents the percentage of area under the curve for each group, with a shared area of 41.08%. The x-axis represents the associative learning recognition scores.
(Dao-Castellana et al., 1998) Stroop errors: ETOH vs Controls

- ETOH
- Controls

n=100, both groups

y axis = (% area of each score band X 100)/2

Stroop errors

(39.66% shared area. 61.41% of ETOH distribution lies beyond +.99 SD of control group)
(Psych. Corp., 1997) Immediate Memory Index: ETOH vs. standardization

n=100, both groups

y axis = % area of each score band X 100/2

WMS-III IMI scores
(92.60% shared area)
(Smith & Burt, 1973) TPT Total Time: ETOH vs control

**Graph Description:**
- The graph compares TPT (Total Propagation Time) between two groups: ETOH and control.
- The x-axis represents TPT total time, and the y-axis represents an index of the area of each score term X 1000.
- The graph shows two curves: one for ETOH and another for control.
- The area under the curve for ETOH is 57.32% of the area under the curve for the control group.
- The graph indicates that ETOH has a higher total time compared to the control group.

**Legend:**
- **ETOH**
- **control**
(Yohnman et al., 1985) Pegboard Non-Dominant Hand: ETOH vs. control

n=100, both groups

y axis = % area of each score band X 100/2

Pegboard time - non-dominant hand
(53.47% shared area)
(Uzzell et al., 1988) WAIS Block Design: Mild-Med CHI with Visual Field Defects vs. Strd

![Graph showing WAIS Block Design scores]

- CH
- strd

n=100, both groups

y axis = (% area of each score band x 100)/2

WAIS BD scores

(90.43% shared area)
Albert, M. S., & Moss, M. B. (1992). The assessment of memory disorders in patients with Alzheimer's disease. In L. R. Squire & N. Butters (Eds.), Neuropsychology of memory (2nd ed., pp. 211-219). New York: Guilford.

Allen, D. N., & Landis, R. K. B. (1997). Neuropsychological correlates of substance use disorders. In P. J. Snyder & P. D. Nussbaum (Eds.), Clinical neuropsychology: A pocket handbook for assessment. (pp. 591-612). Washington, DC: American Psychological Association.

Alzheimer's Association. (1999, March). Understanding Alzheimer's: Statistics/Prevalence, [website]. The Alzheimer's Association. Available: http://www.alz.org. (accessed March 1999).

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. Journal of Consulting and Clinical Psychology, 49, 323-330.

Axelrod, B. N., Vanderploeg, R. D., & Rawlings, D. B. (1999). WAIS-R prediction equations in patients with traumatic brain injury. Journal of Clinical and Experimental Neuropsychology, 21, 368-374.

Axelrod, B. N., Vanderploeg, R. D., & Schinka, J. A. (1999). Comparing methods for estimating premorbid intellectual functioning. Archives of Clinical Neuropsychology, 14, 341-346.
Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology, 32*, 55-65.

Barona, A., & Chastain, R. L. (1986). An improved estimate of premorbid IQ for blacks and whites on the WAIS-R. *International Journal of Clinical Neuropsychology, 8*, 169-173.

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

Barrett, J. J., Haley, W. E., Harrell, L. E., & Powers, R. E. (1997). Knowledge about Alzheimer's disease among primary care physicians, psychologists, nurses, and social workers. *Alzheimer's Disease and Associated Disorders, 11*, 99-106.

Barron, J. H., & Russell, E. W. (1992). Fluidity theory and neuropsychological impairment in alcoholism. *Archives of Clinical Neuropsychology, 7*, 175-188.

Bassett, S. S., & Slater, E. J. (1990). Neuropsychological function in adolescents sustaining mild closed head injury. *Journal of Pediatric Psychology, 15*, 225-236.

Beardsall, L. (1998). Development of the Cambridge Contextual Reading Test for improving the estimation of premorbid verbal intelligence in older persons with dementia. *British Journal of Clinical Psychology, 37*, 229-240.
Beardsall, L., & Brayne, C. (1990). Estimation of verbal intelligence in an elderly community: A prediction analysis using a shortened NART. *British Journal of Clinical Psychology*(29), 83-90.

Beardsall, L., & Huppert, F. (1997). Short NART, CCRT and Spot-the-Word: Comparisons in older and demented persons. *British Journal of Clinical Psychology*, 36, 619-622.

Beardsall, L., & Huppert, F. A. (1994). Improvement in NART word reading in demented and normal older persons using the Cambridge Contextual Reading Test. *Journal of Clinical and Experimental Neuropsychology*, 16, 232-242.

Berkowicz, J., Ewen, R. B., & Cohen, J. (1976). *Introductory statistics for the behavioral sciences* (2nd ed.). New York: Academic Press.

Berry, D. T. R., Carpenter, G. S., Campbell, D. A., Schmitt, F. A., Helton, K., & Lipke-Molby, T. (1994). The New Adult Reading Test-Revised: Accuracy in estimating WAIS-R IQ scores obtained 3.5 years earlier from normal older persons. *Archives of Clinical Neuropsychology*, 9, 239-250.

Bigler, E. D., Hubler, D. W., Cullum, C. M., & Turkheimer, E. (1985). Intellectual and memory impairment in dementia: Computerized axial tomography volume correlations. *Journal of Nervous and Mental Disease*, 173, 347-352.

Bishop, E. G., Dickson, A. L., & Allen, M. T. (1990). Psychometric intelligence and performance on selective reminding. *Clinical Neuropsychologist*, 4, 141-150.
Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist, 3*, 129-136.

Bolter, J., Gouvier, W., Veneklasen, J., & Long, C. J. (1982). Using demographic information to predict premorbid IQ: A test of clinical validity with head trauma patients. *Clinical Neuropsychology, 4*, 171-174.

Bondi, M. W., Salmon, D. P., & Kaszniak, A. W. (1996). The neuropsychology of dementia. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (2nd ed., pp. 164-199.). New York: Oxford.

Bornstein, R. A., & Chelune, G. J. (1988). Factor structure of the Wechsler Memory Scale-Revised. *Clinical Neuropsychologist, 2*, 107-115.

Botwinick, J., Storandt, M., & Berg, L. (1986). A longitudinal, behavioral study of senile dementia of the Alzheimer type. *Archives of Neurology, 43*, 1124-1127.

Brayne, C., & Beardsall, L. (1990). Estimation of verbal intelligence in an elderly community: An epidemiological study using the NART. *British Journal of Clinical Psychology, 29*, 217-223.

Butters, M. A., Salmon, D. P., & Butters, N. (1997). Neuropsychological assessment of dementia. In M. Storandt & G. R. VandenBos (Eds.), *Neuropsychological assessment of dementia and depression in older adults: A clinician's guide* (pp. 33-59). Washington: American Psychological Association.
Carswell, L. M., Graves, R. E., Snow, W. G., & Tierney, M. C. (1997). Predicting verbal IQ of elderly individuals. *Journal of Clinical and Experimental Neuropsychology*, 19, 914-921.

Carswell, L. M., Snow, W. G., & Tierney, M. C. (1992). The NART as a postdictor of IQs in bright normal elderly subjects [Abstract]. *Journal of Clinical and Experimental Neuropsychology*, 15, 65.

Christensen, H., & MacKinnon, A. (1992). Wechsler Intelligence Scale profiles in Alzheimer Type Dementia and healthy ageing. *International Journal of Geriatric Psychiatry*, 7, 241-246.

Chui, H., & Zhang, Q. (1997). Evaluation of dementia: a systematic study of the usefulness of the American Academy of Neurology's Practice Parameters. *Neurology*, 49, 925-935.

Conway, S. C., & O'Carroll, R. E. (1997). An evaluation of the Cambridge Contextual Reading Test (CCRT) in Alzheimer's disease. *British Journal of Clinical Psychology*, 36, 623-625.

Cooper, D., & Fabroni, M. (1988). Relationship between the Wechsler Adult Intelligence Scale-Revised and the Wide Range Achievement Test-Revised in a sample of normal adults. *Educational and Psychological Measurement*, 48, 799-803.
Corrigan, J. D., Agresti, A. A., & Hinkeldey, N. J. (1987). Psychometric characteristics of the Category Test: Replication and extension. *Journal of Clinical Psychology, 43*, 368-376.

Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford & D. M. Parker & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 21-49). Hillsdale, NJ: Lawrence Erlbaum.

Crawford, J. R., Allan, K. M., Cochrane, R. H. B., & Parker, D. M. (1990). Assessing the validity of NART-estimated premorbid IQs in the individual case. *British Journal of Clinical Psychology, 29*, 435-436.

Crawford, J. R., Moore, J. W., & Cameron, I. M. (1992). Verbal fluency: A NART-based equation for the estimation of premorbid performance. *British Journal of Clinical Psychology, 31*, 327-329.

Crawford, J. R., Nelson, H. E., Blackmore, L., Cochrane, R. H. B., & Allan, K. M. (1990). Estimating premorbid intelligence by combining the NART and demographic variables: An examination of the NART standardisation sample and supplementary equations. *Personality and Individual Differences, 11*, 1153-1157.

Crawford, J. R., Obonsawin, M. C., & Allan, K. M. (1998). PASAT and components of WAIS-R performance: Convergent and discriminant validity. *Neuropsychological Rehabilitation, 8*, 255-272.
Crawford, J. R., Parker, D. M., Allan, K. M., Jack, A. M., & Morrison, F. M. (1991). The Short NART: Cross-validation, relationship to IQ and some practical considerations. *British Journal of Clinical Psychology, 30*, 223-239.

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

Crawford, J. R., Parker, D. M., Stewart, L. E., Besson, J. A. O., & De Lacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology, 28*, 267-273.

Crawford, J. R., Stewart, L. E., Garthwaite, P. H., Parker, D. M., & Besson, J. A. O. (1988). The relationship between demographic variables and NART performance in normal subjects. *British Journal of Clinical Psychology, 27*, 181-182.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Dao-Castellana, M. H., Legault, Y. S., Martinot, J. L., Aubin, H. J., Crouzel, C., Feldman, L., Barrucand, D., Rancurel, G., Feline, A., & Syrota, A. (1998). Frontal dysfunction in neurologically normal chronic alcohol subjects: Metabolic and neuropsychological findings. *Psychological Medicine, 28*, 1039-1048.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.
Dikmen, S. S., Temkin, N., & Armsden, G. (1989). Neuropsychological recovery: Relationship to psychosocial functioning and postconcussional complaints. In H. S. Levin & H. M. Eisenberg & A. L. Benton (Eds.), Mild head injury (pp. 229-241). New York: Oxford.

Drug Abuse USA. (1996, November). Getting the facts about alcohol, [website]. Megalynx Web Services, Inc. Available: http://www.drug-abuse.com/. (Accessed April 1999).

Eppinger, M. G., Craig, P. L., Adams, R. L., & Parsons, O. A. (1987). The WAIS-R index for estimating premorbid intelligence: Cross-validation and clinical utility. Journal of Consulting and Clinical Psychology, 55, 86-90.

Faust, D., Guilmette, T. J., Hart, K., Arkes, H. R., Fishbourne, F. J., & Davey, L. (1988). Neuropsychologists' training, experience, and judgment accuracy. Archives of Clinical Neuropsychology, 3, 145-163.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical method for grading the cognitive status of patients for clinicians. The Journal of Psychiatry Research, 12, 189-198.

Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. Archives of Clinical Neuropsychology, 12, 711-738.
Franzen, M. D., Robbins, D. E., & Sawicki, R. F. (1989). Tests of general intelligence. Reliability and validity in neuropsychological assessment (pp. 73-89). New York: Plenum.

Fromm, D., Holland, A. L., Nebes, R. D., & Oakley, M. A. (1991). A longitudinal study of word-reading ability in Alzheimer’s disease: Evidence from the National Adult Reading Test. Cortex, 27, 367-376.

Fromm-Auch, C., & Yeudall, C. T. (1983). Normative data for the Halstead Reitan neuropsychological tests. Journal of Clinical Neuropsychology, 5, 221-228.

Gilliard, J., & Rabins, P. V. (1999). Carer Support. In G. K. Wilcock & R. S. Bucks & K. Rockwood (Eds.), Diagnosis and management of dementia: A manual for memory disorders teams (pp. 279-293). New York: Oxford.

Goldstein, F. C., Gary, H. E., Jr., & Levin, H. S. (1986). Assessment of the accuracy of regression equations proposed for estimating premorbid intellectual functioning on the Wechsler Adult Intelligence Scale. Journal of Clinical and Consulting Psychology, 8, 405-412.

Gouvier, W. D., Bolter, J. F., Veneklasen, J. A., & Long, C. J. (1983). Predicting verbal and performance IQ from demographic data: Further findings with head trauma patients. Clinical Neuropsychology, 5, 119-121.
Graves, R. E., Carswell, L. M., & Snow, W. G. (1999). An evaluation of the sensitivity of premorbid IQ estimators for detecting cognitive decline. *Psychological Assessment, 11*, 29-38.

Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology, 13*, 933-949.

Gronwall, D. (1976). Performance changes during recovery from closed head injury. *Proceedings of the Australian Association of Neurologists*, 13, 143-147.

Gronwall, D., & Wrightson, P. (1974). Delayed recovery of intellectual function after minor head injury. *Lancet, 2*, 605-609.

Gronwall, D., & Wrightson, P. (1981). Memory and information processing capacity after closed head injury. *Journal of Neurology, Neurosurgery, and Psychiatry, 44*, 889-895.

Hawkins, K. A. (1995). Limitations to the validity of the Barona Regression Formula and similar demographically-based methods of estimating pre-injury intellectual functioning. *Behavioral Sciences and the Law, 13*, 491-503.

Hawkins, K. A. (1998). Indicators of brain dysfunction derived from graphic representations of the WAIS-III/WMS-III Technical Manual clinical samples data: A preliminary approach to clinical utility. *The Clinical Neuropsychologist, 12*, 535-551.
Hawkins, K. A., Sledge, W. H., Orleans, J. F., Quinlan, D. M., Rakfeldt, J., & Hoffman, R. E. (1993). Normative implications of the relationship between reading vocabulary and Boston Naming Test performance. *Archives of Clinical Neuropsychology, 8*, 525-537.

Haxby, J. V., Grady, C. L., Koss, E., Horwitz, B., Heston, L., Schapiro, M., Friedland, R. P., & Rapoport, S. I. (1990). Longitudinal study of cerebral metabolic asymmetries and associated neuropsychological patterns in early dementia of the Alzheimer type. *Archives of Neurology, 47*, 753-760.

Hebert, L. E., Scherr, P. A., Beckett, L. A., Albert, M. S., Pilgrim, D. M., Chown, M. J., Funkenstein, H. H., & Evans, D. A. (1995). Age-specific incidence of Alzheimer's disease in a community population. *The Journal of the American Medical Association, 273*, 1354-1359.

Johnstone, B., Hexum, C. L., & Ashkanazi, G. (1995). Extent of cognitive decline in traumatic brain injury based on estimates of premorbid intelligence. *Brain Injury, 9*, 377-384.

Jones, B., & Parsons, O. (1971). Impaired abstraction ability in chronic alcoholics. *Archives of General Psychiatry, 24*, 71-75.

Jones, R. W., & Ferris, S. H. (1999). Age-related memory and cognitive decline. In G. K. Wilcock & R. S. Bucks & K. Rockwood (Eds.), *Diagnosis and management of dementia: A manual for memory disorders teams.* (pp. 211-230). New York: Oxford.
Kareken, D. A., Gur, R. C., & Saykin, A. J. (1995). Reading on the Wide Range Achievement Test-Revised and parental education as predictors of IQ: Comparison with the Barona formula. Archives of Clinical Neuropsychology, 10, 147-157.

Kareken, D. A., & Williams, J. M. (1994). Human judgment and estimation of premorbid intellectual function. Psychological Assessment, 6, 83-91.

Karzmark, P., Heaton, R. K., Grant, I., & Matthews, C. G. (1985). Use of demographic variables to predict full scale IQ: A replication and extension. Journal of Clinical and Experimental Neuropsychology, 7, 412-420.

Kaufman, A. S. (1990). Assessing adolescent and adult intelligence. Boston: Allyn & Bacon.

Kay, T. (1986). The unseen injury: Minor head trauma, Minor head injury: An introduction for professionals (pp. 1-16). Southboro, MA: National Head Injury Foundation.

Kirk, A., & Kertesz, A. (1991). On drawing impairment in Alzheimer's disease. Archives of Neurology, 48, 73-77.

Klesges, R. C., Fisher, L., Vasey, M., & Pheley, A. (1985). Predicting adult premorbid functioning levels: Another look. International Journal of Clinical Neuropsychology, 7, 1-3.
Krull, K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist, 9*, 83-88.

Larrabee, G. J., Haley, J. A., Largen, J. W., & Levin, H. S. (1985). Sensitivity of age-decline resistant ("Hold") WAIS subtests to Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 7*, 497-504.

Law, R., & O'Carroll, R. E. (1998). A comparison of three measures of estimating premorbid intellectual level in dementia of the Alzheimer type. *International Journal of Geriatric Psychiatry, 13*, 727-730.

Lees, A. J., & Smith, E. (1983). Cognitive deficits in the early stages of Parkinson's disease. *Brain, 106*, 257-270.

Levin, H. S., Benton, A. L., & Grossman, R. G. (1982). Cognitive effects, *Neurobehavioral consequences of closed head injury* (pp. 123-139). New York: Oxford.

Levin, H. S., Mattis, S., Ruff, R. M., Eisenberg, H. M., Marshall, L. F., Tabaddor, K., High, W. M., & Frankowski, R. F. (1987). Neurobehavioral outcome following minor head injury: A three-center study. *Journal of Neurosurgery, 66*, 234-243.

Lezak, M. D. (1995). Neuropsychological assessment (3rd ed., pp. 102-109). New York: Oxford.

Long, J. A., & McLachlan, J. F. C. (1974). Abstract reasoning and perceptual-motor efficiency in alcoholics. *Quarterly Journal of Studies on Alcohol, 35*, 1220-1229.
Lynch, J. K., & McCaffery, R. J. (1997). Premorbid intellectual functioning and the determination of cognitive loss. In R. J. McCaffery & A. D. Williams & J. M. Fisher & L. C. Laing (Eds.), The practice of forensic neuropsychology (pp. 91-115). New York: Plenum.

Maddrey, A. M., Cullum, C. M., Weiner, M. F., & Filley, C. M. (1996). Premorbid intelligence estimation and level of dementia in Alzheimer's disease. Journal of the International Neuropsychological Society, 2, 551-555.

Matarazzo, J. D., Daniel, M. H., Prifitera, A., & Herman, D. O. (1988). Inter-subtest scatter in the WAIS-R standardization sample. Journal of Clinical Psychology, 44, 940-957.

McFie, J. (1975). Diagnostic interpretation, Assessment of organic intellectual impairment (pp. 29-47). New York: Academic Press.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology, 34, 939-944.

McLendon, B. M., & Doraiswamy, P. M. (1999). Defining meaningful change in Alzheimer's disease trials: The donepezil experience. Journal of Geriatric Psychiatry and Neurology, 12, 39-48.
Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Microsoft. (1997-98). Excel 98 (Version Macintosh Edition). Redmond, WA: Microsoft Corporation.

Mitrushina, M. N., Boone, K. B., & D'Elia, L. F. (1999). *Handbook of normative data for neuropsychological assessment*. New York: Oxford.

Mortensen, E. L., Gade, A., & Reinisch, J. M. (1991). A critical note on Lezak's 'Best Performance Method' in clinical neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 13*, 361-371.

Nelson, H. E., & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology, 14*, 259-267.

Nelson, H. E., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the new adult reading test. *Cortex, 14*, 234-244.

Nixon, S. J. (1996). Alzheimer's disease and vascular dementia. In R. L. Adams & O. A. Parsons & J. L. Culbertson & S. J. Nixon (Eds.), *Neuropsychology for clinical practice: Etiology, assessment, and treatment of common neurological disorders* (pp. 65-105). Washington, DC: American Psychological Association.

O'Carroll, R. (1995). The assessment of premorbid ability: A critical review. *Neurocase, 1*, 83-89.
O'Carroll, R. E., Baikie, E. M., & Whittick, J. E. (1987). Does the National Adult Reading Test hold in dementia? *British Journal of Clinical Psychology, 26*, 315-316.

O'Carroll, R. E., Curran, S. M., Ross, M., Murray, C., Riddle, W., Moffoot, A. P. R., Ebmeier, K. P., & Goodwin, G. M. (1994). The differentiation of major depression from dementia of the Alzheimer's type using within-subject neuropsychological discrepancy analysis. *British Journal of Clinical Psychology, 33*, 23-32.

Oscar-Berman, M., Clancy, J. P., & Weber, D. A. (1993). Discrepancies between IQ and memory scores in alcoholism and aging. *Clinical Neuropsychologist, 7*, 281-296.

Paolo, A. M., & Ryan, J. J. (1992). Generalizability of two methods of estimating premorbid intelligence in the elderly. *Archives of Clinical Neuropsychology, 7*, 135-143.

Paolo, A. M., Tröster, A. I., Ryan, J. J., & Koller, W. C. (1997). Comparison of NART and Barona demographic equation premorbid IQ estimates in Alzheimer's disease. *Journal of Clinical Psychology, 53*, 713-722.

Paque, L., & Warrington, E. K. (1995). A longitudinal study of reading ability in patients suffering from dementia. *Journal of the International Neuropsychological Society, 1*, 517-524.

Parsons, O. A. (1996). Alcohol abuse and alcoholism. In R. L. Adams & O. A. Parsons & J. L. Culbertson & S. J. Nixon (Eds.), *Neuropsychology for clinical practice: Etiology, assessment, and treatment of common neurological disorders* (pp. 175-201). Washington: American Psychological Association.
Patterson, K., Graham, N., & Hodges, J. R. (1994). Reading in dementia of the Alzheimer type: A preserved ability? *Neuropsychology, 8*, 395-407.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivink, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology, 56*, 303-308.

Porteus, S. D. (1965). *Porteus Maze Test: Fifty years of application.* (pp. 16-29). CA: Palo Alto.

Psychological Corporation. (1997). *WAIS-III WMS-III Technical manual.* San Antonio: Author.

Putnam, S. H., Ricker, J. H., Ross, S. R., & Kurtz, J. E. (1999). Considering premorbid functioning: Beyond cognition to a conceptualization of personality in postinjury functioning. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice* (pp. 38-81). Lisse, Netherlands: Swets & Zeitlinger.

Query, W. T., & Megran, J. (1983). Age-related norms for AVLT in a male patient population. *Journal of Clinical Psychology, 39*, 136-137.

Raguet, M. L., Campbell, D. A., Berry, D. T. A., Schmitt, F. A., & Smith, G. T. (1996). Stability of intelligence and intellectual predictors in older persons. *Psychological Assessment, 8*, 154-160.
Reynolds, C. R. (1997). Postcripts on premorbid ability estimation: Conceptual addenda and a few words on alternative and conditional approaches. *Archives of Clinical Neuropsychology, 12*, 769-778.

Reynolds, C. R., & Gutkin, T. B. (1979). Predicting the premorbid intellectual status of children using demographic data. *Clinical Neuropsychology, 1*, 36-38.

Ritchie, A. J., Lam, M., & Rankin, E. J. (1996). Estimating premorbid intelligence: Comparisons of the Barona Formula, North American Adult Reading Test (NAART), Wide Range Achievement Test-3 Reading Subtest (WRAT-3), and Oklahoma Premorbid Intelligence Estimate (OPIE) [Abstract]. *Archives of Clinical Neuropsychology, 12*, 395.

Russell, E. W. (1972). WAIS factor analysis with brain-damaged subjects using criteria measures. *Journal of Consulting and Clinical Psychology, 39*, 333-339.

Rutherford, W. H. (1989). Postconcussion symptoms: Relationship to acute neurological indices, individual differences, and circumstances of injury. In H. S. Levin & H. M. Eisenberg & A. L. Benton (Eds.), *Mild Head Injury* (pp. 218-228). New York: Oxford.

Ryan, D. H. (1994). Misdiagnosis in dementia: Comparisons of diagnostic error rate and range of hospital investigation according to medical specialty. *International Journal of Geriatric Psychiatry, 9*, 141-147.
Ryan, J. J., & Paolo, A. M. (1992). A screening procedure for estimating premorbid intelligence in the elderly. *The Clinical Neuropsychologist, 6*, 53-62.

Sattler, J. M. (1992). *Assessment of children* (3rd ed.). San Diego: Jerome M. Sattler.

Schinka, J. A., & Vanderploeg, R. D. (1994). Subtest performance and demographic predictors of WAIS-R premorbid ability [Abstract]. *Clinical Neuropsychologist, 8*, 360.

Schinka, J. A., & Vanderploeg, R. D. (2000). Estimating premorbid level of functioning. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment*, (pp. 39-67). Mahwah, NJ: Lawrence Erlbaum.

Schlosser, D., & Ivison, D. (1989). Assessing memory deterioration with the Wechsler Memory Scale, the National Adult Reading Test, and the Schonell Graded Word Reading Test. *Journal of Clinical and Experimental Neuropsychology, 11*, 785-792.

Schlottman, R. S., & Johnsen, D. E. (1991). The Intellectual Correlates Scale and the predication of premorbid intelligence in brain-damaged adults. *Archives of Clinical Neuropsychology, 6*, 363-374.

Schmand, B., Geerlings, M. I., Cees, J., & Lindeboom, J. (1998). Reading ability as an estimator of premorbid intelligence: Does it remain stable in emergent dementia? *Journal of Clinical and Experimental Neuropsychology, 20*, 42-51.
Sellers, A. H., Burns, W. J., & Guyrke, J. S. (1996). Prediction of premorbid intellectual functioning of young children using demographic information. *Applied Neuropsychology, 3*, 21-27.

Shipley, W. C. (1930). *Shipley Institute of Living Scale*. Los Angeles: Western Psychological Services.

Silverstein, A. B. (1987). Accuracy of estimates of premorbid intelligence based on demographic variables. *Journal of Clinical Psychology, 43*, 493-495.

Sivan, A. B. (1992). *Benton Visual Retention Test* (5th ed.). San Antonio: Psychological Corporation.

Smith, J. W., Burt, D. W., & Chapman, R. F. (1973). Intelligence and brain damage in alcoholics: A study in patients of middle and upper social class. *Quarterly Journal of Studies on Alcohol.*, 34, 414-422.

Smith-Seemiller, L., Franzen, M. D., Burgess, E. J., & Prieto, L. R. (1997). Neuropsychologists' practice patterns in assessing premorbid intelligence. *Archives of Clinical Neuropsychology, 12*, 739-744.

Spreen, O., & Strauss, E. (1998). A compendium of neuropsychological tests: Administration, norms, and commentary (2nd ed., pp. 44-50). New York: Oxford.

Spruill, J., & Beck, B. (1986). Relationship between the WAIS-R and Wide Range Achievement Test-Revised. *Educational and psychological measurement, 46*, 1037-1040.
Stebbins, G. T., Wilson, R. S., Gilley, B. A., & Fox, J. H. (1988). Estimation of premorbid intelligence in dementia [Abstract]. *Journal of Clinical and Experimental Neuropsychology, 10*, 63-64.

Stebbins, G. T., Wilson, R. S., Gilley, D. W., Bernard, B. A., & Fox, J. H. (1990). Use of the National Adult Reading Test to estimate premorbid IQ in dementia. *The Clinical Neuropsychologist, 4*, 18-24.

Storandt, M., Botwinick, J., Danziger, W. L., Berg, L., & Hughes, C. P. (1984). Psychometric differentiation of mild senile dementia of the Alzheimer type. *Archives of Neurology, 41*, 497-499.

Storandt, M., Stone, K., & LaBarge, E. (1995). Deficits in reading performance in very mild dementia of the Alzheimer's type. *Neuropsychology, 9*, 174-176.

Sweet, J. J., Moberg, P. J., & Tovain, S. M. (1990). Evaluation of Wechsler Adult Intelligence Scale-Revised premorbid IQ formulas in clinical populations. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 41-44.

Taylor, K. I., Salmon, D. P., Rice, V. A., Bondi, M. W., Hill, R. L., Ernesto, C. R., & Butters, N. (1996). Longitudinal examination of American National Adult Reading Test (AMNART) performance in Dementia of the Alzheimer Type (DAT): Validation and correction based on degree of cognitive decline. *Journal of Clinical and Experimental Neuropsychology,(6)*, 883-891.
Trahan, D. E. (1997). Relationship between facial discrimination and visual neglect in patients with unilateral vascular lesions. *Archives of Clinical Neuropsychology, 12*, 57-62.

Tremont, G., Hoffman, R. G., Scott, J. G., & Adams, R. L. (1998). Effect of intellectual level on neuropsychological test performance: A response to Dodrill (1997). *The Clinical Neuropsychologist, 12*, 560-567.

US Census Bureau. (1998, December). *National estimates: Annual population estimated by age, group, and sex, selected years from 1990 to 1998.* [website]. U.S. Census Bureau, Population Division, Population Projections Branch. Available: http://www.census.gov/. (Accessed April, 1999).

Uzzell, B. P., Dolinskas, C. A., & Langfitt, T. W. (1988). Visual field defects in relation to head injury severity. *Archives of Neurology, 45*, 420-424.

Vanderploeg, R. D., & Schinka, J. A. (1995). Predicting WAIS-R IQ premorbid ability: Combining subtest performance and demographic variable predictors. *Archives of Clinical Neuropsychology, 10*, 225-239.

Vanderploeg, R. D., Schinka, J. A., & Axelrod, B. N. (1996). Estimation of WAIS-R premorbid intelligence: Current ability and demographic data used in best-performance fashion. *Psychological Assessment, 8*, 404-411.
Vanderploeg, R. D., Schinka, J. A., Baum, K. M., Tremont, G., & Mittenberg, W. (1998). WISC-III premorbid prediction strategies: Demographic and best performance approaches. Psychological Assessment, 10, 277-284.

Vogt, A. T., & Heaton, R. K. (1977). Comparison of Wechsler Adult Intelligence Scale indices of cerebral dysfunction. Perceptual and Motor Skills, 45, 607-615.

Watson, C. G., & Klett, W. G. (1974). Are nonverbal IQ tests adequate substitutes for the WAIS. Journal of Clinical Psychology, 30, 55-57.

Watt, K. J., & O'Carroll, R. E. (1999). Evaluating methods for estimating premorbid intellectual ability in closed head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 66, 474-479.

Wechsler, D. (1944). The measurement of intelligence. Baltimore: Williams & Wilkins.

Wechsler, D. (1958). Mental deterioration and its appraisal, Measurement and appraisal of adult intelligence (4th ed., pp. 199-213). Baltimore: Williams & Wilkins.

Weingartner, H., Kaye, W., Smallberg, S. A., Ebert, M. H., Gillin, J. C., & Sitaram, N. (1981). Memory failures in progressive idiopathic dementia. Journal of Abnormal Psychology, 90, 187-196.

Williams, J. M. (1997). The prediction of premorbid memory ability. Archives of Clinical Neuropsychology, 12, 745-756.
Williamson, D., Krull, K. R., & Scott, J. G. (1996). Revision and further validation of the Oklahoma Premorbid Intelligence Estimate [Abstract]. Archives of Clinical Neuropsychology, 12, 430.

Williamson, D. J. G., Scott, J. D., & Adams, R. L. (1996). Traumatic brain injury. In R. L. Adams & O. A. Parsons & J. L. Culbertson & S. J. Nixon (Eds.), Neuropsychology for clinical practice: Etiology, assessment, and treatment of common neurological disorders (pp. 9-64). Washington: American Psychological Association.

Willshire, D., Kinsella, G., & Prior, M. (1991). Estimating WAIS-R IQ from the National Adult Reading Test: A cross-validation. Journal of Clinical and Experimental Neuropsychology, 13, 204-216.

Wilson, B., Kolb, B., Odland, L., & Wishaw, I. Q. (1987). Alcohol, sex, age, and the hippocampus. Psychobiology, 15, 300-307.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., & Whitman, D. (1978). An index of premorbid intelligence. Journal of Consulting and Clinical Psychology, 46, 1554-1555.

Woodruff-Pak, D. S. (1997). The neuropsychology of aging. Malden, MA: Blackwell.

Wrobel, N. H., & Wrobel, T. A. (1996). The problem of assessing brain damage in psychiatric samples: Use of personality variables in prediction of WAIS-R scores. Archives of Clinical Neuropsychology, 11, 625-635.
Yates, A. J. (1956). The use of vocabulary in the measurement of intellectual
deterioration - A review. *Journal of Mental Science*, 102, 409-440.

Yeudall, L. T., Reddon, J. R., Gill, D. M., & Stefanyk, W. O. (1987). Normative
data for the Halstead-Reitan neuropsychological test stratified by age and sex. *Journal of
Clinical Psychology*, 43, 346-367.

Yohman, J. R., Parsons, O. A., & Leber, W. R. (1985). Lack of recovery in male
alcoholics' neuropsychological performance one year after treatment. *Alcoholism:
Clinical and Experimental Research*, 9, 114-117.

Yuspeh, R. L., Vanderploeg, R. N., & Kershaw, D. A. J. (1998). Normative data
on the measure of estimated premorbid abilities as part of a dementia evaluation. *Applied
Neuropsychology*, 5, 149-153.

Zachary, R. A., Crumpton, E., & Spiegel, D. E. (1985). Estimating WAIS-R IQ
from the Shipley Institute of Living Scale. *Journal of Clinical Psychology*, 41, 532-540.