Neuropsychol Rev (2017) 27:174–186
DOI 10.1007/s11065-017-9349-1

REVIEW

Neuropsychological Test Administration by Videoconference: A Systematic Review and Meta-Analysis

Timothy W. Brearly[1,2,3] · **Robert D. Shura**[1,2,3] · **Sarah L. Martindale**[1,2,3] · **Rory A. Lazowski**[4] · **David D. Luxton**[5] · **Brian V. Shenal**[6,7] · **Jared A. Rowland**[1,3,8,9]

Received: 31 October 2016 / Accepted: 8 May 2017 / Published online: 16 June 2017

Abstract The purpose of the current systematic review and meta-analysis was to assess the effect of videoconference administration on adult neurocognitive tests. We investigated whether the scores acquired during a videoconference administration were different from those acquired during on-site administration. Relevant counterbalanced crossover studies were identified according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Twelve studies met criteria for analysis. Included samples consisted of healthy adults as well as those with psychiatric or neurocognitive disorders, with mean ages ranging from 34 to 88 years. Heterogenous data precluded the interpretation of a summary effect for videoconference administration. Studies including particpants with a mean age of 65–75, as well as studies that utilized a high speed network connection, indicated consistent performance across videoconference and on-site conditions, however studies with older participants and slower connections were more variable. Subgroup analyses indicated that videoconference scores for untimed tasks and those allowing for repetition fell 1/10th of a standard deviation below on-site scores. Test specific analyses indicated that verbally-mediated tasks including digit span, verbal

Electronic supplementary material The online version of this article (doi:10.1007/s11065-017-9349-1) contains supplementary material, which is available to authorized users.

✉ Timothy W. Brearly
tbrearly@gmail.com

Robert D. Shura
robert.shura2@va.gov

Sarah L. Martindale
sarah.martindale-supak@va.gov

Rory A. Lazowski
rlazowski@collegeboard.org

David D. Luxton
ddluxton@uw.edu

Brian V. Shenal
brian.shenal@va.gov

Jared A. Rowland
jared.rowland@va.gov

1 VA Mid-Atlantic Mental Illness Research, Education, and Clinical Centers (MIRECC), W. G. "Bill" Hefner Veterans Affairs Medical Center, Salisbury, NC 28144, USA

2 Mental Health and Behavioral Sciences Service Line, W. G. "Bill" Hefner Veterans Affairs Medical Center, Salisbury, NC 28144, USA

3 Department of Psychiatry and Behavioral Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27103, USA

4 The College Board, Yardley, PA 19067, USA

5 Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA 98102, USA

6 Mental Health Service Line, Salem Veterans Affairs Medical Center, Salem, VA 24153, USA

7 Virginia Tech Carilion School of Medicine, Roanoke, VA 24016, USA

8 Research and Education Service Line, W. G. "Bill" Hefner Veterans Affairs Medical Center, Salisbury, NC 28144, USA

9 Department of Neurobiology and Anatomy, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA

fluency, and list learning were not affected by videoconference administration. Scores for the Boston Naming Test fell 1/10th of a standard deviation below on-site scores. Heterogenous data precluded meaningful interpretation of tasks with a motor component. The administration of verbally-mediated tasks by qualified professionals using existing norms was supported, and the use of visually-dependent tasks may also be considered. Variability in previous studies indicates a need for further investigation of motor-dependent tasks. We recommend the development of clinical best practices for conducting neuropsychological assessments via videoconference, and advocate for reimbursement structures that allow consumers to benefit from the increased access, convenience, and cost-savings that remote assessment provides.

Keywords Teleneuropsychology · Telepsychology · Telehealth · Telemedicine · Mhealth · Assessment

Introduction

Videoconference-mediated telehealth is used in diverse patient care settings including primary care, critical care, neurology, behavioral health, and psychiatry, among other specialty areas (Davis et al. 2014; Fortney et al. 2015; Johnson et al. 2015; Lilly et al. 2014; Tensen et al. 2016; Wechsler 2015; Wennergren et al. 2014). Improved access to care, patient satisfaction, convenience, and potential cost-savings are some of the principal benefits of telehealth-based services (Bashshur et al. 2015; Czaja 2016; Davis et al. 2014; Fordyce et al. 2007; Hilty et al. 2007). A strong and growing evidence base demonstrating that modern videoconference-based telehealth does not impede successful clinical intervention or rapport with patients has generally supported its adoption and expansion within health care systems and by individual providers (Acierno et al. 2016; Backhaus et al. 2012; Barak et al. 2008; Gros et al. 2013; Hilty et al. 2013).

The capability to provide neuropsychological assessments remotely is another potential benefit of videoconference-mediated telehealth (Adjorlolo 2015; Backhaus et al. 2012; Grosch et al. 2011; Luxton et al. 2014). Diagnostic agreement between on-site and videoconference screenings for neurocognitive disorders has been demonstrated (Loh et al. 2007; Shores et al. 2004), and recent studies have supported the feasibility of administering neuropsychological measures by means of secure videoconference software in rural, urban, and culturally-diverse healthcare settings (Cullum et al. 2014; Parikh et al. 2013; Vahia et al. 2015). Further, three pilot clinics within the Department of Veteran's Affairs have described the outcomes of videoconference neuropsychological assessments using a variety of existing tests (Barton et al. 2011; Harrell et al. 2014; Turner et al. 2012). Within these clinics, remote evaluations have been offered to adults across the lifespan who cannot attend an on-site neuropsychology appointment due to travel-related difficulties or who prefer a remote appointment for personal convenience. Services have primarily been offered to satellite clinics, most of which employ a technician who configures and orients patients to the videoconference setting and provides or collects test response forms. Noted challenges have included difficulty providing "hands-on" portions of testing or neurobehavioral exams and fewer opportunities for behavioral observation due to limitations of camera angle or variable connection quality. Predominately positive feedback from patients and their families has been consistent across neuropsychological studies where this was assessed, including appreciation for the opportunity to receive specialty care without the logistical hurdles of travel distance (Barton et al. 2011; Harrell et al. 2014; Parikh et al. 2013; Turner et al. 2012).

Despite the benefits of videoconference neuropsychological assessment, whether videoconference administration might influence the reliability and validity of neuropsychological test scores is not fully established. Several studies have found no significant score differences between on-site and videoconference administration of tests (Ciemins et al. 2009; DeYoung et al. 2015; Galusha-Glasscock et al. 2015; McEachern et al. 2008; Menon et al. 2001; Turkstra et al. 2012; Vahia et al. 2015; Vestal et al. 2006). However, others have found differences in scores for some tests including the Clock Drawing Test (Cullum et al. 2006, 2014; Grosch et al. 2015; Hildebrand et al. 2004), Digits Forward (but not Digits Backward) (Wadsworth et al. 2016), the Hopkins Verbal Learning Test-Revised (HVLT-R: Cullum et al. 2006, 2014), Ammon's Quick Test (Kirkwood et al. 2000), Oral Trail Making Part A (Wadsworth et al. 2016), the Boston Naming Test (BNT: Wadsworth et al. 2016), and the Adult Memory and Information Processing Battery (AMIPB) Story Recall task (Kirkwood et al. 2000). These differences are particularly common on the Clock Drawing Test, though it has been noted that this may be due to the extremely limited range of possible scores, making the Clock Drawing Test poorly suited for test-retest comparisons (Cullum et al. 2006). In light of this, some have emphasized that small effect sizes can be statistically significant without being associated with differences in clinical presentation or interpretation (Jacobsen et al. 2003), and that differences in scores are often smaller than what would be expected due to measurement error (Wadsworth et al. 2016). However, others have cautioned against the possible dangers of decreased precision in some settings, such as when testing for mild impairment, interpreting scores close to cutoffs between levels of impairment, or assessing for change by means of serial testing (Hildebrand et al. 2004; Jacobsen et al. 2003; Turkstra et al. 2012).

Differences in results between test administration conditions may be due to factors other than systematic error introduced by videoconference. For instance, studies that have identified differences between modalities note that higher scores are not consistently associated with the on-site condition, and videoconference administration has resulted in higher test scores in some studies (Ball et al. 1993; Cullum et al. 2014; Turkstra et al. 2012; Wadsworth et al. 2016). Variability in testing between conditions, and on the same test between studies, may be partially due to factors such as delay between administrations, network connection speed, display or camera quality, display size, sound quality, computer or transmission device speed, microphone quality, speaker quality, test reliability, familiarity or comfort with videoconference technology, or variations in test administration (Cernich et al. 2007; Grosch et al. 2011). As one example, Cullum et al. (2014) hypothesized that favorable scores obtained by means of videoconference testing may be related to unintentional discrepancies in test administration (e.g., clearer enunciation of words by examiners in an effort to ensure comprehension in the videoconference condition). Others have noted that patients may report less distraction when tested by videoconference due to reduced interpersonal anxiety (Jacobsen et al. 2003; Kirkwood et al. 2000).

Given the limitations of the current literature, the field of clinical neuropsychology has been hesitant to adopt videoconference technology into practice. A frequently cited concern is the reliance of many neuropsychological tests on standardized administration techniques and normative data, and the lack of an empirical basis for applying those norms to telehealth administration. Further, many neuropsychological tests require the physical manipulation of materials, which may not be possible when the examiner is not physically present. Thus, it is important to assess whether, or under what circumstances, neurocognitive tests normed through traditional on-site administration can be administered by means of videoconference in lieu of the development of videoconference specific norms.

To address this need, the purpose of the current systematic review and meta-analysis was to assess the effect of videoconference-mediated administration on adult neurocognitive test results. Specifically, we investigated whether the scores acquired during a videoconference administration of neuropsychological tests were significantly different from those acquired during on-site administration with the same participants. Second, we conducted subgroup analyses by test and test type to investigate whether different tests may vary in their appropriateness for videoconference administration. Finally, ad hoc moderator analyses were conducted to explore sources of heterogeneity across studies.

Methods

Inclusion and Exclusion Criteria

Prior to beginning the systematic literature search, a priori inclusion and exclusion criteria were identified (study protocol not published). Included studies assessed the test performance of adult participants (>17 years old) by utilizing a counterbalanced cross-over design to compare test results between on-site and videoconference conditions. The rigor of a counterbalanced within-subjects cross-over design, where all participants are tested in both conditions with condition order being alternated across the sample, reduces the potential confounds of practice and condition ordering effects, enhances statistical power, and reduces between-participant variation. It was considered acceptable for study participants to receive some assistance with videoconference technology set-up and orientation (e.g., learning how to adjust volume, being introduced to limits of webcam coverage area), but any studies that utilized active involvement by assistants during the testing process were excluded to facilitate the generalizability of findings across clinical settings. Any studies that utilized proprietary software or hardware specifically designed for test administration (e.g., touchscreen kiosks, mobile applications) were also excluded. Finally, any studies that utilized only self-report symptom or personality measures were excluded.

Data Identification and Screening

To objectively consolidate the evidence regarding videoconference neuropsychological assessment, empirical research articles meeting a priori criteria were selected according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. As outlined in Fig. 1, a systematic search for published studies using PubMed, EBSCOhost (PsycArticles, PsycInfo), and ProQuest was conducted on January 9th–11th, 2016 using the search string "(tele OR remote OR video OR cyber) AND cognitive AND (testing OR assessment OR evaluation)" with a date range of 1980–2016 for all databases.

EBSCOhost was used to search abstracts and titles from quantitative studies published in peer reviewed journals, resulting in 294 articles. ProQuest returned 78 results after searching citations and abstracts within the subject areas of "neurosciences," "psychology," and "mental health" using the index terms "Memory OR Mental health OR Attention OR Cognition OR Psychiatry OR Neuroscience OR Learning and memory OR Working memory." Finally, PubMed was used to search titles and abstracts and returned 449 results after five video recordings were removed. The total count of 821 results was reduced to 680 after duplicates were removed. All abstracts were then independently screened by the first and second authors, and any articles that did not

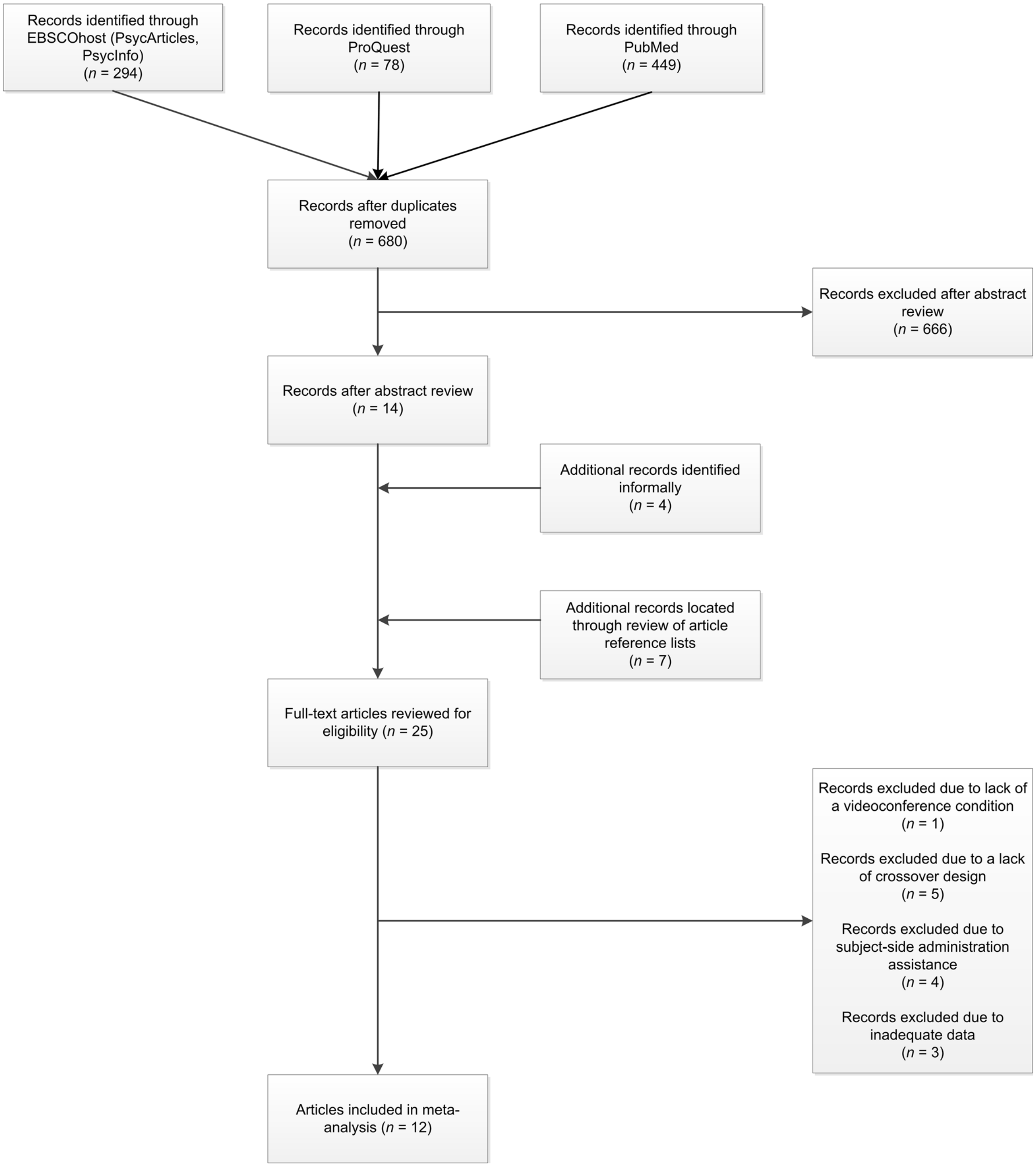


Fig. 1 Flowchart of the study selection process

include data on on-site and videoconference administered cognitive testing excluded. This process resulted in the exclusion of 666 articles leaving 14 articles remaining for full-text review. Four additional studies published in 2015 and 2016 were identified through a review of articles identified through less formal sources (e.g., Google Scholar, Read by QxMD, and relevant e-mail listservs). Finally, the reference lists of all 18 remaining articles were reviewed, resulting in the identification of seven additional studies appropriate for full-text review.

In all, a total of 25 studies were selected for full-text review. The full-text reviews were conducted independently by the first, second, and third authors, with each reviewer rating every study on all inclusion and exclusion criteria. After review, ratings were discussed by the first three authors with any differences being resolved by consensus agreement. This process resulted in four article exclusions due to participant - side assistance with test administration, five due to an absence of cross-over design, and one for not utilizing videoconference administration. For articles that did not report the minimal data required for meta-analysis (e.g., means and standard deviations), contact attempts were made by e-mail and ResearchGate. Four studies by authors that did not respond or who were not able to provide the data required for meta-analysis after a 9-month period were excluded. Twelve studies were ultimately included in analyses.

Data Extraction

The following data, detailed in Table 1, were extracted by the first author from included studies: sample size, population characteristics, mean age, mean education, connection speed, delay between administration conditions, and tests administered. Means and standard deviations for the on-site and videoconference conditions, correlations between administrations, and the direction of the effects were extracted for each test in each study. When data were not reported in this manner, the mean difference between on-site and videoconference administration conditions, the standard deviation of the difference, the correlation between administrations, and the direction of the effects were extracted for each test included in each study. Two studies did not report correlations for all tests administered thus the procedure recommended by the Cochrane Collaboration was followed (Higgins and Green 2011). Specifically, the average correlation for all tests included in the meta-analysis was used to impute a correlation value, $r = .73$, that was lower than values reported for other tests in the study, and was thus considered a conservative estimate. Further, a sensitivity analysis was completed using the lowest, $r = .23$, and highest, $r = .97$, correlation values reported for any test by any included study, and the significance or non-significance of the summary mean effect was consistent across all three values (lowest r, highest r, imputed r) providing further evidence for the appropriateness of this approach.

The importance of differentiating by task requirements when considering the validity of videoconference test administration has been previously noted (Grosch et al. 2011). Arguably, the primary issues in videoconference administration include the loss of information transmission due to audio or video stuttering or loss of quality. Although no more than minimal technology malfunctions were noted during test administration across studies, there was variation in technology quality across studies, and no objective measurements of visual or auditory abilities or comprehension of stimuli by participants were included. Data were not available to allow error analysis comparisons between conditions (e.g., number of requests for repetition of stimuli or responses). Thus, it is possible that tests that may be more sensitive to variations in transmission quality during the testing process may have been more variable across conditions then less administratively restrictive tasks. For these reasons, timed tests and single-trial tests that could be confounded by repetition of information by the examiner or examinee were coded as "synchronous dependent." Untimed tests that were deemed to be robust to breaks or variations in visual or audio quality were coded as "non-synchronous dependent." For example, digit span tests and list-learning tests where coded as synchronous dependent, based on their susceptibility to interference from technological disruption during administration. Tests such as the BNT and visuoconstructional tasks (e.g., Figure Recall) were coded as non-synchronous dependent, because the scores are not timing dependent, and repetition by the examiner or examinee is generally permitted (Table 2). The term synchronous is commonly used in telehealth literature to describe care provided in real-time without breaks in communication (Luxton et al. 2014).

Meta-Analytic Approach

Extracted data were analyzed using Comprehensive Meta-Analysis, version 3.3. The random effects model was chosen for all analyses because it is the most realistic approach when combining real world data from diverse settings compared to the fixed effects model (Borenstein et al. 2010). The random effects model assumes that the effects across studies are not identical, assumes each study has its own population effect size, and most accurately reflects the relative diversity of tests and technology conditions included in the analysis (Cheung and Vijayakumar 2016; Higgins and Green 2011). Further, the random effects model allows researchers to generalize results to studies that are similar but not included in this meta-analysis, whereas this generalization is not permissible with the fixed effects model. Planned analyses included calculation of a summary mean effect, mean effects for synchronous and non-synchronous dependent tests, and a mean effect for each test (when possible). Additional moderator analyses were also conducted to explore sources of heterogeneity across studies.

Given the relatively small sample size in some of the included studies, Hedges' g was the standardized mean difference metric chosen for the calculation of study effect sizes, as recommended by the Cochrane Collaboration (Higgins and Green 2011). Effect sizes were interpreted as small when $g \geq .20$, medium when $g \geq .50$, and large when $g \geq .80$ (Cohen 1988). For studies that reported more than one test score, a pooled effect size of all relevant scores was calculated and used as the unit of analysis to account for dependency in individual test outcomes (Borenstein et al. 2010). Effects in the negative direction indicated that scores on the videoconference condition were lower than those acquired through on-site administration.

To assess homogeneity in the sample(s) of effect sizes, the Q statistic was calculated. When Q is significant, it indicates that the variance in the population of effect sizes is greater than would be expected from sampling error alone. I^2 was also calculated, which indicates the proportion of the variation in effect size estimates that is due to heterogeneity rather than chance (Higgins and Thompson 2002). I^2 is interpreted as a

Table 1 Details of studies included in meta-analysis

Study	Sample size	Population characteristics	Mean age (years)	Mean education (years)	Connection speed	Delay between conditions	Tests administered
Cullum et al. 2006	33	Neurocognitive (MCI, AD)	73	15	High speed	Same day	BNT-15, DS, HVLT, MMSE, PF, SF
Cullum et al. 2014	202	Mixed (MCI, AD, healthy)	69	14	High speed	Same day	BNT-15, CD, DS, HVLT, MMSE, PF, SF
Galusha-Glasscock et al. 2015	18	Mixed (MCI, AD, healthy)	70	14	High speed	Same day	RBANS
Grosch et al. 2015	8	Psychiatric (outpatient)	-	-	384 kbit/s	Same day	CD, DS, MMSE
Hildebrand et al. 2004	29	Healthy	68	13	384 kbit/s	2 to 4 weeks	BTA, CD, WAIS-III MR, PF, WAIS-III VC [d]
Jacobsen et al. 2003	32	Healthy	35	17	384 kbit/s	Same day [b]	BVRT, DS, GP, SDMT, SRT, VOSPS, WAIS-I VC, WMS-R LM
Kirkwood et al. 2000	27	Psychiatric (inpatient, outpatient)	46	-	128 kbit/s	Same day	AMIPB, NART, QT
Loh et al. 2004	20	Mixed	82	-	384 kbit/s [a]	-	MMSE
Loh et al. 2007	20	Mixed	79	-	384 kbit/s	-	MMSE
Montani et al. 1997	14	Mixed (rehabilitation, no psychiatric history)	88	-	Coaxial	8 days	CD, MMSE
Vestal et al. 2006	10	Neurocognitive (AD)	74	-	384 kbit/s	Same day	BNT, MAE, PF
Wadsworth et al. 2016	84	Mixed (dementia, healthy)	65	13	High speed	Same day [c]	BNT, CD, DS, HVLT, MMSE, OT, PF, SF

AMIPB, Adult Memory and Information Processing Battery; *BNT*, Boston Naming Test; *BNT*-15, Boston Naming Test-15 Item; *BTA*, Brief Test of Attention; *BVRT*, Benton Visual Retention Test; *CD*, Clock Drawing Test; *DS*, Digit Span; *GP*, Grooved Pegboard; *HVLT*, Hopkins Verbal Learning Test-Revised; *LM*, Logical Memory; *MAE*, Multilingual Aphasia Examination (Aural Comprehension of Words and Phrases, Token Test); *MCI*, Mild Cognitive Impairment; *MMSE*, Mini-Mental Status Exam; *MR*, Matrix Reasoning; *NART*, National Adult Reading Test; *OT*, Oral Trails A + B; *PF*, Phonemic Fluency; *QT*, Quick Test; *RAVLT*, Rey Auditory Verbal Learning Test; *RBANS*, Repeatable Battery for the Assessment of Neuropsychological Status; *SF*, Semantic Fluency; *SRT*, Seashore Rhythm Test; *SDMT*, Symbol Digit Modalities Test; *VC*, Vocabulary; *VOSPS*, Visual Object and Space Perception Battery Silhouettes; *WAIS*, Wechsler Adult Intelligence Scale; *WMS*, Wechsler Memory Scale

[a] One participant was tested using 128 kbit/s connection

[b] 3 subjects were tested at <= 3 day delay

[c] 2 subjects were tested at longer delays (7 and 14 days)

[d] RAVLT was also administered, however, study authors identified possible problems with RAVLT data (by e-mail) and recommended its exclusion from analyses

Table 2 Test type coding

Test	Synchronous dependent
Single-trial (AMIPB List Learning or Story Recall, DS, HVLT, LM, MMSE, NART, OT, PF, SF, SRT)	Yes
Timed (AMIPB Information Processing, BTA, BVRT, GP, SDMT)	Yes
Untimed (AMIPB Figure Recall, BNT, CD, MAE Aural Comprehension or Token Test, QT, VOSPS, MR, VC)	No

RBANS scores not included due to lack of subtest level data

percentage, with "0%" indicating no observed heterogeneity, and higher percentages indicating increasing heterogeneity. Although caution should be taken when categorizing the measurement of heterogeneity, Higgins et al. (2003) suggests that I^2 values of 25%, 50%, and 75% be interpreted as low, moderate, and high, respectively.

Study Quality and Risk of Bias

The PRISMA guidelines provide recommended criteria for rating study quality and risk of bias. However, these criteria are not well-suited to cross-over studies. A checklist has been proposed by Ding et al. (2015) for the rating of cross-over studies, however, many of the proposed criteria are addressed by a counterbalanced design (e.g., carry-over effects). Further, it is not possible to conceal the presence of videoconference versus on-site conditions from participants or study investigators. For this reason, study inclusion or exclusion criteria were designed to only include properly counterbalanced studies that provided all data necessary to proceed with meta-analysis, thus precluding a formal rating of study quality. Overall, included studies were deemed to be of moderate to high quality, with moderate quality studies neglecting to account for changes in functioning over time by using variable or extended delays between conditions, not clearly documenting the concealment of results from investigators until the completion of both testing conditions, or not reporting important outcome data (thus requiring further direct inquiry with study investigators for inclusion).

Publication Bias

The assumption of publication bias analyses is that studies with large effects will be more likely to be published, however, the literature currently being assessed would arguably be more likely to be published if smaller effects are found. This is because many reseachers publishing in this arena are interested in supporting the use of videoconferencing in neuropsychological assessment, which would be supported by the finding of small effect sizes. Therefore, some of the

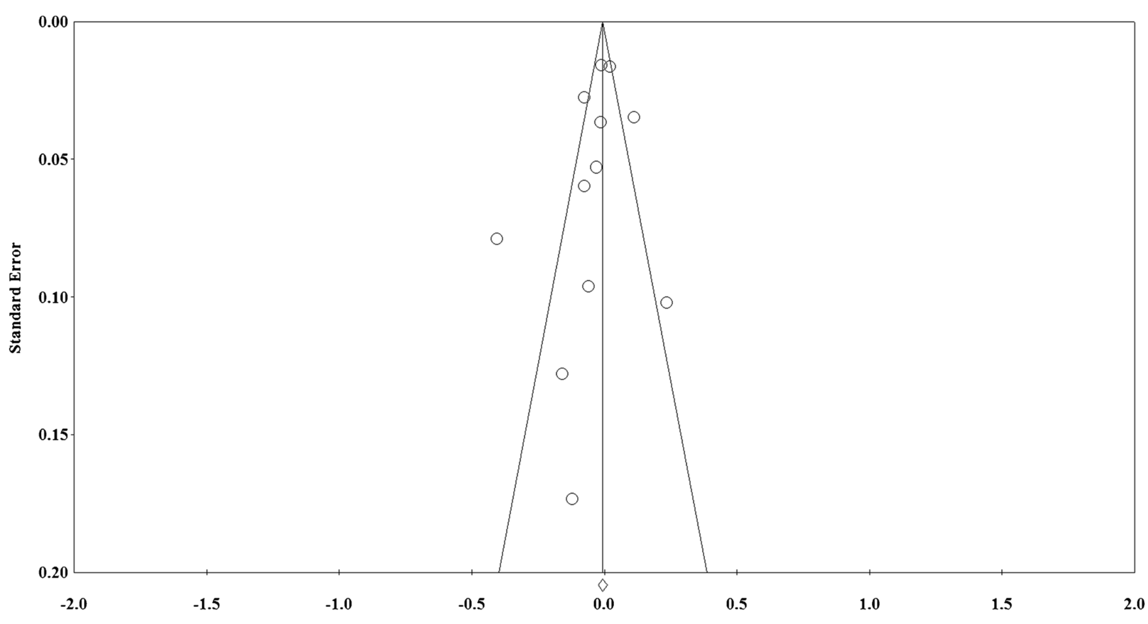


Fig. 2 Funnel plot of individual study effects (Hedge's g) by standard errors in relation to overall summary effect

common statistical tests which assess for the influence of publication bias on a statistically significant effect (Fail-safe N, Trim and Fill) were not relevant.

A funnel plot (Fig. 2) was used to visualize data to ensure that there was not a clustering of studies away from the mean effect size, as this would suggest publication bias (Borenstein et al. 2010).

Inspection of the funnel plot suggested adequate symmetry, with the expected clustering of larger studies at the top and an even distribution of studies to both the left and right of the mean effect. Studies with a larger standard error of measurement did not group on one side of the mean effect, suggesting that neither the size nor the direction of the effect influenced which studies were published. This conclusion was further investigated using Begg and Mazumdar's rank correlation test (Begg and Mazumdar 1994). This test resulted in a Kendall's tau b of −0.227, with a two-sided p-value of .304. Though power was limited due to the relatively small number of studies, this result indicated that there was not a significant relationship between sample size and effect size in the included studies, thus publication bias was not suggested.

Results

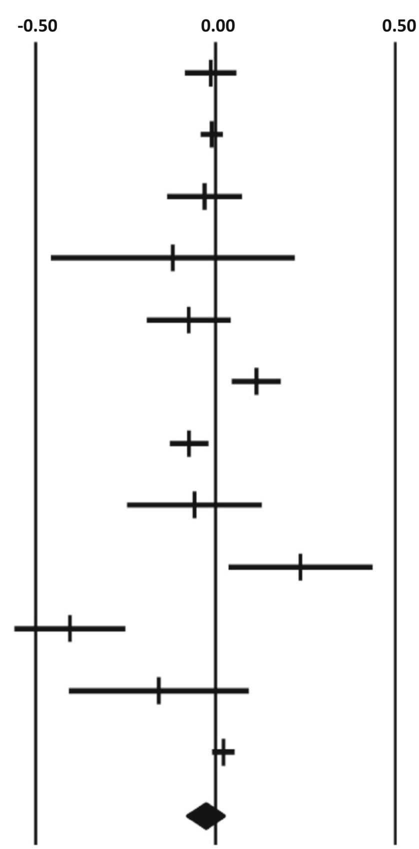
Overall Effect of Videoconference Administration

A total sample of 497 participants across 12 included studies was used to calculate the summary weighted mean effect (Table 3). Patient characteristics were diverse and included healthy participants, participants being treated on a medical unit, participants in inpatient and outpatient psychiatric or substance use treatment, and participants diagnosed with mild cognitive impairment, Alzheimer's disease, or other dementias.

The mean effect size of videoconference administration was small, and videoconference administration of tests did not yield a significant change in test scores, $g = -0.03$; SE = 0.03; 95% CI [−0.08, 0.02], $p = .253$. The magnitude of the effect indicated that overall, videoconference test scores were about 1/33rd of a standard deviation less than on-site scores and were not significantly different. However, as was evident in the included forest plot, there was significant heterogeneity between studies, $Q(11) = 55.67$, $p \leq .001$, $I^2 = 80.24\%$, suggesting the need

Table 3 Primary analysis results

Study	Hedges' g	Standard error	Lower limit	Upper limit	p-Value
Cullum et al., 2006	-0.01	0.04	-0.09	0.06	.708
Cullum et al., 2014	-0.01	0.02	-0.04	0.02	.503
Galusha-Glasscock et al., 2015	-0.03	0.05	-0.14	0.07	.561
Grosch et al., 2015	-0.12	0.17	-0.46	0.22	.490
Hildebrand et al., 2004	-0.08	0.06	-0.19	0.04	.212
Jacobsen et al., 2003	0.11	0.04	0.05	0.18	.001
Kirkwood et al., 2000	-0.07	0.03	-0.13	-0.02	.008
Loh et al., 2004	-0.06	0.10	-0.25	0.13	.541
Loh et al., 2007	0.24	0.10	0.04	0.44	.021
Montani et al., 1997	-0.41	0.08	-0.56	-0.25	<.001
Vestal et al., 2006	-0.16	0.13	-0.41	0.09	.217
Wadsworth et al., 2016	-0.02	0.02	-0.01	0.05	.182
Summary Effect	**-0.03**	**0.03**	**-0.08**	**0.02**	**.253**



for more fine-grained analysis of study subgroups. This need was also evident in the substantial variation in the direction of the effects. Across all 79 scores from included studies, 26 mean scores were higher for the videoconference condition (32.91%), 48 mean scores were higher for the on-site condition (60.76%), and five mean scores were exactly equal between conditions (6.33%).

Effect of Videoconference for Timed and Single-Trial (Synchronous Dependent) Tests

To further explore possible sources of variation contributing to heterogeneity across studies, the mean effect of videoconference administration on synchronous and non-synchronous dependent tests was calculated. As previously discussed, synchronous tests refer to tests for which administration might be negatively affected by disruptions in information transmission. For studies where multiple scores were reported for the same test (e.g. digits forward and digits backward) the mean of the scores was used as the unit of analysis to avoid "double counting" the data from a single sample. The effect for non-synchronous dependent tests was statistically significant but small, $g = -0.10$; SE = 0.03; 95% CI [−0.16, −0.04], $p < .001$. The distribution of effect sizes was considered homogenous, $Q(8) = 12.99$, $p = .112$, $I^2 = 38.43\%$. This result indicates that videoconference administration of non-synchronous dependent tests yielded neurocognitive test scores approximately 1/10th of a standard deviation lower than on-site administration. On the other hand, the mean effect of videoconference administration across synchronous-dependent tests was heterogeneous across studies, supporting the need for a more fine-grained analysis, $Q(10) = 56.42$, $p < .001$, $I^2 = 82.28\%$. To further explore the nature of the heterogeneity present in previous analyses, planned subgroup analyses were conducted for all tests that were utilized in at least three studies.

Effect of Videoconference on Verbally-Mediated Tests

As detailed in Table 4, the mean effect of videoconference administration on digit span tests (i.e., Repeatable Battery for the Assessment of Neuropsychological Status, Wechsler Adult Intelligence Scale, Wechsler Memory Scale), phonemic fluency, category fluency, and total learning on list learning tasks was small and not statistically significant. The distribution of scores was homogenous across studies, with small or moderate non-significant variation. The mean effect sizes indicated that scores acquired through videoconference administration were approximately 1/10th to 1/50th of a standard deviation different from on-site comparison scores.

Effect of Videoconference on Visually and Verbally-Mediated Tests

Videoconference administration of the BNT (15-item and full administration) across studies yielded a statistically significant but small summary effect, reducing the score by about 1/10th of a standard deviation. The distribution of scores was homogenous across studies, indicating that this is a consistent finding. However, the distributions of scores for the Clock Drawing Test and the Mini-Mental State Exam (MMSE) were heterogeneous, with inconsistency between studies precluding meaningful interpretation.

Post hoc Moderator Analyses

Additional subgroup analyses were run for all moderator categories which contained three or more studies in order to parse effects according to study mean age, population, connection speed, and delay between conditions (Table 5). Mean age was converted to a categorical variable, as all but two studies had a mean age of 65 or higher.

Excluding younger participants did not explain excess heterogeneity across studies. However, studies with a mean age between 65 and 75 were homogenous, with no difference

Table 4 Test-specific analyses

Test	Number of studies	Total participants	Hedge's g	Standard error	p	Q	p	I^2(%)
BNT	4	329	−0.12	0.03	<.001	1.76	.624	0.00
Semantic fluency	3	319	−0.08	0.09	.353	5.77	.056	65.34
Clock drawing	5	335	−0.13	0.10	.203	12.60	.013	68.25
Digit Span	5	359	−0.05	0.06	.429	9.38	.052	57.34
List learning (total)	3	313	0.10	0.06	.066	4.59	.101	56.46
MMSE	7	380	−0.04	0.06	.543	30.36	<.001	80.24
Phonemic fluency	5	356	−0.02	0.03	.367	1.41	.842	0.00

List Learning refers to the total learning scores for the HVLT-R and AMIPB List Learning subtest

BNT, Boston Naming Test; *MMSE*, Mini-Mental State Exam

Table 5 Moderator subgroup analyses

Test	Number of studies	Total participants	Hedge's *g*	Standard error	*p*	*Q*	*p*	I^2(%)
Mean age								
Older adult (≥ 65)	9	430	−0.04	0.03	.162	37.54	<.001	78.69
65–75	6	376	0.00	0.01	.778	5.77	.329	13.40
> 75	3	54	−0.08	0.19	.672	25.59	<.001	92.18
Population								
Non-healthy	10	436	−0.05	0.03	.112	42.10	<.001	78.62
Mixed	8	401	−0.04	0.03	.235	36.20	<.001	80.66
Connection speed								
Internet/network connection	11	483	0.00	0.02	.855	29.66	.001	66.28
High speed connection	4	337	0.00	0.01	.847	2.67	.445	0.00
384 kbit/s	6	119	0.01	0.06	.866	15.81	.007	68.38
Same day administration	8	414	0.00	0.02	.842	22.38	.002	68.72

"Mixed" group excludes psychiatric samples. A study that utilized a coaxial cable connection was excluded from the "internet/network connection" group, and a study that utilized a 128 kbit/s connection was excluded from the remaining "connection speed" analyses

between videoconference and on-site administrations. There was significant heterogeneity across studies with an average age over 75, though this subsample included only three studies with a total of 54 participants. Due to limited specificity of sample characterization, as well as the presence of only two studies with healthy participants, subgroup analyses could not rule out the effect of clinical symptoms on consistency between conditions. Studies reporting a high speed network connection were homogenous, with no effect found for videoconference administration. Delay between conditions did not appear to be a primary source of heterogeneity.

Discussion

Before videoconference-mediated neuropsychological testing can be expanded, it is important to know whether the reliability and validity of testing is influenced by this medium of delivery. This was the first systematic review and meta-analysis to assess the effects of videoconference administration on neuropsychological test performance. The current findings did not reveal a clear trend towards inferior performance when tests were administered via videoconference. Consistent differences were found for only one test (BNT) and the effect size was small. Further, among a subset of studies utilizing a high speed connection, there was no effect for videoconference administration. There was, however, some variability across studies associated with specific test characteristics.

Test vulnerability to disruptions in technology (e.g., break in connection, loss of sound) was not a clear source of variation. Specifically, non-synchronous dependent test scores (those deemed robust to videoconference administration due to lack of timing and allowance for repetition) yielded a consistently small negative effect in the videoconference condition, with scores falling approximately 1/10th of a standard deviation below respective on-site counterparts. This finding was consistent for the BNT specifically, which yielded scores 1/10th of a standard deviation, approximately, below those acquired through on-site administration. It should be noted that three of the four BNT studies used the 15-item version, rather than the full BNT. Thus, these findings may not generalize to the full BNT (where errors would have a less substantial effect on the total score). Among synchronous-dependent test scores (i.e., scores that required timing or that prohibited repetition), verbal tasks such as digit span, list learning, and verbal fluency were not affected by videoconference administration, demonstrating consistent scores that were not different between conditions. There was greater variability in the apparent effect of videoconference administration on synchronous-dependent tests requiring interaction with physical objects, such as the MMSE and Clock Drawing Test, precluding interpretation of the effect of videoconference administration on these specific tests.

It was not possible to estimate an effect of videoconference administration on the two available individual tests requiring a motor response (MMSE and Clock Drawing Test), due to the high level of variability between studies. At the individual study level, the motor-dependent test that yielded the largest negative effect for videoconference administration was the Clock Drawing Test. However, although the largest calculated effect for the Clock Drawing Test was $g = -0.69$ (about 7/10ths of a standard deviation below on-site scores (see Grosch et al. 2015)), another study found a null effect for this same test (see Cullum et al. 2014). Though videoconference administration may have been the cause of this variability, another possibility is that study heterogeneity for the Clock Drawing Test and MMSE was due to other uncontrolled sources of error such as variations in connection, display or sound quality, distance of visual information from the camera, variations in scoring approach (e.g., Clock Drawing Test), or the multifactorial nature of the task (e.g., MMSE). In fact, the distribution of higher study mean scores across both on-site and videoconference conditions suggests that factors beyond videoconferencing were likely the cause. As an illustration, problems due to technology were expected to lead to lower scores in videoconference compared to the on-site conditions. However, most scores on the MMSE were higher in the videoconference condition. These unexpected results suggest the need for further studies to better understand the dynamics underlying the variation in these two tests, and perhaps motor-dependent tests more generally.

Heterogeneity was also associated with the moderating factors of age and connection speed. Studies with a mean sample age between 65 and 75 were homogenous with no difference between conditions, suggesting that older adults should not categorically be excluded from videoconference-mediated testing. Though all studies reported a connection adequate for testing, network speed may also have been a source of variation. Specifically, studies with connections identified as high speed were homogenous, with no effect for videoconference administration. Some studies have also indicated that practice effects or test-retest reliability may vary more in clinical populations (see Beglinger et al. 2005), and the role of this variable as a source of heterogeneity could not be ruled out. However, two individual studies included in our analysis looked for differences in order effects and effect sizes between healthy and cognitively compromised participant performances and did not find them to be present (Cullum et al. 2014; Galusha-Glasscock et al. 2015).

Finally, though the current analysis adequately adjusted for practice effects by using a crossover design, the effects of chance and measurement error across administrations cannot be fully ruled out. A well-established cutoff for identifying true differences in scores in a clinical setting is 1.65 standard deviations (Duff 2012). This is well above the differences identified in the current meta-analysis, which did not exceed approximately 1/10th of a standard deviation. This suggests high levels of consistency across conditions and limited relevance for clinical interpretation. It should also be noted that this meta-analysis tested for significance nearly 30 times, thus at least one of the significance findings is likely due to chance (Higgins and Green 2011).

Investigations in the field of videoconference-mediated neuropsychological assessment are still relatively sparse. However, recent advances in secure, HIPAA-compliant videoconference technology have provided new opportunities for investigators who are interested in informing clinical practice through high quality studies. It is recommended that future studies consider a counterbalanced cross-over design using brief, standardized delays, consistently report demographic data, means, standard deviations, and correlations for test scores, and provide detailed documentation of the types of technology used. Future studies should consistently report all scores for each test administered, with particular emphasis on scores important for clinical practice (e.g., single-trial learning, delayed recall of list) to allow for accumulation of results across studies. There is a need for studies that incorporate complex tasks such as those often used to assess processing speed, complex attention, and those that rely on motor and visual abilities. The development of standardized methods for the presentation of visual stimuli, to ensure consistent clarity and visibility of discrete details, would also be prudent. However, care should be taken to account for test reliability when forming conclusions given variations in psychometric quality across tests.

Study limitations include a relatively small number of studies fit for inclusion ($n = 12$), limited variety of tests administered, and variability in study populations with a lack of potentially relevant demographic data in some studies (e.g., level of education). For instance, only two studies used non-cognitively or psychiatrically compromised samples, limiting the value of related subgroup analysis. Another important limitation is our exclusion of studies that used participant-side assistance with test administration. For instance, our current analysis found significant heterogeneity across motor dependent tests, but this finding may not generalize to settings with remote support staff. Individual studies that administered motor-dependent tasks with personal to assist in the presentation of stimuli and test forms, while the qualified examiner remotely provided verbal instructions, have not found any significant differences (Beery-Buktenica Developmental Test of Visual-Motor Integration-IV, RBANS, Wechsler Abbreviated Scale of Intelligence Perceptual Reasoning Index) (Temple et al. 2010; Turkstra et al. 2012). Additional crossover studies were also available, however, diverse reporting of data in the literature prohibited inclusion. Further, included studies were experimental by design, and it is possible that participant performance in a purely clinical context might differ. Strengths of the current analysis include the systematic identification of studies according to PRISMA guidelines, rigorous inclusion or exclusion criteria for study design, analysis of publication bias, and the inclusion of subgroup analyses.

In conclusion, this meta-analytic review provides support for the use of videoconferencing technology in the remote administration of neuropsychological tests, particularly those that rely on verbal responses from participants. Additional research is needed to fully understand the factors that contribute to variable study outcomes on tests requiring a motor response, though current literature suggests that administration assistance from remote support staff provides results comparable to on-site administration. In the future, web and mobile application-based tests will likely facilitate the collection of motor and visually-dependent performance without remote assistance. Finally, we recommend the development of clinical best practices for conducting neuropsychological evaluations by videoconference, and advocate for reimbursement structures that allow consumers to benefit from the increased access, convenience, and cost-savings that remote testing provides. We also remind practitioners of the ethical importance of completing training in the technical, legal, and clinical intricacies of videoconference assessment prior to engaging in remote assessment.

Acknowledgements This research was supported by the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Mental Illness, Research, and Treatment and the Mid-Atlantic Mental Illness Research, Education, and Clinical Center (VISN 6 MIRECC) of the Department of Veterans Affairs Office of Mental Health Services. Special thanks to Larry A. Tupler for his support of this and other related teleneuropsychology investigations within the VISN 6 MIRECC.

References

Acierno, R., Gros, D. F., Ruggiero, K. J., Hernandez-Tejada, B. M. A., Knapp, R. G., Lejuez, C. W., et al. (2016). Behavioral activation and therapeutic exposure for posttraumatic stress disorder: a noninferiority trial of treatment delivered in person versus home-based telehealth. *Depression and Anxiety*. doi:10.1002/da.22476.

Adjorlolo, S. (2015). Can teleneuropsychology help meet the neuropsychological needs of western Africans? The case of Ghana. *Applied Neuropsychology: Adult, 22*(5), 388–398. doi:10.1080/23279095.2014.949718.

Backhaus, A., Agha, Z., Maglione, M. L., Repp, A., Ross, B., Zuest, D., et al. (2012). Videoconferencing psychotherapy: a systematic review. *Psychological Services, 9*(2), 111–131. doi:10.1037/a0027924.

Ball, C. J., Scott, N., McLaren, P. M., & Watson, J. P. (1993). Preliminary evaluation of a low-cost videoconferencing (LCVC) system for remote cognitive testing of adult psychiatric patients. *British Journal of Clinical Psychology, 32*(3), 303–307. doi:10.1111/j.2044-8260.1993.tb01060.x.

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services, 26*(2–4), 109–160. doi:10.1080/15228830802094429.

Barton, C., Morris, R., Rothlind, J., & Yaffe, K. (2011). Video-telemedicine in a memory disorders clinic: evaluation and management of rural elders with cognitive impairment. *Telemedicine and e-Health, 17*(10), 789–793.

Bashshur, R. L., Shannon, G. W., Bashshur, N., & Yellowlees, P. M. (2015). The empirical evidence for telemedicine interventions in mental disorders. *Telemedicine and e-Health*. doi:10.1089/tmj.2015.0206.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088–1101. doi:10.2307/2533446.

Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J., et al. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology, 20*(4), 517–529. doi:10.1016/j.acn.2004.12.003.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111. doi:10.1002/jrsm.12.

Cernich, A. N., Brennana, D. M., Barker, L. M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology, 22*(Suppl 1), S39–S48. doi:10.1016/j.acn.2006.10.004.

Cheung, M. W.-L., & Vijayakumar, R. (2016). A guide to conducting a meta-analysis. *Neuropsychology Review*, 1–8. doi:10.1007/s11065-016-9319-z.

Ciemins, E., Halloway, B., Coon, P. J., McClosky-Armstrong, T., & Min, S. (2009). Telemedicine and the Mini-mental state Examination: Assessment from a distance. *Telemedicine and e-Health*. doi:10.1089/tmj.2008.0144.

Cohen, J. (1988). *Statistical power analysis for the behavioral Sciences*. Hillsdale: L. Erlbaum Associates.

Cullum, C., Weiner, M., Gehrmann, H., & Hynan, L. (2006). Feasibility of telecognitive assessment in dementia. *Assessment, 13*(4), 385–390. doi:10.1177/1073191106289065.

Cullum, C., Hynan, L., Grosch, M., Parikh, M., & Weiner, M. (2014). Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society, 20*(10), 1028–1033. doi:10.1017/S1355617714000873.

Czaja, S. J. (2016). Long-term care services and support systems for older adults: the role of technology. *American Psychologist, 71*(4), 294–301. doi:10.1037/a0040258.

Davis, L. E., Coleman, J., Harnar, J., & King, M. K. (2014). Teleneurology: successful delivery of chronic neurologic care to 354 patients living remotely in a rural state. *Telemedicine and e-Health, 20*(5), 473–477. doi:10.1089/tmj.2013.0217.

DeYoung, N., Belkonen, S., Luci, K., Slater, M., Quigley, T., & Shenal, B. (2015). The reliability and accuracy of administering the Montreal cognitive assessment over telehealth. Presented at the 43rd annual meeting of the International neuropsychological Society, Denver, CO.

Ding, H., Hu, G. L., Zheng, X. Y., Chen, Q., Threapleton, D. E., & Zhou, Z. H. (2015). The method quality of cross-over studies involved in Cochrane systematic reviews. *PloS One, 10*(4). doi:10.1371/journal.pone.0120519.

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. *Archives of Clinical Neuropsychology, 27*(3), 248–261. doi:10.1093/arclin/acr120.

Fordyce, M. A., Chen, F. M., Doescher, M. P., & Hart, L. G. (2007). 2005 physician supply and distribution in rural areas of the United States. Final report, 116. http://depts.washington.edu/uwrhrc/uploads/RHRC%20FR116%20Fordyce.pdf. Accessed 20 July 2016.

Fortney, J. C., Pyne, J. M., Turner, E. E., Farris, K. M., Normoyle, T. M., Avery, M. D., et al. (2015). Telepsychiatry integration of mental health services into rural primary care settings. *International Review of Psychiatry, 27*(6), 525–539. doi:10.3109/09540261.2015.1085838.

Galusha-Glasscock, J. M., Horton, D. K., Weiner, M. F., & Cullum, C. M. (2015). Video teleconference administration of the Repeatable Battery for the assessment of neuropsychological status. *Archives of Clinical Neuropsychology*. doi:10.1093/arclin/acv058.

Gros, D. F., Morland, L. A., Greene, C. J., Acierno, R., Strachan, M., Egede, L. E., et al. (2013). Delivery of evidence-based psychotherapy via video telehealth. *Journal of Psychopathology and Behavioral Assessment, 35*(4), 506–521. doi:10.1007/s10862-013-9363-4.

Grosch, M. C., Gottlieb, M. C., & Cullum, C. M. (2011). Initial practice recommendations for teleneuropsychology. *The Clinical Neuropsychologist, 25*(7), 1119–1133. doi:10.1080/13854046.2011.609840.

Grosch, M. C., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2015). Video teleconference-based neurocognitive screening in geropsychiatry. *Psychiatry Research, 225*(3), 734–735. doi:10.1016/j.psychres.2014.12.040.

Harrell, K. M., Wilkins, S. S., Connor, M. K., & Chodosh, J. (2014). Telemedicine and the evaluation of cognitive impairment: the additive value of neuropsychological assessment. *Journal of the American Medical Directors Association, 15*(8), 600–606. doi:10.1016/j.jamda.2014.04.015.

Higgins, J. P., & Green, S. (2011). Cochrane handbook for systematic reviews of interventions. http://handbook.cochrane.org/front_page.htm. Accessed 8 Apr 2016.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. doi:10.1002/sim.1186.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327*(7414), 557–560.

Hildebrand, R., Chow, H., Williams, C., Nelson, M., & Wass, P. (2004). Feasibility of neuropsychological testing of older adults via videoconference: implications for assessing the capacity for independent living. *Journal of Telemedicine and Telecare, 10*(3), 130–134.

Hilty, D. M., Nesbitt, T. S., Kuenneth, C. A., Cruz, G. M., & Hales, R. E. (2007). Rural versus suburban primary care needs, utilization, and satisfaction with telepsychiatric consultation. *The Journal of Rural Health, 23*(2), 163–165. doi:10.1111/j.1748-0361.2007.00084.x.

Hilty, D. M., Ferrer, D. C., Parish, M. B., Johnston, B., Callahan, E. J., & Yellowlees, P. M. (2013). The effectiveness of telemental health: a 2013 review. *Telemedicine and e-Health, 19*(6), 444–454. doi:10.1089/tmj.2013.0075.

Jacobsen, S. E., Sprenger, T., Andersson, S., & Krogstad, J.-M. (2003). Neuropsychological assessment and telemedicine: a preliminary study examining the reliability of neuropsychology services performed via telecommunication. *Journal of the International Neuropsychological Society, 9*(3), 472–478. doi:10.1017/S1355617703930128.

Johnson, K. A., Meyer, J., Yazar, S., & Turner, A. W. (2015). Real-time teleophthalmology in rural Western Australia. *Australian Journal of Rural Health, 23*(3), 142–149. doi:10.1111/ajr.12150.

Kirkwood, K. T., Peck, D. F., & Bennie, L. (2000). The consistency of neuropsychological assessments performed via telecommunication and face to face. *Journal of Telemedicine and Telecare, 6*, 147–151.

Lilly, C. M., Zubrow, M. T., Kempner, K. M., Reynolds, H. N., Subramanian, S., Eriksson, E. A., et al. (2014). Critical care telemedicine: evolution and state of the art*. *Critical Care Medicine, 42*(11), 2429–2436. doi:10.1097/CCM.0000000000000539.

Loh, P., Ramesh, P., Maher, S., Saligari, J., Flicker, L., & Goldswain, P. (2004). Can patients with dementia be assessed at a distance? The use of telehealth and standardised assessments. *Internal Medicine Journal, 34*(5), 239–242.

Loh, P., Donaldson, M., Flicker, L., Maher, S., & Goldswain, P. (2007). Development of a telemedicine protocol for the diagnosis of Alzheimer's disease. *Journal of Telemedicine and Telecare, 13*(2), 90–94. doi:10.1258/135763307780096159.

Luxton, D. D., Pruitt, L. D., & Osenbach, J. E. (2014). Best practices for remote psychological assessment via telehealth technologies. *Professional Psychology: Research and Practice, 45*(1), 27–35. doi:10.1037/a0034547.

McEachern, W., Kirk, A., Morgan, D. G., Crossley, M., & Henry, C. (2008). Reliability of the MMSE administered in-person and by telehealth. *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques, 35*(5), 643–646. doi:10.1017/S0317167100009458.

Menon, A. S., Kondapavalru, P., Krishna, P., Chrismer, J. B., Raskin, A., Hebel, J. R., & Ruskin, P. E. (2001). Evaluation of a portable low cost videophone system in the assessment of depressive symptoms and cognitive function in elderly medically ill veterans. *The Journal of Nervous and Mental Disease, 189*(6), 399–401.

Montani, C., Billaud, N., Tyrrell, J., Fluchaire, I., Malterre, C., Lauvernay, N., et al. (1997). Psychological impact of a remote psychometric consultation with hospitalized elderly people. *Journal of Telemedicine and Telecare, 3*(3), 140–145. doi:10.1258/1357633971931048.

Parikh, M., Grosch, M. C., Graham, L. L., Hynan, L. S., Weiner, M., Shore, J. H., & Cullum, C. M. (2013). Consumer acceptability of brief videoconference-based neuropsychological assessment in older individuals with and without cognitive impairment. *The Clinical Neuropsychologist, 27*(5), 808–817. doi:10.1080/13854046.2013.791723.

Shores, M. M., Ryan-Dykes, P., Williams, R. M., Mamerto, B., Sadak, T., Pascualy, M., et al. (2004). Identifying undiagnosed dementia in residential care veterans: comparing telemedicine to in-person clinical examination. *International Journal of Geriatric Psychiatry, 19*(2), 101–108. doi:10.1002/gps.1029.

Temple, V., Drummond, C., Valiquette, S., & Jozsvai, E. (2010). A comparison of intellectual assessments over video conferencing and in-person for individuals with ID: preliminary data. *Journal of Intellectual Disability Research, 54*(6), 573–577. doi:10.1111/j.1365-2788.2010.01282.x.

Tensen, E., van der Heijden, J. P., Jaspers, M. W. M., & Witkamp, L. (2016). Two decades of teledermatology: current status and integration in national healthcare systems. *Current Dermatology Reports, 5*(2), 96–104. doi:10.1007/s13671-016-0136-7.

Turkstra, L. S., Quinn-Padron, M., Johnson, J. E., Workinger, M. S., & Antoniotti, N. (2012). In-person versus telehealth assessment of discourse ability in adults with traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 27*(6), 424–432. doi:10.1097/HTR.0b013e31823346fc.

Turner, T. H., Horner, M. D., Vankirk, K. K., Myrick, H., & Tuerk, P. W. (2012). A pilot trial of neuropsychological evaluations conducted via telemedicine in the veterans health administration. *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association, 18*(9), 662–667. doi:10.1089/tmj.2011.0272.

Vahia, I. V., Ng, B., Camacho, A., Cardenas, V., Cherner, M., Depp, C. A., et al. (2015). Telepsychiatry for neurocognitive testing in older rural Latino adults. *The American Journal of Geriatric Psychiatry, 23*(7), 666–670. doi:10.1016/j.jagp.2014.08.006.

Vestal, L., Smith-Olinde, L., Hicks, G., Hutton, T., & Hart Jr., J. (2006). Efficacy of language assessment in Alzheimer's disease: comparing in-person examination and telemedicine. *Clinical Interventions in Aging, 1*(4), 467.

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., et al. (2016). Remote neuropsychological assessment in rural American Indians with and without cognitive impairment. *Archives of Clinical Neuropsychology, acw030*. doi:10.1093/arclin/acw030.

Wechsler, L. R. (2015). Advantages and limitations of teleneurology. *JAMA Neurology, 72*(3), 349–354. doi:10.1001/jamaneurol.2014.3844.

Wennergren, J., Munshi, I., Fajardo, A. D., & George, V. V. (2014). Implementation of clinical video telemedicine (CVT) within a VA medical center is cost effective and well received by veterans. *International Journal of Clinical Medicine, 5*(12), 711–716. doi:10.4236/ijcm.2014.512097.