Published in final edited form as:

J Aging Health. 2015 October ; 27(7): 1199–1222. doi:10.1177/0898264315577590.

A Summary Score for the Framingham Heart Study Neuropsychological Battery

Brian Downer, PhD¹, David W. Fardo, PhD², and Frederick A. Schmitt, PhD²

¹University of Texas Medical Branch, Galveston, USA

²University of Kentucky, Lexington, USA

Abstract

Objective—To calculate three summary scores of the Framingham Heart Study neuropsychological battery and determine which score best differentiates between subjects classified as having normal cognition, test-based impaired learning and memory, test-based multidomain impairment, and dementia.

Method—The final sample included 2,503 participants. Three summary scores were assessed: (a) composite score that provided equal weight to each subtest, (b) composite score that provided equal weight to each cognitive domain assessed by the neuropsychological battery, and (c) abbreviated score comprised of subtests for learning and memory. Receiver operating characteristic analysis was used to determine which summary score best differentiated between the four cognitive states.

Results—The summary score that provided equal weight to each subtest best differentiated between the four cognitive states.

Discussion—A summary score that provides equal weight to each subtest is an efficient way to utilize all of the cognitive data collected by a neuropsychological battery.

Keywords

cognition; dementia; mild cognitive impairment

Introduction

Current interest in the preclinical and prodromal stages of Alzheimer's disease (AD) has made it important that cognitive measures be evaluated for their ability to provide an

Reprints and permissions: sagepub.com/journalsPermissions.nav

Corresponding Author: Brian Downer, University of Texas Medical Branch, Sealy Center on Aging, 301 University Blvd., Galveston, TX 77555, USA. brdowner@utmb.edu.

Authors' Note

The opinions and conclusions are solely of the authors and do not reflect the Framingham Heart Study or National Heart, Lung, and Blood Institute (NHLBI).

Framingham Heart Study data can be accessed by submitting a research proposal to the database of Genotypes and Phenotypes.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

objective marker for these cognitive states, similar to the criteria for mild cognitive impairment (MCI; Karrasch, Sinerva, Gronholm, Rinne, & Laine, 2005; Shankle et al., 2005). As there may be subtle cognitive changes occurring during the earliest stages of AD (Jack et al., 2012; Johnson, Storandt, Morris, & Galvin, 2009; Knopman et al., 2012; Langbaum et al., 2014; Riley et al., 2011), tracking cognition from the preclinical and prodromal stages of AD to MCI would further aid in the design of prevention and secondary treatment trials. Many ongoing cohort studies have collected cognitive data from participants using comprehensive neuropsychological batteries that assess a variety of cognitive domains. Investigators will often use the cognitive data to calculate a summary score that provides equal weight to each subtest or cognitive domain assessed. Another approach limits analysis to subtests that assess specific cognitive domains that meet the objectives of the study. Although these approaches are an efficient method to utilize cognitive data by limiting the number of statistical models employed, research on which approach best differentiates between various cognitive states is lacking.

The purpose of the current study was to calculate a series of three summary scores using data from the Framingham Heart Study (FHS) Offspring Cohort to determine which summary score best differentiated between participants classified as having normal cognition (NC), test-based impaired learning and memory (ILM_{TB}), test-based multidomain impairment (MDI_{TB}), or dementia. Creating a summary score of the FHS neuropsychological battery will enable investigators to assess severity of cognitive impairment among dementia, ILM_{TB}, and MDI_{TB} patients, effectively measure cognitive decline among dementia, ILM_{TB}, and MDI_{TB} patients, aid in the identification of patients who progress from NC to dementia, and help investigators differentiate between normal and abnormal cognitive aging. Also, the design and methodology of the present analysis can be replicated and used by other investigators to create summary scores for the neuropsychological batteries employed in other studies.

Method

FHS Offspring Cohort

This report reflects a secondary analysis of cognitive data collected from participants of the FHS Offspring Cohort. The FHS Offspring Cohort was initiated in 1971 (Feinleib, Kannel, Garrison, McNamara, & Castelli, 1975) and a total of eight clinical examinations have been completed between 1971 and 2008. Beginning in 1999, surviving participants of the Cohort were recruited to participate in a secondary study in which they received a comprehensive neuropsychological battery. A total of 2,557 members of the FHS Offspring Cohort received this test battery between 1999 and 2005, and 1,713 (67.0%) participants attended a follow-up evaluation prior to 2007 in which they received the same tests. Individuals with missing data for one or more subtests of the battery during the baseline examination ($n = 54$, 2.1%) were excluded from the final sample. This was done so the summary scores for each participant would reflect the same combination of cognitive functioning measures. The final sample included 2,503 participants who were randomized to either a training or validation set. The training set was used to estimate the effects of age, sex, and education (<high school, some college, college degree) on the scores for each assessment included in the FHS

for persons who had not received a dementia diagnosis or who had a decline >1.5 standard deviations (*SD*) on one or more assessments between the baseline and postevaluations. The effects of these demographics were determined using a multivariable model that included age, sex, and education as predictors of each test as well as the omnibus score. The regression coefficients from this model were then used to calculate predicted scores for each of the cognitive assessments. The estimated effects of age, sex, and education are presented in Table 1.

FHS Neuropsychological Battery

The FHS neuropsychological battery includes subtests that assess naming and language (Boston Naming Test; Goodglass & Kaplan, 1983), visuoperceptual skills (Hooper Visual Organization Test [HVOT]; Hooper, 1966), premorbid intelligence (Wide Range Achievement Test; Wilkinson, 1993), abstract reasoning (Similarities; Wechsler, 1955), motor speed (Finger Tapping Test; Shimoyama, Ninchoji, & Uemura, 1990), along with visual attention, set switching, and visual-motor skills (Trail Making Test [TMT] A and B; Armitage, 1946), subtests of the Wechsler Memory Scale (Wechsler, 1945) that assess verbal memory (Logical Memory Immediate and Delayed Recall), learning (Paired-Associate Memory Immediate and Delayed Recall), and visual memory (Visual Reproductions Immediate and Delayed Recall). A higher score on each subtest of the neuropsychological battery, with the exception of the TMT, indicates better cognitive function. The TMT A and TMT B are timed assessments that require a participant to connect a random alphabetic sequence (TMT A) and alphanumeric sequence (TMT B). The validity and reliability of these cognitive assessments are well documented (Ashendorf, Jefferson, Green, & Stern, 2009; Boyd, 1981; Gill, Reddon, Stefanyk, & Hans, 1986; Huff, Collins, Corkin, & Rosen, 1986; Russell, 1975; Ryan & Ward, 1999; Salthouse, 2011).

Two additional cognitive scores were derived utilizing the raw scores from the TMT A and TMT B (Trail Making Difference Score) and Logical Memory Immediate and Delayed Recall (Memory Retention). The Trail Making Difference Score was obtained by calculating the minutes to complete the TMT B minus the minutes to complete the TMT A. This difference score accounts for the motor speed and visual scanning components of the TMT A and has been shown to provide a reliable measure of executive function and cognitive flexibility (Sanchez-Cubillo et al., 2009). The Memory Retention Score was obtained by dividing the score for Logical Memory Delayed Recall by the score for Logical Memory Immediate Recall (Logical Memory Delayed / Logical Memory Immediate; Russell, 1975). This score provides a measure of memory retention relative to initial encoding (Mathews et al., 2014).

Case Definitions of Dementia, Impaired Learning and Memory, and Multidomain Impairment

Since 1979, incident cases of dementia in the Offspring Cohort have been recorded on an ongoing basis as part of the Epidemiology of Dementia study (Bachman et al., 1993; Bachman et al., 1992). The Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) has been administered to participants during each clinical exam to identify participants who required further neuropsychological evaluation. Participants who scored

below the education-adjusted cutoff score on the MMSE (score < 22 for < 7 years education, score < 24 for 8 to 11 years education, score < 25 for high school graduate, and score < 26 for any education beyond high school), declined three or more points on the MMSE since the previous examination, or were referred by themselves, a family member, family physician, or FHS physician, were suspected of having dementia and underwent further evaluation by an expert panel that included at least one neurologist and one neuropsychologist. This panel evaluated all available health records, neurologic and neuropsychological examinations, and reports of all hospitalizations to determine if the reported cognitive impairment was likely due to dementia. Participants were considered as having dementia by the panel if they met the following three criteria: (a) independent diagnosis of cognitive impairment consistent with dementia, defined as impairment in three or more cognitive domains, including language, memory, visuospatial skills, personality/behavior, and cognition, as well as intact consciousness (Cummings & Benson, 1986), by a neurologist and neuropsychologist; (b) participant has experienced cognitive impairment for at least 1 year; and (c) definitive cognitive decline based on participant's preexisting level of cognitive functioning. For cases of dementia, the panel assessed dementia severity (mild, moderate, or severe) according to criteria consistent with the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.; *DSM-III-R*; American Psychiatric Association [APA], 1987) dementia subtype (AD, AD with stroke, vascular dementia [VaD], AD with VaD [mixed dementia], or other), according to the criteria described in Bachman et al. (1992).

In this study, 36 participants who had received a diagnosis of dementia received a neuropsychological battery between 1999 and 2005, but seven cases ($n = 5$ AD; $n = 2$ other dementia) had missing data for one or more subtests. These seven cases were excluded from this analysis. For the present analysis, the subtests included in the neuropsychological battery and two derived cognitive scores were used to classify nondemented participants as NC, ILM_{TB}, or test-based multidomain impaired (MDI_{TB}) based on the approach used by Abner and colleagues (2012). Impaired performance on a subtest was defined as a score >1.5 *SD* below what would be expected given the participant's age, sex, and level of education. This definition is consistent with the criteria provided by the National Institute on Aging–Alzheimer's Association (NIA-AA) workgroup (Albert et al., 2011). A classification of ILM_{TB} was given to participants with no dementia and who had impaired performance on one or more subtests of learning and memory or memory retention but nonimpaired performance on all other subtests. A classification of MDI_{TB} was given to participants with no dementia and who had impaired performance on one or more subtests in the neuropsychological battery, which could include subtests of learning and memory, and memory retention. A classification of NC was given to participants who were absent of dementia and did not have an impaired performance on any subtests or the two derived cognitive scores.

Summary Scores

A total of three summary scores were created: (a) composite subtests, (b) learning and retention, and (c) composite domains. The composite subtests summary score was calculated by first transforming the raw score of each subtest into a *z* score by subtracting a

participant's score (x) minus the sample mean (μ) and dividing by the SD (σ) [z score = $(x - \mu) / \sigma$]. The z scores for TMT A, TMT B, and Trail Making Difference Score were multiplied by -1 so that higher scores represented more intact cognition. Once a z score was calculated for each subtest, the average of the z scores was calculated to obtain the composite subtests summary score.

The learning and retention summary score was comprised of the six sub-tests from the Wechsler Memory Scale that assessed learning and memory (verbal memory [Logical Memory Immediate and Delayed Recall], learning [Paired-Associate Memory Immediate and Delayed Recall], and visual memory [Visual Reproductions Immediate and Delayed Recall]) and the derived Memory Retention Score. The summary score was obtained by first transforming each raw score into a z score followed by calculating the average of the z scores.

The cognitive domains summary score was comprised of the specific cognitive domains assessed by the FHS neuropsychological battery that were identified in the present analysis by conducting a factor analysis followed by an orthogonal (varimax) rotation. A total of three factors were extracted (eigenvalue >1) and a fourth factor with an eigenvalue of 0.9 was also included (remaining eigenvalues <0.71). The proportion of variance explained by the three factors and four factors was 0.61 and 0.79, respectively. The following subtests loaded onto the four factors based on a rotated factor pattern score above 0.5: *visual/spatial memory* (Visual Memory Immediate Recall [0.90], Visual Memory Delayed Recall [0.9], HVOT [0.5]), *verbal memory* (Logical Memory Immediate Recall [0.93], Logical Memory Delayed Recall [0.90]), *new learning* (Paired Associates Immediate Recall [0.88], Paired Associates Delayed Recall [0.9]), and *attention/concentration* (TMT A [0.87], TMT B [0.77]). The TMT Difference Score was added to the *attention/concentration* domain and Memory Retention was added to the *verbal memory* domain as these domains included the individual assessments used to derive these additional cognitive measures. Once these cognitive domains were identified, the next step was to sum the raw scores for the subtests to create a total score for each specific cognitive domain. These total scores for each cognitive domain were then transformed into z scores and the average of these z scores was calculated to obtain the composite domain summary score.

Statistical Analysis

Receiver operator characteristic (ROC) analysis was conducted utilizing data from the validation sample to obtain ROC curves for each of the summary scores. The area under the curve (AUC) is interpreted as the probability that a randomly selected participant who is a case has a lower summary score than a randomly selected participant who is a control. To control for the effects of age, sex, and education, the residuals of the summary scores were obtained by calculating the difference between the observed summary score and predicted summary score given a participant's age, sex, and education (see Table 1 for estimated effects of age, sex, and education on the three summary scores). ROC analysis was also used to determine the sensitivity (proportion of correctly identified cases), specificity (proportion of correctly identified controls), positive predictive value (ppv; probability that participants who test *positive* are true cases), negative predictive value (npv; the probability that

participants who test *negative* are true controls), and diagnostic accuracy (proportion of cases and controls who are correctly identified) for the optimal cut point of the residuals for each summary score. The optimal cut points were determined according to the “closest top left” criteria in which the point that is closest to perfect differentiation is identified and the corresponding cut point is deemed as optimal (Coffin & Sukhatme, 1997).

Results

Characteristics of FHS Offspring Cohort

The demographic characteristics of participants included in the validation sample according to cognitive status are provided in Table 2. Among participants who completed a neuropsychological battery, 564 (45.1%) were classified as NC, 269 (21.5%) were classified as ILM_{TB}, 399 (31.9%) were classified MDI_{TB}, and 19 (1.5%) participants had received a diagnosis of dementia ($n = 16$ AD; $n = 1$ AD with stroke; $n = 1$ AD with vascular dementia; $n = 1$ other). There were significant differences for age ($p < .01$), education ($p = .01$), and *APOE* e4 allele status ($p < .01$) according to cognitive status. There were no differences in the proportion of males and females according to cognitive status. All three summary scores were highly correlated and the Pearson correlation coefficients were as follows: (a) composite subtests and composite domains $r = .97$, (b) composite subtests and learning and memory $r = .92$, (3) composite domains and learning and memory $r = .94$.

Effects of Age, Sex, and Education on Cognition

There were significant differences in subtest performance, two derived scores, and the three summary scores according to age, sex, and education (Table 1). Performance for all cognitive measures decreased significantly with age. Women had significantly higher performance on Logical Memory Immediate and Delayed Recall, Paired-Associate Memory Immediate and Delayed Recall, TMT A, and all three summary scores. Participants with a college degree had significantly higher performance on all cognitive measures compared with participants with a high school degree or less. Participants who completed some college had significantly higher performance on Logical Memory Immediate and Delayed Recall, Visual Recall Immediate and Delayed Recall, Similarities, Boston Naming Test, and all three summary scores compared with participants with a high school degree or less.

Differentiating Between Cognitive Groups

Dementia versus NC—All three summary scores were highly accurate when differentiating between NC and demented participants. The AUC (the probability that a randomly selected dementia case has a lower score than a randomly selected NC control) among the three summary scores were all greater than 0.97 (Table 3 and Figure 1). The optimal cut point for the composite subtest summary score had significantly higher specificity and overall test accuracy compared with the composite domains summary score but not compared with the learning and retention summary score (Figure 1). These results provide evidence for the use of the composite subtests summary score for differentiating between NC and dementia participants.

ILM_{TB} versus NC—There were no significant differences in AUCs among the three summary scores (Table 4 and Figure 2). The optimal cut point for the composite domains summary score had the greatest AUC, sensitivity, npv, and overall test accuracy and the learning and retention summary score had the greatest specificity. However, these differences were not statistically significant compared with the other summary scores. Based on these results, there are no apparent differences between the three summary scores in their ability to differentiate between participants classified as ILM_{TB} and NC.

MDI_{TB} versus NC—The composite subtests summary score had significantly greater AUC, and overall test accuracy, compared with the composite domains and learning and retention summary scores (Table 5 and Figure 3). In addition, the composite domains summary score had a significantly greater AUC and overall test accuracy compared with the learning and retention summary score. The composite subtest summary score had significantly higher sensitivity, specificity, npv, and ppv compared with the learning and retention summary score (Table 5). Based on this evidence, the composite subtests summary score is best able to differentiate between participants classified as normal and MDI_{TB} and a summary score that is limited to assessments of learning and retention should not be used to differentiate between these cognitive groups.

ILM_{TB} versus dementia—The composite subtests summary score had the greatest AUC of the three summary scores, but the wide confidence interval due to the small number of participants diagnosed with dementia prevented any significant differences in AUC according to cognitive status from being detected (Table 6 and Figure 4). The optimal cut points for the composite sub-tests and composite domains summary scores both had significantly higher specificity and overall test accuracy compared with the learning and retention summary score. There were no significant differences in the ability of the composite subtests and composite domains summary scores when differentiating between dementia and ILM_{TB} participants. Based on this evidence, the learning and retention summary score is not as effective in differentiating between dementia and ILM_{TB} participants in the FHS Offspring Cohort.

MDI_{TB} versus dementia—There were no significant differences between the AUCs for any of the three summary scores (Table 7 and Figure 5). The optimal cut point for the learning and retention summary score had significantly higher specificity and overall test accuracy compared with the composite subtests and composite domains summary scores. The optimal cut point for the learning and retention score also had the highest ppv among the three summary scores, but this difference was not statistically significant. Based on this evidence, the learning and retention score is the most effective for differentiating between MDI_{TB} and dementia participants in the FHS Offspring Cohort.

ILM_{TB} and MDI_{TB}—The composite subtest summary score had significantly greater AUC compared with composite domains summary score and learning and retention summary score (Table 8 and Figure 6). Most noticeably, the learning and retention summary score was not able to differentiate between participants classified as ILM_{TB} and MDI_{TB} (AUC = 0.51 [0.46, 0.55]). This was expected as both cognitive groups could have impaired

functioning on subtests of learning, memory, and retention. The AUC for the composite domains summary score was also significantly greater than the learning and retention summary score. Furthermore, the optimal cut point for the learning and retention summary score had significantly lower specificity, ppv, npv, and overall test accuracy compared with the other two summary scores. Finally, the optimal cut point for the composite subtests summary score had significantly higher sensitivity compared with the learning and retention summary score. Based on this evidence, the composite subtest score is best able to differentiate between ILM_{TB} and MDI_{TB} due to the significantly higher AUC and sensitivity compared with the composite domains summary score.

Discussion

Summary

This study sought to calculate a series of three summary scores for the FHS neuropsychological battery and conduct a ROC analysis to determine which summary score was best able to differentiate between participants classified as having normal cognitive function, ILM_{TB}, MDI_{TB}, and dementia. The relatively high values for AUC, sensitivity, and specificity indicate that the optimal cut points for all three summary scores were able to accurately differentiate between normal and dementia participants, and between normal and MDI_{TB} participants. However, the optimal cut point for the composite subtests summary score performed best when differentiating between these groups based on the specificity, sensitivity, ppv, npv, and overall test accuracy. This suggests that not providing equal weight to the cognitive domains of the FHS neuropsychological battery and including subtests that assess additional cognitive domains other than learning and retention is beneficial when differentiating between normal and dementia, and normal and MDI_{TB} participants. All three summary scores accurately differentiated between MDI_{TB} and dementia participants based on the respective AUCs, but the optimal cut point for the learning and retention summary score had significantly higher specificity and overall test accuracy compared with the other two summary scores. These significant differences may be due to the optimal cut point being much lower than the composite subtests and domains summary score. Therefore, the learning and memory score was able to correctly identify more participants classified as MDI_{TB}, whereas the other two summary scores were better able to identify participants with dementia. This is reflected by the lower sensitivity of the learning and retention score compared with the other two summary scores.

The characteristics of the cognitive assessments included in the neuropsychological battery and how these assessments were utilized in each summary score may explain the observed differences in the ability of the three summary scores to differentiate between the four cognitive groups. Six of the 11 cognitive assessments included in the FHS neuropsychological battery and derived Memory Retention Score measured cognitive domains within memory. Several diagnostic criteria for dementia, including *DSM-III-R*, *DSM-IV* (APA, 1994), and NIA-AA workgroup (McKhann et al., 2011), require a patient to have impaired memory to receive a diagnosis of dementia. The method used to obtain the composite subtests summary score gave equal weight to each assessment included in the summary score while maintaining the emphasis on learning and memory domains.

Therefore, the emphasis placed on memory in the diagnosis of dementia and greater weight placed on learning and memory in the composite subtests score may explain why this summary score outperformed the composite domains summary score, which provided equal weight to all cognitive domains, when differentiating between participants classified as NC versus dementia. However, the inclusion of additional cognitive domains in the composite subtests and composite domains summary scores appeared to be beneficial, especially when differentiating between participants classified as dementia versus ILM_{TB}, and ILM_{TB} versus MDI_{TB}. This may be because ILM_{TB} participants were defined in this study as having impaired performance on assessments that measured learning and memory, but normal performance on all other cognitive assessments. Therefore, the higher performance by participants classified as ILM_{TB} on assessments that measured cognitive domains other than learning and memory compared with MDI_{TB} and dementia participants allowed for the composite subtests and composite domains summary scores to outperform the learning and retention summary score. The only instance in which the learning and retention summary score outperformed the other summary scores was when differentiating between MDI_{TB} versus dementia. This suggests that the inclusion of additional cognitive domains in the summary score does not improve the utility of the summary score, as was observed when differentiating between dementia versus ILM_{TB}, and ILM_{TB} versus MDI_{TB}.

There are unique cognitive characteristics that differentiate dementia subtypes (Desmond, 2004; Kraybill et al., 2005; Weintraub, Wicklund, & Salmon, 2012). These differences make it likely that the utility of the summary scores derived in the present analysis to differentiate between dementia and the other cognitive groups vary for dementia subtypes. The majority of dementia cases included in the final sample were attributed to AD, but the final sample also included cases of mixed dementia (AD and vascular dementia), AD with stroke, and dementia cases classified as other. As a result of the limited number of dementia cases in the FHS Offspring Cohort, the present study was unable to assess the utility of the summary scores to differentiate between each dementia subtype and other cognitive groups, or differentiate between different subtypes of dementia. As impaired learning and memory are symptoms of AD, it is likely that the ability of the summary scores employed in the current study to differentiate between AD and the other cognitive groups would be similar to the observed results when differentiating between dementia and the other cognitive groups. Future research should assess the ability of the summary scores utilized in the current study to differentiate between dementia subtypes and how results may differ according to dementia subtype.

Many population-based cohort studies assess cognitive functioning using a neuropsychological battery comprised of a collection of cognitive assessments that are included for reasons linked to the specific goals of the study. But there are considerable differences between studies in the number and type of cognitive assessments included in the neuropsychological battery. This makes it challenging to compare the results of the summary scores obtained in the present analysis to those obtained from other studies. A total score for the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological battery has been reported to accurately differentiate between older adults with NC, MCI, and AD (Chandler et al., 2005). Compared with the results reported by Chandler et al. (2005), the composite subtests summary score calculated from the FHS

study had a similar AUC, sensitivity, and specificity when differentiating between normal, cognitively impaired, and dementia participants. This suggests that a summary score that is the sum of the subtests included in the neuropsychological battery has similar utility compared with the composite subtest method used in the current study.

A total score was not calculated in the current study because this method does not take into account the differential weighting that each subtest has toward the total score due to the variation in the minimum and maximum possible scores for each subtest. As a result, a participant's total score may not accurately represent his or her overall level of cognitive functioning because the total score may be highly influenced by performance on only one or two subtests. Furthermore, a total score approach does not account for the different scales in which some cognitive assessments are scored. The FHS neuropsychological battery includes the TMT A and TMT B, which are timed assessments measured in either minutes or seconds. The relative weight of the TMT A and TMT B scores in a total summary score would vary substantially depending on if the time to complete the assessments was measured in seconds or minutes. A potential resolution would be to not include the TMT A and TMT B assessments in a total summary score, but these assessments are sensitive to the presence of dementia (Rasmusson, Zonderman, Kawas, & Resnick, 1998) and MCI (Ashendorf et al., 2008), and not including these assessments would potentially decrease the utility of the summary score.

Limitations

There are limitations of this study that should be acknowledged. First is the low number of participants who had received the neuropsychological battery and were diagnosed with dementia. A total of 1,251 participants in the validation sample completed a neuropsychological battery, but only 19 of these participants had received a diagnosis of dementia. The low prevalence of dementia is due to the relatively young age of the Offspring Cohort. This limited the ability to detect statistically significant differences between the three summary scores when classifying between dementia and the other cognitive groups. The low power was reflected by the wide confidence intervals for AUC, sensitivity, specificity, npv, ppv and overall test accuracy in the analyses that included participants with dementia. The low prevalence of dementia must also be taken into consideration when interpreting the ppv and npv for the summary scores because these metrics are influenced by disease prevalence (Altman & Bland, 1994). The ppv of a diagnostic measure increases as the prevalence of the disease increases, whereas the npv decreases as the prevalence of the disease increases. This means that in a population where the prevalence of a disease is low, a relatively high number of people will be incorrectly classified as having the disease compared with the number of people who test positive and truly have the disease. It should also be noted that the dementia cases included in the final sample are likely to be limited to participants in the mild stages of dementia as cases needed to have survived long enough postdiagnosis and be healthy enough to receive the neuropsychological battery. Therefore, these dementia cases may be biased and not include moderate or severe dementia cases.

A second limitation is that calculating specificity, sensitivity, ppv, npv, and overall accuracy of the three summary scores required comparing the classification results from the summary scores with those obtained from a predetermined gold standard. The gold standard for classifying participants as MDI_{TB} and ILM_{TB} was consistent with the diagnostic criteria for MCI established by the NIA-AA workgroup (Albert et al., 2011). The NIA-AA work-group emphasizes that persons diagnosed with MCI are able to maintain their ability to independently perform activities of daily living (ADLs), such as paying bills, cooking a meal, or shopping. Although the FHS includes measures for ADLs, this criterion was not included in the classification of MDI_{TB} and ILM_{TB} because ADL measures were not collected concurrently with when a participant received the neuropsychological battery. Not including the ability to independently perform ADLs in the diagnostic criteria raises the possibility that some participants classified as ILM_{TB} or MDI_{TB} may have dementia if their level of cognitive impairment limited their ability to independently perform ADLs.

A third limitation is the validity three summary scores could not be assessed. The validity of each summary score could be determined by correlating these scores with another cognitive assessment such as the MMSE. The MMSE is commonly used in clinical settings to assess cognitive functioning and screen for dementia. However, we were unable to correlate the summary scores with the MMSE because this assessment was not included as part of the neuropsychological battery. Also, we were unable to assess the test-retest reliability of the summary scores. The final sample included participants who received the neuropsychological battery during a follow-up examination, but no participant was reevaluated within 1 year.

Finally, the results from the ROC analysis may overestimate the ability of the summary scores to differentiate between the cognitive subgroups because the cognitive assessments used to classify participants as normal, MDI_{TB}, and ILM_{TB} were also used to create the three summary scores. We minimized biases that may have arisen from using this approach by defining NC, MDI_{TB}, and ILM_{TB} according to performance on individual cognitive assessments as opposed to using the summary scores. Also, we estimated normative scores for each cognitive assessment according to age, educational attainment, and gender using a random subsample of the FHS Offspring Cohort who received the neuropsychological battery.

Conclusion

To summarize, all three summary scores were able to accurately differentiate between participants classified as cognitively normal and dementia, ILM_{TB} and dementia, MDI_{TB} and dementia, and with the exception of the learning and retention summary score, ILM_{TB}, and MDI_{TB}. The learning and retention summary score performed best when differentiating between MDI_{TB} and dementia. Collectively, these findings suggest that composite subtests summary score has the highest utility when differentiating between normal and dementia participants, normal and MDI_{TB} participants, and ILM_{TB} and dementia participants, but the learning and retention summary score should be employed when differentiating between MDI_{TB} and dementia participants.

Acknowledgments

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute on Aging (K25AG043546, P30AG028383, 5R01AG019241, 5R01AG038651). The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University.

References

- Abner EL, Kryscio RJ, Cooper GE, Fardo DW, Jicha GA, Mendiondo MS, Schmitt FA. Mild cognitive impairment: Statistical models of transition using longitudinal clinical data. *International Journal of Alzheimer's Disease*. 2012:Article 291920.
- Albert MS, Dekosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Phelps CH. 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: The Journal of the Alzheimer's Association*. 2011; 7:270–279.
- Altman DG, Bland JM. Diagnostic tests 3: Receiver operating characteristic plots. *British Medical Journal*. 1994; 309:188. [PubMed: 8044101]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3. Washington, DC: Author; 1987. rev
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 1994.
- Armitage SG. An analysis of certain psychological tests used for the evaluation of brain injury. *Psychological Monographs*. 1946; 60(1):i–48.
- Ashendorf L, Jefferson AL, Green RC, Stern RA. Test-retest stability on the WRAT-3 reading subtest in geriatric cognitive evaluations. *Journal of Clinical and Experimental Neuropsychology*. 2009; 31:605–610. [PubMed: 18821160]
- Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*. 2008; 23(2):129–137. [PubMed: 18178372]
- Bachman DL, Wolf PA, Linn R, Knoefel JE, Cobb J, Belanger A, White LR. Prevalence of dementia and probable senile dementia of the Alzheimer type in the Framingham Study. *Neurology*. 1992; 42:115–119. [PubMed: 1734291]
- Bachman DL, Wolf PA, Linn RT, Knoefel JE, Cobb JL, Belanger AJ, D'Agostino RB. Incidence of dementia and probable Alzheimer's disease in a general population: The Framingham Study. *Neurology*. 1993; 43(Pt. 1):515–519. [PubMed: 8450993]
- Boyd JL. A validity study of the Hooper Visual Organization Test. *Journal of Consulting and Clinical Psychology*. 1981; 49:15–19. [PubMed: 7217472]
- Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, Cullum CM. A total score for the CERAD neuropsychological battery. *Neurology*. 2005; 65:102–106. [PubMed: 16009893]
- Coffin M, Sukhatme S. Receiver operating characteristic studies and measurement errors. *Biometrics*. 1997; 53:823–837. [PubMed: 9333348]
- Cummings JL, Benson DF. Dementia of the Alzheimer type. An inventory of diagnostic clinical features. *Journal of the American Geriatrics Society*. 1986; 34(1):12–19. [PubMed: 3941239]
- Desmond DW. The neuropsychology of vascular cognitive impairment: Is there a specific cognitive deficit? *Journal of the Neurological Sciences*. 2004; 226(1–2):3–7.10.1016/j.jns.2004.09.002 [PubMed: 15537510]
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Preventive Medicine*. 1975; 4:518–525. [PubMed: 1208363]
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]

- Gill DM, Reddon JR, Stefanyk WO, Hans HS. Finger tapping: Effects of trials and sessions. *Perceptual & Motor Skills*. 1986; 62:675–678. [PubMed: 3503282]
- Goodglass, H.; Kaplan, E. *The assessment of aphasia and related disorders*. 2. Philadelphia, PA: Lea & Febiger; 1983.
- Hooper, HE. *The Hooper Visual Organization Test*. Torrance, CA: Western Psychological Services; 1966.
- Huff FJ, Collins C, Corkin S, Rosen TJ. Equivalent forms of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology*. 1986; 8:556–562. [PubMed: 3805252]
- Jack CR Jr, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, Petersen RC. An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. *Annals of Neurology*. 2012; 71:765–775.10.1002/ana.22628 [PubMed: 22488240]
- Johnson DK, Storandt M, Morris JC, Galvin JE. Longitudinal study of the transition from healthy aging to Alzheimer disease. *Archives of Neurology*. 2009; 66:1254–1259. [PubMed: 19822781]
- Karrasch M, Sinerva E, Gronholm P, Rinne J, Laine M. CERAD Test performances in amnesic mild cognitive impairment and Alzheimer's disease. *Acta Neurologica Scandinavica*. 2005; 111:172–179.10.1111/j.1600-0404.2005.00380.x [PubMed: 15691286]
- Knopman DS, Jack CR Jr, Wiste HJ, Weigand SD, Vemuri P, Lowe V, Petersen RC. Short-term clinical outcomes for stages of NIA-AA preclinical Alzheimer disease. *Neurology*. 2012; 78:1576–1582.10.1212/WNL.0b013e3182563bbe [PubMed: 22551733]
- Kraybill ML, Larson EB, Tsuang DW, Teri L, McCormick WC, Bowen JD, Cherrier MM. Cognitive differences in dementia patients with autopsy-verified AD, Lewy body pathology, or both. *Neurology*. 2005; 64:2069–2073.10.1212/01.WNL.0000165987.89198.65 [PubMed: 15985574]
- Langbaum JB, Hendrix SB, Ayutyanont N, Chen K, Fleisher AS, Shah RC, Reiman EM. An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. 2014; 10:666–674.
- Mathews M, Abner E, Kryscio R, Jicha G, Cooper G, Smith C, Schmitt FA. Diagnostic accuracy and practice effects in the National Alzheimer's Coordinating Center Uniform Data Set neuropsychological battery. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. 2014; 10:675–683.
- McKhann G, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Phelps CH. 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: The Journal of the Alzheimer's Association*. 2011; 7:263–269.
- Rasmusson XD, Zonderman AB, Kawas C, Resnick SM. Effects of age and dementia on the trail making test. *The Clinical Neuropsychologist*. 1998; 12:169–178.
- Riley KP, Jicha GA, Davis D, Abner EL, Cooper GE, Stiles N, Schmitt FA. Prediction of preclinical Alzheimer's disease: Longitudinal rates of change in cognition. *Journal of Alzheimer's Disease*. 2011; 25:707–717.
- Russell EW. A multiple scoring method for the assessment of complex memory functions. *Journal of Consulting and Clinical Psychology*. 1975; 43:800–809.
- Ryan JJ, Ward LC. Validity, reliability, and standard errors of measurement of two seven subtest short forms of the Weschler Adult Intelligence Scale-III. *Psychological Assessment*. 1999; 11:207–211.
- Salthouse TA. What cognitive abilities are involved in trail-making performance? *Intelligence*. 2011; 39:222–232. [PubMed: 21789028]
- Sanchez-Cubillo I, Perianez JA, Adrover-Roig D, Rodriguez-Sanchez JM, Rios-Lago M, Tirapu J, Barcelo F. Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*. 2009; 15:438–450. [PubMed: 19402930]
- Shankle WR, Romney AK, Hara J, Fortier D, Dick MB, Chen JM, Sun X. Methods to improve the detection of mild cognitive impairment. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:4919–4924.10.1073/pnas.0501157102 [PubMed: 15781874]
- Shimoyama I, Ninchoji T, Uemura K. The Finger-Tapping Test. A quantitative analysis. *Archives of Neurology*. 1990; 47:681–684. [PubMed: 2346396]

- Wechsler D. A Standardized Memory Scale for clinical use. *The Journal of Psychology: Interdisciplinary and Applied*. 1945; 19:87–95.
- Wechsler, D. Wechsler Adult Intelligence Scale. New York, NY: Psychological; 1955.
- Weintraub S, Wicklund AH, Salmon DP. The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*. 2012; 2(4):a006171.10.1101/cshperspect.a006171 [PubMed: 22474609]
- Wilkinson, GS. Wide Range Achievement Test 3—Administration manual. Wilmington, DE: Jastak; 1993.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
ROC curves differentiating between cognitively normal and demented participants.
Note. ROC = receiver operator characteristic.

Figure 2.

ROC curves differentiating between cognitively normal and ILM_{TB} participants.

Note. ROC = receiver operator characteristic; ILM_{TB} = test-based impaired learning and memory.

Figure 3.

ROC curves differentiating between cognitively normal and MDI_{TB} participants.

Note. ROC = receiver operator characteristic; MDI_{TB} = test-based multidomain impairment.

Figure 4.

ROC curves differentiating between demented and ILM_{TB} participants.

Note. ROC = receiver operator characteristic; ILM_{TB} = test-based impaired learning and memory.

Figure 5.

ROC curves differentiating between demented and MDI_{TB} participants.

Note. ROC = receiver operator characteristic; MDI_{TB} = test-based multidomain impairment.

Figure 6.

ROC curves differentiating between ILM_{TB} and MDI_{TB} participants.

Note. ROC = receiver operator characteristic; ILM_{TB} = test-based impaired learning and memory; MDI_{TB} = test-based multidomain impairment.

Table 1

Effects of Age, Sex, and Education on Cognitive Performance.

Cognitive measure	Age	Sex	Some college	College degree
LMI	-0.041 [-0.065, -0.017]	1.18 [0.73, 1.62]	0.90 [0.31, 1.48]	2.21 [1.69, 2.74]
LMD	-0.050 [-0.075, -0.025]	1.36 [0.89, 1.82]	1.09 [0.49, 1.70]	2.29 [1.74, 2.83]
PASI	-0.076 [-0.10, -0.052]	1.55 [1.10, 2.00]	0.44 [-0.15, 1.02]	1.30 [0.78, 1.83]
PASD	-0.033 [-0.043, -0.023]	0.60 [0.40, 0.79]	0.062 [-0.19, 0.31]	0.42 [0.20, 0.65]
VRI	-0.099 [-0.12, -0.077]	-0.052 [-0.46, 0.36]	0.85 [0.31, 1.38]	1.53 [1.05, 2.01]
VRD	-0.10 [-0.12, -0.072]	0.04 [-0.40, 0.48]	0.95 [0.38, 1.53]	1.75 [1.23, 2.27]
TMT A	-0.0056 [-0.0071, -0.0041]	0.038 [0.0098, 0.065]	0.031 [-0.0048, 0.067]	0.045 [0.013, 0.078]
TMT B	-0.022 [-0.026, -0.018]	0.028 [-0.051, 0.11]	0.084 [-0.019, 0.19]	0.24 [0.15, 0.33]
HVOT	-0.12 [-0.14, -0.095]	0.36 [-0.04, 0.76]	0.19 [-0.33, 0.71]	0.68 [0.21, 1.15]
Similarities	-0.068 [-0.091, -0.045]	0.074 [-0.36, 0.50]	1.37 [0.81, 1.92]	2.86 [2.36, 3.37]
BNT	-0.050 [-0.066, -0.034]	-0.20 [-0.50, 0.10]	0.96 [0.57, 1.34]	1.73 [1.38, 2.09]
TMT (B-A)	-0.017 [-0.020, -0.013]	-0.0096 [-0.077, 0.058]	0.053 [-0.035, 0.14]	0.20 [0.12, 0.27]
Memory retention score	<i>-0.0015</i> [-0.0029, -1.9e ⁻⁴]	0.015 [-0.010, 0.040]	0.022 [-0.011, 0.054]	0.023 [-0.0063, 0.054]
Composite subtest	-0.019 [-0.022, -0.016]	0.13 [0.076, 0.19]	0.17 [0.097, 0.25]	0.38 [0.31, 0.45]
Composite domains	-0.026 [-0.030, -0.021]	0.23 [0.15, 0.31]	0.19 [0.083, 0.29]	0.45 [0.36, 0.54]
Learning and retention composite	-0.019 [-0.023, -0.014]	0.23 [0.15, 0.30]	0.19 [0.088, 0.29]	0.41 [0.32, 0.50]

Note. Estimates for the effects of age, sex, and education were obtained using data from participants included in the training set ($n = 1,252$). Referent category for sex is male. Referent category for education was high school degree or less; LMI = Logical Memory, Immediate Recall; LMD = Logical Memory, Delayed Recall; PASI = Paired Associates, Immediate Recall; PASD = Paired Associates, Delayed Recall; VRI = Visual Reproductions, Immediate Recall; VRD = Visual Reproductions, Delayed Recall; TMT A (B) = Trail Making Test A (B); HVOT = Hooper Visual Organization Test; BNT = Boston Naming Test; TMT A, B, and TMT B-A are timed assessments; higher score indicates lower performance. Scores for TMT A, TMT B, and TMT B-A were multiplied by -1 so that higher scores reflected higher cognition. **Bold** $p < .01$, *italics* $p < .05$, but not $p < .01$.

Demographic Characteristics of Normal, Impaired, and Demented Participants Who Completed the Neuropsychological Battery.

Table 2

Characteristic	Normal (<i>n</i> = 564)	ILM (<i>n</i> = 269)	MDI (<i>n</i> = 399)	Dementia (<i>n</i> = 19)	<i>p</i> value
Age (<i>SD</i>)	60.3 (9.2)	60.5 (8.7)	63.5 (9.8)	73.5 (5.6)	<.01
Education (%)					.01
<High school	170 (30.1)	91 (33.8)	164 (41.1)	8 (42.1)	
Some college	144 (25.5)	77 (28.6)	103 (25.8)	6 (31.6)	
College degree	250 (44.3)	101 (37.5)	132 (33.1)	5 (26.3)	
Sex					.45
Male	260 (46.1)	119 (44.2)	197 (49.4)	7 (36.8)	
Female	304 (53.9)	150 (55.8)	202 (50.6)	12 (63.2)	
<i>APOE</i>					<.01
e4−	197 (79.1)	97 (77.6)	171 (79.5)	3 (17.6)	
e4+	52 (20.9)	28 (22.4)	44 (20.5)	14 (82.4)	

Note. Results based on participants included in the validation set. Analysis of covariance and chi-square tests used to assess differences in continuous and categorical variables, respectively. ILM = Impaired learning and memory; MDI = Multidomain impairment; *APOE* = apolipoprotein E.

Bold font indicates statistical significance $p < .05$.

Table 3

Differentiating Between Normal and Dementia Participants.

	Composite subtest	Composite domains	Learning and retention
AUC	0.98 [0.95, 1.00]	0.97 [0.94, 1.00]	0.97 [0.94, 1.00]
Cutoff	−0.16	−0.13	−0.07
Sensitivity	0.89 [0.67, 0.99]	0.89 [0.67, 0.99]	0.89 [0.67, 0.99]
Specificity	0.96 [0.94, 0.97] ^{a,b}	0.92 [0.90, 0.94]	0.88 [0.85, 0.91]
PPV	0.41 [0.26, 0.58]	0.28 [0.17, 0.41]	0.20 [0.12, 0.30]
NPV	1.00 [0.99, 1.00]	1.00 [0.99, 1.00]	1.00 [0.99, 1.00]
Test accuracy	0.96 [0.94, 0.97] ^{a,b}	0.92 [0.90, 0.94]	0.88 [0.85, 0.91]

Note. AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

^aSignificant difference compared with composite domains summary score.

^bSignificant difference compared with learning and retention summary score.

Table 4Differentiating Between Normal and ILM_{TB}.

	Composite subtest	Composite domains	Learning and retention
AUC	0.76 [0.72, 0.80]	0.80 [0.76, 0.83]	0.78 [0.74, 0.81]
Cutoff	0.17	0.21	0.15
Sensitivity	0.72 [0.67, 0.78]	0.74 [0.68, 0.79]	0.71 [0.65, 0.76]
Specificity	0.69 [0.65, 0.73]	0.74 [0.70, 0.77]	0.75 [0.71, 0.78]
PPV	0.53 [0.47, 0.58]	0.57 [0.52, 0.62]	0.57 [0.52, 0.62]
NPV	0.84 [0.80, 0.87]	0.86 [0.82, 0.89]	0.84 [0.81, 0.87]
Test accuracy	0.70 [0.67, 0.73]	0.74 [0.71, 0.77]	0.73 [0.70, 0.76]

Note. AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

^a Significant difference compared with composite domains summary score.

^b Significant difference compared with learning and retention summary score.

Table 5Differentiating Between Normal and MDI_{TB} Participants.

	Composite subtest	Composite domains	Learning and retention
AUC	0.90 [0.88, 0.92] ^{a,b}	0.86 [0.83, 0.88] ^b	0.79 [0.76, 0.82]
Cutoff	0.03	0.13	0.16
Sensitivity	0.78 [0.74, 0.82] ^b	0.75 [0.71, 0.79]	0.69 [0.64, 0.74]
Specificity	0.85 [0.82, 0.88] ^b	0.79 [0.76, 0.83]	0.74 [0.71, 0.78]
PPV	0.79 [0.74, 0.83] ^b	0.72 [0.68, 0.76]	0.66 [0.61, 0.70]
NPV	0.85 [0.82, 0.88] ^b	0.82 [0.78, 0.85]	0.77 [0.74, 0.81]
Test accuracy	0.82 [0.80, 0.85] ^{a,b}	0.78 [0.75, 0.80] ^b	0.72 [0.69, 0.75]

Note. MDI_{TB} = test-based multidomain impairment; AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

^aSignificant difference compared with composite domains summary score.

^bSignificant difference compared with learning and retention summary score.

Table 6Differentiating Between ILM_{TB} and Dementia Participants.

	Composite subtest	Composite domains	Learning and retention
AUC	0.92 [0.84, 1.00]	0.90 [0.80, 1.00]	0.84 [0.74, 0.95]
Cutoff	−0.41	−0.61	−0.34
Sensitivity	0.84 [0.60, 0.97]	0.84 [0.60, 0.97]	0.79 [0.54, 0.94]
Specificity	0.90 [0.86, 0.93] ^b	0.92 [0.88, 0.95] ^b	0.70 [0.64, 0.75]
PPV	0.37 [0.23, 0.53]	0.43 [0.27, 0.61]	0.16 [0.09, 0.24]
NPV	0.99 [0.96, 1.00]	0.99 [0.97, 1.00]	0.98 [0.95, 0.99]
Test accuracy	0.90 [0.85, 0.93] ^b	0.92 [0.88, 0.95] ^b	0.70 [0.65, 0.76]

Note. ILM_{TB} = test-based impaired learning and memory; AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

^aSignificant difference compared with composite domains summary score.

^bSignificant difference compared with learning and retention summary score.

Table 7Differentiating Between MDI_{TB} and Dementia Participants.

	Composite subtest	Composite domains	Learning and retention
AUC	0.85 [0.73, 0.97]	0.86 [0.74, 0.98]	0.84 [0.74, 0.94]
Cutoff	−0.70	−1.03	−1.14
Sensitivity	0.79 [0.54, 0.94]	0.79 [0.54, 0.94]	0.63 [0.38, 0.84]
Specificity	0.85 [0.81, 0.89]	0.89 [0.85, 0.92]	0.96 [0.94, 0.98] ^{a,b}
PPV	0.20 [0.12, 0.31]	0.25 [0.15, 0.38]	0.46 [0.27, 0.67]
NPV	0.99 [0.97, 1.00]	0.99 [0.97, 1.00]	0.98 [0.96, 0.99]
Test accuracy	0.85 [0.81, 0.88]	0.88 [0.85, 0.91]	0.95 [0.92, 0.97] ^{a,b}

Note. MDI_{TB} = test-based multidomain impairment; AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

^aSignificant difference compared with composite domains summary score.

^bSignificant difference compared with learning and retention summary score.

Table 8Differentiating Between ILM_{TB} and MDI_{TB} Participants.

	Composite subtest	Composite domains	Learning and retention
AUC	0.70 [0.67, 0.74] ^{a,b}	0.62 [0.58, 0.67] ^b	0.51 [0.46, 0.55]
Cutoff	−0.15	−0.23	−0.08
Sensitivity	0.62 [0.57, 0.67] ^b	0.53 [0.48, 0.58]	0.51 [0.46, 0.56]
Specificity	0.69 [0.63, 0.74] ^b	0.67 [0.61, 0.72] ^b	0.54 [0.47, 0.60]
PPV	0.75 [0.70, 0.79] ^b	0.70 [0.65, 0.75] ^b	0.62 [0.56, 0.67]
NPV	0.55 [0.49, 0.60] ^b	0.49 [0.44, 0.54] ^b	0.42 [0.37, 0.48]
Test accuracy	0.65 [0.61, 0.68] ^b	0.59 [0.55, 0.62] ^b	0.52 [0.48, 0.56]

Note. ILM_{TB} = test-based impaired learning and memory; MDI_{TB} = test-based multidomain impairment; AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

^aSignificant difference compared with composite domains summary score.

^bSignificant difference compared with learning and retention summary score.