Journal of Alzheimer's Disease 53 (2016) 1107–1114
DOI 10.3233/JAD-160248
IOS Press

Measurement Error, Reliability, and Minimum Detectable Change in the Mini-Mental State Examination, Montreal Cognitive Assessment, and Color Trails Test among Community Living Middle-Aged and Older Adults

Joanne Feeney[a,b,1,*], George M. Savva[b,c,1], Claire O'Regan[b], Bellinda King-Kallimanis[b], Hilary Cronin[b] and Rose Anne Kenny[b]
[a]*Centre for Public Health, Queens University Belfast, Royal Victoria Hospital, Belfast, UK*
[b]*The Irish Longitudinal Study on Ageing, Lincoln Gate, University of Dublin, Trinity College, Dublin, Ireland*
[c]*School of Health Sciences, University of East Anglia, Norwich Research Park, Norwich, UK*

Accepted 27 April 2016

Abstract.
Background: Knowing the reliability of cognitive tests, particularly those commonly used in clinical practice, is important in order to interpret the clinical significance of a change in performance or a low score on a single test.
Objective: To report the intra-class correlation (ICC), standard error of measurement (SEM) and minimum detectable change (MDC) for the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), and Color Trails Test (CTT) among community dwelling older adults.
Methods: 130 participants aged 55 and older without severe cognitive impairment underwent two cognitive assessments between two and four months apart. Half the group changed rater between assessments and half changed time of day.
Results: Mean (standard deviation) MMSE was 28.1 (2.1) at baseline and 28.4 (2.1) at repeat. Mean (SD) MoCA increased from 24.8 (3.6) to 25.2 (3.6). There was a rater effect on CTT, but not on the MMSE or MoCA. The SEM of the MMSE was 1.0, leading to an MDC (based on a 95% confidence interval) of 3 points. The SEM of the MoCA was 1.5, implying an MDC95 of 4 points. MoCA (ICC = 0.81) was more reliable than MMSE (ICC = 0.75), but all tests examined showed substantial within-patient variation.
Conclusion: An individual's score would have to change by greater than or equal to 3 points on the MMSE and 4 points on the MoCA for the rater to be confident that the change was not due to measurement error. This has important implications for epidemiologists and clinicians in dementia screening and diagnosis.

Keywords: Aging, cognition, measurement error, reliability

INTRODUCTION

Measures of cognitive function are not perfectly reliable, and this has implications for their application in clinical and research settings. In clinical practice it is important to know how likely it is that an observed change in an individual is due to chance, or

[1]These authors contributed equally to this work.

[*]Correspondence to: Dr. Joanne Feeney, Centre for Public Health, Queens University Belfast, Royal Victoria Hospital, Belfast BT12 6BA, UK. Tel.: +44 353 1 8963738; E-mail: joanne.feeney@qub.ac.uk.

what the clinical relevance is of a low score on a single test.

In research settings, unreliable measures can lead to regression dilution bias or false positive associations when testing predictors of cognitive change, when using cognitive scores as predictors, or to control for confounding [1]. Methods to correct for measurement error rely on having good estimates of the standard error of measurement (SEM) [2], that is the standard deviation (SD) of the within-person variation in each measure. From this we can derive the reliability of the measure by comparison with the variability between individuals and the minimum detectable change (MDC) for an individual, which is the largest difference that could reasonably be attributed to a random fluctuation in performance.

The Mini-Mental State Examination (MMSE) [3] and the Montreal Cognitive Assessment (MoCA) [4] are two of the most commonly used short cognitive tests, in clinical assessment and diagnosis, to determine eligibility for treatment and as inclusion criteria and outcome measures for clinical trials and epidemiological studies.

The Color Trails Test (CTT) is a culturally neutral version of Trail Making Task used to measure aspects of executive function and attention. As with the Trail Making Task, two CTT tasks are applied. CTT1 assesses visuo-motor skills and attention, while CTT2 assesses sequencing and mental flexibility, in addition to the skills required for CTT1 [5].

The test-retest reliability of the MMSE has been reported in many populations. The MoCA is more recently developed, and its reliability has only previously been tested in small convenience samples of patients, or with non-English language versions of the instrument. While the reliability of the TMT has been investigated in various populations, there is little known about the reliability of CTT generally, and even less in older adults specifically.

Here we report practice effects, SEM, intra-class correlation, minimum detectable change, inter-rater reliability, and effect of the time between assessments on reliability for of MMSE, MoCA, CTT1, and CTT2 among a representative sample of the middle-aged and older Irish adult population.

MATERIALS AND METHODS

Population and sample

The Survey of Health Ageing and Retirement in Europe/Irish Longitudinal Study on Ageing (SHARE/TILDA) collaboration was established to estimate the measurement properties of a comprehensive health assessment among nationally representative samples of the European population. The SHARE-Ireland study has been described in detail previously [6], but in brief, a population representative sample of community-living people aged 50 years and older was recruited, each of whom completed a structured interview with a trained interviewer in 2007 (total $N = 1119$).

The extant SHARE-Ireland cohort at 2010 ($n = 827$) was contacted and offered the opportunity to undergo a comprehensive health assessment within the dedicated TILDA health assessment center at Trinity College Dublin. Initial contact was made by post and followed-up by telephone between September 2011 and March 2012, with 377 participants consenting to be directly contacted regarding the study by mail within a ten day period. Of these, 253 agreed to an initial health assessment. Ethical approval was obtained from the Faculty of Health Sciences Ethics Committee, Trinity College. All participants gave informed consent.

Assessments

The health assessment was the same as that administered to participants of TILDA and was conducted in the TILDA health assessment center at Trinity College Dublin by one of two experienced research nurses [7]. The assessment took approximately 3 h, was conducted in either a morning or an afternoon session, and included a large battery of cognitive, cardiovascular, gait, and vision measures including the MMSE, MoCA, and CTT administered by research nurses.

A reduced version of the TILDA computer-aided personal interview was adapted into a self-completed electronic form and administered before the health assessment. This included questions on chronic disease, disability, employment, social and financial circumstances.

Repeat assessment

On completing their first health assessment, 180 participants were invited to attend an identical repeat assessment, scheduled to be conducted approximately one month (for 50% of participants) or three months (for the remainder) after the initial assessment. In total 128 participants agreed to the repeat assessment giving a response rate of 71%

(25 refused; 27 could not be scheduled within the required timeframe).

Among 50% of participants the nurse conducting the assessment was changed, while the other 50% had the same assessor at both interviews. Time of day (morning versus afternoon) was also switched between assessments for 50%. This approach was taken in order to identify within-person variation from variation caused by changing time of day or rater. Change of nurse, change of time of day, and delay between assessments were randomised using a minimisation routine designed to achieve balance between all combinations of these covariates, the age group, and sex of the participants.

Analysis

The outcomes of interest were MMSE and MoCA scores, time to complete CTT1 and CTT2 (in seconds) and the difference between the two (delta CTT). Analysis was restricted to include only participants who completed a repeat assessment. Missing data was very rare in the sample, with a maximum of three participants missing scores for any of the tests at any occasion and so complete case analysis was used. Mean scores were compared between the initial and repeat assessments, between each rater, and time of day using paired *t-tests* or Mann-Whitney U tests as appropriate. Repeat and follow-up scores for all measures were plotted to illustrate the distribution of differences between pairs of assessments. CTT1 and CTT2 times are not normally distributed and have a variance strongly related to average scores, so findings for these are presented both on the natural scale for ease of interpretation and on a log scale on which they have a homoscedastic normal distribution. Comparisons of mean scores between baseline and repeat measurements were made using paired *t-tests*, restricted to relevant subgroups, for example those who changed rater or changed time of day between assessments.

Mixed effects multilevel linear regression models were used to the between- and within-person variance components. The variation in the underlying true cognitive scores between individuals was assumed to be normally distributed with standard deviation $SD_{Between}$. The variation between the true underlying score and each observed score was assumed to be constant across individuals with standard deviation SD_{Within}. The minimum detectable change is $\sqrt{2} \times Z \times SD_{Within}$, where Z=1.96 for the 95% limit (that is, 95% of observed differences between pairs of observations will be within this limit given no true difference) and $Z = 1.65$ for the 90% limit. The ICC is the proportion of total variance not accounted for by within person variation, that is $\text{ICC} = \frac{SD^2_{Between}}{SD^2_{Between}+SD^2_{Within}}$.

Rater, time of day, and whether the observation was a baseline or repeat were added as fixed effects to estimate the effects of these factors on mean scores and to remove their contribution from estimates of within person variation. A term corresponding to the effect of rater within participants was added in a further analysis to test whether the test-retest difference was significantly higher among those who changed rater.

Each reliability statistic is presented with a 95% confidence interval, and all analyses were conducted using Stata 12.1.

RESULTS

Participant characteristics

A total of 128 participants underwent repeat assessments (58 men). The median age of the sample was 71 y (range 55–94 y, IQR 66–76 y). The median delay between assessments was 88 days (range 28–141 days, IQR 70–104 days). The median baseline MMSE of the sample was 29 (range 20–30, IQR 28–29). The median MoCA was 25 (range 13–30, IQR 23–28). Sixty participants were allocated to change time of day between assessments, while 70 were not. Sixty one participants changed rater, while 69 did not.

Mean cognitive scores across repeats, raters and time of day

Table 1 shows the scores for each cognitive test at baseline and repeat assessments, for each rater, and by time of day. The rater and time of day scores are only shown for participants who changed each respective factor at the repeat. MMSE (0.3 points; $p = 0.020$), MoCA (0.41 points; $p = 0.038$), CTT1 time (3.1 s; $p = 0.031$), and CTT2 (3.9 s; $p = 0.029$) times all improved slightly between the baseline and repeat assessments, indicating a likely practice effect. The Delta CTT did not improve. There was no significant effect of time of day on any measure, although CTT1 and CTT2 times were slower in the morning than in the afternoon among participants who switched time of day in their assessments; these

Table 1
MoCA and MMSE scores in baseline and repeat assessments with associated variance and reliability statistics

	MMSE (points)	MoCA (points)	CTT1		CTT2	
			Natural scale (seconds)	Log scale	Natural scale (seconds)	Log scale
Mean	28.1	24.8	53.7	3.91	109.4	4.64
Baseline (SD)	(2.07)	(3.61)	(21.7)	(0.38)	(37.4)	(0.32)
Mean Repeat	28.4	25.2	50.6	3.84	105.5	4.59
(SD)	(2.07)*	(3.56)*	(21.1)*	(0.40)*	(40.8)*	(0.35)*
Mean Rater 1[a]	28.0	24.8	50.2	3.85	103.2	4.58
(SD)	(2.22)	(3.67)	(19.0)	(0.35)	(36.8)	(0.32)
Mean Rater 2[a]	28.2	25.1	52.0	3.89	108.1	4.64
(SD)	(2.26)	(3.97)	(20.4)	(0.36)	(36.7)*	(0.30)*
Mean Test AM[b]	28.5	25.3	49.8	3.84	100.1	4.56
(SD)	(1.60)	(3.30)	(19.6)	(0.37)	(32.2)	(0.30)
Mean Test PM[b]	28.6	25.8	46.5	3.76	96.8	4.52
(SD)	(1.98)	(3.12)	(18.6)	(0.39)	(32.9)	(0.33)
SD between individuals	1.77	3.22	20.7	0.35	41.7	0.34
(95% CI)	(1.54–2.05)	(2.81–3.69)	(17.9–23.9)	(0.31–0.41)	(36.6–47.5)	(0.30–0.38)
SEM	1.02	1.52	11.3	0.19	13.7	0.12
(95% CI)	(0.91–1.16)	(1.34–1.72)	(9.9–12.8)	(0.17–0.22)	(12.1–15.6)	(0.10–0.13)
ICC	0.75	0.81	0.77	0.77	0.90	0.89
(95% CI)	(0.67–0.82)	(0.75–0.87)	(0.69–0.84)	(0.70–0.84)	(0.86–0.93)	(0.85–0.92)
90% MDC	2.38	3.54	26.2	0.45	32.0	0.27
(95% CI)	(2.11–2.69)	(3.13–4.00)	(23.1–29.8)	(0.39–0.51)	(28.2–36.2)	(0.24–0.30)
95% MDC	2.84	4.21	31.2	0.53	38.1	0.32
(95% CI)	(2.51–3.21)	(3.73–4.76)	(27.6–35.5)	(0.47–0.60)	(33.6–43.2)	(0.29–0.36)

*scores at retest were statistically significantly improved compared to scores at baseline ($p < 0.05$ using a paired *t-test*) for all tests. CTT2 times were significantly different between raters ($p < 0.05$ using a paired *t-test*). [a]Rater scores are calculated only among participants who changed rater at the repeat assessment. [b]Time of day scores are calculated only among participants who changed time of day at the repeat assessment.

differences were not statistically significant. Mean CTT2 time was significantly different between raters (difference = 4.8 s; $p = 0.046$), but the rater did not affect any other test score. The effect of rater did not vary within participants for any measure suggesting that there is no additional variance beyond within-participant variation introduced by changing rater.

Reliability of MMSE and MoCA

The SEM for MMSE is around 1 point and for MoCA is around 1.5 points. This leads to estimates of minimum detectable change of 2.38 points at the 90% level and 2.84 at the 95% level for MMSE, and 3.54 and 4.21 at the 90% and 95% level for MoCA, respectively. Figure 1 indicates regions defined by the 95% minimum detectable change, outside which differences between test scores could indicate a genuine change. Comparing the between and within person standard error leads to estimates of reliability for MMSE of 0.75 and for MoCA of 0.81 in this setting. There was no difference in reliability when assessments were greater or less than 88 days apart (data not shown).

Each of these estimates of reliability is reasonably precise as indicated by small 95% confidence intervals in Table 1.

Reliability of Color Trails Tests (CTT)

On the natural scale the SEM of CTT1 is around 11.3 s, and for CTT2 is 13.7 s. However the magnitude of this error is strongly related to the baseline time (Spearman's rank correlation rho = 0.39; $p < 0.001$), whereas on the log scale the error is unrelated to baseline score (rho = 0; $p = 1$). This indicates that CTT errors are multiplicative and not additive, and variance estimates should be reported on the log scale to be applicable across the ability range. On the log scale the SEM for CTT1 is 0.19 corresponding to a 95% MDC of 0.45. This means that in 95% of cases we would expect a repeat CTT1 time to be between 63% and 157% of the first CTT1 time (since log (1.57)=0.45). The corresponding range for CTT2 is 76% to 131%. The SEM and MDC values shown in Table 1 can be used to derive similar ranges at different confidence levels. Figure 2 displays the average CTT scores at baseline and repeat on both the natural and log scale. Estimates of the reliability of CTT1 and

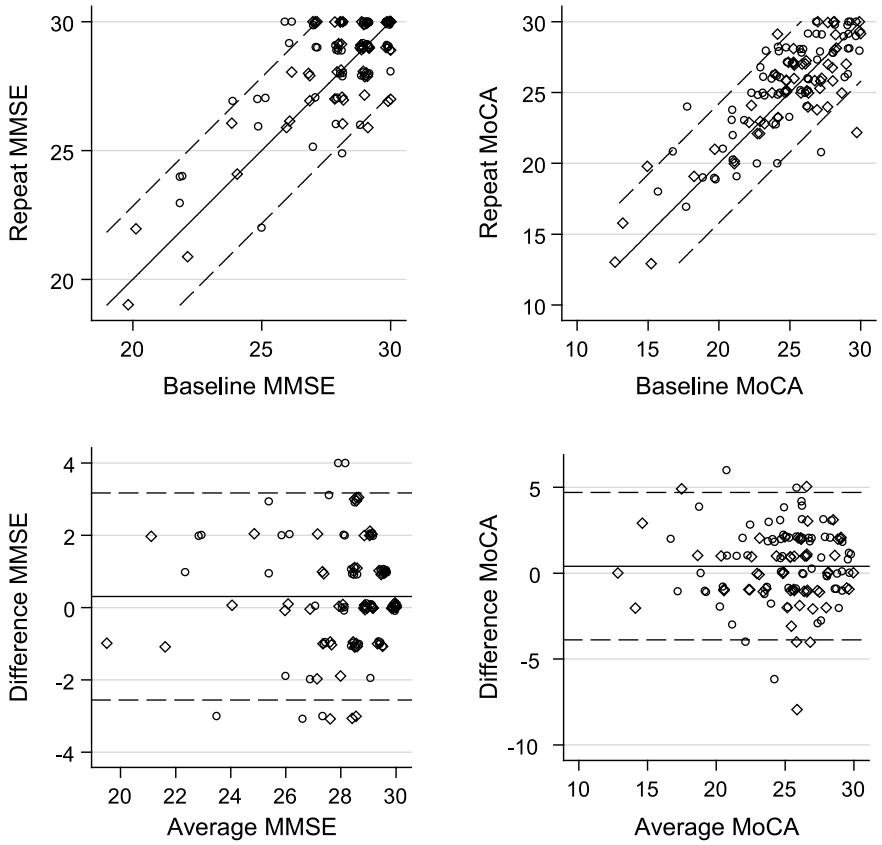

Fig. 1. (Above) baseline and repeat scores for MoCA and MMSE with lines of equality and dashed lines indicating the 95% minimum detectable difference. (Below) Bland-Altman plots showing the difference versus baseline for MoCA and MMSE, with dashed lines representing the 95% limits of agreement, and higher differences indicated better scores at the repeat assessment. Circles represent individuals assessed by the same rater at both assessments, diamonds represent individuals assessed by different raters at each assessment.

CTT2 are not affected by log transformation. Despite Delta CTT being less susceptible to practice effects, it is considerably less reliable than CTT2 (ICC = 0.75; 95% CI = 0.66–0.82 compared to ICC = 0.90; 95% CI = 0.86–0.93). Both CTT2 and delta CTT are similarly correlated with MMSE and MoCA scores (data not shown).

DISCUSSION

We have estimated the SEM, test-retest reliability, and minimum detectable change of three commonly used cognitive tests, within a sample representative of a relatively healthy population aged 55 years and older, when tests are given 2–4 months apart. These estimates should be used when assessing changes in cognitive test scores in individuals or when using methods to adjust for measurement error in research studies when measurements are taken months or years apart. Cognitive tests are typically interpreted in clinical or epidemiological contexts to reflect underlying cognitive function net of any day-to-day fluctuation caused by transient acute illness. Reliability estimated using a 2–4 month lag shows the ability of cognitive tests to measure this and the variation in cognition likely to be attributable to chance when measures are taken at this interval or longer, as is typical in clinical and research contexts. Reliability estimates using shorter time periods would better reflect the pure measurement properties of the instrument, but may overestimate the reliability likely to be observed in most practical settings. Equally, a longer lag period between testing would not be advisable when investigating reliability because of the increasing likelihood that reliability estimates would be influenced by onset of disease processes [8] or other medium to long-term changes in health impacting cognitive state. Hence when interpreting long term changes in cognitive function, when tests are

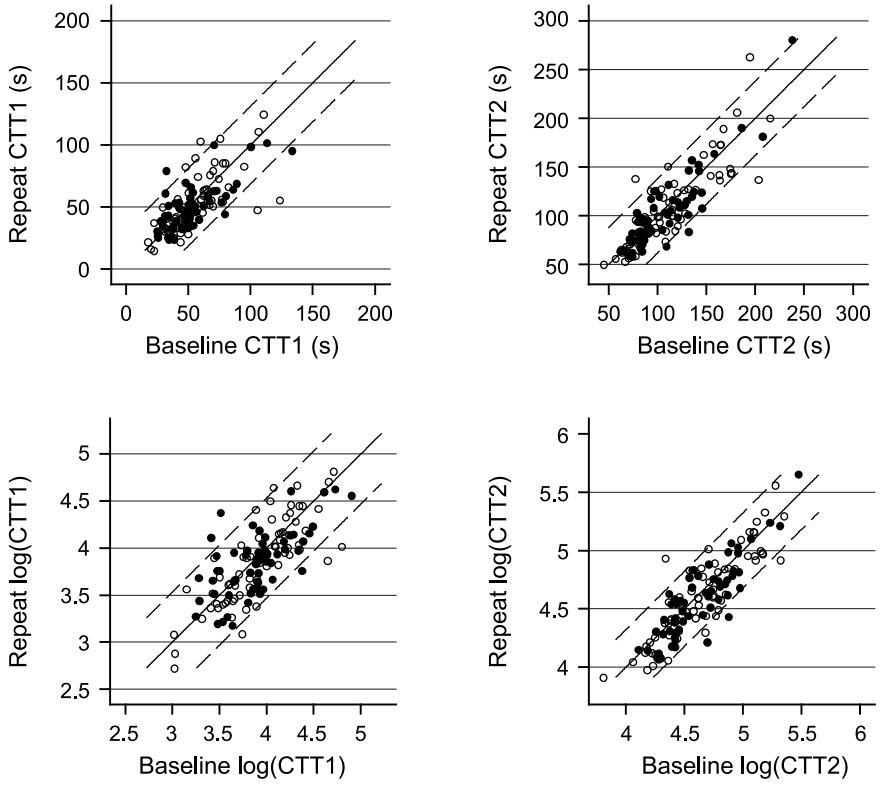

Fig. 2. Baseline and repeat CTT times presented on (above) the natural scale and (below) a logarithmic scale. Hollow circles represent individuals assessed by the same rater at both assessments, filled circles represent individuals assessed by different raters at each assessment.

repeated over months or years, reliability estimates using a 2–4 month period are likely to be the most appropriate.

For the MMSE, the SEM is around 1 point, meaning that a difference of 3 points between two observations should be considered beyond that expected by chance. For the MoCA the SEM is 1.5 points, meaning that a difference of 3-4 points could be expected by chance depending on the confidence level used. The test-retest reliability of the MMSE has been widely reported in different populations [reviewed in 8], with findings broadly in agreement with our study. That the MMSE in our sample has similar reliability to other populations suggests that the reliability estimates of other tests are likely to be generalizable. Despite the increasing use of the MoCA as a clinical and research tool, its test-retest reliability is not widely reported. Nasreddine et al. discovered a SD of 2.5 points in test-retest differences among 23 members of a convenience sample during the development of the instrument [4], which corresponds closely to our SEM of 1.5 points since the corresponding SD from our results is $\sqrt{2} \times 1.5 =$ 2.1 points. Previous estimates of reliability from validation studies of non-English language translations of MoCA report higher values for the ICC than we have done, although the ICC depends on the distribution of scores within the sample it is estimated in and so is typically not comparable between settings and populations [9–12]. A study in China among 111 cerebrovascular disease patients also indicated good inter-rater reliability [13].

The authors of the CTT report correlation coefficients of 0.64 and 0.79 for CTT1 and CTT2, respectively, in a sample of young adults [5]. Additional investigations of reliability are lacking, particularly among older adults, although Dong et al. recently reported fair to good (0.57 and 0.67) test-retest correlations for CTT1 and CTT2 time, in a sample of 30 older adults attending a memory clinic [14]. In addition, Tavakoli and colleagues showed very high correlation coefficients (>0.9) for the re-test reliability of both CTT1 and CTT2 in an Iranian normative sample of adults spanning a wide age range [15]. Our results add to this literature by providing reliability information for a population sample of 128

older adults, and including estimates of absolute reliability (SEM) as well as relative reliability (ICC). The findings are in line with previous studies in showing CTT2 to be more reliable than CTT1. The reliability of the Trail Making Task is much more widely studied [16].

Strengths and weaknesses of our study

Our study used a relatively large sample of middle aged and older adults with cognitive function in a generally healthy range. While the initial SHARE Ireland sample was selected to be representative of the Irish population, the selection process into the current sample may have led to relatively healthy sample compared with the older population. There was no indication that the reliability of any test varied with baseline cognitive function when examined on the appropriate scale and so our results are likely to be generalizable to people with mild to moderate cognitive impairment including clinical samples. We have reported confidence intervals for all of our estimates of reliability, allowing their precision to be assessed. Most studies of reliability are conducted using very small samples meaning that estimates are likely to be imprecise, although precision is rarely reported. Our estimates of reliability are precise, and remove rater, time of day and practice effects using a mixed effects multilevel model. The time delay between assessments varied from one month to nearly five months, but no differences in reliability were observed across that range suggesting that neither drift nor day-to-day fluctuation in underlying cognitive function in our group during this time was a significant problem.

Implications

The cognitive tests we have explored have limited reliability when being used to compare individual scores at two time points, suggesting that on their own they may not be useful measures of change. The poor reliability of these cognitive tests also has implications for their use in determining thresholds for, e.g., individual eligibility for treatment or for assessments of driving safety. For example a SEM of 1.5 points means that an individual with a 'true' underlying MoCA score of around 26 could reasonably give a test score of between 23 and 29. This also has implications for planned screening programs that trigger referrals to diagnostic services based on single cognitive test scores.

Our findings are important for researchers using cognitive tests as outcome measures. Estimates of within-person variability are important to inform power calculations and to correct for regression dilution bias caused by misclassification in exposures or confounding variables. Our findings also suggest that the Color Trails Tests should be analyzed on logarithmic scales to ensure homoscedasticity because, when analyzed on a natural scale, those with lower performance had higher variation in CTT completion times.

Our results suggest that many previously detected effects of therapies or risk factors that affect cognitive function might not be detectable at the individual level using the tests we have examined. However it is important to note that our estimates of detectable change do not reflect clinically important change, and so small effects on cognitive function as seen in many research studies may still represent clinically important differences if confirmed in sufficiently large samples.

As well as estimates of reliability, our estimates of rater effects and practice effects are also important for research and clinical practice. Clinicians and researchers using all cognitive assessments should be aware of the practice effects. In addition, every effort should be made to maintain the same rater between administrations, in order to eliminate the impact of inter-rater variability observed for CTT2.

CONCLUSION

Data on the reliability of cognitive tests is sparse, with estimates typically derived from small subsets of clinical samples and presented without any measures of precision. Cognitive tests have substantial with-person variation when repeated over periods of weeks suggesting that they may not precisely reflect the underlying cognitive function of an individual, instead being influenced by day-to-day fluctuation or measurement error. Robust reliability studies should be performed in different populations to inform the use and interpretation of tests.

ACKNOWLEDGMENTS

This research was supported by the U.S. National Institute of Aging (NIA) grant number AG040387 and The Division of Ageing Research and Development at the Institute of Public Health (IPH), Ireland (formally the Centre for Ageing and Research Development in Ireland).

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/16-0248r1).

REFERENCES

[1] Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM (2005) When is baseline adjustment useful in analyses of change? an example with education and cognitive change. *Am J Epidemiol* **162**, 267-278.

[2] Frost C, Thompson SG (2000) Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *J R Stat Soc Ser* **163**, 173-189.

[3] Folstein MF, Folstein SE, McHugh PR (1975) Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[4] Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005) The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* **53**, 695-699.

[5] D'Elia LF, Satz P, Uchiyama CL, White T (1996) *Color Trails Test: Professional manual*. Psychological Assessment Resources, Inc., Odessa, FL.

[6] Börsch-Supan A, Brugiavini A, Jürges H, Mackenbach J, Siegrist J, Weber G (2005) Health, ageing and retirement in Europe - First results from the Survey of Health, Ageing and Retirement in Europe. Available from: http://www.share-project.org/uploads/tx_sharepublications/SHARE_FirstResultsBookWave1.pdf

[7] Cronin H, O'Regan C, Finucane C, Kearney P, Kenny RA (2013) Health and aging: Development of the irish longitudinal study on ageing health assessment. *J Am Geriatr Soc* **61**, S269-S278.

[8] Tombaugh TN, McIntyre NJ (1992) The Mini-Mental State Examination: A comprehensive review. *J Am Geriatr Soc* **40**, 922-935.

[9] Gómez F, Zunzunegui M, Lord C, Alvarado B, García A (2013) Applicability of the MoCA-S test in populations with little education in Colombia: MoCA-S in populations with little education. *Int J Geriatr Psychiatry* **28**, 813-820.

[10] Duro D, Simões MR, Ponciano E, Santana I (2010) Validation studies of the Portuguese experimental version of the Montreal Cognitive Assessment (MoCA): Confirmatory factor analysis. *J Neurol* **257**, 728-734.

[11] Chu LW, Ng KH, Law AC, Lee AM, Kwan F (2015) Validity of the Cantonese Chinese Montreal Cognitive Assessment in Southern Chinese. *Geriatr Gerontol Int* **15**, 96-103.

[12] Chen X, Zhang R, Xiao Y, Dong J, Niu X, Kong W (2015) Reliability and validity of the Beijing version of the Montreal Cognitive Assessment in the evaluation of cognitive function of adult patients with OSAHS. *PLoS One* **10**, e0132361.

[13] Tu Q, Jin H, Ding B, Yang X, Lei Z, Bai S, Zhang YD, Tang XQ (2013) Reliability, validity, and optimal cutoff score of the Montreal Cognitive Assessment (Changsha version) in ischemic cerebrovascular disease patients of Hunan Province, China. *Dement Geriatr Cogn Disord EXTRA* **3**, 25-36.

[14] Dong Y, Thompson CL, Huey Joanne Tan S, Ben Swie Lim L, Pang W, Li-Hsian Chen C (2013) Test-retest reliability, convergent validity and practice effects of the RBANS in a memory clinic setting: A pilot study. *Open J Med Psychol* **2**, 11-16.

[15] Tavakoli M, Barekatain M, Emsaki G (2015) An Iranian normative sample of the Color Trails Test. *Psych Neurosci* **8**, 75-81.

[16] Wagner S, Helmreich I, Dahmen N, Lieb K, Tadic A (2011) Reliability of three alternate forms of the trail making tests a and B. *Arch Clin Neuropsychol* **26**, 314-321.