Feasibility of an Online Consensus Approach for the Diagnosis of Cognitive Impairment and Dementia in Rural South Africa

Darina T Bassil[1*], Meagan T Farrell[1], Albert Weerman[2], Muqi Guo[1], Ryan G Wagner[3], Adam M Brickman[4,5,6], M Maria Glymour[7], Kenneth M Langa[8,9,10,11], Jennifer J Manly[4,5,6], Brent Tipping[12], India Butler[12], Stephen Tollman[3] and Lisa F Berkman[1,13]

1. Harvard Center for Population and Development Studies, Harvard University, Cambridge, MA, USA, 02138.
2. Center for Economic and Social Research, University of Southern California, Los Angeles, CA, USA, 90089.
3. MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, 2193.
4. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA, 10032.
5. Gertrude H. Sergievsky Center, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA, 10032.
6. Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA, 10032.
7. Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA, 94158.
8. Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, USA, 48109.
9. Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA, 48109.
10. Institute for Social Research, University of Michigan, Ann Arbor, MI , USA, 48109.
11. Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI, USA, 48109.
12. Division of Geriatric Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, 2050.
13. Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA, 02115.

* Corresponding Author: Darina T Bassil, Harvard Center for Population and Development Studies, Harvard University, 9 Bow Street, Cambridge, MA, USA, 02138. Email: dbassil@hsph.harvard.edu

Conflict of Interest: JJM received grants from NIH, NIA, NIND, NHLBI (institutional payment) and support from Alzheimer's association for conference attendance. AMB received consulting fees from Cognition Therapeutics, holds patents (including US Patent US9867566B2 and U.S. provisional patent application no.: 63/146,958) and serves on the advisory boards for Albert Einstein College of Medicine and Cogstate (SAB). MMG received grants from NIH/NIA and the Robert Wood Johnson Foundation, royalties from the Oxford University Press, and participates in the SWAN (Study of Women Across the Nation) data safety monitoring board. KML received grants from the NIA and Alzheimer's Association (institutional payment), consulting fees from Harvard, University of Pennsylvania, University of Minnesota, University of Colorado, Dartmouth, University of Southern California, and payment for expert testimony on a DSMB for a clinical trial at Indiana University. BT received consulting fees from Aurai Senior Living and the University of the Witwatersrand Donald Gordon Medical Centre, as well as payments for lectures from the Department of Family Medicine – University of Pretoria and Astellas pharmaceutical. BT also serves on the executive committees of the Faculty of Consulting Physicians of South Africa (FCPSA), South African Geriatrics Society (SAGS), Wits Donald Gordon, and the Department of Internal Medicine of the University of the Witwatersrand. All reported conflicts of interest are outside the scope of this published work. The remaining authors have no conflicts of interest to report.

Consent Statement: All participants gave their informed consent to participate in the HAALSI Dementia Study, in accordance with the ethical standards set by the relevant institutional review boards.

Abstract

INTRODUCTION: We describe the development and feasibility of using an online consensus approach for diagnosing cognitive impairment and dementia in rural South Africa.

METHODS: Cognitive assessments, clinical evaluations, and informant interviews from Cognition and Dementia in the Health and Aging in Africa Longitudinal Study (HAALSI Dementia) were

reviewed by an expert panel using a web-based platform to assign a diagnosis of cognitively normal, mild cognitive impairment (MCI), or dementia.

RESULTS: 635 participants were assigned a final diagnostic category, with 298 requiring adjudication conference calls. Overall agreement between each rater's independent diagnosis and final diagnosis (via the portal or consensus conference) was 78.3%. A moderate level of agreement between raters' individual ratings and the final diagnostic outcomes was observed (average κ coefficient= 0.50).

DISCUSSION: Findings show initial feasibility in using an online consensus approach for the diagnosis of cognitive impairment and dementia in remote, rural and low-resource settings.

Keywords:

Harmonized Cognitive Assessment Protocol; consensus; dementia diagnosis, mild cognitive impairment; clinical decision-making; South Africa; population-based research

1 Introduction

Valid and reliable assessments of dementia are essential to advance research, yet difficult to implement in many settings where research is most urgently needed. To improve dementia

diagnostic reliability and accuracy, many research studies recommend a consensus panel approach for the diagnosis of dementia.[1-4] In this approach, a panel of expert clinicians meet to review and discuss a study participant's cognitive and clinical profile to adjudicate the diagnosis of dementia,[3, 5] using standardized criteria such as the National Institute on Aging-Alzheimer's Association diagnostic guidelines for Alzheimer's disease (NIA-AA)[6] and the Diagnostic and Statistical Manual of Mental Disorders (DSM-5).[7] Consensus panels could be especially useful in low- and middle-income countries (LMICs) where population-based research studies using detailed neuropsychological tests remain scarce, with limited to no availability of comprehensive normative population data.[8-10] However, in-person consensus conferences can be very costly, time consuming and impractical, especially in remote and rural areas. Few studies have evaluated the feasibility, validity, and reliability of using a web-based consensus approach for the diagnosis of cognitive impairment and dementia.[3, 11]

In this paper, we use data from our Cognition and Dementia in the Health and Aging in Africa Longitudinal Study of an INDEPTH community in South Africa (HAALSI Dementia),[12] a harmonized sister study to the United States Health and Retirement Study (HRS) Harmonized Cognitive Assessment Protocol (HCAP) Project,[13] and implement a rigorous and robust web-based online consensus system for remote capture of diagnostic cognitive outcomes. We built our web interface based on similar systems deployed in the Longitudinal Aging Study in India- Diagnostic Assessment of Dementia (LASI-DAD)[11] and the Monogahela-Youghiogheny Health Aging Team study (MYHAT).[3] We first describe detailed information on the web interface design, diagnostic criteria for dementia ascertainment, and the consensus expert panel. We also report baseline characteristics and distribution of final diagnostic outcomes for our cohort, assess the feasibility of using a multidisciplinary web-based consensus conference approach in rural South Africa, and identify the key factors responsible for diagnostic variability among raters. A successful web-based consensus

panel could be a novel, efficient and less expensive approach to facilitate the diagnosis of cognitive impairment and dementia, especially in LMICs.

2 Methods

2.1 Study Cohort

HAALSI Dementia is an ongoing prospective cohort study on a stratified sub-sample of 635 HAALSI participants aged 50 and older in Agincourt, a rural South African Community. The study design and methods have been previously described.[12] In brief, HAALSI Dementia collects detailed neuropsychological and functional assessments, informant interviews, and neurological and clinical evaluations to enable cross-cultural comparison and cross-calibration with international HCAP and HRS studies. Furthermore, HAALSI Dementia is strategically oversampled for respondents categorized as higher risk for dementia to allow future estimation of dementia prevalence in the parent HAALSI cohort.

The first wave of data collection was completed in January 2020. Data were reviewed by an expert consensus panel to assign diagnosis of cognitively normal, mild cognitive impairment (MCI), or dementia, based on the NIA-AA criteria.[6]

This study was approved by University of the Witwatersrand Human Research Ethics Committee (ref.M190443), Harvard T.H. Chan School of Public Health Institutional Research Ethics Board (ref.18–1459,19–1396) and Mpumalanga Provincial Research and Ethics Committee.

2.2 Online Consensus Interface – Design

Our online consensus website was modelled after LASI-DAD[11] and MYHAT[3] web portals. Both studies showed the online consensus system to be an efficient, feasible and valid approach for the clinical diagnosis of dementia. Following the same format and structure as LASI-DAD, our web interface entailed two main pages: assessments and ratings.

Raters independently reviewed all the data on the assessments page, including demographic, cognitive, neurological and informant measures (Supplemental Figure 1). Next, raters navigated to the ratings page and worked through the NIA-AA criteria to submit their final diagnosis (Supplemental Figure 2).

Research participants were anonymous and identified only by a unique case ID. Raters were provided with unique rater IDs to access the web portal securely. The web portal randomly selected and assigned cases for each rater to review. Cases were reviewed one at a time, and any selections made by raters on the web page were autosaved. All recorded data from the site were autogenerated in a database and securely exported to STATA for data analysis and management.

2.2.1 Assessments: Cognitive battery, informant interview and clinical examination

All assessments were administered and captured via tablets using Computer-Assisted Personal Interviewing (CAPI) software. The respondents' demographic information (age, sex, education) and limited health history were made available on the website. Any narrative text or remarks recorded by the nurses or fieldworkers during visits were translated and displayed on the page. Data on administered cognitive measures, informant scales and neurological findings were cleaned and managed in STATA and imported securely to the online assessments page (Figure 1). Full details on individual cognitive, informant and neurological measures have been published.[12]

Cognitive battery: HAALSI Dementia cognitive measures and scores were shown under the respondent's assessments column and were categorized by their principal domain as follows: (a) General cognitive status: Mini Mental State Examination (MMSE),[14] days of the week,[15] Telephone Interview for Cognitive Status (TICS),[16] Community Screening Inventory for Dementia (CSID)[17] and Clinical Dementia Rating (CDR) semi-structured interview;[18] (b) Episodic memory: Immediate and delayed word recall,[19] word list recognition,[19] logical memory (story recall) immediate,[20] delayed and recognition;[20] (c) Executive function and attention: Motor sequences,[21] go no/go,[22] similarities and differences[23] and raven's standard progressive matrices;[24] (d) Visuospatial/Spatial memory: Constructional praxis,[19] constructional praxis recall,[19] spatial working memory,[25] symbol cancellation[26] and clock draw;[27] (e) Language: Boston Naming Test,[28] semantic fluency,[29] Token Test[30] and phoneme (letter) fluency.[31] The respondents' assessments column also included a "mood" sub-heading which outlined the respondent's self-reported measure of depression based on the Center for Epidemiologic Studies Depression (CES-D) scale.[32]

In the absence of normative population data for our cohort and in an effort to aid the raters with their diagnostic decisions, the following items were also shown for each cognitive measure: (1) weighted means to provide a crude estimate of the participants' performance relative to the average performance of individuals aged 50 and over across the parent HAALSI cohort; (2) percent of the sample missing the cognitive test; and (3) percent of the sample who performed at minimum and maximum score range. To assess cognitive decline over time, raters were also shown cognitive scores from the word recall task administered during parent HAALSI waves 1 (2014-2015) and 2 (2018-2019). For cognitive tasks that required participants to draw or write on the tablets (constructional praxis, clock draw and MMSE), raters were able to see and evaluate the actual drawings completed by participants (Supplemental Figure 3).

As part of the cognitive assessment, respondents also completed a reading assessment to assess basic literacy (letter and number recognition) through secondary level reading skills. Literacy scores were displayed alongside the self-reported education level to assist raters with the interpretation of cognitive test results.

Informant Interview: The informant section summarized key demographic and background information (relationship to respondent, years known, frequency informant sees respondent) and presented all informant scales including Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE),[33] Blessed Dementia Rating Scale,[34] CSI-D Cognitive Activities Questionnaire,[17] 10/66 Dementia Research Group Informant Questionnaire,[35] HRS Activities Questionnaire[13] and CDR semi-structured interview.[18] To help raters navigate through the informant data, different elements from the aforementioned scales were organized under the following headings: (a) Respondent's mental status for information on changes in respondent's cognitive abilities, judgement, problem solving and personal care activities; (b) Respondent's activities for changes in Activities of Daily Living (ADLs), Instrumental ADLs (IADLs), leisure activities as well as a general overview of respondent's social engagement and community affairs; and (c) Respondent's behavior for any behavioral and mood changes reported by the informant (Supplemental Figure 1).

Clinical and Neurological Examination: All clinical data were reported under the health history column on the assessments page. Clinical findings were summarized as normal/ abnormal for each of the following evaluated neurological exam components: face (muscle weakness, swallow, and tongue mobility), speech, upper and lower limbs mobility, balance, gait, vision, hearing and any other observations noted by the nurse. Raters were also able to click on any of the neurological exam components for a more detailed overview of clinical findings. Other reported information

included blood pressure, pulse, grip strength and timed up and go readings as well as information on relevant medical history and medications (Supplemental Figure 1).

2.2.2 Ratings: Diagnostic Criteria for Dementia Ascertainment

The assessments page was followed by the ratings page which showed a series of questions to determine the presence and severity of dementia. The primary diagnostic outcome was determined using the NIA-AA criteria,[6] whereby raters had to answer yes/no to impairment for each NIA-AA criteria item (Supplemental Figure 2). Raters could move freely between the assessments and ratings pages during their review, but were not allowed to move to another case before submitting their rating.

Once the rater responded to all criteria items, the web portal applied the responses to a preprogrammed algorithm designed by our team, to automatically generate the final diagnosis. Using the NIA-AA algorithm, diagnostic outcomes were developed based on the following rules:

- *Cognitively normal:* No evidence of cognitive or functional impairment based on objective testing or informant report.
- *MCI:* Evidence of cognitive impairment according to objective testing or informant report, but not meeting criteria for dementia, as described in the published work from the Aging, Demographics, and Memory Study (ADAMs) research team[36-39].
- *Dementia:* Evidence of cognitive impairment according to both the objective tests and informant report and additionally, the participant had 1) evidence that the cognitive impairment interfered with social and/or occupational function, 2) had experienced cognitive decline based on previous parent HAALSI waves, 3) had cognitive

impairments that could not be explained by delirium or psychiatric disorders, and 4) had cognitive impairments in at least two cognitive domains.

Prior to submitting the generated final diagnosis for each case, raters additionally: 1) provided an overall CDR score (0, 0.5, 1, 2, 3) to rate the severity of cognitive and functional impairment, 2) rated the presence of Aphasia, Agnosia, and Apraxia to enable diagnostic comparisons with other studies which used the DSM-5 criteria for dementia diagnosis, and 3) provided their level of confidence in the rating (not at all certain, only slightly certain, reasonably certain, very little doubt, absolutely certain).

2.3 Consensus Panel

The raters panel consisted of U.S. and South African clinicians with expertise in evaluating individuals with possible cognitive impairment or dementia. Experts on the panel included neurologists, geriatricians, neuropsychologists as well as researchers with specialized training in the field of dementia. Selected raters had expertise in the clinical and research settings of cognition and dementia, in diverse populations, as well as experience with the neuropsychological assessments and diagnostic criteria employed in the study.

All raters completed an online training session on how to navigate the web portal and submit ratings. During training, raters were also required to independently submit ratings for a number of pilot cases and provide feedback on their experience with the portal. Revisions to the web portal were made during the training phase only. Additionally, a detailed manual was provided to all raters , including guidance on the consensus portal, detailed descriptions of cognitive tests and other measures, and an overview of the diagnostic criteria used.

For the online consensus review, each case was reviewed by three raters, randomly chosen at the time of review. Cases were completed at raters' own pace, so some raters completed more cases than the others. All three raters needed to agree on the final NIA-AA diagnosis (cognitively normal, MCI, or dementia). Cases for which there was disagreement on the final diagnostic category among the three independent reviewers were discussed during biweekly consensus conference calls. Consensus conference calls, which were led by a moderator, were open to all raters on the committee, not just raters who originally reviewed the case, and typically had between 3 to 6 raters in attendance. For each discrepant case, the moderator presented all relevant information including the neuropsychological testing profile, informant interview and clinical observations from the neurological exam to the consensus panel. Members of the consensus panel were also able to log in and see the details of the case currently under discussion. The moderator then focused the discussion on the specific item(s) that led to disagreement between the three reviewers and allowed all consensus members to discuss the case and agree on a final diagnosis (Figure 1).

2.4 Statistical Analyses

Distributions of diagnostic outcomes (cognitively normal, MCI, dementia) by age and sex were summarized with frequency counts and percentages. Differences in cognitive measures between diagnostic outcomes were assessed using Chi-square tests and one-way analyses of variance (ANOVA), where appropriate. Overall concordance rate was measured as overall agreement between each rater's initial diagnoses (the ratings first entered in the online portal) divided by the final diagnoses (based on consensus agreement via the portal or consensus conference).

An unweighted Cohen's kappa coefficient (κ) was calculated to account for agreement expected to occur by chance alone. [40, 41] Cohen's kappa coefficient (κ) was also used to assess the level of

agreement among raters by key NIA-AA diagnostic criterions such as impairment based on objective testing and/or informant report, and impairment in social and occupational function, as well as by NIA-AA diagnostic cognitive domains (memory, executive, language, visuospatial functions and behavioral changes).

Kappa statistics were interpreted using the Landis and Koch method where strength of agreement was defined as poor (< 0.00), slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) or almost perfect (0.81–1.00).[42]

Concordance between rater's diagnosis and final diagnostic outcomes were also compared by rater specialty. Specifically, using the final consensus diagnosis as the gold standard outcome, we coded the diagnosis rating as 1 "cognitively normal", 2 "MCI", 3 "dementia" and calculated the average difference score between each rater's assigned rating and the final consensus rating.

All analyses were performed using STATA 16 package (StataCorp, College Station, Texas, USA) software by authors M.G and D.T.B.

3 Results

3.1 Participants' diagnostic outcomes, baseline characteristics and cognitive profile.

A total of 635 cases were reviewed by a panel of 11 consensus members (6 U.S. and 5 South African) and assigned a diagnosis of cognitively normal, MCI or dementia. From July 2020 through October 2021, each case was randomly assigned to three consensus members for independent review. Consensus panel members were compromised of neurologists (n=5), neuropsychologists (n=2), geriatricians (n=2) and dementia researchers (n=2) with a mean experience of 12 years (range, 6-25 years) and an equal number of men (n=6) and women (n=5) on the panel. On average, it took raters

between 3 to 5 minutes to review each case, with more complex cases taking approximately 10-15 minutes to complete. After the independent individual ratings, 298 cases (47%) showed disagreement in the final diagnosis between the three raters. Discordant cases were resolved during biweekly conference calls and were mostly decisions between cognitively normal and early MCI, or late MCI and dementia.

The mean age in our cohort was 69.87 (SD=11.55) and the majority of study participants had no formal education (55 %). Final diagnostic categories included 314 cognitively normal (49%), 184 MCI (29%), and 137 dementia (22%) cases, which reflects our initial sampling framework to oversample for respondents categorized as higher risk for dementia. The distribution of diagnoses by age and sex is presented in Table 1.

We also compared mean cognitive scores by diagnostic group for 631 participants (4 participants were missing the cognitive battery). Overall, participants diagnosed with MCI and dementia performed worse on all cognitive measures (Table 2). Differences in mean cognitive scores between groups were significantly different from zero for all cognitive tests constituting the memory, executive, language and visuospatial domains (Table 2). Visuospatial tasks especially those requiring drawing had high rates of missingness, hence interpretation of these results may be tenuous.

3.2 Rater reliability: Agreement between individual ratings and final diagnoses

As each case was reviewed by three raters, a total of 1905 individual diagnostic ratings were made. Overall agreement between each rater's initial diagnosis and the final diagnosis (based on the consensus agreement via the portal or consensus conference) was 78.3% (1494/1905). The average κ statistic was 0.50, suggesting a moderate level of agreement between raters' individual ratings and

the final diagnostic outcomes (Table 3). Upon investigating agreement level by diagnostic outcome (at final determination), we observed a higher level of overall rater agreement with cognitively normal (k = 0.61) and dementia (k = 0.57) diagnoses and conversely lower agreement with MCI (k = 0.34). Kappas for individual raters ranged between 0.34-0.96, but measurement of inter-rater reliability was not possible due to differences in number of cases reviewed by each rater (Table 3).

To better understand the reasons for disagreement between raters, we investigated the percent agreement between raters for main NIA-AA diagnostic criterions (cognitive impairment and social and occupational functional impairment) among cases that triggered consensus conference calls. In general, of those cases that were adjudicated between MCI and dementia, we found that discordance between raters was mostly due to difficulty in assessing social and occupational functioning (Figure 2A). Specifically, the three raters almost always agreed on whether the participant had cognitive impairment (91%), but disagreement on the social and occupational functioning criterion was responsible for over 50 % of cases sent to adjudication. Another reason for discordance among raters was the absence of normative data, cases between cognitively normal and MCI were sent to adjudication conference calls mainly because of disagreement on the presence of cognitive impairment (over 85 % of cases) (Figure 2A). Upon further exploration of the cognitive impairment criterion by cognitive domain, we found the highest level of disagreement (74%) between raters particularly for tests within the visuospatial domain, followed by the memory domain (70%) (Figure 2B).

We additionally compared variations in diagnosis of cognitive impairment by rater's professional specialty. We classified raters into two categories: neurologists/geriatricians (7 raters) and neuropsychologists/ dementia research psychologists (4 raters). An average score of zero indicated congruency between the rater's individual rating and the final consensus rating, while an average

score above zero indicated over-diagnosis of MCI/dementia and an average score below zero indicated under-diagnoses of MCI/ dementia. Neurologists/ geriatricians were more likely to over-diagnose MCI/dementia, averaging positive difference scores when compared to neuropsychologists/ dementia research psychologists. One neuropsychologist had an average difference score below zero, indicating leniency towards reporting normal cognitive function for study participants who had a final assigned consensus diagnosis of MCI (Figure 3).

4 Discussion

This study is, to the best of our knowledge, the first in South Africa and sub-Saharan Africa, to evaluate the feasibility of using a web-based consensus approach for the diagnosis of cognitive impairment and dementia. We assembled a panel of 11 expert U.S. and South African neurologists, neuropsychologists, geriatricians and researchers with clinical and research experience in dementia, to ascertain dementia diagnoses to the entire HAALSI Dementia cohort (n=635). All members of the consensus panel found the web-based portal user-friendly and well-structured in terms of presentation of relevant demographic, cognitive, clinical and functional information for clinical decision-making. The web-based consensus approach proved to be a flexible and efficient methodology as raters were able to securely access the web portal and review cases from anywhere and at any time. The consensus panel offered an improvement in diagnostic confidence and reliability, especially in the absence of normative population data and the observed variability in diagnoses trends by rater specialty.

In this web-based consensus study, where each case was assigned a diagnostic category by three independent raters, we found an overall diagnostic concordance rate of 78.3% and a moderate level of agreement (k=0.50) between the overall independent ratings and final diagnosis obtained via consensus. Almost half of the cases required adjudication, with the majority being borderline between normal cognitive function and early MCI or between late MCI and early dementia. Consensus panel members found adjudication discussions on late MCI and early dementia particularly challenging given the heterogenous nature of MCI, which is traditionally viewed as a transitional period between normal aging and dementia.[43] We observed the lowest level of overall rater agreement for the MCI category (k=0.34). The MCI/dementia conundrum mainly stemmed from discordance between raters on whether the participant had cognitive impairment that affected social and occupational function. Contextual factors-- such as the high unemployment rates in South Africa, low educational attainment, and cultural norms whereby it is common for younger family members to take over household responsibilities and social activities even if the individual is not cognitively impaired --made it difficult for raters to assess some participants' true level of social and occupational impairment.[44,45] Our findings build on the existing web-based consensus work reported by our sister study in India, LASI-DAD, which observed significant rater differences in the assessment of "social and community activities" and "home and hobbies" CDR specific domains.[11]

Another reason for the observed discordance in raters' diagnoses is the difference in levels of importance that each rater assigned to the available cognitive and informant measures. This is not surprising in the absence of normative data, especially given the low educational exposure and limited knowledge of psychometric properties within our cohort. Disagreement between raters on impairment were especially low for the visuospatial domain (74%), mainly attributable to the high

levels of missingness for measures that required participants to draw (constructional praxis and clock draw). Consequently, it was difficult for raters to assess the visuospatial domain, specifically to decide on impairment for participants with missing data as compared to participants who performed poorly on these measures. We have previously reported on missingness in the visuospatial domain and have linked it to the unfamiliarity/ uneasiness with holding a pen (or tablet stylus) among participants with low literacy, along with a high prevalence of visual impairment.[12, 46] For these reasons, the consensus panel often deemphasized these measures during the diagnostic process. Our in-depth investigation into the causes for discordance among raters will help inform and improve measures for the visuospatial and social/occupation domains for future waves of the longitudinal HAALSI Dementia study.

Potential limitations of our study include that we were not able to estimate inter-rater reliability, due to differences in number of cases reviewed by each rater. However, we computed the average kappa and overall rater reliability and showed reasonable level of agreement between raters' assigned diagnosis and the final diagnostic outcomes. We did not validate our findings against an "external" reference standard such as pathological results or an in-person clinically diagnosed sample. Our future work includes using magnetic resonance imaging (MRI) and biomarker sub-studies to further validate and support the validity of this web-based approach.

Overall, this study further complements the consensus work completed by LASI-DAD, and demonstrates a successful novel web-based consensus panel approach for the diagnosis of cognitive impairment and dementia in a rural South African population with low education level.

This web-based consensus methodology is especially promising for LMICs, as we were able to implement this approach in a very remote, low-literacy, low-resource setting with limited access to healthcare services and no access to neuropsychologists and/ or dementia specialized professionals.

Finally, we will use the assigned consensus diagnoses to develop a diagnostic algorithm to assign dementia probability scores in the larger parent HAALSI cohort.

Funding

This work was supported by the National Institute on Aging [5R01AG054066]. HAALSI Dementia is a collaboration between Harvard Center for Population and Development Studies from Harvard T.H. Chan School of Public Health & MRC/Wits Rural Public Health and Health Transitions Research Unit from School of Public Health at the University of the Witwatersrand in South Africa.

References

1. Bertens LC, Broekhuizen BD, Naaktgeboren CA, et al. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS Med* 2013; **10**: e1001531.

2. Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 2009; **62**: 797-806.

3. Weir DR, Wallace RB, Langa KM, et al. Reducing case ascertainment costs in U.S. population studies of Alzheimer's disease, dementia, and cognitive impairment-Part 1. *Alzheimers Dement* 2011; **7**: 94-109.

4. Handels RL, Wolfs CA, Aalten P, et al. Optimizing the use of expert panel reference diagnoses in diagnostic studies of multidimensional syndromes. *BMC Neurol* 2014; **14**: 190.

5. Magaziner J, Zimmerman SI, German PS, et al. Ascertaining dementia by expert panel in epidemiologic studies of nursing home residents. *Ann Epidemiol* 1996; **6**: 431-7.

6. McKhann GM, Knopman DS, Chertkow H, et al. 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* 2011; **7**: 263-9.

7. AP A. Diagnostic And Statistical Manual Of Mental Disorders: DSM-5. . *Washington, DC: American Psychiatric Association* 2013.

8. Razzouk D, Sharan P, Gallo C, et al. Scarcity and inequity of mental health research resources in low-and-middle income countries: a global survey. *Health Policy* 2010; **94**: 211-20.

9. Prince M, Acosta D, Chiu H, Scazufca M, Varghese M, Dementia Research G. Dementia diagnosis in developing countries: a cross-cultural validation study. *Lancet* 2003; **361**: 909-17.

10. Ferri CP, Jacob KS. Dementia in low-income and middle-income countries: Different realities mandate tailored solutions. *PLoS Med* 2017; **14**: e1002271.

11. Lee J, Ganguli M, Weerman A, et al. Online Clinical Consensus Diagnosis of Dementia: Development and Validation. *J Am Geriatr Soc* 2020; **68 Suppl 3**: S54-S9.

12. Bassil DT, Farrell MT, Wagner RG, et al. Cohort Profile Update: Cognition and dementia in the Health and Aging in Africa Longitudinal Study of an INDEPTH community in South Africa (HAALSI dementia). *International journal of epidemiology* 2021.

13. Langa KM, Ryan LH, McCammon RJ, et al. The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. *Neuroepidemiology* 2020; **54**: 64-74.

14. 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**: 189-98.

15. Ganguli M, Ratcliff G, Chandra V, et al. A hindi version of the MMSE: The development of a cognitive screening instrument for a largely illiterate rural elderly population in india. *International Journal of Geriatric Psychiatry* 1995; **10**: 367-77.

16. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol* 1988; **1**: 111-7.

17. Hall KS, Gao S, Emsley CL, Ogunniyi AO, Morgan O, Hendrie HC. Community screening interview for dementia (CSI 'D'); performance in five disparate study sites. *International journal of geriatric psychiatry* 2000; **15**: 521-31.

18. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International psychogeriatrics* 1997; **9**: 173-6.

19. Morris JC, Heyman A, Mohs RC, et al. The consortium to establish a registry for Alzheimer's disease (CERAD): I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989.

20. Drozdick LW, Raiford SE, Wahlstrom D, Weiss LG. The Wechsler Adult Intelligence Scale—Fourth Edition and the Wechsler Memory Scale—Fourth Edition. 2018.

21. Torralva T, Roca M, Gleichgerrcht E, Lopez P, Manes F. INECO Frontal Screening (IFS): A brief, sensitive, and specific tool to assess executive functions in dementia–CORRECTED VERSION. *Journal of the International Neuropsychological Society* 2009; **15**: 777-86.

22. Tripathi R, Kumar JK, Bharath S, Marimuthu P, Varghese M. Clinical validity of NIMHANS neuropsychological battery for elderly: A preliminary report. *Indian journal of psychiatry* 2013; **55**: 279.

23. Morris J. Current vision and scoring rules the clinical dementia rating (CDR). *Neurology* 1993; **43**: 2412-4.

24. Raven J. The Raven's progressive matrices: change and stability over culture and time. *Cognitive psychology* 2000; **41**: 1-48.

25. Wechsler D. *WMS-R: Wechsler memory scale-revised*: Psychological Corporation; 1987.

26. Glosser G, Wolfe N, Albert ML, et al. Cross-cultural cognitive examination: validation of a dementia screening instrument for neuroepidemiological research. *Journal of the American Geriatrics Society* 1993; **41**: 931-9.

27. Freedman M, Leach L, Kaplan E, Shulman K, Delis DC. *Clock drawing: A neuropsychological analysis*: Oxford University Press, USA; 1994.

28. Borod JC, Goodglass H, Kaplan E. Normative data on the Boston diagnostic aphasia examination, parietal lobe battery, and the Boston naming test. *Journal of Clinical and Experimental Neuropsychology* 1980; **2**: 209-15.

29. Lezak MD, Howieson DB, Loring DW, Fischer JS. *Neuropsychological assessment*: Oxford University Press, USA; 2004.

30. De Renzi A, Vignolo LA. Token test: A sensitive test to detect receptive disturbances in aphasics. *Brain: a journal of neurology* 1962.

31. Benton AL, deS K, Sivan AB. *Multilingual aphasia examination*: AJA associates; 1994.

32. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* 1977; **1**: 385-401.

33. Jorm AF, Rodgers B, Henderson AS, et al. Occupation type as a predictor of cognitive decline and dementia in old age. *Age Ageing* 1998; **27**: 477-83.

34. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br J Psychiatry* 1968; **114**: 797-811.

35. Prince M, Acosta D, Ferri CP, et al. A brief dementia screener suitable for use by non-specialists in resource poor settings--the cross-cultural derivation and validation of the brief Community Screening Instrument for Dementia. *Int J Geriatr Psychiatry* 2011; **26**: 899-907.

36. Gure TR, Langa KM, Fisher GG, Piette JD, Plassman BL. Functional Limitations in Older Adults Who Have Cognitive Impairment Without Dementia. *Journal of Geriatric Psychiatry and Neurology* 2013; **26**: 78-85.

37. Plassman BL, Langa KM, Fisher GG, et al. Prevalence of cognitive impairment without dementia in the United States. *Annals of internal medicine* 2008; **148**: 427-34.

38. Plassman BL, Langa KM, Fisher GG, et al. Prevalence of Dementia in the United States: The Aging, Demographics, and Memory Study. *Neuroepidemiology* 2007; **29**: 125-32.

39. Plassman BL, Langa KM, McCammon RJ, et al. Incidence of dementia and cognitive impairment, not dementia in the united states. *Annals of Neurology* 2011; **70**: 418-26.

40. Jacob C. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; **XX**.

41. Fleiss JL, Cohen, Jacob , Everitt, B.S. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* 1969; **72**: 323–7.

42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159-74.

43. Narasimhalu K, Ang S, De Silva DA, et al. Severity of CIND and MCI predict incidence of dementia in an ischemic stroke cohort. *Neurology* 2009; **73**: 1866-72.

44. Klasen S W, I. Surviving Unemployment Without State Support: Unemployment and Household Formation in South Africa. *Journal of African Economics* 2009; **18**.

45. Bohman DM, van Wyk NC, Ekman SL. Tradition in transition--intergenerational relations with focus on the aged and their family members in a South African context. *Scand J Caring Sci* 2009; **23**: 446-55.

46. Humphreys GW, Duta MD, Montana L, et al. Cognitive Function in Low-Income and Low-Literacy Settings: Validation of the Tablet-Based Oxford Cognitive Screen in the Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI). *J Gerontol B Psychol Sci Soc Sci* 2017; **72**: 38-50.

Figures

Figure 1: Flow of HAALSI Dementia activities leading to web-based expert panel diagnoses

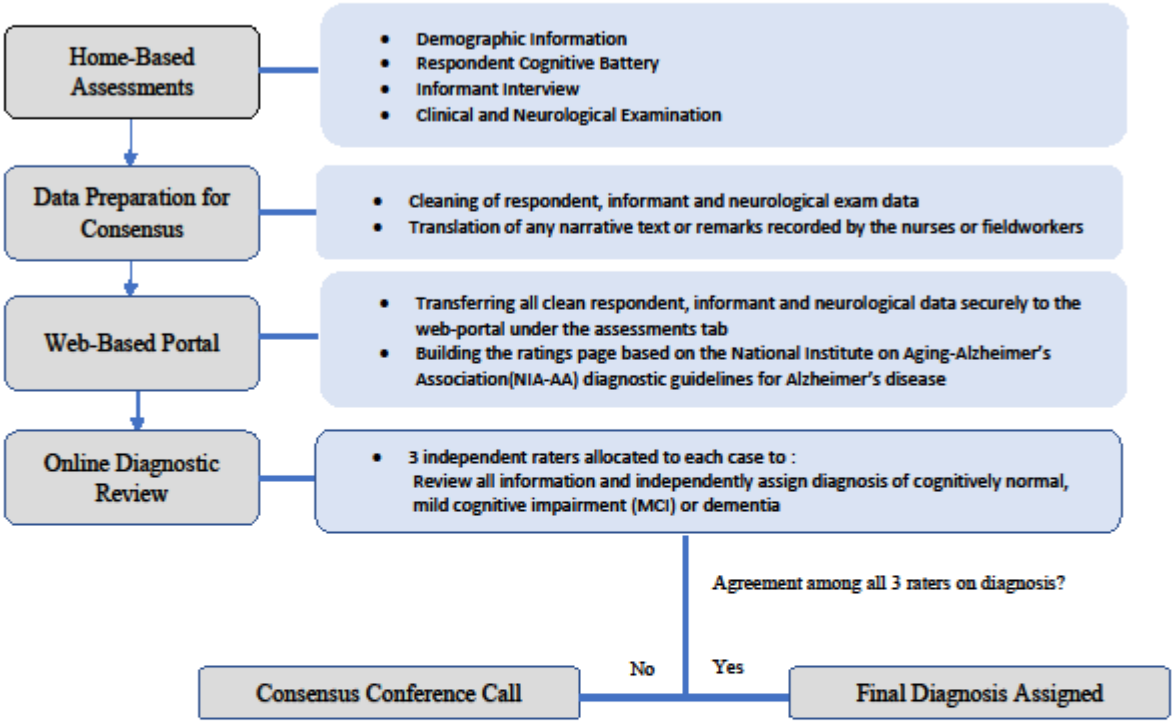


Figure 2: A. Percent of cases discussed in adjudication conference calls (cognitively normal and MCI, MCI and dementia) due to disagreement in key NIA-AA diagnostic criterions (n=272)

Figure 2: B. Disagreement in cognitive impairment between raters by NIA-AA cognitive domain for adjudicated sample (n=272)

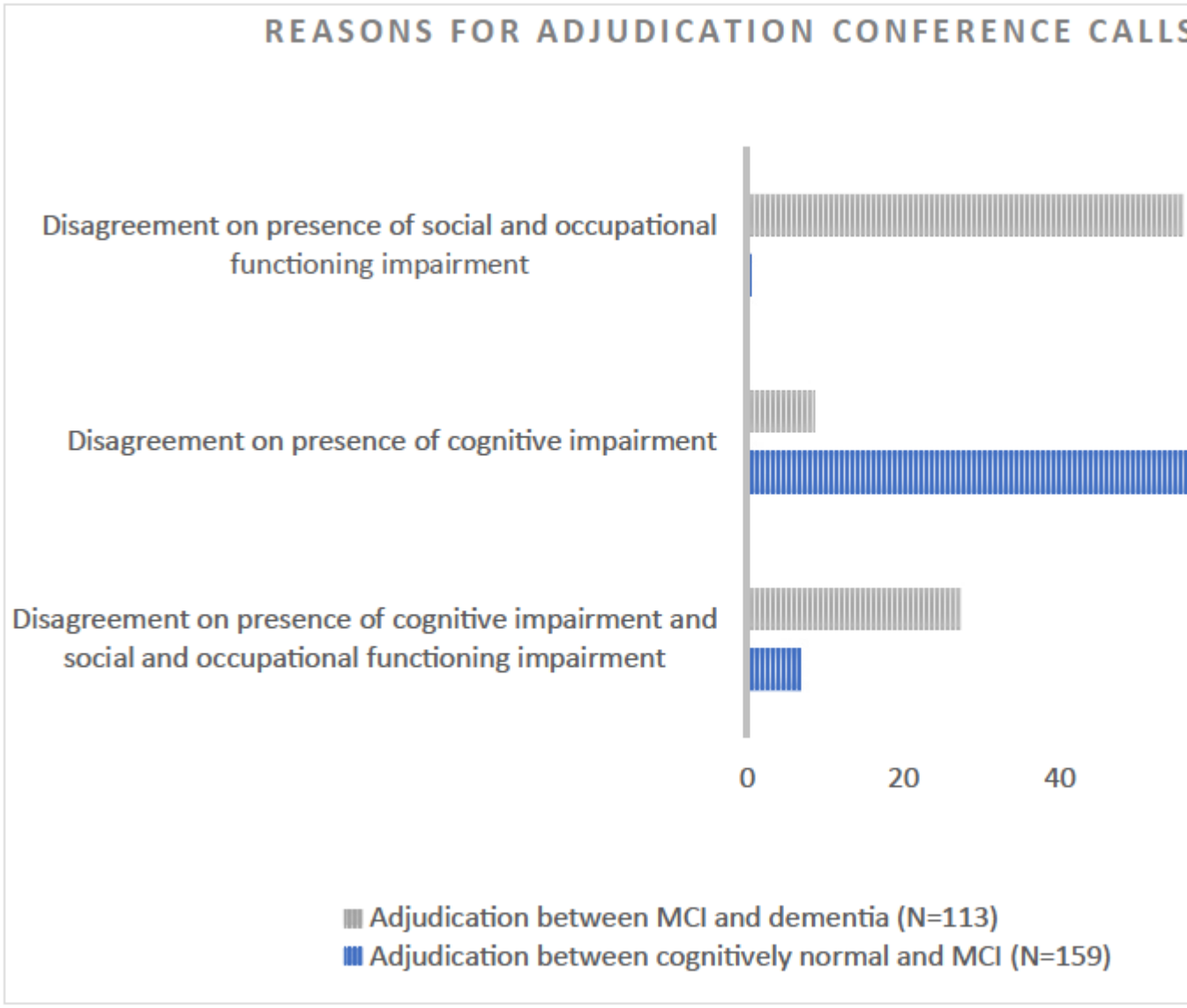
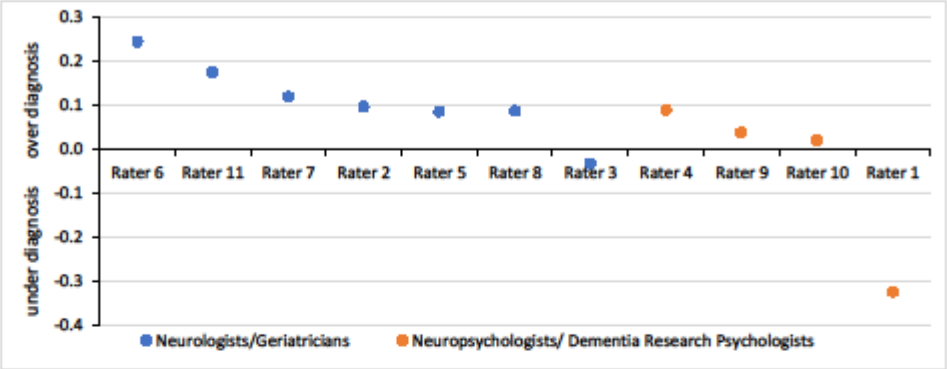


Figure 3. Overall concordance between individual raters' diagnoses and final consensus diagnoses (reference) by rater specialty.



Tables

Table 1: Distribution of diagnoses (cognitively normal, MCI, dementia) by age and sex

Table 2: Mean score (SD) for cognitive tests and IQCODE scale by diagnostic outcomes (cognitively normal, MCI, dementia)

Table 3: Individual and overall levels of agreement between raters and consensus diagnoses

Table 1. Distribution of diagnoses (cognitively normal, MCI, dementia) by age and sex

	Diagnosis (N/%)			
	Cognitively Normal	MCI[a]	Dementia	Observations

Total	314 (49.5)	184 (29.0)	137 (21.6)	635
Age groups				
50-54	41 (91.1)	2 (4.4)	2 (4.4)	45
55-59	80 (75.5)	20 (18.9)	6 (5.7)	106
60-64	55 (64.7)	23 (27.1)	7 (8.2)	85
65-69	51 (57.3)	23 (25.8)	15 (16.9)	89
70-74	38 (46.3)	28 (34.2)	16 (19.5)	82
75+	49 (21.5)	88 (38.6)	91 (39.9)	228
Gender				
Female	182 (46.6)	119 (30.4)	90 (23.0)	391
Male	132 (54.1)	65 (26.6)	47 (19.3)	244

[a]MCI – Mild Cognitive Impairment

Table 2. Mean score (SD) for cognitive tests and IQCODE scale by diagnostic outcomes (cognitively normal, MCI, dementia)

	Diagnosis				
	Cognitively Normal (n=313)	MCI[a] (n=183)	Dementia (n=135)	p-value	% missing
TICS[b]	3.2(0.8)	2.6(0.8)	2(0.9)	<0.001	0.6
Symbol Cancellation	20.3(12.8)	10.1(8.2)	8.5(8.6)	<0.001	15.9
CSID[c]	3.9(0.4)	3.7(0.6)	3.3(0.9)	<0.001	0.6
Similarities & Differences	3.4(0.7)	3.2(0.8)	2.8(1.0)	<0.001	0.6
Days of the Week	13.2(1.6)	11.0(2.7)	8.8(3.3)	<0.001	0.6
Go/No Go	16.0(3.5)	11.8(4.6)	8.4(5.0)	<0.001	1.1
Motor Sequence	2.8(0.4)	2.4(0.7)	1.9(1.0)	<0.001	0.8
Spatial Working Memory	1.4(0.9)	1.0(0.9)	1.0(0.8)	<0.001	6.1
Token Test	4.2(1.4)	3.0(1.6)	2.5(1.5)	<0.001	0.6
Boston Naming Test	14.4(1.7)	13.3(2.8)	11.7(4.1)	<0.001	0.6
Raven's Progressive Matrices	8.6(2.5)	6.6(2.4)	5.6(2.7)	<0.001	5.0
Total Immediate Word Recall	14.7(3.2)	10.8(3.8)	8.1(3.8)	<0.001	0.6
Total Delayed Word Recall	4.8(1.7)	3.0(1.6)	2.0(1.5)	<0.001	0.6
Word Recognition	16.6(2.4)	14.7(2.8)	12.9(3.0)	0.0073	0.6
Logical Memory Immediate—Story 1	10.3(3.5)	7.7(3.8)	5.6(4.1)	<0.001	0.6
Logical Memory Immediate—Story 2	6.9(1.7)	5.4(2.2)	4.1(2.3)	<0.001	0.6
Logical Memory Delayed—Story 1	8.7(3.8)	6.0(3.9)	4.1(3.6)	<0.001	2.4
Logical Memory Delayed—Story 2	5.9(2.1)	4.3(2.4)	3.1(2.4)	<0.001	2.4
Logical Memory—Recognition	9.5(1.7)	8.7(1.9)	8.1(1.9)	<0.001	0.6
Constructional Praxis	6.3(2.3)	4.3(2.5)	3.5(2.1)	<0.001	24.4
Constructional Praxis Recall	3.9(2.8)	1.9(2.1)	1.4(1.8)	<0.001	23.9

MMSE[d]	22.2(2.9)	16.5(3.5)	12.8(4.3)	<0.001	0.6
Semantic Fluency	12.2(6.8)	10.0(3.7)	8.7(3.4)	<0.001	0.6
Phoneme Fluency	2.9(3.0)	1.3(1.9)	0.8(1.3)	<0.001	2.2
CESD Total Score[e]	10.2(7.9)	15.3(9.3)	17.7(12.2)	<0.001	0.6
Clock Draw	1.7(0.8)	1.0(0.6)	0.9(0.4)	<0.001	43.1
IQCODE[f]	3.1(0.2)	3.3(0.3)	3.9(0.6)	<0.001	0.9

[a] MCI -Mild Cognitive Impairment. SD - Standard Deviation. [b]TICS - Telephone Interview for Cognitive Status. [c]CSID - Community Screening Inventory for Dementia. [d]MMSE - Mini Mental Status Exam. [e]CESD - Centers for the Epidemiological Study Depression, [f]IQCODE- Informant Questionnaire on Cognitive Decline in the Elderly

Table 3. Individual and overall levels of agreement between raters and consensus diagnoses

Rater	Kappa	Z	Prob>Z	Number of cases reviewed
1	0.346	2.88	0.002	37
2	0.553	12.56	<0.001	262
3	0.774	16.38	<0.001	240
4	0.3873	4.34	<0.001	45
5	0.6881	19.81	<0.001	420
6	0.3864	4.37	<0.001	41
7	0.6346	9.17	<0.001	101
8	0.5618	14.06	<0.001	288
9	0.7852	21.07	<0.001	373
10	0.9634	9.03	<0.001	50
11	0.664	6.49	<0.001	46
Cognitively Normal	0.6076			
MCI[a]	0.3377			
Dementia	0.5693			
Overall	0.5006			

[a] MCI- Mild Cognitive Impairment