DOI: 10.1111/jgs.17190

CLINICAL INVESTIGATION

Journal of the American Geriatrics Society

Diagnostic accuracy of virtual cognitive assessment and testing: Systematic review and meta-analysis

Jennifer A. Watt MD PhD[1,2,3] | Natasha E. Lane MD PhD[3,4] | Areti Angeliki Veroniki PhD[1,5,6] | Manav V. Vyas MBBS MSc[7] | Chantal Williams MSc[1] | Naveeta Ramkissoon MPH[1] | Yuan Thompson PhD[1] | Andrea C. Tricco PhD[1,8] | Sharon E. Straus MD MSc[1,2,8] | Zahra Goodarzi MD MSc[9,10,11]

[1]Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital-Unity Health Toronto, Toronto, Ontario, Canada

[2]Division of Geriatric Medicine, Department of Medicine, University of Toronto, Toronto, Ontario, Canada

[3]ICES, G1 06, Toronto, Ontario, Canada

[4]Department of Medicine, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

[5]Department of Primary Education, School of Education, University of Ioannina, Ioannina, Greece

[6]Institute of Reproductive and Developmental Biology, Department of Surgery & Cancer, Faculty of Medicine, Imperial College, London, UK

[7]Division of Neurology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada

[8]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[9]Department of Medicine, University of Calgary, Foothills Medical Centre – North Tower, Calgary, Alberta, Canada

[10]Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada

[11]O'Brien Institute of Public Health, University of Calgary, Calgary, Alberta, Canada

Correspondence
Jennifer A. Watt, Li Ka Shing Knowledge Institute, St. Michael's Hospital-Unity Health Toronto 209 Victoria Street, East

Abstract

Background/Objectives: Virtual (i.e., telephone or videoconference) care was broadly implemented because of the COVID-19 pandemic. Our objectives were to compare the diagnostic accuracy of virtual to in-person cognitive assessments and tests and barriers to virtual cognitive assessment implementation.

Design: Systematic review and meta-analysis.

Setting: MEDLINE, EMBASE, CDSR, CENTRAL, PsycINFO, and gray literature (inception to April 1, 2020).

Participants and interventions: Studies describing the accuracy or reliability of virtual compared with in-person cognitive assessments (i.e., reference standard) for diagnosing dementia or mild cognitive impairment (MCI), identifying virtual cognitive test cutoffs suggestive of dementia or MCI, or describing correlations between virtual and in-person cognitive test scores in adults.

Measurements: Reviewer pairs independently conducted study screening, data abstraction, and risk of bias appraisal.

Results: Our systematic review included 121 studies (15,832 patients). Two studies demonstrated that virtual cognitive assessments could diagnose dementia with good reliability compared with in-person cognitive assessments: weighted kappa 0.51 (95% confidence interval [CI] 0.41–0.62) and 0.63 (95% CI 0.4–0.9), respectively. Videoconference-based cognitive assessments were 100% sensitive and specific for diagnosing dementia compared with in-person cognitive assessments in a third study. No studies compared telephone with in-person cognitive assessment accuracy. The Telephone Interview for Cognitive Status (TICS; maximum score 41) and modified TICS (maximum score 50) were the only virtual cognitive tests compared with in-person cognitive assessments in >2 studies with extractable data for meta-analysis. The optimal TICS cutoff suggestive of dementia ranged from 22 to 33, but it was 28 or 30 when testing was conducted in English (10 studies; 1673 patients). Optimal modified TICS cutoffs suggestive of MCI ranged from 28 to 31 (3 studies; 525 patients).

Building, Room 723, Toronto, Ontario, Canada, M5B 1W8.
Email: jennifer.watt@utoronto.ca

Funding information
St. Michael's Hospital Foundation

Sensory impairment was the most often voiced condition affecting assessment. **Conclusion:** Although there is substantial evidence supporting virtual cognitive assessment and testing, we identified critical gaps in diagnostic certainty.

KEYWORDS
diagnostic test accuracy, meta-analysis, systematic review, telemedicine, virtual care

INTRODUCTION

In an era of physical distancing measures to prevent the spread of COVID-19, we are encouraged by health authorities to facilitate virtual medical encounters (i.e., telephone or videoconference) for older adults. As physical distancing restrictions relax, some patients may return to their doctor's office for in-person assessments. Others will receive ongoing care virtually: many believe the popularity of virtual medical care will continue.[1] Cognitive assessments (e.g., based on Diagnostic and Statistical Manual of Mental Disorders [DSM] criteria for dementia and mild cognitive impairment [MCI]), which consist of identifying cognitive domain deficits (e.g., delayed recall, visuospatial skills) and associated functional impairments, have always been a critical component of in-person assessments for older adults. As part of a cognitive assessment, cognitive testing (e.g., Montreal Cognitive Assessment [MoCA], Telephone Interview for Cognitive Status [TICS]) can support a diagnosis of MCI or dementia through standardized identification of cognitive domain deficits and enable clinicians to develop holistic treatment plans that help older adults to complete basic and instrumental activities of daily living.[2–5] Now, cognitive assessments and cognitive testing will continue to play an important role in virtual assessments, but our approach must be adapted to the virtual environment.

The recent rapid adoption of virtual cognitive assessments into clinical practice has created an immediate need to better understand the accuracy of virtual cognitive assessments and testing. If we are to continue providing evidence-informed care for older adults in the COVID-19 era, clinicians caring for older adults must become skilled at conducting virtual cognitive assessments and testing. We completed a systematic review of diagnostic accuracy studies to fill this critical knowledge gap. In our systematic review, we describe the accuracy and reliability of virtual compared with in-person cognitive assessments for diagnosing dementia or MCI, virtual cognitive test cutoffs suggestive of dementia or MCI compared with an in-person cognitive assessment, and correlations between virtual and in-person cognitive test scores. To enable rapid implementation of virtual cognitive assessment and testing into routine clinical practice, we synthesized barriers and facilitators to their use as described by study authors of included studies.

Key Points

- The recent rapid adoption of virtual (i.e., telephone or videoconference) cognitive assessment into clinical practice has created an urgent need to better understand its accuracy.
- We conducted a systematic review and meta-analysis (121 studies [15,832 patients]) to describe the accuracy and reliability of virtual compared with in-person cognitive assessments for diagnosing dementia or mild cognitive impairment (MCI), identify virtual cognitive test cutoffs suggestive of dementia or MCI compared with an in-person cognitive assessment, and highlight barriers and facilitators to implementing virtual cognitive assessments and testing.
- Although three studies comparing videoconference to in-person cognitive assessments (based on Diagnostic and Statistical Manual of Mental Disorders [DSM] criteria) demonstrated good reliability and accuracy of virtual cognitive assessments in diagnosing dementia, we did not identify any studies comparing the accuracy of telephone to in-person cognitive assessments—this represents a critical knowledge gap.

Why Does this Paper Matter?

We highlight diagnostic uncertainty associated with virtual cognitive assessment and testing, an absence of studies comparing the accuracy of telephone to in-person cognitive assessment, and key barriers and facilitators to implementing virtual cognitive assessments and tests in clinical practice.

METHODS

We published our protocol on Open Science Framework and registered it with PROSPERO (CRD42020186290). Our systematic review was written in accordance with the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies.[6]

Data sources and searches

We searched MEDLINE, EMBASE, CDSR, CENTRAL, and PsycINFO for citations published in any language from inception until April 1, 2020 (see File S1 for MEDLINE search strategy). We also searched the gray literature (File S2) and reviewed reference lists of included studies and related systematic reviews.

Study selection

We included studies assessing the accuracy or reliability of virtual compared with in-person cognitive assessments for diagnosing dementia or MCI, identifying virtual cognitive test (e.g., TICS, MoCA) cutoffs suggestive of dementia or MCI compared with an in-person cognitive assessment, or describing the correlation between virtual and in-person cognitive test scores in adults.[4,7] Our population of interest included adults (1) without cognitive impairment, (2) diagnosed with MCI, or (3) diagnosed with dementia as per established criteria (e.g., Petersen, Diagnostic and Statistical Manual of Mental Disorders [DSM], National Institute on Aging-Alzheimer's Association).[3,8–10] After reviewer pairs (ZG, NL, and JW) reached at least 80% agreement in a pilot screening exercise, they independently screened (1) all citations and (2) full-text articles (of citations retained from step #1) to establish inclusion eligibility. Discrepancies regarding study inclusion were resolved by deliberation within reviewer pairs.

Data abstraction and quality assessment

Pairs of reviewers (ZG, NL, NR, YT, MV, JW, and CW) independently abstracted data from each included full-text article and appraised each article's quality with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[11] We rated studies at high risk of bias from study flow and timing if virtual and in-person cognitive tests were conducted >6 months apart. Where reported by study authors, we abstracted study and study-level participant characteristics from each study (i.e., authorship, year of publication, degree of cognitive impairment [i.e., no cognitive impairment, MCI, dementia], average age of study population, and proportion of women in each study population); details describing how cognitive assessments and tests were implemented (i.e., method of virtual cognitive testing delivery [e.g., telephone, videoconference], language of cognitive testing, barriers to implementing virtual cognitive testing); and outcome data comparing test scores from cognitive assessments and tests conducted virtually compared with in-person (i.e., correlation coefficients, sensitivity, specificity, true positives [TP], true negatives [TN], false positives [FP], false negatives [FN], and area under the receiver operating characteristics [AUC] curve). Discrepancies regarding data abstraction and quality assessment were resolved by deliberation within reviewer pairs or by a third reviewer.

Quantitative data synthesis

We derived summary receiver operating characteristic curves (SROC), optimal cutpoint thresholds (maximum value of the weighted sum of sensitivity and specificity), and AUC values at optimal cutpoint thresholds. We pooled TP, TN, FP, and FN values at cognitive test cutoffs using restricted maximum likelihood estimation from inverse variance weighted linear mixed effects models.[12] These meta-analytic models can incorporate TP, TN, FP, and FN values from multiple cutpoints in each included study and estimate the optimal cutpoint threshold by incorporating all data points from included studies into a single meta-analytic model.[12] SROC curves were presented with confidence regions.[12] We derived meta-analytic effect estimates from studies comparing a virtual cognitive test to an in-person reference standard, which we defined as assessments conducted by clinicians using established diagnostic criteria (e.g., DSM or Petersen criteria).[3,8] We presented positive (LR+) and negative (LR−) likelihood ratios to facilitate interpretation of meta-analytic estimates.[13] The TICS is administered by telephone; thus, we did not conduct a subgroup analysis based on modality of administration. We did a post hoc subgroup analysis of studies where the mean years of formal education completed was >8, which roughly corresponds to at least an elementary school level of education. We performed a sensitivity analysis by excluding studies at high risk of bias from conduct or interpretation of the reference standard.[11] We were unable to complete a sensitivity analysis by excluding studies at

high risk of bias from patient flow through the study and timing of reference and index tests because there were too few remaining studies. Analyses were performed in *R*, version 4.0.0.[12]

Qualitative data synthesis

We conducted a qualitative meta-synthesis to understand potential barriers and facilitators associated with implementing virtual cognitive assessments and testing.[14,15] Two reviewers (JAW and ZG) coded barriers and facilitators inductively (independently and in duplicate) and categorized them into themes. JAW and ZG categorized codes by test or assessment timing (i.e., before, immediately preceding, or during) and deliberated discrepancies within the reviewer pair.

RESULTS

Characteristics of included studies

We screened 9200 article titles and abstracts and 417 full-text articles, which resulted in 121 included studies (15,832 participants) (Figure 1). We identified 84 virtual cognitive tests (some of which have different versions or contain subtests). Most studies implemented virtual

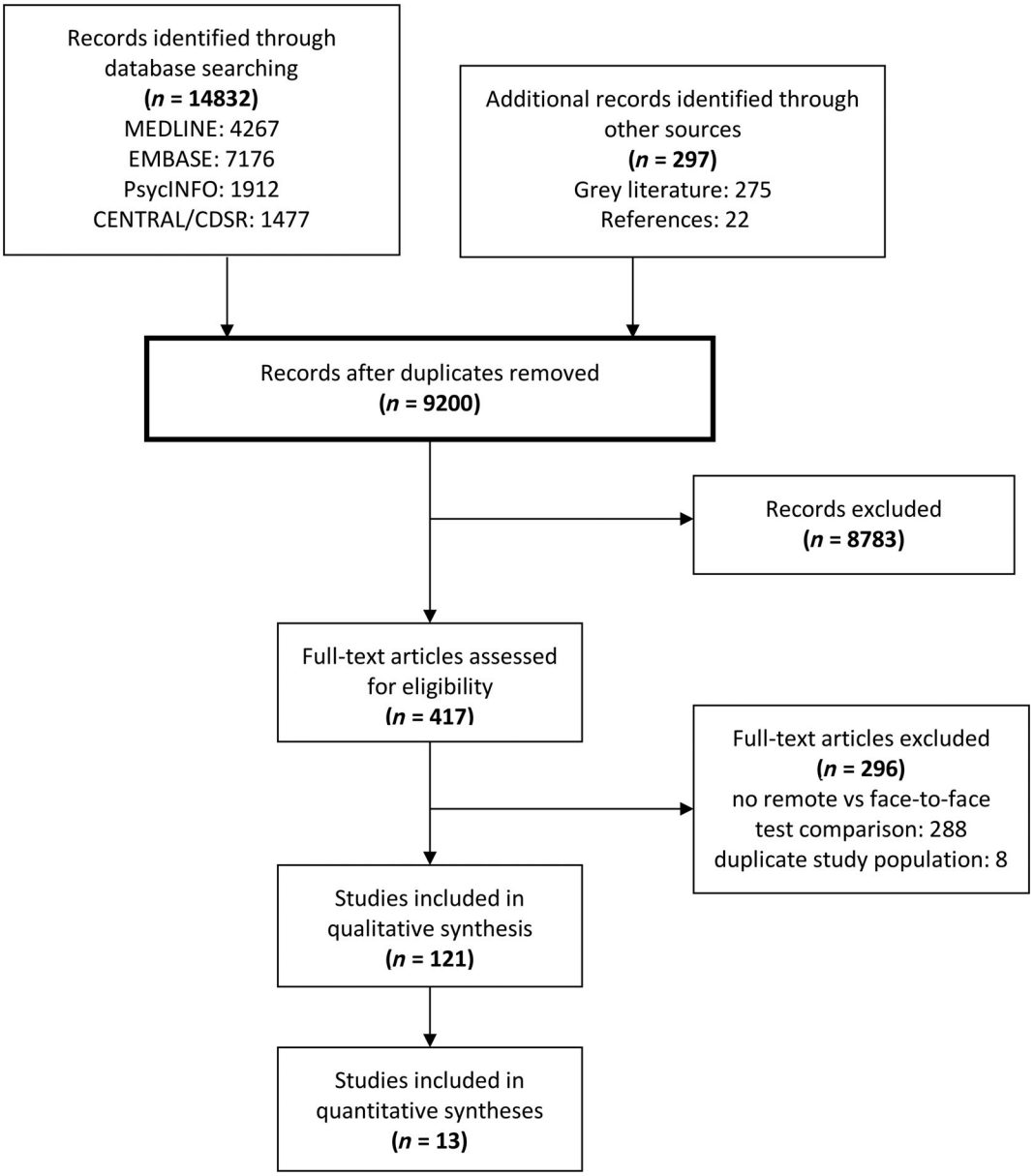

FIGURE 1 PRISMA flow diagram

TABLE 1 Characteristics of included studies

Study characteristic	Number of studies (%)
Mean age	
≤60	16 (14.0)
61–69	23 (19.8)
70–79	62 (51.2)
≥80	11 (9.1)
Not reported	7 (0.6)
Proportion females	
≥50%	69 (57.0)
<50%	38 (31.4)
Not reported	14 (11.6)
Degree of cognitive impairment	
No cognitive impairment	12 (9.9)
Mixed (no cognitive impairment, MCI, dementia)	73 (60.3)
MCI/dementia	2 (1.7)
MCI	1 (0.8)
Dementia	6 (5)
Other	5 (4.1)
Not reported	22 (18.2)
Most commonly reported virtual cognitive tests	
TICS or modified TICS	41 (33.9)
MMSE or modified MMSE	22 (18.2)
MoCA or modified MoCA	9 (7.4)
Language of cognitive assessment or cognitive testing	
English	28 (23.1)
Other[a]	36 (30.6)
Not reported	56 (46.3)
Modality of virtual cognitive testing	
Telephone	85 (69.4)
Videoconference	37 (30.6)

Abbreviations: MCI, mild cognitive impairment; MMSE, Mini Mental State Examination; MoCA, Montreal Cognitive Assessment; TICS, telephone interview for cognitive status.

[a]Other: Chinese (number [n] = 5), Dutch (n = 3), English or Spanish (n = 2), French (n = 2), Finnish (n = 1), German (n = 5), Hebrew (n = 1), Italian (n = 4), Japanese (n = 3), Korean (n = 2), Portuguese (n = 4), Spanish (n = 4), Turkish (n = 1).

cognitive testing via telephone (69.4%) (Table 1 and Table S1). The TICS or modified TICS were the most utilized virtual cognitive tests (reported in 33.9% of studies) (Table 1). Virtual cognitive tests were studied in at least 13 languages (Table 1). Sixty-nine (57.0%) of the studies reported correlations between virtual and in-person cognitive test scores (Table S2). Risk of bias from patient flow through the study and timing of reference and index tests was high in 46.3% of studies and represented the greatest threat to study validity (Table S3 and Figure S1).

Accuracy and reliability of virtual compared with in-person cognitive assessment for diagnosing dementia

We found three studies describing the accuracy or reliability of videoconference compared with in-person cognitive assessments based on DSM-IV criteria.[16–18] In a prospective cohort study of 16 Washington State Veterans' Home residents who did not have a history of dementia, geriatric psychiatrists had 100% agreement (i.e., 100% sensitivity and specificity because the in-person assessment was the reference standard) between videoconference and in-person assessments in diagnosing 12 residents with dementia (type not specified).[16] However, this study was conducted in a highly selected patient population: residents had a high probability of having dementia based on the 7-Minute Screen.[16] Further, all residents had a MMSE completed prior to in-person and videoconference cognitive assessments, and MMSE results were reviewed by the geriatric psychiatrists. In a prospective study of older adults (aged 50 years and older) with undiagnosed cognitive problems who were referred to a memory disorder clinic, videoconference-based cognitive assessments demonstrated good reliability compared with in-person cognitive assessments when completed by specialist physicians experienced in diagnosing dementia (weighted kappa 0.51, 95% CI 0.41–0.62).[17] Further, in a separate study comparing reliability between videoconference and in-person assessments for diagnosing dementia (type not specified) in a cohort of older adults (aged 50 years and older) referred to a memory disorder clinic, Martin-Khan et al., reported a weighted kappa of 0.63 (95% CI 0.4–0.9) between videoconference and in-person assessments conducted by a geriatrician and psycho-geriatrician.[18] Both geriatricians were provided with cognitive testing results before their assessments, which were completed by a clinic nurse. However, it was unclear if these two geriatricians were blinded to each other's findings.[18] We did not find any studies comparing the diagnostic accuracy of a telephone compared with in-person cognitive assessment based on DSM criteria.

Virtual cognitive test cut-offs suggestive of dementia or mild cognitive impairment and correlations between virtual and in-person cognitive test scores

We describe the three most reported virtual cognitive tests in our systematic review: the TICS, MMSE, and

MoCA (Table 1). Only the TICS (maximum score of 41) and modified TICS (maximum score of 50) were compared with a reference standard in >2 studies and had extractable effect estimates that could be meta-analyzed (study-specific estimates of diagnostic accuracy for all cognitive tests are found in Table S2).

Telephone interview for cognitive status (TICS)

We included 10 studies (1673 participants) in our meta-analysis to derive the optimal TICS (maximum score of 41) cutoff value supportive of a diagnosis of dementia (Table 2).[5,19–27] The optimal TICS cutoff suggestive of dementia ranged from 22 to 33, but it was 28 or 30 when the TICS was conducted in English. At the optimal cutoff value of 26, the pooled sensitivity was 0.80 (95% CI 0.61–0.92), the pooled specificity was 0.90 (95% CI 0.72–0.98), and the AUC was 0.92 (Figure 2(A)). In a subgroup of studies where cognitive testing was conducted in English, the optimal cutoff value was 29; the pooled sensitivity at this cutoff value was 0.72 (95% CI 0.40–0.92), the pooled specificity was 0.85 (95% CI 0.62–0.96), and AUC was 0.86. In a subgroup of studies where the mean years of formal education was >8, the optimal cutoff value was 27; the pooled sensitivity at this cutoff value was 0.83 (95% CI 0.58–0.96), the pooled specificity at this cutoff value was 0.92 (95% CI 0.67–0.99), and AUC was 0.94. In a sensitivity analysis where we removed studies at high risk of bias from the conduct or interpretation of the reference standard, the optimal cutoff value was 25; the pooled sensitivity at this cutoff value was 0.78 (95% CI 0.59–0.91), the pooled specificity at this cutoff value was 0.92 (95% CI 0.75–0.98), and AUC was 0.92. Two studies reported cutoffs for the TICS supportive of a diagnosis of MCI: Manly et al. determined the optimal cutoff was 29 and Seo et al. determined the optimal cutoff was 28.[23,25]

TABLE 2 Diagnostic accuracy of TICS and modified TICS

Cutoff	Pooled sensitivity at cutoff (95% CI)	Pooled specificity at cutoff (95% CI)	Positive likelihood ratio	Negative likelihood ratio
DEMENTIA				
TICS (maximum score of 41): 10 studies (1673 participants)[5,19–27]				
25	0.76 (0.56 to 0.90)	0.93 (0.79 to 0.99)	10.9	0.3
26*	0.80 (0.61 to 0.92)	0.90 (0.72 to 0.98)	8	0.2
27	0.84 (0.67 to 0.94)	0.85 (0.64 to 0.96)	5.6	0.2
28	0.87 (0.72 to 0.96)	0.80 (0.55 to 0.94)	4.4	0.2
29	0.90 (0.76 to 0.97)	0.72 (0.46 to 0.90)	3.2	0.1
30	0.92 (0.80 to 0.98)	0.64 (0.37 to 0.86)	2.6	0.1
Modified TICS (maximum score of 50): 3 studies (435 participants)[25,28,29]				
25	0.80 (0.53 to 0.95)	0.93 (0.75 to 0.99)	11.4	0.2
26*	0.85 (0.61 to 0.96)	0.89 (0.68 to 0.98)	7.7	0.2
27	0.89 (0.68 to 0.98)	0.85 (0.61 to 0.96)	5.9	0.1
28	0.92 (0.74 to 0.99)	0.79 (0.53 to 0.94)	4.4	0.1
29	0.94 (0.79 to 0.99)	0.73 (0.44 to 0.92)	3.5	0.1
30	0.96 (0.84 to 0.99)	0.66 (0.36 to 0.88)	2.8	0.1
MILD COGNITIVE IMPAIRMENT				
Modified TICS (maximum score of 50): 3 studies (525 participants)[25,28,30]				
25	0.26 (0.12 to 0.46)	0.95 (0.88 to 0.99)	5.2	0.8
26	0.36 (0.18 to 0.58)	0.93 (0.82 to 0.98)	5.1	0.7
27	0.47 (0.26 to 0.69)	0.89 (0.74 to 0.96)	4.3	0.6
28	0.58 (0.35 to 0.78)	0.83 (0.66 to 0.94)	3.4	0.5
29	0.68 (0.48 to 0.86)	0.77 (0.56 to 0.90)	3	0.4
30*	0.77 (0.55 to 0.91)	0.68 (0.45 to 0.86)	2.4	0.3

Abbreviations: CI, confidence interval; *, optimal cutoff; TICS, telephone interview for cognitive status.

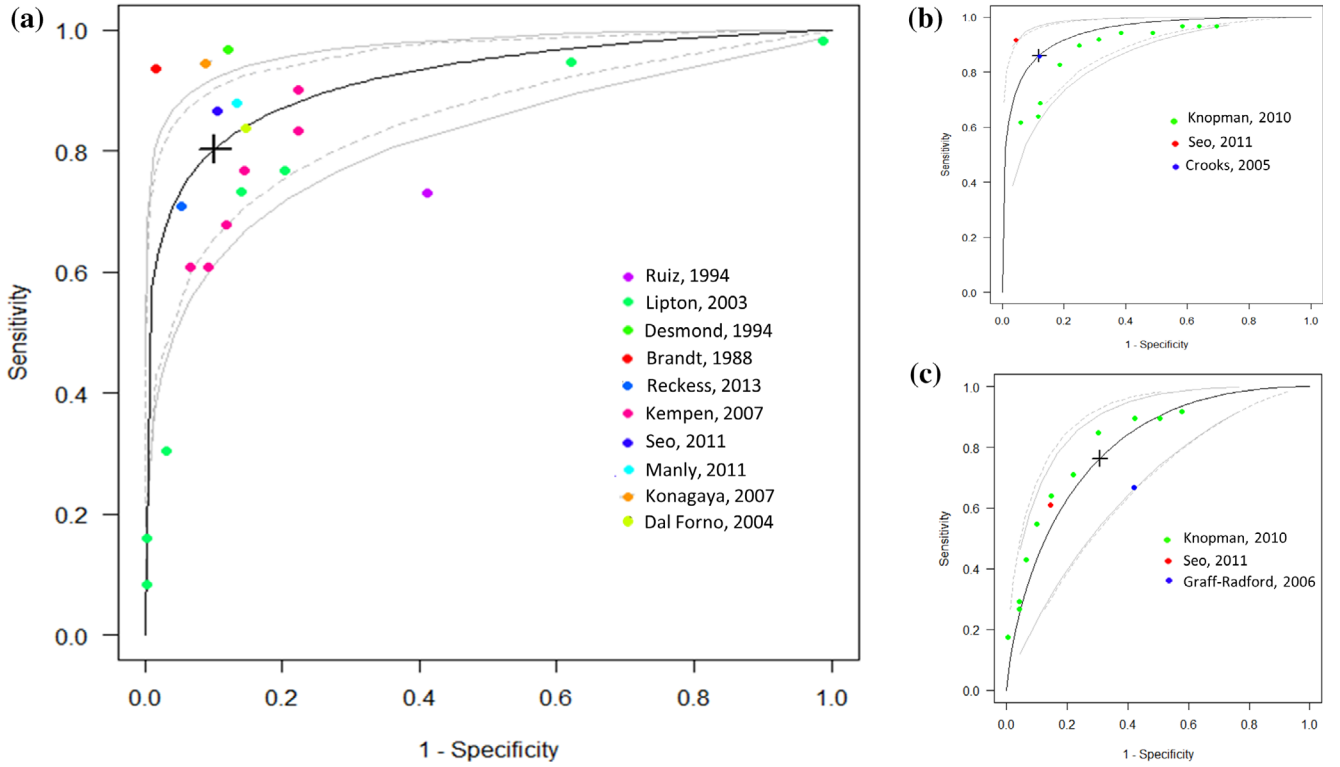

FIGURE 2 Summary receiver operating characteristic curves for telephone interview for cognitive status (TICS) and modified TICS. Panel A is the summary receiver operating characteristic curve (SROC) associated with the optimal cutoff threshold for the Telephone Interview for Cognitive Status (TICS) (maximum score 41) suggestive of dementia. Panel B is the SROC associated with the optimal cutoff threshold for the modified TICS (maximum score 50) suggestive of dementia. Panel C is the SROC associated with the optimal cutoff threshold for the modified TICS suggestive of mild cognitive impairment. SROC curves are indicated by the solid line. The optimal cutoff threshold is indicated by the '+'. Individual data points from each study are indicated by uniquely colored circles. The 95% confidence interval (CI) for sensitivity, given specificity, is indicated by the solid gray line. The 95% CI for specificity, given sensitivity, is indicated by the interrupted gray line

Modified telephone interview for cognitive status (TICS)

We included three studies (435 participants) in our meta-analysis to derive the optimal modified TICS (maximum score of 50) cutoff value supportive of a diagnosis of dementia (Table 2).[25,28,29] At the optimal cutoff value of 27, the pooled sensitivity was 0.85 (95% CI 0.61–0.96), the pooled specificity was 0.89 (95% CI 0.68–0.98), and the AUC was 0.94 (Figure 2 (B)). We included three studies (525 participants) in our meta-analysis to derive the optimal modified TICS cutoff value supportive of a diagnosis of MCI (Table 2).[25,28,30] At the optimal cutoff value of 30, the pooled sensitivity was 0.77 (95% CI 0.55–0.91), the pooled specificity was 0.68 (95% CI 0.45–0.86), and the AUC was 0.80 (Figure 2(C)).

Mini-mental state examination (MMSE)

Two studies reported sensitivity and specificity values, compared with a reference standard, for modified versions of the Mini-Mental State Examination (MMSE) supportive of a diagnosis of dementia (Table S2).[31,32] At a cutoff of 13 (maximum possible score 22), the sensitivity of the MMSE conducted on the telephone in Brazilian Portuguese for diagnosing dementia was 0.9 and specificity 1.[31] At a cutoff of 13 (maximum possible score 26), the sensitivity of the MMSE conducted on the telephone in Cantonese for diagnosing dementia was 1 and specificity was 0.67.[32] One study reported sensitivity and specificity values, compared with a reference standard, for a modified version of the MMSE supportive of with a diagnosis of MCI: at a cutoff of 89 (maximum possible score 100), the sensitivity of the Telephone Modified MMSE in German was 0.83 and specificity was 1.[33] Pearson correlations for telephone compared with in-person MMSE testing ranged from 0.54 to 0.88.[34–36] Pearson correlations for videoconference compared with in-person MMSE testing ranged from 0.89 to 0.92.[37–39]

Montreal-cognitive assessment (MoCA)

Two studies reported sensitivity and specificity values, compared with a reference standard, for modified

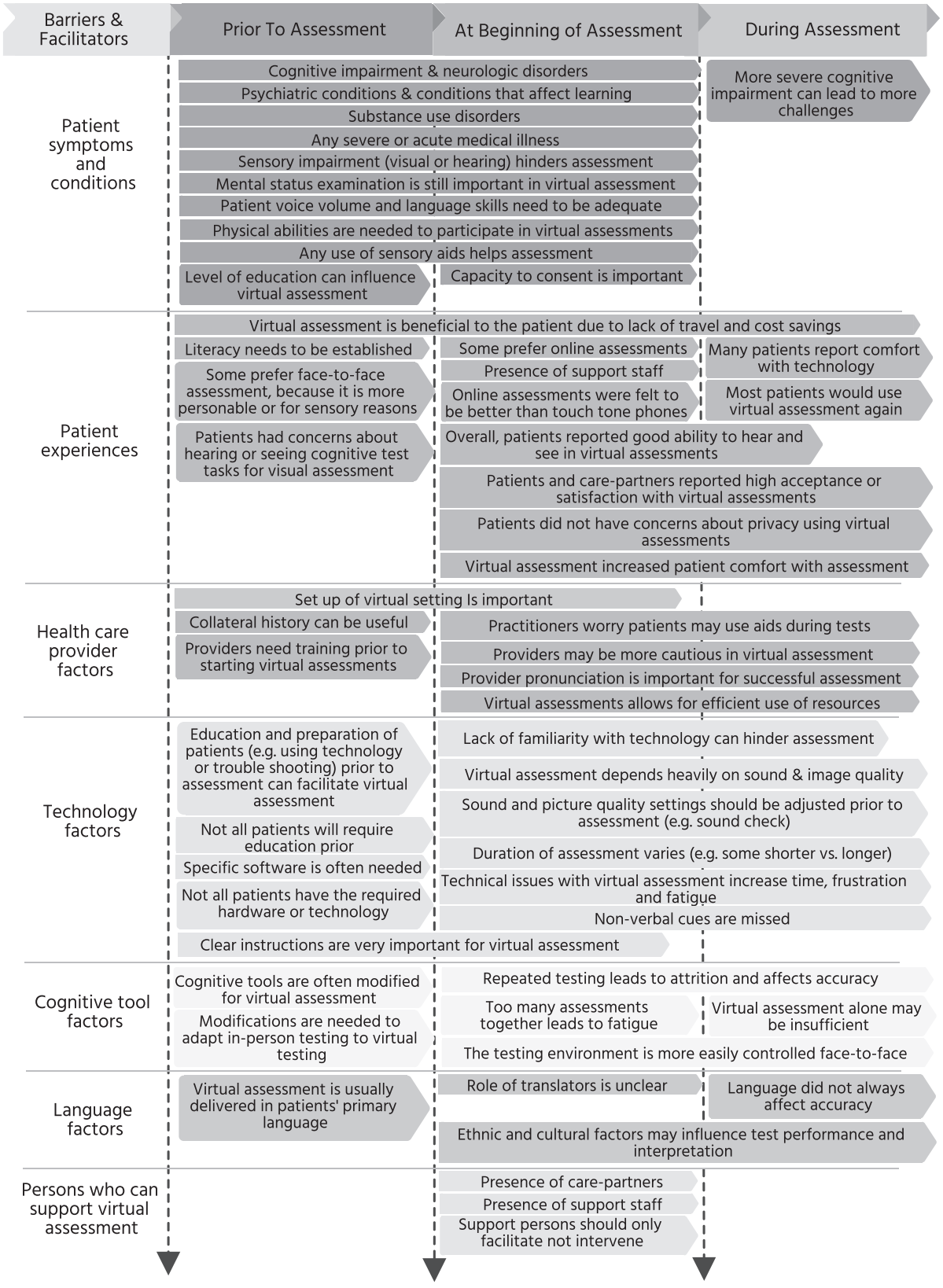

FIGURE 3 Synthesis of barriers and facilitators to virtual cognitive assessment and virtual cognitive testing

versions of the MoCA supportive of a diagnosis of MCI (Table S2).[40,41] At a cutoff of 7 (maximum possible score 12), the sensitivity of the short version of the telephone MoCA for diagnosing MCI was 0.41 and specificity was 0.88.[40] At a cutoff value of 20 (maximum possible score 22), the sensitivity of the MoCA conducted on the telephone in German was 0.9 and specificity was 0.54.[41] Intraclass correlation coefficients for videoconference compared with in-person MoCA testing ranged from 0.59 to 0.99.[42–45]

Barriers and facilitators to virtual cognitive assessment and testing

We derived 56 unique codes and seven major themes in our meta-synthesis of barriers and facilitators associated with virtual cognitive assessment and virtual cognitive testing (Figure 3 and File S3). We derived codes describing barriers and facilitators associated with access to (i.e., not all patients have the required hardware or technology, remote assessment is beneficial to the patient due to lack of travel and cost savings) and implementation of (i.e., clinician pronunciation is important for successful assessment, remote assessment depends heavily on sound and image quality) virtual cognitive assessment and cognitive testing (Figure 3). We identified patient symptoms or conditions affecting assessment, patient and care partner experiences, health care provider factors, cognitive test factors, technology factors, language factors, and presence of persons who could support virtual care as major themes. Sensory (i.e., visual or hearing) impairment was the most often voiced *patient symptom or condition affecting assessment* by patients and clinicians. Overall, patients reported a high degree of acceptance and comfort associated with virtual care. However, a lack of access to and familiarity with technology required for videoconference-based assessments caused frustration and fatigue for patients. To overcome this barrier, some studies had care partners or support staff setup necessary videoconferencing equipment.

DISCUSSION

Our systematic review is a comprehensive synthesis of studies comparing the diagnostic accuracy of videoconference to in-person cognitive assessments and virtual cognitive test cutoffs suggestive of dementia or MCI. Further, we qualitatively synthesized barriers and facilitators to virtual cognitive assessment. We found three studies demonstrating that virtual cognitive assessments have good accuracy compared with in-person cognitive assessments, but we did not identify any studies comparing the accuracy of telephone with in-person cognitive assessments (based on established criteria such as the DSM)—this is an important knowledge gap given that two-thirds of older adults are receiving virtual care via the telephone during the COVID-19 pandemic.[46] However, our systematic review and meta-analysis identified thresholds suggestive of cognitive impairment based on the TICS (the most studied telephone-based cognitive test in our systematic review). Cognitive tests such as the TICS that are coupled with appropriate functional ability inquiry could aid clinicians in completing telephone-based cognitive assessments, but the baseline prevalence of dementia or MCI (i.e., nursing home, memory clinic, or primary care setting) will impact the post-test probability of diagnosis. Although scores from other virtual cognitive tests (i.e., MMSE, MoCA) demonstrated moderate-to-high correlation with those conducted in-person, there was substantial variability across studies and relatively few studies reported cutoffs consistent with dementia or MCI. Identified barriers and facilitators (e.g., hearing impairment, presence of a caregiver to support technology use) to virtual cognitive assessment and testing or the length of time between in-person and virtual cognitive assessments and testing may explain some of this variability. Our results are timely and important given renewed and growing interest in conducting virtual cognitive assessments and testing: there is a large evidence base supporting virtual cognitive testing and assessments, but there are also important knowledge gaps to be filled.

Our systematic review highlights the diagnostic accuracy of a cognitive test that may be unfamiliar to clinicians—the TICS.[5] The TICS was adapted from and correlates highly with the MMSE (Pearson correlation 0.94, $p < 0.0001$).[5,7] The TICS is composed of 11 tasks designed to assess the cognitive domains of orientation, memory, attention/calculation, and language.[5] In the initial TICS validation study, Brandt et al., enrolled 100 patients with dementia (mean TICS 35.79 [SD 1.75], range 0–31) and 33 control patients (mean TICS 13.2 [SD 8.53], range 31–39).[5] They chose a cutoff of fewer than 31 points as supportive of cognitive impairment, which corresponded to a sensitivity of 94% and specificity of 100% for identifying dementia.[5] However, subsequent studies have enrolled patient populations where (1) the difference in mean TICS scores between those with and without dementia was smaller, (2) there was greater variability in TICS scores within groups of patients with and without dementia, and (3) the mean TICS scores in persons with dementia were lower than that in the study by Brandt et al.[21,23,25] Two subgroup analyses in our systematic review and meta-analysis suggest that language and

education may be important determinants of optimal TICS cutoffs supportive of a diagnosis of dementia: 27 was the optimal cutoff threshold where the mean years of formal education was >8 and 29 was the optimal cutoff threshold where cognitive testing was conducted in English. Greater diversity in study patient populations may at least partially explain the lower optimal cutoff threshold for the TICS identified in our meta-analysis.

Barriers and facilitators described by study authors illustrate key considerations for clinicians who are adapting their practices to incorporate virtual cognitive assessments and virtual cognitive testing. Clinicians and patients engaged in virtual cognitive testing experienced some barriers similar to those present in face-to-face cognitive assessments and testing (e.g., the impact of culture, education, and language on the conduct and interpretation of cognitive testing). Other barriers are unique to the virtual environment: lack of access to or familiarity with videoconferencing technology, loss of certain nonverbal cues that suggest potential cognitive impairment, and a lesser ability to prevent disruptions that might occur outside of the clinician's office. Barriers and facilitators associated with telemedicine use by patients and frontline staff have been previously described (e.g., involvement of support persons to facilitate assessment, internet or phone availability, and lack of training).[34,35] We have added to this evidence by highlighting unique considerations for conducting cognitive testing in persons who may have cognitive or sensory impairment.[47,48] Clinicians will need to tailor their approach to virtual cognitive testing to ensure patients' performance is not compromised by challenges imposed by the testing environment rather than their cognitive abilities.

Our systematic review and meta-analysis has several limitations. First, important considerations in understanding the validity and reliability of each cognitive test's diagnostic accuracy, including the training of persons administering tests or the baseline prevalence of cognitive impairment, were not always reported. Second, in many cases, the type of dementia or MCI being assessed was not specified or multiple types were assessed concurrently; we could not comment on the diagnostic accuracy of the TICS in persons with Alzheimer's disease or vascular dementia, specifically. Lastly, identified barriers and facilitators associated with virtual cognitive testing describe only those reported by study authors in their clinical setting, which may not generalize to the experiences of patients, caregivers, and clinicians across different care settings.

In conclusion, although there is evidence supporting virtual cognitive testing and assessments, important knowledge gaps related to telephone assessments, in particular, must be filled because many older adults have not been able to access videoconference assessments during the COVID-19 pandemic.[46,49] The TICS and modified TICS are more extensively validated than other virtual cognitive tests, but virtual cognitive tests familiar to clinicians (i.e., MMSE and MoCA) demonstrated moderate-to-high correlation with in-person test versions and should be studied further to better understand variability in study-specific estimates. Patients, caregivers, clinicians, researchers, and policy-makers must consider both the diagnostic accuracy of virtual cognitive assessments and barriers faced by patients, caregivers, and clinicians in accessing and using these assessments to continue supporting patient needs during the COVID-19 pandemic and beyond.

ACKNOWLEDGMENTS

This study was funded by the St. Michael's Hospital Foundation.

AUTHOR CONTRIBUTIONS

JAW, ZG, AAV, and SES contributed to the design of this study. JAW, ZG, MV, CW, NR, NEL, and YT contributed to article screening and data abstraction. JAW conducted data analyses, and all authors had access to data. JAW drafted the first version of the manuscript and all authors contributed to its revision and interpretation of findings.

CONFLICT OF INTEREST

ACT is funded by a Tier 2 Canada Research Chair in Knowledge Synthesis. SES is funded by a Tier 1 Canada Research Chair in Knowledge Translation. AAV is funded by a European Union's Horizon 2020 grant [No. 754936]. All authors have no conflict of interest to declare.

SPONSOR'S ROLE

This study was funded by the St. Michael's Hospital Foundation, which had no role in the conduct and reporting of this study or decision to submit this manuscript for publication.

ORCID

Jennifer A. Watt https://orcid.org/0000-0002-5296-6013
Natasha E. Lane https://orcid.org/0000-0003-1400-9068
Manav V. Vyas https://orcid.org/0000-0002-2953-2540

TWITTER

Jennifer A. Watt @JennAnnWatt

REFERENCES

1. Weigel G, Ramaswamy A, Sobel L, Salganicoff A, Cubanski J, Freed M. In: Kaiser Family Foundation, ed. Opportunities and Barriers for Telemedicine in the U.S. During the COVID-19 Emergency and Beyond. 2020. https://www.kff.org/womens-

health-policy/issue-brief/opportunities-and-barriers-for-telemedicine-in-the-u-s-during-the-covid-19-emergency-and-beyond/.
2. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *The Gerontologist*. 1969;2(3):179-186.
3. Petersen RC. Doody R, kurz a, *et al*. current concepts in mild cognitive impairment. *Arch Neurol*. 2001;58:1985-1992.
4. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53:695-699.
5. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatr Neuropsychol Behav Neurol*. 1988;1:111-117.
6. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319:388-396.
7. 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.
8. APA. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
9. Roman GC, Tatemichi TK, Erkinjuntti T, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN international workshop. *Neurology*. 1993;43:250-260.
10. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *Lancet Neurol*. 2007;6:734-746.
11. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-536.
12. Steinhauser S, Schumacher M, Rucker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016;16:97.
13. Simel DL, Rennie D. *The Rational Clinical Examination: Evidence-Based Clinical Diagnosis*. JAMA, USA: McGraw-Hill; 2009.
14. Kastner M, Antony J, Soobiah C, Straus SE, Tricco AC. Conceptual recommendations for selecting the most appropriate knowledge synthesis method to answer research questions related to complex evidence. *J Clin Epidemiol*. 2016;73:43-49.
15. Saini M, Shlonsky A. *Systematic Synthesis or Qualitative Research*. USA: OUP; 2012.
16. Shores MM, Ryan-Dykes P, Williams RM, et al. Identifying undiagnosed dementia in residential care veterans: comparing telemedicine to in-person clinical examination. *Int J Geriatr Psychiatry*. 2004;19:101-108.
17. Martin-Khan M, Flicker L, Wootton R, et al. The diagnostic accuracy of telegeriatrics for the diagnosis of dementia via video conferencing. *J Am Med Dir Assoc*. 2012;13:487e419-487e424.
18. Martin-Khan M, Varghese P, Wootton R, Gray L. Successes and failures in assessing cognitive function in older adults using video consultation. *J Telemed Telecare*. 2007;13:60-62.
19. Desmond DW, Tatemichi TK, Hanzawa L. The telephone interview for cognitive status (TICS): reliability and validity in a stroke sample. *Int J Geriatr Psychiatry*. 1994;9:803-807.
20. Dal Forno G, Chiovenda P, Bressi F, et al. Use of an Italian version of the telephone interview for cognitive status in Alzheimer's disease. *Int J Geriatr Psychiatry*. 2006;21:126-133.
21. Konagaya Y, Washimi Y, Hattori H, Takeda A, Watanabe T, Ohta T. Validation of the telephone interview for cognitive status (TICS) in Japanese. *Int J Geriatr Psychiatry*. 2007;22:695-700.
22. Lipton RB, Katz MJ, Kuslansky G, et al. Screening for dementia by telephone using the memory impairment screen. *J Am Geriatr Soc*. 2003;51:1382-1390.
23. Manly JJ, Schupf N, Stern Y, Brickman AM, Tang M-X, Mayeux R. Telephone-based identification of mild cognitive impairment and dementia in a multicultural cohort. *Arch Neurol*. 2011;68:607-614.
24. Reckess GZ, Brandt J, Luis CA, et al. Screening by telephone in the Alzheimer's disease anti-inflammatory prevention trial. *J Alzheimers Dis*. 2013;36:433-443.
25. Seo EH, Lee DY, Kim SG, et al. Validity of the telephone interview for cognitive status (TICS) and modified TICS (TICSm) for mild cognitive impairment (MCI) and dementia screening. *Arch Gerontol Geriatr*. 2011;52:e26-e30.
26. Ruiz RG, Calvo Mauri JF, Carrasco Lopez FJ. Spanish version and pilot study of a telephonic test of cognitive status to evaluate and detect oversights in the assessment and follow-up of dementia. *Aten Primaria*. 1994;13:61-66.
27. Kempen GI, Meier AJ, Bouwens SF, van Deursen J, Verhey FR. The psychometric properties of the Dutch version of the telephone interview cognitive status (TICS). *Tijdschr Gerontol Geriatr*. 2007;38:34-40.
28. Knopman DS, Roberts RO, Geda YE, et al. Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. *Neuroepidemiology*. 2010;34:34-42.
29. Crooks VC, Clark L, Petitti DB, Chui H, Chiu V. Validation of multi-stage telephone-based identification of cognitive impairment and dementia. *BMC Neurol*. 2005;5:8.
30. Graff-Radford NR, Ferman TJ, Lucas JA, et al. A cost effective method of identifying and recruiting persons over 80 free of dementia or mild cognitive impairment. *Alzheimer Dis Assoc Disord*. 2006;20:101-104.
31. Camozzato AL, Kochhann R, Godinho C, Costa A, Chaves ML. Validation of a telephone screening test for Alzheimer's disease. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 2011;18:180-194.
32. Wong SS, Fong KN. Reliability and validity of the telephone version of the Cantonese mini-mental state examination (T-CMMSE) when used with elderly patients with and without dementia in Hong Kong. *Int Psychogeriatr*. 2009;21:345-353.
33. Alexopoulos P, Perneczky R, Cramer B, Grimmer T, Kurz A. Validation of a short telephone test (T3MS) for the diagnosis of cognitive impairment. *Fortschr Neurol Psychiatr*. 2006;74:329-336.
34. Norton MC, Tschanz JA, Fan X, et al. Telephone adaptation of the modified mini-mental state exam (3MS). The Cache County study. *Neuropsychiatr Neuropsychol Behav Neurol*. 1999;12:270-276.
35. Metitieri T, Geroldi C, Pezzini A, Frisoni GB, Bianchetti A, Trabucchi M. The ITEL-MMSE: An Italian telephone version of the mini-mental state examination. *Int J Geriatr Psychiatry*. 2001;16:166-167.
36. Newkirk LA, Kim JM, Thompson JM, Tinklenberg JR, Yesavage JA, Taylor JL. Validation of a 26-point telephone version of the mini-mental state examination. *J Geriatr Psychiatry Neurol*. 2004;17:81-87.

37. 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 standardised assessments. *Intern Med J*. 2004; 34:239-242.
38. Ball CJ, Scott N, McLaren PM, Watson J. Preliminary evaluation of a low-cost VideoConferencing (LCVC) system for remote cognitive testing of adult psychiatric patients. *Br J Clin Psychol*. 1993;32:303-307.
39. Cullum C, Weiner MF, Gehrmann HR, Hynan LS. Feasibility of telecognitive assessment in dementia. *Assessment*. 2006;13:385-390.
40. Pendlebury ST, Welch SJ, Cuthbertson FC, Mariz J, Mehta Z, Rothwell PM. Telephone assessment of cognition after transient ischemic attack and stroke: modified telephone interview of cognitive status and telephone Montreal cognitive assessment versus face-to-face Montreal cognitive assessment and neuropsychological battery. *Stroke*. 2013;44:227-229.
41. Zietemann V, Kopczak A, Muller C, Wollenweber FA, Dichgans M. Validation of the telephone interview of cognitive status and telephone montreal cognitive assessment against detailed cognitive testing and clinical diagnosis of mild cognitive impairment after stroke. *Stroke*. 2017;48:2952-2957.
42. DeYoung N, Shenal BV. The reliability of the Montreal cognitive assessment using telehealth in a rural setting with veterans. *J Telemed Telecare*. 2019;25:197-203.
43. Lindauer A, Seelye A, Lyons B, et al. Dementia care comes home: patient and caregiver assessment via telemedicine. *Gerontologist*. 2017;57:e85-e93.
44. Iiboshi K, Yoshida K, Yamaoka Y, et al. A validation study of the remotely administered Montreal cognitive assessment tool in the elderly Japanese population. *Telemed J E Health*. 2019;26:1-9.
45. Abdolahi A, Bull MT, Darwin KC, et al. A feasibility study of conducting the Montreal cognitive assessment remotely in individuals with movement disorders. *Health Informatics J*. 2016; 22:304-311.
46. Liu L, Goodarzi Z, Jones A, Posno R, Straus SE, Watt JA. Factors associated with virtual care access in older adults: a cross-sectional study. *Age Ageing*. 2021, In Press. https://doi.org/10.1093/ageing/afab021.
47. Brewster L, Mountain G, Wessels B, Kelly C, Hawley M. Factors affecting front line staff acceptance of telehealth technologies: a mixed-method systematic review. *J Adv Nurs*. 2014;70:21-33.
48. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. *J Med Internet Res*. 2020;22:e16407.
49. Joy M, McGagh D, Jones N, et al. Reorganisation of primary care for older adults during COVID-19: a cross-sectional database study in the UK. *Br J Gen Pract*. 2020;70:e540-e547.

Editor's Note

Watt et al. offer a thoughtful, relevant and rigorous review of the current state of knowledge regarding remotely administered neurocognitive diagnostic testing, and the news is encouraging: there is some evidence that results of video and telephone administration can be comparable to in-person administration. Watt et al confine their review to multi-domain diagnostic tests, or 'mini-batteries', not simple, first-stage cognitive impairment screening tools such as those that might be used in Annual Wellness Visits. In addition, the tests best studied for comparability of remote and in-person administration are proprietary, supporting the need to establish accurate remote testing protocols for freely-available assessments. Clinicians should also understand that remote administration itself can impose limitations related to the use of and comfort with technology, and may exacerbate (or at least not diminish) the cultural, linguistic, educational, sensory and other contextual biases and barriers that can be encountered with existing in-person tests.

Soo Borson, MD

TWITTER

Jennifer A. Watt @JennAnnWatt

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

Appendix File S1: MEDLINE Literature Search Strategy
File S2: Gray Literature Databases Searched
Table S1: Characteristics of Included Studies
Table S2: Study-Specific Measures of Diagnostic Test Accuracy
Table S3: Study-Specific Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)
Figure S1: Summary of QUADAS-2 Quality Assessment Across Studies Included in this Systematic Review
File S3: Barriers and Facilitators Abstracted from Studies Comparing Virtual to In-Person Cognitive Testing and Assessment
Supplementary References

How to cite this article: Watt JA, Lane NE, Veroniki AA, et al. Diagnostic accuracy of virtual cognitive assessment and testing: Systematic review and meta-analysis. *J Am Geriatr Soc*. 2021;1–12. https://doi.org/10.1111/jgs.17190