Published in final edited form as:

Alzheimers Dement. 2014 November ; 10(6): 760–768.e1. doi:10.1016/j.jalz.2014.02.008.

A Comparison of Neuropsychological Performance Between US and Russia: Preparing for a Global Clinical Trial

Kathleen M. Hayden^{a,*}, Oksana A. Makeeva^b, L. Kristin Newby^{c,d}, Brenda L. Plassman^a, Valentina V. Markova^b, Ashley Dunham^c, Heather R. Romero^a, Zarui A. Melikyan^a, Cassandra M. Germain^a, Kathleen A. Welsh-Bohmer^a, and Allen D. Roses^e on behalf of on the behalf of the TOMSK-DUKE Study Group

^aJoseph and Kathleen Bryan ADRC, Department of Psychiatry, Duke University Medical Center, Durham, NC, U.S.

^bCenter for Clinical Trials, Nebbiolo LLC, Tomsk, Russia

^cDuke Clinical Research Institute Durham, NC, U.S.

^dDepartment of Medicine, Duke University Medical Center, Durham, NC, U.S.

^eZinfandel Pharmaceuticals, Durham, NC, U.S.

Abstract

Background—Understanding regional differences in cognitive performance is important for interpretation of data from large multinational clinical trials.

Methods—Data from Durham and Cabarrus Counties in North Carolina, U.S. and Tomsk, Russia ($n=2,972$) were evaluated. The Montreal Cognitive Assessment (MoCA), Trail Making Test Part B (Trails B), CERAD Word List Memory Test delayed recall (WLM), and self-report ADCS Mail-In Cognitive Function Screening Instrument (MCFISI) were administered at each site. Multilevel modeling measured the variance explained by site and predictors of cognitive performance.

Results—Site differences accounted for 11% of the variation on the MoCA; 1.6% on Trails B; 1.7% on WLM; and 0.8% in MCFISI scores. Prior memory testing was significantly associated with WLM. Diabetes and stroke were significantly associated with Trails B and MCFISI.

Conclusions—Sources of variation include cultural differences, health conditions, and exposure to test stimuli. Findings highlight the importance of local norms to interpret test performance.

© 2014 Elsevier Inc. All rights reserved

*Kathleen M. Hayden Joseph and Kathleen Bryan ADRC DUMC Box 3503 Durham, NC 27710 khayden@duke.edu 919-668-1584. Tomsk-Duke Study Group Investigators: Larisa I. Minaycheva,^b Stephan V. Buikin,^b Irina A. Zhukova,^b Natalia G. Zhukova,^b Yuka Maruyama,^c Kent Anglin,^a Pat Connelly,^a Pete Gaskell,^a Ann Herrin,^a Charles Loughlin,^a Heather MacDonald,^a Michelle McCart,^a Ron Nelson,^a and Lawrence Whitley.^a

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

cross-cultural comparison; neuropsychological tests; Alzheimer's disease; primary prevention
delay of onset of MCI due to Alzheimer's Disease

1. Introduction

In the last decade, the number of multinational clinical trials has increased [1,2]. Global clinical trials present challenges including: harmonizing procedures across countries, finding culturally appropriate diagnostic tools, and identification or creation anew of qualified research sites[3]. In recent years, AD researchers have turned to primary prevention rather than secondary or tertiary treatments, thereby pushing the clinical diagnosis back from onset of AD to onset of mild cognitive impairment (MCI) due to AD[4]. As a result, clinical trials will necessarily be longer and will require more sensitive and well-calibrated tools. Issues specific to Alzheimer's disease (AD) make these trials even more difficult as an AD diagnosis is essentially based upon subjective clinical information rather than purely objective measures. Neuropsychological test performance can dictate whether or not a participant is called upon for more detailed and expensive clinical assessments. If the performance threshold is too high, false positives will incur additional costs in assessment. If the threshold is too low, cases may be missed, the trial may be prolonged, and investigational therapies could inappropriately fail. Imprecision in these measures and their interpretation can add noise, reduce power, and increase costs. As the field moves into global clinical trials for prevention, it will be critically important to identify culturally appropriate neuropsychological tests and establish country specific normative values to support generalization of study results.

The process of diagnosing MCI due to AD is difficult to standardize in one language across multiple sites and even more-so across multiple languages and cultures. Neuropsychological testing can facilitate standardization of protocols so that clinicians review identical measures at each site. The selection of neuropsychological tests to differentiate normal cognition from mild impairment requires careful consideration. Unfortunately, some of the most commonly used tests in AD clinical practice in the U.S. either have not been translated into other languages or their reliability and validity have not been evaluated across multiple countries.

The literature on cognitive performance across different countries and cultures is mixed. Some studies have shown that there are differences in cognitive performance across countries[5–9]. These differences are thought to stem from cultural factors[7,10,11], educational differences[6,7,11], variation in the prevalence of health conditions that may affect performance, and varying levels of experience with timed tests or testing in general. These differences are often evident on both verbal and non-verbal tests [7] and the extent of the difference varies by test [12]. One multinational study using the ADAS-Cog[13] showed that differences could be minimized when the instruments were harmonized across countries[9]. Others have found that differences across countries are minimal and not clinically meaningful. One study found slightly more variation in older adults[14], but still concluded that there were no meaningful differences across countries. A second study found

“support for the hypothesis of commonality of some neuropsychological functions across cultures” [15]. Finally, a large World Health Organization (WHO) Neurobehavior Core Test Battery (NCTB) study of >2300 individuals aged 16–65 across 10 countries and found general consistency in performance across countries[12].

To compare performance across countries and cultures, some level of equivalence in measurement[16] must be established. Interpretive equivalence refers to similarities in meaning of concepts and constructs, whereas procedural equivalence refers to the structural aspects of the instruments themselves [17,18]. Herein we sought to evaluate interpretive equivalence in performance on standard cognitive testing using data from three research studies that administered the same protocols: two in North Carolina, U.S. (Durham and the city of Kannapolis and its surrounding county, Cabarrus County), and one in Tomsk, Russia. The Russian versions of the instruments were harmonized with the U.S. versions, and we hypothesized that the U.S. sites would be more similar to each other than to the Russian site. Further, we theorized that factors that are known to affect cognition (*i.e.*, age, sex, educational attainment, and cardiovascular health) would similarly affect performance across sites.

2. Methods

2.1 Participants

Data were collected from three community studies that were engaged in building participant registries for a study to delay the onset of MCI due to AD: 1) the PREPARE Study, Durham, North Carolina, U.S. 2) the Measurement to Understand Reclassification of Disease of Cabarrus/Kannapolis (MURDOCK) Community Registry and Biorepository Study[19], Kannapolis, North Carolina, U.S. and 3) Evaluation of Neuropsychological Status in Elderly Population in Russia Study, Tomsk, Russia. Approval for each study was obtained from local Institutional Review Boards (United States) or Ethical Committees (Russia). Participants provided informed consent for all procedures. Identical cognitive testing protocols were administered at each site.

Participants in the PREPARE Study at the Joseph and Kathleen Bryan Alzheimer's Disease Research Center (Bryan ADRC) at Duke University in Durham, North Carolina were recruited primarily from the Bryan ADRC Alzheimer's Disease Prevention Registry. This registry included individuals aged 55 and older who did not have a diagnosis of dementia. Registrants were recruited *via* community outreach events, newsletters, health fairs, family reunions, civic groups, church groups, and educational presentations. Participants were also recruited from the registry of Center for the Study of Aging and Human Development at Duke University and through the Bryan ADRC web site. Individuals who signed up for the registry agreed to be contacted for future studies; thus, they were contacted for participation in the PREPARE Study. PREPARE was designed to characterize participants' physical and cognitive health and determine their willingness to participate in an upcoming clinical trial. Participants were specifically asked if they would consider participating in a pharmacological trial for the prevention of AD. Administration of the PREPARE protocol took approximately one hour and comprised participants' medical history, vital signs, a blood sample, and a brief cognitive battery.

The MURDOCK Study Community Registry and Biorepository recruits participants from the adult residents of Cabarrus County and the city of Kannapolis, North Carolina. The study is designed to capture a large portion of area residents in a representative population, with a goal of launching translational medicine studies that focus on –omics assessment of disease[19]. The MURDOCK study has a number of sub-studies designed to facilitate and foster further study of the cohort. One such sub-study, titled the Memory Health Study, is a collaboration between the Bryan ADRC and the MURDOCK study to identify and collect cognitive performance data from MURDOCK Study participants aged 55 and older. In addition to collection of data for the main MURDOCK Study Registry and Biorepository, participants undergo a cognitive screening protocol that is identical to the Durham PREPARE Study protocol.

The third study was the Nebbiolo Center for Clinical Trials study, “The Evaluation of Neuropsychological Status in Elderly Population in Russia,” which took place in Toms, Russia. This study included a pool of volunteers aged 56 to 89 years, who were identified using outreach activities and a primary care centralized medical system in the city of Toms. Subjects of the target age group were contacted by phone to participate in the study. Participants were administered the Russian translation of a cognitive screening protocol identical to that of the Durham PREPARE Study. The data collection application used in the PREPARE Study was translated for use in Russian to ensure that data were collected in a consistent manner.

2.2 Measures

Each participant provided basic demographic information including age, sex, marital status, race, ethnicity, and education level. Education levels were defined as: less than high school, high school, some college or Associate's degree, Bachelor's degree, Master's or higher professional degree. Questions about health included family medical history, and participant history of health problems including: diabetes, hypertension, stroke, history of depression, and cardiovascular disease (CVD). We defined CVD as a dichotomous variable indicating the presence or absence of atrial fibrillation, heart attack, pacemaker, or valve replacement. Diabetes, hypertension, and stroke were coded separately as each has been associated with cognitive performance in later life[20–23]. Height and weight were measured and body mass index (BMI) was calculated as $\text{weight (lbs)}/\text{height(in)}^2 \times 703$ for US participants and $\text{weight(kg)}/\text{height(m)}^2$ for the Russian participants. Obesity was defined as $\text{BMI} \geq 30$. Participants were also asked about their prior experience with cognitive testing. If prior testing was reported, they were asked why they were tested and how often.

2.3 Cognitive assessments

Cognitive assessment comprised four measures, the Montreal Cognitive Assessment (MoCA)[24], the Trail Making Test Part B (Trails B)[25], the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Delayed Recall Word List Memory Task (WLM)[26], and the Alzheimer's Disease Cooperative Studies (ADCS) Mail-In Cognitive Function Screening Instrument (MCFISI)[27]. The MoCA is a global cognitive function screening tool, developed for the detection of MCI[24]. It has a higher level of sensitivity (90%) and specificity (100%) for detecting MCI than the Mini Mental State

Examination[28]. The MoCA covers a number of cognitive domains including: attention, executive, language, memory, orientation, and visuospatial. Scores range from 0–30 with higher scores indicating better cognitive function. A cut-point of 26 (after education correction) is recommended for cognitive disorders and 16 for AD patients in the U.S.[24]. We therefore considered scores falling below 16 to reflect severe cognitive impairment and individuals scoring below this cut point were excluded from the present study. The MoCA has been translated into many languages including Russian.

The Trail Making Test Part B is widely used in clinical practice[25]. It is a timed test of executive function and requires set switching as participants are required to connect alternating numbers and letters in order. There is some debate over whether Trails B is a culturally fair test and alternatives have been developed, such as the Color Trails Test[29]. However, some have suggested that the alternative Color Trails Test is not an equivalent measure[30]. The CERAD WLM task was developed as part of the CERAD protocol for the assessment of dementia and has proven to be a sensitive test for mild impairment as well as dementia[26,31,32]. The 10-item word list has three learning trials and delayed recall is tested after a brief delay. Scores range from zero to ten. The CERAD battery has been translated into many languages and is available upon request <http://cerad.mc.duke.edu/Default.htm>. The ADCS MCFISI[27] is a self-report memory questionnaire that was developed for the ADCS Prevention Instrument Project specifically for detecting cognitive change in large prevention trials[33]. This brief 14-item questionnaire was derived from questions that are commonly asked during a standard clinical assessment[27]. Although designed as a mail-in questionnaire, the measure was completed during the clinical assessment.

2.4 Translations

All measures were translated and adapted for use with Russian-speaking older adults by a bilingual Russian-English speaking neuropsychologist with work experience both in Russia and the US with the help of investigators from the Tomsk, Russia site. When available, the official Russian version of the test was used. Procedures for translation and adaptation of measures were consistent with procedures outlined by the International Test Commission Guidelines for Translating and Adapting Tests[34].

Tests translated for this study, Trails B and the ADCS MCFISI, were back translated into English to check the accuracy of Russian translation. In an effort to achieve interpretive equivalence, Russian translations were reviewed and revised by Russian specialists in psychology and medicine to ensure all the items were appropriate, clearly worded, and could be readily understood by potential study participants. On the ADCS MCFISI, items such as driving (Russian elderly frequently use public transportation rather than drive) and managing personal finances (Russian elderly more commonly use cash rather than credit card and/or checks) were adapted for the Russian lifestyle. As noted above, the available Russian translations for the MoCA (<http://www.mocatest.org/>) and CERAD WLM (<http://www.scarletline.com/aglezerman/CERAD.html>) were used.

The translated measures were pilot-tested on individuals within the study's target age range to see if any further changes to the instructions or questionnaires were needed. To ensure

uniformity of administration of the protocol across all three sites, the MURDOCK and Russian research teams completed training and certification procedures administered by the Durham neuropsychology team. Quality control procedures for data included 100% paper form checks, 10% audit of testing sessions (most sessions were recorded), and personnel re-certification every three months.

2.5 Analytical methodology

Differences between the three sites were evaluated with GLM or χ^2 as appropriate. Total scores for the MoCA, Trails B, WLM, and MCFSI were normalized using a Tukey transformation[35]. Scores were then standardized by using pooled sample means (across all three sites) and standard deviations (SD) to calculate z-scores for each participant. Multilevel modeling was used to measure the variance explained by each site and to evaluate the influence of covariates. Additional mixed models were used to explore differences by site and best predictors of cognitive performance for each test. Data were re-analyzed using country specific means and standard deviations (SD) to estimate z-scores (one set of means and SD for the U.S. and one set for Russia). By using country specific means and SD, we attempted to simulate comparisons that might be made if each site were to use country-specific norms. Finally, we combined the two U.S. sites to see if a better model fit could be achieved by simply comparing performance in the U.S. to Russia. SAS version 9.3 was used for all analyses[36].

3. Results

A total of 3,363 volunteers were administered the questionnaires and testing. Due to missing data, including test scores (n=69), demographic information (n=4), or health status information (n=208), we excluded data from 281 participants. Another 30 participants were excluded because they did not complete the examination in their native language (which may have affected the test validity), and 80 individuals were excluded due to low MoCA scores (scores <16 may reflect significant cognitive impairment). The final analytic sample included 2,972 participants, ranging in age from 55 to 93 years (median age 68 [55, 93]). The population comprised 70.3% women, and education levels ranged from less than high school to Master's degree or higher. Participants were mostly white (90.8%), with the Durham site having the highest proportion of African Americans (14.8%). Specific details of the three samples are listed in Table 1. We compared demographic characteristics of included *versus* excluded participants where such information was available. Participants who were excluded (n=391) were significantly older (p<0.0001), more often male (p=0.01), had lower levels of education (p<0.0001), and higher prevalence of self-reported CVD (p<0.0001), hypertension (p<0.0001), and depression (p<0.05). The two U.S. sites differed significantly in their distributions of sex, race, education, and prevalence of CVD, depression, and diabetes.

Raw test scores by site are shown in Table 2. Initial multilevel modeling results using pooled means showed site differences accounting for 35% of the variation on the MoCA; 38% on Trails B; 5% on the WLM; and 20% in MCFSI scores. Covariates including demographic and health variables, education, and prior exposure to memory testing

accounted for 19% of site variation on the MoCA; 20% on Trails B; 62% on the WLM delayed recall; and 3% in MCFSI scores (Figure 1a). When these models were re-run after applying country specific z-scores, the differences were attenuated. Differences between the three sites accounted for 11% of variation on the MoCA; 1.6% of variation on Trails B; 1.7% of WLM; and 0.8% of MCFSI scores (Figure 1b). The variance due to covariates accounted for a greater proportion of the variance in the revised model.

To determine more specifically which covariates were associated with performance on each measure, we used mixed models to evaluate differences by site. Models shown in Table 3 were derived using country specific z-scores and represent standard deviation units. For all cognitive measures, increasing age was associated with slightly poorer performance and higher education was associated with better performance. Performance on the MoCA was associated with site, age, sex, race, and education. In the analysis that employed pooled means across countries, Tomsk participants performed >1 SD unit slower than Durham on Trails B (see supplemental Table 4). When country specific z-scores were used, there was no significant site difference. Stroke and diabetes were associated with slowed performance on Trails B across sites in both the pooled and country specific analyses. WLM delayed recall performance was associated with demographic characteristics and a history of prior memory testing and these results were similar in the analysis using country specific z-scores. Scores on the ADCS MCFSI were associated with history of stroke, diabetes, prior memory testing, and depression in addition to demographic characteristics. The effects of sex and education varied across sites and across tests. When we collapsed the two US sites together and re-ran the analysis (data not shown), the results were similar; furthermore, comparisons of the fit statistics showed significantly better fit in models that considered Durham and Kannapolis separately.

4. Discussion

Regional differences in cognitive performance are important to understand in large multinational clinical trials. Potential sources of variation include cultural differences in attitudes towards time and timed tests, previous experience with cognitive testing, education, and the prevalence of vascular disease and other health conditions.

There were significant differences in performance across sites (percent variance attributable to site differences ranged from 5.2% to 37.6%) when pooled means were used. However, after application of country-specific z-scores, site differences were reduced to a range of 0.8%–11%. Of particular interest was the large difference in raw score performance between the US sites and the Russia site on the Trail Making Test Part B. Such differences have been noted previously[6] and are not surprising in this sample for a number of reasons. First, performance on Trails B is predicated on the notion that it draws upon over-learned serial stimuli[30]. In Russia, the alphabet is not overlearned, nor is it learned in the same manner as it is in the U.S., *i.e.*, there is no alphabet song, it is not used for organizational purposes. Second, the Russian concept of the importance of time is not the same as in the U.S., nor have they been as acculturated to timed tests as residents of the U.S. The concept that *speeded* cognitive performance is as important as *accurate* cognitive performance is not as valued a cultural concept in certain cultures as it is in the U.S. [37]. In a comparison of 50

U.S. and 50 Russian participants' performance on timed tasks and attitudes toward testing, Agranovich et al[37] found that Russians performed slower on the Color Trails Test[29], which is similar to the Trail Making Test Part B[30] but thought to be more culturally fair. Participants completed an assessment of attitudes toward time, the Culture of Time Inventory-33 (COTI-33). There were significant differences between groups on two of five COTI-33 factors (planning and punctuality), which mediated performance differences on two timed tests, implying complex relationships between attitudes about time and test performance. In the current study, differences in performance on Trails B were largely attenuated once country specific z-scores were applied.

Previous experience with cognitive testing or “test-wiseness” [38] is another consideration and is associated with exposure to cognitive tests in educational, clinical, or research settings and refers to the ability to understand the characteristics of a test and use strategies to improve performance. This concept is especially relevant as clinical trials spread to countries that have not had much exposure to neuropsychological testing[39]. Indeed, in our sample, 33.5% of Durham participants and 14.1% of Kannapolis/Cabarrus participants reported having taken a memory test before. Only 6.5% of Tomsk participants reported having done so. Participants in Kannapolis/Cabarrus and Tomsk were more likely to have experience with memory tests due to medical conditions rather than for research purposes as was the case for most of the Durham site. This prior memory testing was associated with better performance on Trails B and the WLM in Durham, indicating a relationship between test-wiseness and performance. There were small but significant differences between Durham and Kannapolis on WLM and MCFSI attributable to prior testing, which may reflect the reason for prior testing at those sites (medical conditions vs. research participation).

The effects of education are a common theme in cross-cultural studies and it has been noted that education and culture can be confounded [7,40]. However, it is possible that cultural or ethnic differences in cognitive performance associated with education are greater within countries than across countries. In international studies, comparisons by years of education may not be appropriate [41]. Rather, it may be more appropriate to compare groups by degree earned due to variability in correspondence between years of education and degree obtained [37]. In this study, we grouped individuals by the highest education level attained. The Russian site had a relatively high level of education, as Tomsk is the home to six universities whereas the Kannapolis site had the lowest average level of education.

Variation in cognitive performance may be also due to non-cultural factors such as differences in the prevalence and treatment of health conditions. Consistent with results of a prior study comparing health in older adults in the US and Russia [42], there were significant differences between sites on most health factors. Differences in the prevalence of various health factors may influence the average cognitive performance by country. The Russian group had significantly higher rates of CVD, hypertension, obesity, and stroke, consistent with prior public health reports from that region[43,44]. Reasons for the high rates of CVD and CVD risk factors have been attributed to reductions in public health expenditures in the 1990s, high rates of smoking, and poor eating habits. Further, the emergence of high and low socioeconomic groups resulted in reduced access to care for the

lower socioeconomic segment of Russian society[44]. In the current study, the effects of diabetes and stroke differed across tests but differences in prevalence rates did not result in differences across sites once country specific z-scores were applied.

The Russian group had the lowest prevalence of self-reported depression of the three groups. However, Russians more frequently reported feeling depressed and/or hopeless in the past 7 days than the other two groups. This apparent inconsistency is in line with reports of high rates of depressive symptoms[45] and the stigma Russians place on depression[46].

4.1 Limitations

There are a number of limitations to note including the fact that this analysis is cross-sectional and individuals' performance at a single time point can be influenced by many factors. Sites used different methods of recruitment although all strove to identify cohorts that were representative of the local population. Nonetheless, there may have been some bias in sample selection. It should be noted that more individuals from Tomsk were excluded from this analysis due to poor MoCA scores (<16). Volunteers for cognitive testing studies of this sort may not be representative and thus results may not be generalizable to the larger population. Another limitation is the use of self-reported health status. Cardiovascular health variables are self-reported and may be prone to recall bias or misclassification. However, studies have shown that self-reported medical history of well-defined conditions tends to be accurate[47–49].

4.2 Conclusions

Our goal was to establish interpretive equivalence across measures and we hypothesized that the two U.S. sites would perform more similarly to each other than to the Russian site. Initial findings showed significant differences across sites, but when country specific z-scores were applied, the three sites performed similarly across tests. This suggests that our efforts to adapt the Russian versions of the tests achieved some measure of interpretive equivalence, yet the application of country-specific z-scores were necessary to achieve further equivalence among sites. As a result, our findings concur with those that suggest similar performance across countries when harmonized instruments are used [9], and further highlight the importance of using country-specific z-scores or norms.

Our results point to the importance of appropriate normative values to guide interpretation of test performance for different countries and cultures in both clinical and research contexts. Performance must be measured against some metric (*i.e.*, normative data). If the norms are based on a population that is not well matched to the population under study, for example different age ranges, cultures, education levels, health conditions, or test-wiseness, then misclassification is likely to occur. For the most accurate measure of performance, the normative values should be derived from a group that most closely represents the sample under study. Thus, indicators of normality or abnormality will be aligned with the population's cultural nuances. The lack of existing regional norms presents a challenge as the development of appropriate and well-validated normative standards is time consuming, complicated, and expensive. However, as clinical trials to delay the onset of MCI and AD

become more international, it is important to consider the differences and the influences of culture when designing neuropsychological batteries for clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding provided by Zinfandel Pharmaceuticals, Durham, NC; Takeda Pharmaceuticals, Deerfield, IL; grants from the National Institute on Aging (K01-AG029336 and P30-AG028377); the MURDOCK Study and Duke University's CTSA grant UL1-RR024128-01 from the National Center for Advancing Translational Sciences (NCATS).

References

1. Karlberg JPE. Globalization of sponsored clinical trials. *Nat Rev Drug Discov.* 2008; 7:458–458.
2. Thiers FA, Sinskey AJ, Berndt ER. Trends in the globalization of clinical trials. *Nature Reviews Drug Discovery.* 2008; 7:13–14.
3. Ibia E, Binkowitz B, Saillot JL, Talerico S, Koerner C, Ferreira I, et al. Ethical considerations in industry-sponsored multiregional clinical trials. *Pharm Stat.* 2010; 9:230–241. [PubMed: 20824884]
4. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. 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. *Alzheimers Dement.* 2011; 7:270–279. [PubMed: 21514249]
5. Agranovich AV, Puente AE. Do Russian and American normal adults perform similarly on neuropsychological tests? Preliminary findings on the relationship between culture and test performance. *Arch Clin Neuropsychol.* 2007; 22:273–282. [PubMed: 17331697]
6. Fernandez AL, Marcopulos BA. A comparison of normative data for the Trail Making Test from several countries: Equivalence of norms and considerations for interpretation. *Scandinavian Journal of Psychology.* 2008; 49:239–246. [PubMed: 18419589]
7. Rosselli M, Ardila A. The impact of culture and education on non-verbal neuropsychological measurements: a critical review. *Brain and Cognition.* 2003; 52:326–333. [PubMed: 12907177]
8. Skirbekk V, Loichinger E, Weber D. Variation in cognitive functioning as a refined approach to comparing aging across countries. *Proc Natl Acad Sci U S A.* 2012; 109:770–774. [PubMed: 22184241]
9. Verhey FR, Houx P, Van Lang N, Huppert F, Stoppe G, Saerens J, et al. Cross-national comparison and validation of the Alzheimer's Disease Assessment Scale: results from the European Harmonization Project for Instruments in Dementia (EURO-HARPID). *Int J Geriatr Psychiatry.* 2004; 19:41–50. [PubMed: 14716698]
10. Ardila A. Cultural values underlying psychometric cognitive testing. *Neuropsychol Rev.* 2005; 15:185–195. [PubMed: 16395623]
11. Romero HR, Lageman SK, Kamath VV, Irani F, Sim A, Suarez P, et al. Challenges in the neuropsychological assessment of ethnic minorities: summit proceedings. *Clin Neuropsychol.* 2009; 23:761–779. [PubMed: 19530035]
12. Anger WK, Cassitto MG, Liang YX, Amador R, Hooisma J, Chrislip DW, et al. Comparison of performance from three continents on the WHO-Recommended Neurobehavioral Core Test Battery. *Environ Res.* 1993; 62:125–147. [PubMed: 8325258]
13. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry.* 1984; 141:1356–1364. [PubMed: 6496779]
14. Paul RH, Gunstad J, Cooper N, Williams LM, Clark CR, Cohen RA, et al. Cross-cultural assessment of neuropsychological performance and electrical brain function measures: additional validation of an international brain database. *Int J Neurosci.* 2007; 117:549–568. [PubMed: 17365135]

15. Levav M, Mirsky AF, French LM, Bartko JJ. Multinational neuropsychological testing: Performance of children and adults. *Journal of Clinical and Experimental Neuropsychology*. 1998; 20:658–672. [PubMed: 10079042]
16. Liang, J. Assessing Cross-Cultural Comparability in Mental Health Among Older Adults. In: Skinner, JH.; Teresi, JA.; Holmes, D.; Stahl, SM.; Stewart, AL., editors. *Multicultural Measurement in Older Populations*. Springer Publishing Company; New York, NY: 2002.
17. Johnson TP. Methods and frameworks for crosscultural measurement. *Med Care*. 2006; 44:S17–20. [PubMed: 17060823]
18. Pedraza O, Mungas D. Measurement in cross-cultural neuropsychology. *Neuropsychology review*. 2008; 18:184–193. [PubMed: 18814034]
19. Bhattacharya S, Dunham AA, Cornish MA, Christian VA, Ginsburg GS, Tenenbaum JD, et al. The Measurement to Understand Reclassification of Disease of Cabarrus/Kannapolis (MURDOCK) Study Community Registry and Biorepository. *Am J Transl Res*. 2012; 4:458–470. [PubMed: 23145214]
20. Hochstenbach J, Mulder T, van Limbeek J, Donders R, Schoonderwaldt H. Cognitive Decline Following Stroke: A Comprehensive Study of Cognitive Decline Following Stroke*. *Journal of clinical and experimental neuropsychology*. 1998; 20:503–517. [PubMed: 9892054]
21. Knopman D, Boland L, Mosley T, Howard G, Liao D, Szklo M, et al. Cardiovascular risk factors and cognitive decline in middle-aged adults. *Neurology*. 2001; 56:42–48. [PubMed: 11148234]
22. Launer LJ, Hughes T, Yu B, Masaki K, Petrovitch H, Ross GW, et al. Lowering midlife levels of systolic blood pressure as a public health strategy to reduce late-life dementia: perspective from the Honolulu Heart Program/Honolulu Asia Aging Study. *Hypertension*. 2010; 55:1352–1359. [PubMed: 20404223]
23. McDowd JM, Filion DL, Pohl PS, Richards LG, Stiers W. Attentional abilities and functional outcomes following stroke. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2003; 58:P45–P53.
24. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005; 53:695–699. [PubMed: 15817019]
25. Reitan, R. Trail Making Test: Manual for Administration and Scoring. Reitan Neuropsychological Laboratory; Tucson, AZ: 1986.
26. Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989; 39:1159–1165. [PubMed: 2771064]
27. Walsh SP, Raman R, Jones KB, Aisen PS. ADCS Prevention Instrument Project: the Mail-In Cognitive Function Screening Instrument (MCFSI). *Alzheimer disease and associated disorders*. 2006; 20:S170–178. [PubMed: 17135810]
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975; 12:189–198. [PubMed: 1202204]
29. D'Elia, LF.; Satz, P.; Uchiyama, CL.; White, T. Psychological Assessment Resources. 1994. Color Trails Test: Professional Manual.
30. Dugbartey AT, Townes BD, Mahurin RK. Equivalence of the color trails test and trail making test in nonnative English-speakers. *Archives of Clinical Neuropsychology*. 2000; 15:425–431. [PubMed: 14590218]
31. Welsh K, Butters N, Hughes J, Mohs R, Heyman A. Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Arch Neurol*. 1991; 48:278–281. [PubMed: 2001185]
32. Woodard JL, Dorsett ES, Cooper JG, Hermann BP, Sager MA. Development of a brief cognitive screen for mild cognitive impairment and neurocognitive disorder. *Aging, Neuropsychology, and Cognition*. 2005; 12:299–315.
33. Ferris SH, Aisen PS, Cummings J, Galasko D, Salmon DP, Schneider L, et al. ADCS Prevention Instrument Project: overview and initial results. *Alzheimer Disease & Associated Disorders*. 2006; 20:S109–S123. [PubMed: 17135805]

34. Hambleton, RK.; Merenda, P.; Spielberger, C., editors. Adapting educational and psychological tests for cross-cultural assessment. Lawrence S. Erlbaum Publishers; Hillsdale, NJ: 2005.
35. Freeman MF, Tukey JW. Transformations Related to the Angular and the Square Root. *The Annals of Mathematical Statistics*. 1950; 21:607–611.
36. SAS/STAT. 9.2 ed. Cary, NC: 2011.
37. Agranovich AV, Panter AT, Puente AE, Touradji P. The culture of time in neuropsychological assessment: exploring the effects of culture-specific time attitudes on timed test performance in Russian and American samples. *J Int Neuropsychol Soc*. 2011; 17:692–701. [PubMed: 21554785]
38. Scruggs TE, Lifson SA. Current conceptions of testwiseness: Myths and realities. *School Psychology Review*. 1985; 14:339–350.
39. Petryna A. Clinical trials offshored: on private sector science and public health. *BioSocieties*. 2007; 2:21–40.
40. Perez-Arce, P.; Puente, A. Neuropsychological assessment of ethnic minorities. In: Sbordone, RJ.; Long, CJ., editors. *Ecological validity of neuropsychological tests*. GR Press; Delray Beach: 1996.
41. Manly JJ, Jacobs DM, Sano M, Bell K, Merchant CA, Small SA, et al. Effect of literacy on neuropsychological test performance in nondemented, education-matched elders. *J Int Neuropsychol Soc*. 1999; 5:191–202. [PubMed: 10217919]
42. Jogerst GJ, Daly JM, Hesli V, Saha C. Comparison of health and effective functioning in Russia and the United States. *Clin Interv Aging*. 2006; 1:189–196. [PubMed: 18044115]
43. Nichols, M.; Townsend, N.; Luengo-Fernandez, R.; Leal, J.; Gray, A.; Scarborough, P., et al. *European Cardiovascular Disease Statistics 2012*. European Society of Cardiology, Sophia Antipolis; Brussels: 2012.
44. Petrukhin IS, Lunina EY. Cardiovascular Disease Risk Factors and Mortality in Russia: Challenges and Barrier. *Public Health Reviews*. 2011; 33:436–449.
45. Jogerst GJ, Zheng S, Frolova EV, Kim MY. Late-life depressive symptoms: an international study. *Family Practice*. 2012; 29:407–415. [PubMed: 22147242]
46. Turvey CL, Jogerst G, Kim MY, Frolova E. Cultural differences in depression-related stigma in late-life: a comparison between the USA, Russia, and South Korea. *Int Psychogeriatr*. 2012; 24:1642–1647. [PubMed: 22621852]
47. Baumeister H, Kriston L, Bengel J, Harter M. High agreement of self-report and physician-diagnosed somatic conditions yields limited bias in examining mental-physical comorbidity. *J Clin Epidemiol*. 2010; 63:558–565. [PubMed: 19959329]
48. Kehoe R, Wu S-Y, Leske MC, Chylack LT. Comparing Self-reported and Physician-reported Medical History. *American Journal of Epidemiology*. 1994; 139:813–818. [PubMed: 8178794]
49. Schneider ALC, Pankow JS, Heiss G, Selvin E. Validity and Reliability of Self-reported Diabetes in the Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology*. 2012; 176:738–743. [PubMed: 23013620]

Research in Context

Systematic Review

The MeSH terms, "cross-cultural comparison," "neuropsychological tests," "internationality," and "clinical trials as topic" were used to identify studies. Few have evaluated the international impact of cultural differences on cognitive performance specifically in older populations[9, 41] although the general topic has gained more attention in recent years. Verhey et al.[9] performed a cross-national comparison and harmonization of the Alzheimer's Disease Assessment Scale (ADAS) in 283 patients with probable or possible AD. Differences in translations and administration across countries were found. The authors noted that there had been few prior attempts to harmonize instruments cross-culturally and concluded that instruments need to be harmonized across countries for multinational studies and the process should be centrally coordinated.

Interpretation

The current study similarly emphasizes the need to consider cross-cultural differences in cognitive performance and their potential effects on clinical trials. To our knowledge, this is the largest multinational study to quantitatively compare cross-cultural cognitive test performance in cognitively normal older adults. Results show that variation in performance can be attributed to several factors and suggest that some of the difference can be addressed with the use of local norms.

Future Directions

Future studies should evaluate the cross-cultural equivalence of tests used in multinational clinical trials. Some variation in performance can be addressed with local normative values. However, tests that are more culture-bound may require more detailed study in order to harmonize measures across countries.

Abbreviations: Trails B=Trail Making Test Part B; MoCA=Montreal Cognitive Assessment; MCFSI= Alzheimer's Disease Cooperative Studies Mail-In Cognitive Function Screening Instrument; WLM= Word List Memory Test delayed recall.
Covariates included: age, sex, race, education, cardiovascular disease (including: atrial fibrillation, heart attack, pacemaker, or valve replacement), diabetes, blood pressure, obesity, stroke, and exposure to prior memory testing.

Figure 1.

Table 1

Demographic characteristics of analytic sample (n=2,972)

Characteristic	Durham	Kannapolis/Cabarrus	Toms	Total
n(%)	1206 (40.6)	886 (29.8)	880 (29.6)	2972
Mean Age(SD) ^{df}	67.8 (7.8)	67.5 (7.9)	70.6 (5.4)	68.6 (7.3)
Sex(female) ^{a,df}	837 (69.4)	572 (64.6)	681 (77.4)	2090 (70.3)
Race ^{b,df}				
White	1007 (83.5)	812 (91.7)	880 (100)	2699 (90.8)
Afr American	178 (14.8)	58 (6.6)	-	236 (7.9)
Other	21 (1.7)	16 (1.7)	-	37 (1.2)
Education Level ^{b,df}				
1 <High School	8 (0.7)	25 (2.8)	67 (7.6)	100 (3.4)
2 High School	87 (7.2)	177 (20.0)	64 (7.3)	328 (11.0)
3 Some College	245 (20.3)	362 (40.9)	287 (32.6)	894 (30.1)
4 Bachelor's or higher	866 (71.8)	322 (36.3)	462 (52.5)	1650 (55.5)
Health				
CVD (1 or more) ^{b,df}	151 (12.5)	173 (19.5)	392 (44.6)	716 (24.1)
Depression ^{a,df}	350 (29.0)	205 (23.1)	45 (5.1)	600 (20.2)
Diabetes ^{a,c}	158 (13.1)	167 (18.9)	162 (18.4)	487 (16.4)
Hypertension ^{df}	578 (47.9)	442 (49.9)	716 (81.4)	1736 (58.4)
Obesity ^c	369 (30.6)	297 (33.5)	320 (36.4)	986 (33.2)
Stroke ^{d,e}	37 (3.1)	41 (4.6)	68 (7.7)	146 (4.9)
Exposure to prior cognitive testing ^{b,df}	408 (33.9)	123 (13.9)	55 (6.3)	586 (19.8)

Results shown are number(percent) unless otherwise indicated as mean(standard deviation). Categorical comparisons were made with χ^2 tests; comparisons of continuous variables were made with general linear models.

Abbreviations: SD=Standard deviation; CVD indicates the presence of one or more of the following conditions: atrial fibrillation, congestive heart failure, pacemaker, valve replacement, or heart attack. MoCA=Montreal Cognitive Assessment.

Significance Comparisons:

Durham vs Kannapolis:

Durham vs Toms:

Kannapolis vs Toms:

^a p<.05;

^b p<.001

^c p<.05;

^d p<.001

^e p<.05;

f
 $p < .001$

Table 2

Mean raw cognitive test scores by site

Test	Site			
	Durham	Cabarrus/Kannapolis	Tomsk	Total
MoCA ^{b,df}	26.9 (2.5)	24.9 (3.1)	22.7 (2.9)	25.0 (3.3)
Score ≥ 26, n(%)	915 (75.9)	411 (46.4)	170 (19.3)	1496 (50.3)
Score < 26, n(%)	291 (24.1)	475 (53.6)	710 (80.7)	1476 (49.7)
MCFSI ^{b,df}	2.5 (2.5)	3.0 (2.6)	5.1 (3.2)	3.4 (2.9)
Trails B ^{b,df}	95.9 (50.4)	111.4 (62.9)	183.98 (72.1)	126.6 (71.9)
WLM ^{b,df}	7.0 (2.1)	6.5 (2.1)	6.1 (2.0)	6.5 (2.1)

Values represent mean(standard deviation) unless noted as number (%).

Abbreviations: SD=Standard deviation; MoCA=Montreal Cognitive Assessment. MCFSI=Alzheimer's Disease Cooperative Studies Mail-In Cognitive Function Screening Instrument. Trails B=Trail Making Test Part B. WLM= Word List Memory Test delayed recall.

General linear model significance comparisons:

Durham vs Kannapolis:

Durham vs Tomsk:

Kannapolis vs Tomsk:

^a p<.05;

^b p<.001

^c p<.05;

^d p<.001

^e p<.05;

^f p<.001

Table 3
Mixed linear regression models evaluating the influence of demographic characteristics on test performance, based on country-specific z-scores.

Variable	MoCA			Trails B			Word List			MCFSI		
	Durham	Kannapolis	Tonsk	Durham	Kannapolis	Tonsk	Durham	Kannapolis	Tonsk	Durham	Kannapolis	Tonsk
Site	Reference	0.49 (−0.40, 1.37)	1.75 (0.72, 2.79)	Reference	0.66 (−0.23, 1.54)	0.84 (−0.20, 1.87)	Reference	0.61 (−0.33, 1.55)	2.09 (0.99, 3.18)	Reference	0.17 (−0.80, 1.14)	1.76 (0.62, 2.89)
Age (years)	−0.03 (−0.03, −0.02)	−0.01 (−0.02, −0.003)	−0.02 (−0.03, −0.01)	−0.04 (−0.05, −0.04)	−0.01 (−0.02, −0.001)	−0.01 (−0.02, 0.006)	−0.03 (−0.04, −0.03)	−0.01 (−0.02, 0.004)	−0.02 (−0.03, −0.01)	−0.01 (−0.02, −0.01)	−0.003 (−0.01, 0.01)	−0.02 (−0.03, −0.005)
Sex	0.21 (0.10, 0.31)	0.09 (−0.07, 0.25)	−0.15 (−0.32, 0.03)	0.14 (0.03, 0.24)	0.04 (−0.12, 0.21)	−0.20 (−0.37, −0.02)	0.55 (0.43, 0.66)	−0.05 (−0.22, 0.13)	−0.35 (−0.54, −0.16)	−0.003 (−0.12, 0.12)	0.002 (−0.17, 0.18)	−0.18 (−0.37, 0.01)
Education	0.38 (0.31, 0.46)	−0.09 (−0.19, 0.01)	−0.11 (−0.21, −0.01)	0.34 (−0.26, 0.41)	−0.01 (−0.12, 0.09)	0.02 (−0.08, 0.12)	0.19 (0.11, 0.27)	−0.05 (−0.16, 0.06)	−0.03 (−0.14, 0.07)	0.24 (0.16, 0.33)	−0.03 (−0.14, 0.09)	−0.08 (−0.19, 0.03)
Race	−0.55 (−0.65, −0.46)	-	-	−0.50 (−0.60, −0.41)	-	-	−0.14 (−0.24, −0.05)	-	-	−0.25 (−0.35, −0.14)	-	-
Diabetes	−0.14 (−0.23, −0.06)	-	-	−0.14 (−0.22, −0.05)	-	-	0.05 (−0.11, 0.07)	-	-	−0.14 (−0.23, −0.04)	-	-
Stroke	−0.11 (−0.26, 0.03)	-	-	−0.32 (−0.47, −0.18)	-	-	0.01 (−0.14, 0.17)	-	-	−0.32 (−0.48, −0.16)	-	-
Depression	−0.01 (−0.08, 0.09)	-	-	−0.07 (−0.15, 0.01)	-	-	−0.06 (−0.14, 0.17)	-	-	−0.51 (−0.60, −0.42)	-	-
History of Mem Test	−0.03 (−0.14, 0.07)	0.02 (−0.18, 0.22)	−0.001 (−0.26, 0.26)	0.12 (0.01, 0.22)	−0.17 (−0.36, 0.03)	−0.17 (−0.43, 0.09)	0.24 (0.12, 0.35)	−0.32 (−0.53, −0.11)	−0.22 (−0.50, 0.05)	−0.04 (−0.16, 0.08)	−0.25 (−0.46, −0.03)	−0.15 (−0.43, 0.13)

Estimate (95% confidence interval); statistically significant items are bolded.

Durham is the reference category, thus all comparisons are to be made to the Durham site. In instances where there is no significant difference between sites (Race, Diabetes, Stroke, and Depression) estimates and 95% confidence intervals are listed under Durham only.

Abbreviation: Mem Test=Memory Test.