Cognitive Assessment by Telemedicine: Reliability and Agreement between Face-to-Face and Remote Videoconference-Based Cognitive Tests in Older Adults Attending a Memory Clinic

Herb Howard C. Hernandez[1], Poh Ling Ong[2], Philomena Anthony[2], Siew Ling Ang[2], Nur Bazilah Mohd Salim[2],
Pey Ying Suzanne Yew[2], Noorhazlina Bte Ali[1], Jun Pei Lim[1,3], Wee Shiong Lim[1,3], Justin Chew[1,3]

[1]Department of Geriatric Medicine, Tan Tock Seng Hospital, Singapore
[2]Dementia Care, Nursing Specialty, Tan Tock Seng Hospital, Singapore
[3]Institute of Geriatrics and Active Ageing, Tan Tock Seng Hospital, Singapore

Corresponding Author:
Herb Howard C. Hernandez, MD, GDGRM
Department of Geriatric Medicine, Tan
Tock Seng Hospital, 11 Jalan Tan Tock
Seng Singapore 308433
E-mail: herb_howard_hernandez@ttsh.
co m.sg
ORCID:
https://orcid.org/0000-0002-7205-859X

Received: January 16, 2022
Revised: February 22, 2022
Accepted: February 26, 2022

Background: The coronavirus disease 2019 (COVID-19) pandemic has spurred the rapid adoption of telemedicine. However, the reproducibility of face-to-face (F2F) versus remote videoconference-based cognitive testing remains to be established. We assessed the reliability and agreement between F2F and remote administrations of the Abbreviated Mental Test (AMT), modified version of the Chinese Mini-Mental State Examination (mCMMSE), and Chinese Frontal Assessment Battery (CFAB) in older adults attending a memory clinic. **Methods:** The participants underwent F2F followed by remote videoconference-based assessment by the same assessor within 3 weeks. Reliability was evaluated using intraclass correlation coefficients (ICC; two-way mixed, absolute agreement), the mean difference between remote and F2F-based assessments using paired-sample t-tests, and agreement using Bland-Altman plots. **Results:** Fifty-six subjects (mean age, 76±5.4 years; 74% mild; 19% moderate dementia) completed the AMT and mCMMSE, of which 30 completed the CFAB. Good reliability was noted based on the ICC values—AMT: ICC=0.80, 95% confidence interval [CI] 0.68–0.88; mCMMSE: ICC=0.80, 95% CI 0.63–0.88; CFAB: ICC=0.82, 95% CI 0.66–0.91. However, remote AMT and mCMMSE scores were higher compared to F2F—mean difference (i.e., remote minus F2F): AMT 0.3±1.1, p=0.03; mCMMSE 1.3±2.9, p=0.001. Significant differences were observed in the orientation and recall items of the mCMMSE and the similarities and conflicting instructions of CFAB. Bland–Altman plots indicated wide 95% limits of agreement (AMT -1.9 to 2.6; mCMMSE -4.3 to 6.9; CFAB -3.0 to 3.8), exceeding the a priori-defined levels of error. **Conclusion:** While the remote and F2F cognitive assessments demonstrated good overall reliability, the test scores were higher when performed remotely compared to F2F. The discrepancies in agreement warrant attention to patient selection and environment optimization for the successful adaptation of telemedicine for cognitive assessment.

Key Words: Telemedicine, Telehealth, Dementia

INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic hastened the shift to telemedicine to maintain continuity of care while mitigating the risks of exposure.[1,2] However, despite the growing presence of telehealth services, limited data exists regarding the validity of telemedicine-based cognitive testing, particularly in cognitively impaired older adults.

Previous studies have described the utility of telemedicine for the diagnosis of dementia.[3] Others have evaluated the reliability of the remotely administered Mini-Mental State Exam (MMSE)[4] and the Montreal Cognitive Assessment tool (MoCa); however, these studies were predominantly conducted in younger individuals or specific clinical conditions such as post-stroke.[5] Moreover, few studies have evaluated a battery of telemedicine-based cognitive tests in the vernacular of Asian populations, particularly in older adults with cognitive impairment. In this regard, it is important to establish the reliability and agreement between telemedicine-based and face-to-face (F2F) cognitive assessments to support the validity of remote cognitive testing in older adults in preparation for future public health emergencies[6] or in other settings where distance limits access to timely healthcare.

Therefore, we aimed to determine the reliability and agreement between F2F and remote videoconference-based assessments of three commonly used cognitive screening tools; namely, the Abbreviated Mental Test (AMT), the modified version of the Chinese MMSE (mCMMSE), and the Chinese Frontal Assessment Battery (CFAB), among older adults with known or suspected cognitive impairment.

MATERIALS AND METHODS

Participants and Setting
We recruited 60 community-dwelling older adults presenting with known or suspected cognitive impairment using a convenience sample of patients attending a tertiary hospital memory clinic. The ethics committee of the National Healthcare Group Domain Specific Review Board reviewed and approved this study (No. 2020/00609).

Inclusion and Exclusion Criteria
We included participants aged 65 years and older who could understand English or Mandarin and could independently use WhatsApp Messenger video calls (https://whatsapp.com), or had caregivers to assist them. We excluded individuals with severe hearing or visual impairments or those with severe behavioral and psychological symptoms precluding assessment.

Data Collection
The participants completed two visits, an F2F visit followed by a remote assessment, scheduled 2–3 weeks after the F2F visit. The AMT and mCMMSE were performed by trained nurses specializing in cognition and memory disorders, followed by an assessment of CFAB by a physician running the memory clinic. For each participant, the same nurse and physician performed the F2F and remote assessments. All raters underwent standardization training before the study.

Upon consenting to the study, all participants and their caregivers (if present) were briefed on the conditions under which videoconferencing would occur. An information sheet was provided that described a standardized setting with adequate lighting; absence of visual orientation cues such as clocks, watches, or calendars; and a quiet environment.

We collected baseline demographic data (age, sex, education level, and first language). Dementia diagnosis using the Diagnostic and Statistical Manual of Mental Disorders fourth edition (DSM-IV) criteria and severity using the locally validated Clinical Dementia Rating (CDR)[7] were rated by the participants' physicians.

Cognitive Assessment
Various items on the cognitive tests were adapted for telemedicine, including clarifying the phrasing of questions and accommodating the different locations of the participants and assessors during remote assessment (Table 1). Modifications were also made to the three-stage command and the "read and obey" items of the mCMMSE to avoid participant responses outside of camera view. For the CFAB, the final item, "environmental autonomy," was omitted because it necessitates physical contact between the assessor and the participant. This was also supported from the psychometric standpoint, as this item loaded poorly and, when removed, im-

Table 1. Cognitive tests adapted for videoconferencing

	Face-to-face	Videoconference-based
AMT	Where are we now?	Where are *you* now?
mCMMSE	What floor are we on now?	What floor are *you* on now?
	In which estate are we?	In which estate are *you now*?
	Three-stage command: "Take this piece of paper, fold it in half, and put it on the floor."	"Take this piece of paper, fold it in half, and *hold it in front of you*."
	"Read and obey: Raise your hands."	"Read and obey: *Close your eyes*."
CFAB	Question 6 "Prehension behavior: Do not take my hands?"	Removed Question 6.

AMT, Abbreviated Mental Test; mCMMSE, modified version of the Chinese Mini-Mental State Examination; CFAB, Chinese Frontal Assessment Battery.

proved the internal consistency of FAB.[8] Thus, the final scores for both the F2F and remote CFAB excluded this item.

Statistical Analysis

For a hypothesized intra-class correlation (ICC) between remote and F2F assessments of 0.80[9] against a null value of 0.60 and an alpha value of 0.05, a minimum sample size of 50 participants was required to achieve a power of 0.80.

We analyzed the reliability of the remote and F2F assessments based on ICC values (two-way, mixed, absolute agreement).[10] Values < 0.5 indicated poor reliability, 0.5–0.75 moderate reliability, 0.75–0.9, good reliability; and > 0.9 excellent reliability.[10] Differences between F2F and remote cognitive scores were also examined using the paired samples t-tests. We then evaluated the agreement between the F2F and remote scores using Bland-Altman plots. These plots illustrated the agreement between F2F and remotely administered measures of each cognitive test by plotting the differences between F2F and remote scores against the mean. The two horizontal dotted lines indicate the 95% limits of agreement, which were estimated by the mean difference ± 1.96 times the standard deviation of the differences.

The a priori-defined acceptable limits of agreement were ± 1, ± 2, and ± 2 for AMT, mCMMSE, and CFAB, respectively. These limits were based on previous data demonstrating a minimal clinically important difference (MCID) of > 1 for AMT[11]; for the MMSE, the reported MCID ranges from 1 to 3[12]; thus, an average of 2 was used. Limited information exists on the MCID for CFAB; therefore, a consensus was reached to define a score difference of ≥ 2 as having a significant effect on clinical outcomes.

The data were analyzed using IBM SPSS Statistics for Windows, version 27.0. (IBM Corp., Armonk, NY, USA) and MedCalc for Windows, version 20.013 (MedCalc Software, Ostend, Belgium).

RESULTS

Of the 60 participants who consented to participate in this study, 56 (93.3%) completed both the F2F and remote assessments. Four participants were unable to complete the remote assessment—change of mind by participant and family (n=2), caregiver unable to commit to assisting the participant (n=1), dental condition (n=1). Thirty participants completed both the F2F and remotely administered CFAB. The mean ± standard deviation duration between F2F and remote assessments was 17.7 ± 3.2 days. Thirty-eight participants (68%) required assistance from their caregivers for the remote assessment.

Table 2 shows the demographic and clinical characteristics of the study population. Most of the participants were female and of Chinese ethnicity. The mean education level was 8.38 ± 4.2 years, corresponding to a secondary school level. Cognitive tests were conducted in English for 30 participants (53.6%) and Mandarin Chinese for 26 (46.4%). Almost half of the participants had a pre-existing diagnosis of dementia, with Alzheimer's dementia (AD) the primary etiology in 21 participants (78%). Dementia was rated based on the CDR scale, with most cases of mild severity.

Table 3 shows the mean differences between F2F and remotely administered AMT, mCMMSE, and CFAB, with their respective ICC values. Participants scored higher during remote testing than during F2F for AMT and mCMMSE, with AMT significantly higher by 0.3 ± 1.1 (p=0.029) and mCMMSE by 1.3 ± 2.9 (p=0.001). No significant differences were observed between F2F and remotely administered CFAB mean scores.

All three assessments demonstrated good to excellent levels of

Table 2. Demographics and baseline characteristics

Characteristic	Value
Age (y)	76.0 ± 5.4
Years of education	8.38 ± 4.20
Global CDR	0.78 ± 0.45
CDR sum of boxes	3.2 ± 2.8
Sex, female	31 (55.4)
Ethnicity	
Chinese	51 (91.1)
Malay	1 (1.8)
Indian	3 (5.4)
Others	1 (1.8)
Language	
English	30 (53.6)
Mandarin	26 (46.4)
Educational level	
No formal education	3 (5.4)
Primary	19 (33.9)
Secondary	21 (37.5)
Tertiary	13 (23.2)
Dementia diagnosis	27 (48.0)
Primary etiology of dementia	
Alzheimer's dementia	23 (85.0)
Vascular dementia	2 (7.4)
Mixed Alzheimer's dementia with stroke disease	1 (3.7)
Others	1 (3.7)
Dementia severity	
Mild	20 (74.0)
Mild-moderate	2 (7.0)
Moderate	5 (19.0)
Advanced	0 (0)

Values are presented as mean±standard deviation or number (%).
CDR, Clinical Dementia Rating.

Table 3. Mean differences and ICCs for each cognitive test

	Face-to-face	Videoconference	Mean difference	p-value	ICC (95% CI)
AMT (n = 56)	8.1 ± 1.9	8.5 ± 1.8	0.3 ± 1.1	0.029	0.80 (0.68–0.88)
mCMMSE (n = 56)	20.1 ± 4.9	21.4 ± 4.7	1.3 ± 2.9	0.001	0.80 (0.63–0.88)
CFAB (n = 30)	10.8 ± 2.8	11.2 ± 3.0	0.4 ± 1.7	0.220	0.82 (0.65–0.91)

Values are presented as mean±standard deviation.
AMT, Abbreviated Mental Test; mCMMSE, modified version of the Chinese Mini-Mental State Examination; CFAB, Chinese Frontal Assessment Battery; ICC, intra-class correlation coefficient; CI, confidence interval.

reliability, with ICC values of 0.80 (95% confidence interval [CI] 0.68–0.88), 0.80 (95% CI 0.63–0.88), and 0.82 (95% CI 0.65–0.91) for AMT, mCMMSE, and CFAB, respectively.

Table 4 shows the differences in the F2F versus remote mCMMSE and CFAB scores by domain. For mCMMSE, the participants scored 0.8 ± 1.5 (p < 0.001) and 0.6 ± 1.0 points (p < 0.001) higher during remote assessment in the orientation and recall domains, respectively. For CFAB, participants scored 0.5 ± 0.9 points higher (p = 0.006) for the similarities item and 0.3 ± 0.7 points higher (p = 0.026) for the conflicting instructions item.

Bland-Altman plots for AMT, mCMMSE, and CFAB are shown in Fig. 1A, 1B, and 1C, respectively. Almost all individual plots were within the 95% limit of agreement for all three cognitive tests. We observed evidence of systematic bias (remote minus F2F scores), with the overestimation of remote mCMMSE (bias = 1.3, 95% CI -4.3 to 6.9) and AMT scores (bias = 0.3, 95% CI -1.9 to 2.6), as shown in Table 3. The 95% limits of agreement were wide, ranging between -1.9 to 2.6 for AMT, -4.3 to 6.9 for mCMMSE, and -3.0 to 3.8 for CFAB, exceeding the a priori-defined levels of error. Notably, there were five outliers (test scores that exceeded the 95% limits of agreement) for AMT, three outliers for mCMMSE, and one outlier for CFAB. Though not reaching statistical significance, when compared to non-outliers, outliers showed a trend towards older age (78.9 ± 4.1 vs. 75.5 ± 5.5 years, p = 0.60), greater severity of cognitive impairment (CDR global scores 1.1 ± 0.7 vs. 0.7 ± 0.4, p = 0.82; CDR sum of boxes scores 5.1 ± 3.3 vs. 2.8 ± 2.6, p = 0.71), and lower educational levels (7.3 ± 5.9 vs. 8.6 ± 3.8 years, p = 0.35).

DISCUSSION

The present study adds to the growing body of evidence examining the validity of telemedicine for cognitive assessment in older adults. To our knowledge, this is the first study evaluating remote CFAB assessment. Specifically, remote videoconferencing-based administration of AMT, mCMMSE, and CFAB showed good reliability but only fair agreement with the F2F assessment. A small but significant bias was observed for AMT and mCMMSE be-

tween both assessment modalities, with remote scores higher than those of the F2F-based assessment. We also found wide limits of agreement for all three cognitive tests, exceeding our predefined limits for maximum acceptable differences. These were, in part, driven by outliers with extreme differences, particularly for the mCMMSE. When analyzed by cognitive domains, participants demonstrated higher scores via remote testing in the orientation and recall items of the mCMMSE and the similarities and conflicting instructions items of the CFAB.

Our findings demonstrating good reliability between F2F and remote cognitive testing are consistent with those of prior telehealth studies. Remote MMSE showed an excellent ICC of 0.905[9] and a high correlation (r = 0.90) with F2F administration.[13] However, Loh, et al.[13] also found wide 95% limits of agreement, ranging from -3.9 to 4.5, a finding also consistent with ours. Furthermore, we observed higher remote AMT and mCMMSE scores than that of those administered F2F. The possible explanations for this discrepancy include practice effects, which cannot be eliminated entirely. To mitigate this, we chose an interval of 2–3 weeks between F2F and remote assessments, as reported previously.[14] This time interval sought to balance the possibility of practice effects if the second visit was scheduled too close to the first and to avoid longitudinal changes in test scores if repeat tests were spaced too far apart.[14,15] In support of this time interval, a previous study demonstrated the stability of the MMSE for up to 6 weeks.[16] However, future studies may counterbalance the order of F2F and remote assessments to minimize practice effects.

The higher scores observed during the remote assessment may also be attributed to the cues or prompts provided by the caregivers in our study. To preempt this, we conducted briefings before the remote assessments to ensure a quiet and distraction-free environment for videoconferencing. Nonetheless, for individuals with the largest discrepancies between the remote and F2F assessments, we observed that caregivers frequently prompted participants outside the camera field of view. In addition, the presence of environmental cues (clocks and calendars) may be another plausible reason for the higher remote testing scores, as reflected in the significantly better performance in the orientation domain when admin-

Table 4. Differences in face-to-face versus remote mCMMSE and CFAB scores by domain

Domain	Face-to-face	Videoconference	Mean difference	p-value
mCMMSE				
Orientation	5.5 ± 2.2	6.3 ± 1.7	0.8 ± 1.5	< 0.001
Registration	3.0 ± 0.1	3.0 ± 0	0.2 ± 0.1	0.320
Attention	2.9 ± 1.8	3.1 ± 1.7	0.3 ± 1.3	0.130
Recall	1.1 ± 1.1	1.6 ± 1.2	0.6 ± 1.0	< 0.001
Language	7.0 ± 1.2	6.8 ± 1.2	-0.2 ± 1.2	0.200
Visuospatial	0.7 ± 0.5	0.5 ± 0.5	-0.1 ± 0.6	0.110
CFAB				
Similarities	1.7 ± 0.9	2.2 ± 1.0	0.5 ± 0.9	0.006
Category fluency	2.3 ± 0.6	2.2 ± 0.7	-0.2 ± 0.6	0.170
Motor series	2.6 ± 0.9	2.5 ± 0.8	-0.1 ± 1.0	0.480
Conflicting instructions	2.3 ± 1.0	2.6 ± 0.9	0.3 ± 0.7	0.026
Go-no-go	1.9 ± 1.0	1.8 ± 1.0	-0.1 ± 0.8	0.650

Values are presented as mean±standard deviation.

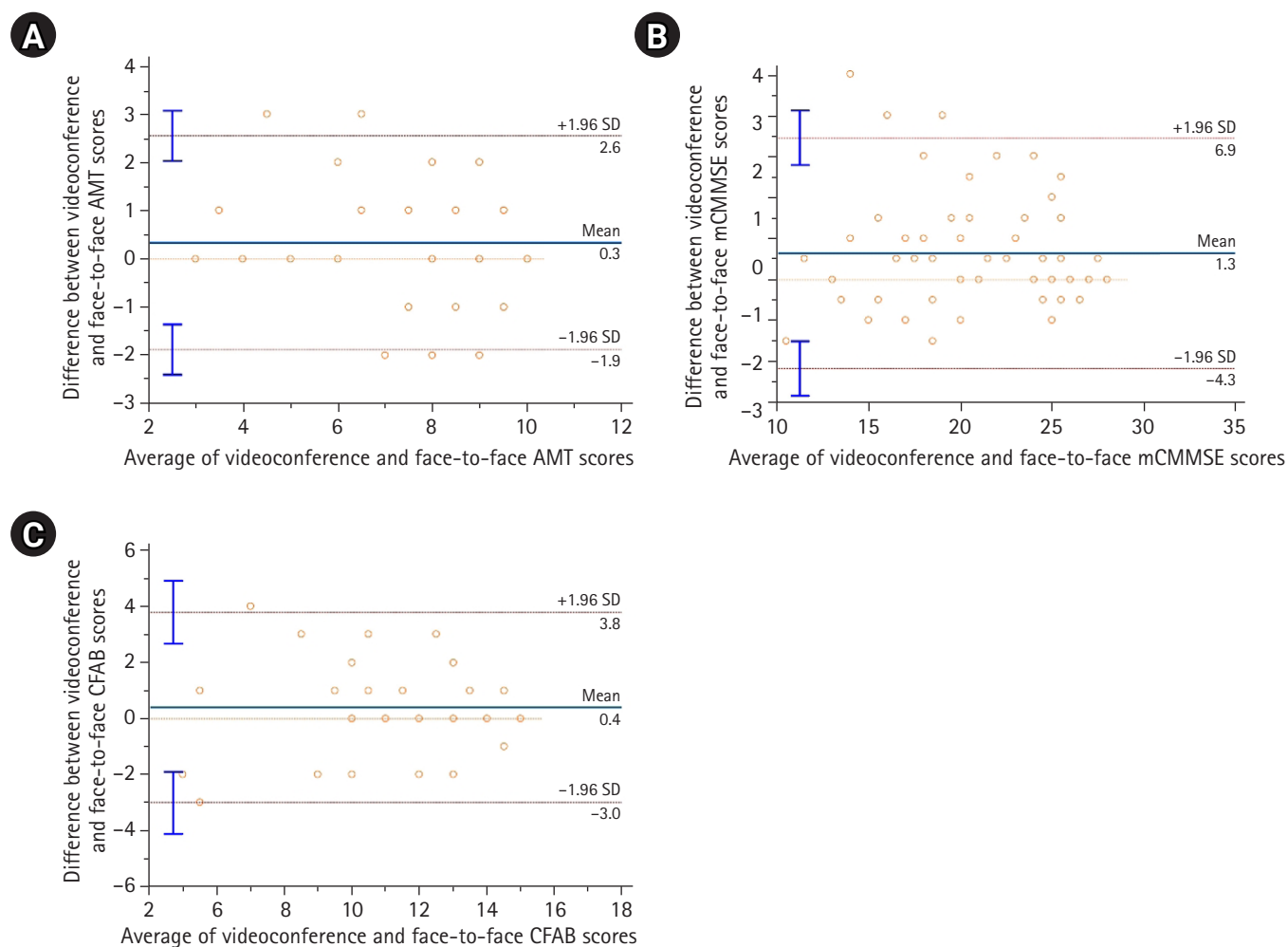mCMMSE, modified version of the Chinese Mini-Mental State Examination; CFAB, Chinese Frontal Assessment Battery.

Fig. 1. Bland-Altman plots for (A) the Abbreviated Mental Test (AMT), (B) modified version of the Chinese Mini-Mental State Exam (mC-MMSE), and (C) Chinese Frontal Assessment Battery (CFAB).

istered remotely. These findings underscore the need for an optimal environment for valid telehealth assessment.

The results of our study highlight the importance of employing various measures of reliability and agreement for the comprehensive evaluation of validity. While many studies have reported good correlations between remote and F2F assessments, high correlations are not synonymous with a good agreement and may fail to detect systematic bias, as observed in our study. Moreover, interpreting the ICC remains challenging owing to its inherent characteristics, which are largely determined by the heterogeneity of the sample such that when variance is high, the ICC is likely to be high, and vice versa.[17] In our study, the wide range of mCMMSE scores may in part explain the high ICC estimates but do not necessarily reflect reliability and agreement between remote and F2F assessments. In contrast, the Bland-Altman plots provided a visual assessment of bias and agreement, enabling the analysis of individual data points and identifying outliers with large degrees of disagreement. Identifying outliers also allowed for further analysis to elucidate the reasons for the large discrepancies between remote and F2F assessments.

The present study evaluated the CFAB adapted for telemedicine, which incorporates motor tasks, including finger tapping and copying a series of hand movements. While the 95% limits of agreement for CFAB exceeded the a priori defined levels, our results still indicated that cognitive tests with motor components might be feasibly completed in a telehealth setting. To adapt the CFAB for remote administration, we omitted the final "environmental autonomy" item, which involved placing the examiner's hands out and instructing the patient not to touch them, and observing for abnormal behavior such as imitation, utilization, and prehension behavior. Omitting this item is unlikely to significantly affect the validity of the CFAB, as demonstrated in a study revealing its limited utility in early cognitive impairment due to a ceiling effect present from normal to early dementia.[8]

The strengths of this study include its use of various measures of reliability and agreement in a single study to evaluate validity with a sample size adequately powered for the primary objective. We also used consistent raters between subjects and standardized the testing procedures before commencing the study to minimize variability. We did not require the use of additional equipment beyond the participants' smartphones. However, the limitations of our study include the lack of data on hearing and visual impairments and their impact on our results. Factors such as mood or behavior that may have influenced remote cognitive testing were not assessed in this study. Furthermore, our results are not generalizable to older adults from community-based populations with normal cognition or at moderate to advanced stages of dementia. More-

over, as our sample included individuals with access to smartphone devices, stable network connectivity, or caregivers who were available to assist (38 of the 56 participants needed caregivers), we were also unable to generalize our results to older adults across the spectrum of socioeconomic status and familiarity with technology.

Our study adds to the body of evidence evaluating the validity of telemedicine-based cognitive assessment, particularly in older adults with cognitive impairment. We also provide results from a "real-world" implementation of telemedicine in a clinical setting during the COVID-19 pandemic. Given the potential clinical and medicolegal ramifications of cognitive testing results, our results suggest that providers should cautiously adopt telemedicine-based cognitive assessments, with careful attention paid to ensure a conducive environment in which remote testing can occur. Nevertheless, during a pandemic that has disproportionately affected older adults, telemedicine serves an important need to maintain continuity of care in settings that face disruption of essential medical services. Further studies are needed to establish the validity of telemedicine for dementia diagnosis and treatment in a larger sample, evaluate the acceptance of telehealth in older adults, and increase access to telehealth services.

REFERENCES

1. Khairat S, Meng C, Xu Y, Edson B, Gianforcaro R. Interpreting

COVID-19 and virtual care trends: cohort study. JMIR Public Health Surveill 2020;6(2):e18811.

2. Cuffaro L, Di Lorenzo F, Bonavita S, Tedeschi G, Leocani L, Lavorgna L. Dementia care and COVID-19 pandemic: a necessary digital revolution. Neurol Sci 2020;41:1977-9.

3. Costanzo MC, Arcidiacono C, Rodolico A, Panebianco M, Aguglia E, Signorelli MS. Diagnostic and interventional implications of telemedicine in Alzheimer's disease and mild cognitive impairment: a literature review. Int J Geriatr Psychiatry 2020;35:12-28.

4. Ciemins EL, Holloway B, Coon PJ, McClosky-Armstrong T, Min SJ. Telemedicine and the mini-mental state examination: assessment from a distance. Telemed J E Health 2009;15:476-8.

5. Chapman JE, Cadilhac DA, Gardner B, Ponsford J, Bhalla R, Stolwyk RJ. Comparing face-to-face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke. J Telemed Telecare 2021;27:484-92.

6. Koh ZY, Law F, Chew J, Ali N, Lim WS. Impact of coronavirus disease on persons with dementia and their caregivers: an audit study. Ann Geriatr Med Res 2020;24:316-20.

7. Lim WS, Chin JJ, Lam CK, Lim PP, Sahadevan S. Clinical dementia rating: experience of a multi-racial Asian population. Alzheimer Dis Assoc Disord 2005;19:135-42.

8. Goh WY, Chan D, Ali NB, Chew AP, Chuo A, Chan M, et al. Frontal assessment battery in early cognitive impairment: psychometric property and factor structure. J Nutr Health Aging 2019;23:966-72.

9. Munro Cullum C, Hynan LS, Grosch M, Parikh M, Weiner MF. Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. J Int Neuropsychol Soc 2014;20:1028-33.

10. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63.

11. Burleigh E, Reeves I, McAlpine C, Davie J. Can doctors predict patients' abbreviated mental test scores. Age Ageing 2002;31:303-6.

12. Andrews JS, Desai U, Kirson NY, Zichlin ML, Ball DE, Matthews BR. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. Alzheimers Dement (N Y) 2019;5:354-63.

13. Loh PK, Ramesh P, Maher S, Saligari J, Flicker L, Goldswain P. Can patients with dementia be assessed at a distance? The use of Telehealth and standardized assessments. Intern Med J 2004;34:239-42.

14. Carotenuto A, Rea R, Traini E, Ricci G, Fasanaro AM, Amenta F. Cognitive assessment of patients with Alzheimer's disease by telemedicine: pilot study. JMIR Ment Health 2018;5:e31.

15. Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. J Clin Exp Neuropsychol 1998;20:339-52.

16. Thal LJ, Grundman M, Golden R. Alzheimer's disease: a correlational analysis of the Blessed Information-Memory-Concentration Test and the Mini-Mental State Exam. Neurology 1986;36:262-4.

17. Ten Cate DF, Luime JJ, Hazes JM, Jacobs JW, Landewe R. Does the intraclass correlation coefficient always reliably express reliability? Comment on the article by Cheung et al. Arthritis Care Res (Hoboken) 2010;62:1357-8.