



Published in final edited form as:

Assessment. 2021 April ; 28(3): 955–963. doi:10.1177/1073191119864646.

## The association between the number of neuropsychological measures and the base rate of low scores

Javier Ultra-Cucarella<sup>1</sup>, Miriam Sánchez-SanSegundo<sup>1</sup>, María Rubio-Aparicio<sup>1</sup>, Juan Carlos Arango-Lasprilla<sup>2</sup>, Rosario Ferrer-Cascales<sup>2</sup>

<sup>1</sup>Department of Health Psychology, University of Alicante, Alicante, Spain

<sup>2</sup>BioCruces Health Research Institute. Cruces University Hospital. Ikerbasque. Basque Foundation for Science, Bilbao, Spain

### Abstract

Obtaining one or more low scores, or scores indicative of impairment, is common in neuropsychological batteries that include several measures even among cognitively normal individuals. However, the expected number of low scores in batteries with differing number of tests is unknown. Using 10 neuropsychological measures from the National Alzheimer's Coordinating Center database, 1,023 permutations were calculated from a sample of 5,046 cognitively normal individuals. The number of low scores (i.e., z-score  $\geq 1.5$ ) varied for the same number of measures and among different number of measures, and did not increase linearly as the number of measures increased. According to the number of low scores shown by fewer than 10% of the sample, cognitive impairment should be suspected for one or more, two or more, and three or more in batteries with up to two measures, 3–9 measures, and 10 measures respectively. These results may increase the identification of mild cognitive impairment.

### Keywords

assessment; cognition; dementia; low scores; neuropsychological battery; permutation; tests

Neuropsychological assessment is essential to determine cognitive functioning both in healthy individuals and in individuals with different neurological diseases (Looi & Sachdev, 1999; McKhann et al., 2011). Clinicians routinely administer a number of neuropsychological tests to assess several cognitive domains, and cognitive impairment is suspected when any individual score in the battery is lower than a predefined cut-off point (e.g., z-score  $\geq 1.5$ , z-score  $\geq 2$ ). However, in some neurological diseases with cognitive impairments such as Mild Cognitive Impairment (MCI), the utility of this approach has been debated (Albert et al., 2011; Winblad et al., 2004) due to the probability of obtaining one or more low scores when a battery of tests is administered. A recent meta-analysis reported that studies analyzing the risk of progression from MCI to Alzheimer's disease (AD) included on average 9.94 tests (Ultra-Cucarella, Ferrer-Cascales, et al., 2018). Because the likelihood of

Corresponding author: Miriam Sánchez-SanSegundo, Ph.D., Department of Health Psychology, University of Alicante, Alicante, Spain, miriam.sanchez@ua.es, Phone: +34 965 903400.

Financial disclosures: none

obtaining one or more low scores increases as the number of measures in the battery increase (Binder, Iverson, & Brooks, 2009; Brooks, Iverson, & White, 2007), the use of a single low score may increase the number of false positive MCI diagnoses (Bondi et al., 2014; Edmonds et al., 2015; Jak et al., 2009; Oltra-Cucarella, Sánchez-SanSegundo, et al., 2018).

One of the main limitations of using multiple neuropsychological measures in the assessment of individuals with neurological diseases is related to the concept of independence of these measures (Crawford, Garthwaite, & Gault, 2007). For instance, if the Trail Making Test (A and B), a test of processing speed and divided attention (Reitan, 1958; Strauss, Sherman, & Spreen, 2006), is administered in isolation, a normal distribution of scores could be expected for each variable in the absence of ceiling or floor effects. The standard procedure, however, is to administer part B after part A (Strauss et al., 2006), and thus independence of performance on part B cannot be assumed. Yet, normative data do not calculate transformed scores on part B taking account of performance on part A. Performance on parts A and B are treated as independent when, in fact, a moderate correlation ( $r = .31-.6$ ) between these measures has been reported (Strauss et al., 2006, p. 668).

A second concern, related to independence, is the number of low scores in the battery associated with normal cognitive variability. It has been reported that the number of healthy individuals obtaining one or more low scores is expected to increase as the number of measures in the battery increases, and the cut-off defining low scores increases. Previous research showed a number of low scores in batteries used to obtain normative data (Brooks et al., 2007), in healthy children and adolescents (Brooks, Iverson, Sherman, & Holdnack, 2009; Brooks, Sherman, & Iverson, 2010), in healthy older adults (Palmer, 1998), and through Monte Carlo simulation (Crawford et al., 2007). For instance, Palmer et al. (1998) examined 132 neurologically and psychiatrically healthy older adults using a battery of 10 measures from 5 memory tests, and found that nearly 40% had one or more low scores and nearly 17% had two or more low scores (i.e., 1.28 SD below the mean). Brooks et al. (2007) analyzed performance of 742 healthy community-dwelling adults from the Neuropsychological Assessment Battery standardization sample. When the Memory module including ten measures from four tests was analyzed, the base rate of low scores approximated the expected values based on the normal distribution. Again, the percentage of individuals obtaining one or more low scores increased considerably for all the cut-off points used to define abnormality, a finding that was replicated in different age groups.

To analyze the influence that the correlation among measures in a battery may have on the base rate of low scores, Crawford et al. (2007) programmed a Monte Carlo simulation and used four different correlations among ten measures. They found mixed results. The probability of obtaining one or more low scores (i.e., z-score  $-1.64$ ) decreased as the correlation among measures increased. Thus, for a battery with 10 measures, 40.10% would obtain one or more low scores for  $r=0$ , whereas 18.55% would obtain one or more low scores for  $r=0.7$ . However, for two or more low scores the direction of this effect reversed. The probability of obtaining two or more low scores were 1.40%, 3.29% and 8.64% for batteries including 4, 6 and 10 measures respectively for  $r=0$ , and raised to 11% for batteries

including ten measures for  $r > 0$ . These results clearly show that the correlations between tests are one of the factors that need to be taken into account when considering the base rate of low scores.

For these reasons, the use of a single low score to identify cognitive impairment may increase the number of misdiagnoses. Thus, knowing the base rate of one or more low scores when several measures are administered would help to identify cognitive functioning in a more realistic manner. To allow for normal variability in cognitive testing, several approaches have analyzed the inclusion of more than one low score to define abnormality in neuropsychological assessment (Bondi et al., 2014; Jak et al., 2009; Oltra-Cucarella, Sánchez-SanSegundo, et al., 2018). Bondi et al. (2014) analyzed three cognitive domains using 6 measures, Jak et al. (2009) analyzed five cognitive domains using 24 measures, and Oltra-Cucarella et al. (2018) analyzed the number of low scores shown by the worst performing 10% of the sample. Another major difference between these procedures is the cut-off point used to define low scores, which varied from 1SD below the mean in Jak et al. (2009) and Bondi et al. (2014) to 1.5SD below the mean in Oltra-Cucarella et al. (2018).

The approaches by Jak/Bondi and by Oltra-Cucarella et al. have shown how the use of standard criteria may incorrectly classify a proportion of cognitively normal individuals as cognitively impaired, and have proven to be more effective to identify those who progress from MCI to AD. However, some questions remain unanswered:

1. Does the number of low scores differ with differing number of measures in the battery?
2. Does the number of low scores increase linearly as the number of measures increase?
3. Does the number of low scores relate to the correlation among measures?

The answer to these questions will likely help to interpret performance on measures within a cognitive battery, in order to identify an abnormal cognitive functioning based on a number of low scores above what should be expected according to normal variability. Therefore, the aim of this study is to analyze the expected number of low scores for comprehensive batteries with different numbers of measures in order to help clinicians identify an unusual performance on cognitive tests.

## METHODS

### Sample

Data from the National Alzheimer's Coordinating Center Uniform Data Set (NACC UDS) (Morris et al., 2006; Weintraub et al., 2009) recruited at 35 Alzheimer's Disease Centers (ADC) from June 2005 to October 2017 were analyzed. Individuals were 65 years-old or older and had been categorized as cognitively normal after meeting all of the following criteria: they had a Mini-Mental State Examination score  $> 23$ , a Clinical Dementia Rating scale (CDR)=0, no previous history of psychiatric or neurologic disorders (e.g., depression, Parkinson's disease, traumatic brain injury, stroke), with English as primary language. Individuals were excluded if they required some assistance in basic activities of daily living

or were completely dependent, and if they had missing data for any of the variables used in the analyses.

Details of recruitment procedures at the participating ADC can be found in Weintraub et al. (2009). Participants were volunteers referred by a relative or a friend, by clinicians or memory clinics, by ADC solicitation, by non-ADC media appeal or by other unknown sources. Per protocol, all participants were accompanied by an informant to corroborate information about functioning in daily living activities. Both participants and informants signed informed consent at each participating ADC.

## Instruments

The NACC UDS database includes ten measures from six neuropsychological tests: Logical Memory immediate and delayed recall scores (LM-imm, LM-del) from the Wechsler Memory Scale-Revised, Digit Span Forward (DSF) and Backward (DSB), semantic fluency (animals and vegetables), the Boston Naming Test 30-item (BNT-30), the Trail Making Test parts A (TMT-A) and B (TMT-B), and the Digit Symbol (DS) Coding subtest from the Wechsler Adult Intelligence Scale-Revised (Shirk et al., 2011). Although three different versions of the neuropsychological battery were used in the NACC UDS, only Versions 1 and 2 were analyzed because Version 3 included a different set of neuropsychological tests (Weintraub et al., 2018).

The NACC UDS database was deemed appropriate to analyze the number of low tests because of the distribution of the data, and also because it includes a number of measures close to the average number of measures used in research on MCI (Oltra-Cucarella, Ferrer-Cascales, et al., 2018). As reported by Weintraub et al. (2009) and Shirk et al. (2011), normative data were calculated from a sample of  $n=3,268$  participants using multiple linear regressions controlling for sex, age and years of education. Such a large dataset ensures that the standardized residuals approximate the normal distribution, which is required for regression coefficients to be unbiased, consistent, efficient and with trustworthy confidence intervals (Williams, Gómez Grajales, & Kurkiewicz, 2013).

## Statistical analysis

The following procedure was used to calculate the number of low scores:

1. Raw scores on each measure for each individual were converted to corrected z-scores using the web-based calculator reported by Shirk et al. (2011). Corrected z-scores  $-1.5$  were considered indicative of low performance, as commonly used for the diagnosis of MCI.
2. The ten measures included in the neuropsychological battery were permuted without replacement with the formula  $\frac{n!}{k!(n-k)!}$ , where  $n$  is the number of measures and  $k$  is the number of measures combined (Levine, 1981). Thus, for example, the number of combinations without replacement of six measures in a battery with 10 measures is 210. The percentage of individuals obtaining  $j$  low scores for each combination of  $k$  measures was averaged across permutations of the same number of  $k$  measures.

3. For each combination of  $k$  measures ( $k=2, k=3, \dots, k=i$ ), Pearson's bivariate correlations among the  $k$  measures were averaged. For example, the correlation of the combination with tests 1, 5 and 8 was calculated by averaging  $r_{15}$ ,  $r_{18}$ , and  $r_{58}$ . Next, a bivariate correlation was calculated between the probability of obtaining  $j$  or more low scores and the average correlation among the measures. Following Crawford et al. (2007), the probability of obtaining  $j$  or more low scores should increase as the correlation among the measures increases when  $j \geq 2$ .
4. The number of low scores shown by fewer than 10% of the sample of cognitively healthy individuals was used to define an unusual number of low scores (Binder et al., 2009; Mistridis et al., 2015; Oltra-Cucarella, Sánchez-SanSegundo, et al., 2018). Thus, if less than 10% of the sample obtain 3 or more low tests in a battery with up to 9 tests, normal variability is interpreted as having low scores in up to two tests (Oltra-Cucarella, Sánchez-SanSegundo, et al., 2018). The selection of a cut-off point to define an unusual number of low scores is not straightforward. Binder et al. (2009) considered as uncommon a number of low scores shown by less than 20% of the sample, and as unusual a number of low scores shown by less than 10% of the sample. When analyzing a sample of children, however, Brooks et al. (2010) considered as uncommon a number of low scores shown by less than 10% of the sample. Although this cut-off can be regarded as arbitrary, it is in line with the cumulative percentages used to define abnormality of any individual test score in MCI (i.e., less than or equal to 7% for  $-1.5SD$ ), and have been used to define *probable memory impairment* when 8 measures from the Wechsler Memory Scale-III were analyzed simultaneously (Brooks, Iverson, Feldman, & Holdnack, 2009).
5. A linear regression was used to analyze the association between the number of low scores with age, sex (male=0, female=1), years of education, and MMSE scores. All possible permutations and statistical analyses were done with IBM Statistical Package for Social Sciences (SPSS) v23. A syntax file for permutations is available at request.

## RESULTS

Data from 12,260 healthy individuals aged 65 or older at baseline were used. Of these, 7,214 were excluded (61 because of need for some assistance in basic activities of daily living, 37 with  $MMSE < 24$ , 393 with  $CDR > 0$ , 14 with unknown years of education, 2,850 with a history of psychiatric or neurological disorders, 102 who had a primary language other than English, and 3,757 with missing data for any of the neuropsychological measures analyzed). The final sample included 5,046 individuals (65.7% females) aged 75.45 ( $SD=7.05$ ) and with 15.61 ( $SD=2.92$ ) years of education. Means and SDs for each of the ten measures are shown in table 1.

The average proportion of individuals obtaining one low score (i.e., z-score  $-1.5$ ) when analyzing each of the ten measures separately was 4.64% ( $SD=3.46$ ), slightly lower than expected according to the normal distribution, with values ranging from 0.46% in the TMT-

B to 10.94% in the BNT. The great majority (68.41%) of individuals had no low scores. The percentages of individuals showing only 1, only 2, and three or more low scores were 21.01%, 7.67% and 2.91% respectively.

The average proportions of  $j$  low scores for different combinations of  $k$  measures are shown in supplementary material. Figure 1 shows the average probability of obtaining  $j$  or more low scores in batteries with  $k$  number of measures. Using the average number of low scores shown by fewer than 10% of the sample, the unusual number of low scores would be identified by one or more low scores in batteries with up to two measures, by two or more low scores in batteries with 3–9 measures, and by three or more low scores in batteries with 10 measures.

Table 2 shows the correlations between the probability of obtaining  $j$  or more low scores and the correlation among measures. Thus, for example, in a battery with 3 measures, the correlation between the probability of obtaining one or more low scores and the average correlation of the three measures (i.e.,  $r_{12}$ ,  $r_{13}$ ,  $r_{23}$ ) is 0.384. It can be seen from table 2 that the probability of obtaining  $j$  or more low scores increases as the average correlation among  $k$  measures increases.

The number of low scores increased with increasing years of education. They decreased for females and for increasing MMSE scores (table 3). There was no statistically significant association between the number of low scores and age. The model with age, sex and years of education explained an additional 0.07% of the variance in the number of low scores compared to the model including MMSE scores in isolation.

## DISCUSSION

The aim of present study was to analyze the unusual number of low scores in batteries with one to 10 measures. It was found that 31.59% of healthy older adults showed one or more low scores (i.e., z-score  $-1.5$ ), and 10.58% had two or more. In addition, it was found that the number of low scores increased with increasing numbers of measures and was related to the correlation among the measures in the battery.

The results of the present study show that obtaining one or more low scores in batteries including several measures is common, and must be taken into consideration when assessing cognitive functioning in healthy individuals or in individuals with neurological diseases. The percentages of individuals obtaining one or more low scores in the present study are close to, although slightly lower than, the ones reported by Palmer et al. (1998) when low scores were defined as 1.28SD below the mean. Also, the results reported here are in line with Crawford et al. (2007), and show that around 90% of cognitively normal individuals are expected to obtain up to two z-scores  $-1.5$  when ten measures are administered, with 31.6% obtaining at least one low score. Contrary to Crawford et al. (2007), it was found that the probability of obtaining  $j$  or more low scores increased as the correlation among measures increased even for one or more low scores. However, the main results of the present study are in agreement with the results reported by Crawford et al. (2007): the number of low scores shown by fewer than 10% of the sample in batteries with ten measures is 3 or more. This



agreement is probably due to the very similar cut-off point used in the present study (i.e., z-score  $-1.5$ ) and in the study by Crawford et al. (i.e., z-score  $-1.64$ ).

When using tests in neuropsychological assessment, clinical neuropsychologists must take into account that the dependence between measures might be a limitation for the interpretation of low scores in cognitive performance. Davis and Wall (2014) analyzed the performance of 76 individuals with brain injury (e.g., traumatic brain injury, cerebrovascular compromise, encephalopathy) on the California Verbal Learning Test, and reported a correlation of 0.91 between short-delay and long-delay free recall measures. Thus, individuals obtaining low scores on the short-delay measure are likely to obtain low scores on the long-delay measure as well. The results reported here suggest that the probability of obtaining one or more low scores may be common (e.g., 19.76%) when highly correlated measures are used, and that this probability may turn uncommon (e.g., 2.08%) if the correlation between measures is low. Taken together, these results support that the use of only one low score may increase the probability of erroneously identifying cognitive impairment. More importantly, relying on two low scores obtained on highly correlated measures could also be problematic. As the results obtained on these two measures are not truly independent of one another, low scores on both measures may also be erroneously indicating cognitive impairment. McGrath (2008) showed that adding variables might not increase the sensitivity of the statistical model over using only one predictor when highly correlated variables are used to predict a dichotomous outcome (e.g., being cognitively impaired).

In the present study, the number of low scores was inversely related to MMSE scores, which could reflect a higher general cognitive functioning. Conversely, the association with age, sex and education was very small (i.e., an additional 0.07% of the variance explained). The surprising results of very low coefficients (e.g., years of education) being statistically significant seem to be accounted for by a high power due to the large sample size. One may expect to find a negative association between the number of low scores and the years of education, as previously reported (Binder et al., 2009). However, z-scores in the present study were calculated using residuals of a multivariate linear regression including age, sex, and years of education, and thus the effects of education are controlled for in the normative data. In line with this, when analyzing the number of low scores on a memory index including 9 measures, Brooks et al. (2007) found that the base rate of low scores did not vary by age or education, a finding that was associated with the use of corrected normative scores.

Large-scale studies have found controversial results regarding the association of formal education and socioeconomic status with cognitive functioning and cognitive decline. Zahodne, Stern and Manly (2015) found a slower cognitive decline in individuals with high education, whereas Wilson et al. (2009) found no association between level of education and rate of cognitive decline. When controlling for childhood socioeconomic conditions, Aartsen et al. (2019) found a faster rate of decline in individuals with more advantaged socioeconomic conditions even after controlling for adulthood socioeconomic conditions, levels of physical activity, depressive symptoms and partner status. These controversial findings suggest potential value in including measures of cultural capital, social capital, and

socioeconomic status other than educational level when interpreting performance on cognitive tests in order to better assess cognitive functioning during aging.

The main implication of our results is their potential use in both clinical and research settings to define normal cognitive variability, and hence cognitive impairment. As shown in figure 1, clinicians can expect a different number of low scores in batteries including 5 measures in comparison with batteries including 10 measures. These results partially agree with previous works that included several low scores to define objective cognitive impairment. Oltra-Cucarella et al. (2018) reported that less than 10% of the sample in a battery with 9 measures obtained three or more scores at or below  $-1.5SD$ , which is in line with the results reported in this study and support their use for defining abnormal performance in cognitive tests. In line with their findings of a more accurate identification of progressors to AD using three or more test to define objective cognitive impairment in comparison with standard criteria, the results of this study suggest that the base rate of low scores approach should be replicated in studies with a higher number of measures in the battery.

Another implication of these results is the need to analyze the base rate of low scores associated with different cut-off points for the identification of individuals at the greatest risk of AD. Other researchers have used a different approach to define cognitive impairment based on several low scores, although the range and type of tests, as well as the number of cognitive domains, differed among studies (Bondi et al., 2014; Clark et al., 2013; Jak et al., 2009, 2016). Taking the verbal memory domain as an example, it has not been empirically established the probability of obtaining two measures at 1SD or higher below the mean for a battery with four measures (Clark et al., 2013) compared to a battery including two measures (Jak et al., 2016). Also, it must be established whether the probability of obtaining two low scores differ for batteries including a delayed recall and recognition subtests from a story recall test compared to a list-learning test (Bondi et al., 2014; Jak et al., 2016).

One limitation of the present study is the use of permutations to obtain the expected number of low scores for batteries with different number of measures. Ideally, a number of batteries with different number of measures would have allowed analyzing real rather than simulated data. However, each study would have contributed a sample with different demographic characteristics and different measures (and different correlations among measures). Related to this, the variables used in permutations may not follow a normal distribution, as is the case for some neuropsychological variables (Crawford et al., 2007). For example, most individuals succeed in the majority of items of several cognitive tests (Strauss et al., 2006), making the distribution of scores negatively skewed. In this case, the percentile ranks might not follow a normal distribution. However, a recent work by Piovesana and Senior (2018) found that samples greater than 85 provided stable means and standard deviations regardless of the level of skewness. Thus, future studies will have to analyze the probability of obtaining  $j$  or more low scores on batteries with different  $k$  number of measures including several skewed measures.

Although with these limitations, this work has important strengths, as it replicates the procedure commonly used in neuropsychological assessment, that is, to treat normative data



for performance on each cognitive test as independent of the performance on other tests. For example, although it must be assumed that performance on the Rey-Osterrieth Complex figure (ROCF) delayed recall task will be affected by the ROCF copy task (Strauss et al., 2006), normative data for the ROCF delayed recall task from which percentiles are obtained are generally calculated without including the ROCF copy scores, treating copy and delayed recall scores as independent. An average score on the delayed recall task might be frequent for individuals obtaining average scores on the copy tasks, but not for individuals obtaining high or very high scores on the copy task. However, the same score on the delayed recall task (e.g.,  $x=15$ ) will be transformed to the same percentile (e.g.,  $P_c=45$ ) for an individual scoring at the 90<sup>th</sup> percentile and for another individual scoring at the 30<sup>th</sup> percentile in the copy task.

Our simulation is probably closer than previous works to the procedures used in neuropsychological assessments and to the distribution of low scores, especially for a high number of tests in a battery. A possible way to improve the expected number of low scores in a battery with several measures could be mixing the methodology used by Crawford et al. (2007) and Jak/Bondi (Bondi et al., 2014; Jak et al., 2009). The probability of obtaining one or more scores within a domain, or one low score per domain, could be calculated for different number of domains and different number of tests within domains by varying the correlation among measures. However, the earlier step should be to empirically state that measures correlate within domains and not across domains, to eliminate potential statistical artifacts. This sort of factorial analysis would, however, limit its applicability to batteries of tests that correlate within domains but not across domains, which is uncommon in neuropsychological assessment where tests measuring attention or executive functioning correlate with tests measuring verbal memory and other domains (Strauss et al., 2006).

In conclusion, using several measures is the rule and not the exception in batteries used to analyze the full range of cognitive abilities in neuropsychological assessment. The results of the present study add to the existing literature by providing the expected number of low scores (i.e.,  $z < -1.5$ ) for different numbers of measures in the battery. Unusual performance should be suspected for one or more low scores in batteries with up to two measures, two or more low scores in batteries with 3 to 9 measures, and three or more low scores in batteries with 10 measures. In these cases, neuropsychologists should perform a much more careful evaluation in order to confirm or rule out the presence of cognitive deficits in their clients. The results reported here are expected to improve the identification of true cognitive impairment by reducing the number of false positive diagnoses, which will likely increase the precision of the estimates of progression of cognitive deficits in a variety of neurological diseases.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements:

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott

Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

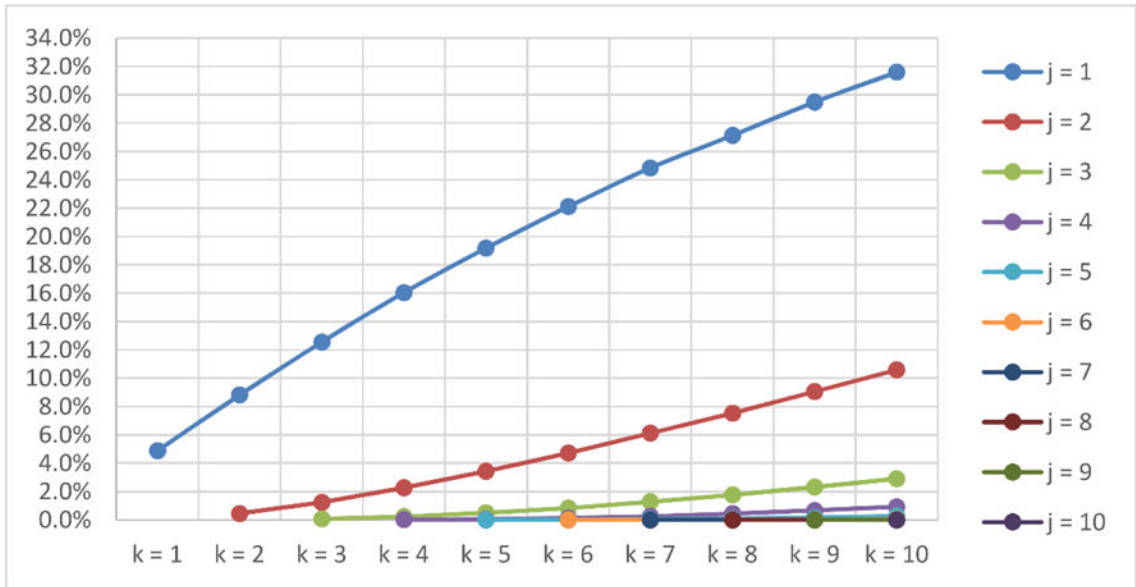
NACC UDS database can be accessed at <https://www.alz.washington.edu/WEB/landingRequest.html>

## REFERENCES

- Aartsen MJ, Cheval B, Sieber S, Van der Linden BW, Gabriel R, Courvoisier DS, ... Cullati S (2019). Advantaged socioeconomic conditions in childhood are associated with higher cognitive functioning but stronger cognitive decline in older age. *Proceedings of the National Academy of Sciences*, 116(12), 5478–5486. 10.1073/pnas.1807679116
- Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, ... Phelps CH (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 270–279. 10.1016/j.jalz.2011.03.008
- Binder LM, Iverson GL, & Brooks BL (2009). To err is human: "abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, 24(1), 31–46. 10.1093/arclin/acn001 [PubMed: 19395355]
- Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, ... Salmon DP (2014). Neuropsychological Criteria for Mild Cognitive Impairment Improves Diagnostic Precision, Biomarker Associations, and Progression Rates. *Journal of Alzheimer's Disease*, 42(1), 275–289. 10.3233/JAD-140276
- Brooks BL, Iverson GL, Feldman HH, & Holdnack JA (2009). Minimizing Misdiagnosis: Psychometric Criteria for Possible or Probable Memory Impairment. *Dementia and Geriatric Cognitive Disorders*, 27(5), 439–450. 10.1159/000215390 [PubMed: 19401611]
- Brooks BL, Iverson GL, Sherman EMS, & Holdnack JA (2009). Healthy children and adolescents obtain some low scores across a battery of memory tests. *Journal of the International Neuropsychological Society*, 15(04), 613. 10.1017/S1355617709090651 [PubMed: 19573280]
- Brooks BL, Iverson G, & White T (2007). Substantial risk of "Accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 13(03), 490–500. 10.1017/S1355617707070531 [PubMed: 17445298]
- Brooks BL, Sherman EMS, & Iverson GL (2010). Healthy Children Get Low Scores Too: Prevalence of Low Scores on the NEPSY-II in Preschoolers, Children, and Adolescents. *Archives of Clinical Neuropsychology*, 25(3), 182–190. 10.1093/arclin/acq005 [PubMed: 20179013]
- Clark LR, Delano-Wood L, Libon DJ, McDonald CR, Nation DA, Bangen KJ, ... Bondi MW (2013). Are empirically-derived subtypes of mild cognitive impairment consistent with conventional subtypes? *Journal of the International Neuropsychological Society*, 19(6), 635–45. 10.1017/S1355617713000313 [PubMed: 23552486]
- Crawford JR, Garthwaite PH, & Gault CB (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology*, 21(4), 419–430. 10.1037/0894-4105.21.4.419 [PubMed: 17605575]
- Davis JJ, & Wall JR (2014). Examining Verbal Memory on the Word Memory Test and California Verbal Learning Test-Second Edition. *Archives of Clinical Neuropsychology*, 29(8), 747–753. 10.1093/arclin/acu030 [PubMed: 25064762]

- Edmonds EC, Delano-Wood L, Clark LR, Jak AJ, Nation DA, McDonald CR, ... Bondi MW (2015). Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimer's & Dementia*, 11(4), 415–424. 10.1016/j.jalz.2014.03.005
- Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, & Delis DC (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *American Journal of Geriatric Psychiatry*, 17(5), 368–75. 10.1097/JGP.0b013e31819431d5
- Jak AJ, Preis SR, Beiser AS, Seshadri S, Wolf PA, Bondi MW, & Au R (2016). Neuropsychological Criteria for Mild Cognitive Impairment and Dementia Risk in the Framingham Heart Study. *Journal of the International Neuropsychological Society*, 22, 1–7. 10.1017/S1355617716000199 [PubMed: 26714882]
- Levine G (1981). *Introductory statistics for psychology: the logic and the methods* New York: Academic Press.
- Looi JCL, & Sachdev PS (1999). Differentiation of vascular dementia from AD on neuropsychological tests. *Neurology*, 53(4), 670–670. 10.1212/WNL.53.4.670 [PubMed: 10489025]
- McGrath RE (2008). Predictor combination in binary decision-making situations. *Psychological Assessment*, 20(3), 195–205. 10.1037/a0013175 [PubMed: 18778156]
- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, ... Phelps CH (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 263–9. 10.1016/j.jalz.2011.03.005
- Mistridis P, Egli SC, Iverson GL, Berres M, Willmes K, Welsh-Bohmer KA, & Monsch AU (2015). Considering the base rates of low performance in cognitively healthy older adults improves the accuracy to identify neurocognitive impairment with the Consortium to Establish a Registry for Alzheimer's Disease-Neuropsychological Assessment Battery (CERAD-). *European Archives of Psychiatry and Clinical Neuroscience*, 265(5), 407–417. 10.1007/s00406-014-0571-z [PubMed: 25555899]
- Morris JC, Weintraub S, Chui HC, Cummings J, DeCarli C, Ferris S, ... Kukull WA (2006). The Uniform Data Set (UDS): Clinical and Cognitive Variables and Descriptive Data From Alzheimer Disease Centers: *Alzheimer Disease & Associated Disorders*, 20(4), 210–216. 10.1097/01.wad.0000213865.09806.92 [PubMed: 17132964]
- Oltra-Cucarella J, Ferrer-Cascales R, Alegret M, Gasparini R, Díaz-Ortiz LM, Ríos R, ... Cabello-Rodríguez L (2018). Risk of progression to AD for different neuropsychological Mild Cognitive Impairment subtypes. A hierarchical meta-analysis of longitudinal studies. *Psychology and Aging*, 33(7), 1007–1021. 10.1037/pag0000294 [PubMed: 30284855]
- Oltra-Cucarella J, Sánchez-SanSegundo M, Lipnicki DM, Sachdev PS, Crawford JD, Pérez-Vicente JA, ... Ferrer-Cascales R (2018). Using the base rate of low scores helps to identify progression from amnesic MCI to AD. *Journal of the American Geriatrics Society*, 66(7), 1360–1366. 10.1111/jgs.15412 [PubMed: 29745971]
- Palmer B (1998). Base Rates of "Impaired" Neuropsychological Test Performance Among Healthy Older Adults. *Archives of Clinical Neuropsychology*, 13(6), 503–511. 10.1016/S0887-6177(97)00037-1 [PubMed: 14590634]
- Piovesana A, & Senior G (2018). How Small Is Big: Sample Size and Skewness. *Assessment*, 25(6), 793–800. 10.1177/1073191116669784 [PubMed: 27655971]
- Reitan RM (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*, 8(3), 271–276. 10.2466/pms.1958.8.3.271
- Shirk SD, Mitchell MB, Shaughnessy LW, Sherman JC, Locascio JJ, Weintraub S, & Atri A (2011). A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimer's Research & Therapy*, 3(6), 32. 10.1186/alzrt94
- Strauss EH, Sherman EHS, & Spreen O (2006). *A Compendium of Neuropsychological Tests. Administration, Norms and Comments* New York: Oxford University Press.
- Weintraub S, Besser L, Dodge HH, Teylan M, Ferris S, Goldstein FC, ... Morris JC (2018). Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). *Alzheimer Disease & Associated Disorders*, 32(1), 10–17. 10.1097/WAD.0000000000000223 [PubMed: 29240561]

- Weintraub S, Salmon D, Mercaldo N, Ferris S, Graff-Radford NR, Chui H, ... Morris JC (2009). The Alzheimer's Disease Centers' Uniform Data Set (UDS): The Neuropsychologic Test Battery. *Alzheimer Disease & Associated Disorders*, 23(2), 91–101. 10.1097/WAD.0b013e318191c7dd [PubMed: 19474567]
- Williams MN, Gómez Grajales CA, & Kurkiewicz D (2013). Assumptions of Multiple Regression: Correcting Two Misconceptions. *Practical Assessment, Research & Evaluation*, 18(11), 1–14.
- Wilson RS, Hebert LE, Scherr PA, Barnes LL, Mendes de Leon CF, & Evans DA (2009). Educational attainment and cognitive decline in old age. *Neurology*, 72(5), 460–465. 10.1212/01.wnl.0000341782.71418.6c [PubMed: 19188578]
- Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund L-O, ... Petersen RC (2004). Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, 256(3), 240–6. 10.1111/j.1365-2796.2004.01380.x [PubMed: 15324367]
- Zahodne LB, Stern Y, & Manly JJ (2015). Differing effects of education on cognitive decline in diverse elders with low versus high educational attainment. *Neuropsychology*, 29(4), 649–57. 10.1037/neu0000141 [PubMed: 25222199]



**Fig. 1. Number of low scores according to the number of measures.**  
 k = number of measures. j = number of low scores or more

**Table 1.**

Descriptive statistics for the neuropsychological measures (n = 5,046)

	<b>Mean</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
MMSE	28.86	1.32	24	30
LM-immediate	13.21	3.86	0	25
LM-delayed	11.94	4.15	0	24
Digits Forward	6.69	1.07	1	8
Digits Backward	4.83	1.22	0	7
Animals	19.54	5.48	1	49
Vegetables	14.45	4.11	2	32
TMT-A	36.07	15.38	12	150
TMT-B	96.25	51.83	12	300
Digit Symbol	47.07	14.16	2	97
BNT-30	26.90	3.36	0	30

MMSE: Mini-Mental State Examination. LM: Logical Memory test. Digits: longest series. TMT: Trail Making Test. BNT: Boston Naming Test – 30 items.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 2.** Correlation between bivariate correlations of  $\gamma$  measures and the probability of obtaining  $j$  or more low scores

	<b>k = 2</b>	<b>k = 3</b>	<b>k = 4</b>	<b>k = 5</b>	<b>k = 6</b>	<b>k = 7</b>	<b>k = 8</b>	<b>k = 9</b>
<b>j = 1</b>	0.340	0.384	0.391	0.394	0.382	0.369	0.355	0.339
<b>j = 2</b>	0.670	0.691	0.697	0.703	0.701	0.698	0.694	0.690
<b>j = 3</b>		0.691	0.759	0.783	0.786	0.787	0.783	0.769
<b>j = 4</b>			0.579	0.713	0.766	0.788	0.802	0.825
<b>j = 5</b>				0.400	0.587	0.691	0.735	0.729
<b>j = 6</b>					0.236	0.470	0.685	0.854

**Table 3.**

Linear regression models predicting the number of low scores

	Variable	b	b Std. error	R	R Std. error
Model 1	MMSE	-0.16	.008	0.264	.78
Model 2	MMSE	-0.17	.009	0.277	.78
	Education	0.02	.004		
	Age	0.00	.002		
	Sex	-0.05	.024		

MMSE: Mini-Mental State Examination. Std. error: standard error.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript