Journal of the International Neuropsychological Society (2023), **29**, 193–204
doi:10.1017/S1355617722000777

Concurrent validity and reliability of at-home teleneuropsychological evaluations among people with and without HIV

Maulika Kohli[1,2], Arin Fisher[2], Ni Sun-Suslow[2], Anne Heaton[2], Matthew S. Dawson[2], Jennifer Marquie[2], Donald R. Franklin[2], Maria Marquine[2,3], Jennifer E. Iudicello[2,3], Robert K. Heaton[2,3] and David J. Moore[2,3]

[1]San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, USA, [2]HIV Neurobehavioral Research Program, University of California San Diego, San Diego, CA, USA and [3]Department of Psychiatry, University of California San Diego, San Diego, CA, USA

(Received 1 November 2021; final revision 14 August 2022; accepted 2 September 2022; First Published online 13 December 2022)

Abstract

Objective: To determine the reliability of teleneuropsychological (TNP) compared to in-person assessments (IPA) in people with HIV (PWH) and without HIV (HIV−). **Methods:** Participants included 80 PWH ($M_{age} = 58.7$, $SD_{age} = 11.0$) and 23 HIV− ($M_{age} = 61.9$, $SD_{age} = 16.7$). Participants completed two comprehensive neuropsychological IPA before one TNP during the COVID-19 pandemic (March–December 2020). The neuropsychological tests included: Hopkins Verbal Learning Test-Revised (HVLT-R Total and Delayed Recall), Controlled Oral Word Association Test (COWAT; FAS-English or PMR-Spanish), Animal Fluency, Action (Verb) Fluency, Wechsler Adult Intelligence Scale 3rd Edition (WAIS-III) Symbol Search and Letter Number Sequencing, Stroop Color and Word Test, Paced Auditory Serial Addition Test (Channel 1), and Boston Naming Test. Total raw scores and sub-scores were used in analyses. In the total sample and by HIV status, test-retest reliability and performance-level differences were evaluated between the two consecutive IPA (i.e., IPA1 and IPA2), and mean in-person scores (IPA-M), and TNP. **Results:** There were statistically significant test-retest correlations between IPA1 and IPA2 (r or $\rho = .603–.883$, $ps < .001$), and between IPA-M and TNP (r or $\rho = .622–.958$, $ps < .001$). In the total sample, significantly lower test-retest scores were found between IPA-M and TNP on the COWAT (PMR), Stroop Color and Word Test, WAIS-III Letter Number Sequencing, and HVLT-R Total Recall ($ps < .05$). Results were similar in PWH only. **Conclusions:** This study demonstrates reliability of TNP in PWH and HIV−. TNP assessments are a promising way to improve access to traditional neuropsychological services and maintain ongoing clinical research studies during the COVID-19 pandemic.

Keywords: human immunodeficiency virus; COVID-19; validation study; neuropsychology; cognition; telehealth

Introduction

The use of telehealth (i.e., audio and videoconferencing to deliver healthcare) is rapidly growing, especially since the coronavirus disease (COVID-19) pandemic (Wosik et al., 2020). Within the field of neuropsychology, an advocacy team established by the Inter Organizational Practice Committee has been providing up-to-date recommendations and guidelines on the use of teleneuropsychological assessments (TNP) to clinicians (Bilder et al., 2020). Meanwhile, a host of training resources and virtual seminars/workshops have been disseminated across neuropsychological organizations, highlighting the rapid and immense need for knowledge within the context of neuropsychological research in this relatively new territory.

Telehealth allows for increased access to health care services, especially among persons with chronic illnesses and disabilities (Lillicrap et al., 2019). It decreases barriers to accessing appropriate care (e.g., lack of transportation, financial constraints, stigmatization) and can allow for patients needing specialty care to access providers despite geographical location (Gajarawala & Pelkowski, 2021; Moffatt & Eley, 2010; Speedie et al., 2008). These advantages are especially pertinent given that already vulnerable populations (e.g., people with HIV; PWH) could be more susceptible to contracting COVID-19 and/or face more adverse health outcomes once infected (Mirzaei et al., 2020).

Telehealth is generally well-accepted by patients, providers, and families (Parikh et al., 2013; Parsons et al., 2021; Shore, 2013), with patient reports of up to 98% satisfaction with videoconferencing and little concerns regarding privacy. Furthermore, some patients found videoconferencing more enjoyable and less anxiety-inducing in their naturalistic environment (Parikh et al., 2013). Despite the potential for decreased emotional connection compared to in-person assessments (IPA), videoconferencing has been found to provide similar personal interactions and possibly more frequent appointments (Bloem et al., 2020; Marra et al., 2020). With regard to TNP, patients and clinicians have found TNP evaluations acceptable and feasible during the COVID-19 pandemic and reported several favorable features, including saved travel time, reduced risk of COVID-19 exposure, and reduced concentration difficulties (Parsons et al., 2021).

Corresponding author: David J. Moore, email: djmoore@health.ucsd.edu

Cite this article: Kohli M., Fisher A., Sun-Suslow N., Heaton A., Dawson M.S., Marquie J., Franklin D.R., Marquine M., Iudicello J.E., Heaton R.K., & Moore D.J. (2023) Concurrent validity and reliability of at-home teleneuropsychological evaluations among people with and without HIV. *Journal of the International Neuropsychological Society* **29**: 193–204, https://doi.org/10.1017/S1355617722000777

Applications of TNP suggest a strong agreement between TNP and IPA across a variety of populations (e.g., older adults, patients with multiple sclerosis, cognitive impairment, psychiatric conditions, cerebrovascular accident) (Barcellos et al., 2021; Cullum et al., 2014; Marra et al., 2020; Matchanova et al., 2020; Tailby et al., 2020; Wadsworth et al., 2018). Among cognitively impaired and non-impaired participants, scores on neuropsychological measures across domains (i.e., memory, attention, verbal fluency, language, executive function) between TNP (administered in-clinic) and in-person conditions were highly concordant (Cullum et al., 2014). Furthermore, a recent systematic review (19 studies) of TNP validity concluded no significant effect related to video-administration of certain cognitive screeners (i.e., MMSE, MoCA), language tests (i.e., Boston Naming Test, Letter Fluency), attention/working memory tasks (i.e., Digit Span Total), and memory tests (Hopkins Verbal Learning Test-Revised) (Brearly et al., 2017; Marra et al., 2020). Majority of these studies conducted the TNP assessments over desktop/laptop (Marra et al., 2020). According to The Inter Organizational Practice Committee, Pearson recommends a display size of at least 9.75 on the patient side, although few studies have examined differences in TNP validity across different device types and sizes. Passell et al. (2021) found an association between device group and reaction time such that measures with more complex stimuli and responses (e.g., Trails A & B) were most affected by screen size. Together, these studies suggest that TNP is reliable and valid for various populations, and validity across various device types remains to be established.

To our knowledge, studies have yet to examine the reliability and comparability of remote TNP assessments versus traditional IPA among PWH. Previous TNP validations studies support the administration of TNP across multiple domains (e.g., memory, attention, executive function) that may be particularly relevant to assess among PWH (Becker et al., 1995; Heaton et al., 2015). Therefore, the cognitive effects of HIV may be amenable to TNP assessment.

Considering the COVID-19 pandemic, there is an immediate need for psychometrically sound neuropsychological assessments that can be administered while patients are at home, to maintain clinical neuropsychological care and ongoing research studies. However, several questions need to be answered before neuropsychologists can confidently use and interpret results from TNP evaluations. The current study will address two primary questions: How comparable are test scores obtained by TNP versus IPA? Is test reliability affected by mode of administration?

The specific purposes of this study are to (1) compare test-retest reliabilities between PWH and those without HIV (HIV−) at the participants' two most recent IPA (IPA1 and IPA2), and between in-person assessments and at-home TNP assessments, and (2) assess performance-level differences in neuropsychological assessment scores between IPA1 and IPA2, and between in-person and at-home TNP assessments. We hypothesized strong test-retest correlations and minimal performance-level differences in our samples of PWH and HIV−. We will additionally investigate the potential effects of technical aspects related to the remote assessment (e.g., participant device type, interruptions to testing) environment on raw scores at the TNP evaluation.

Methods

Participants

Participants included 80 PWH ($M_{age} = 58.7$, $SD_{age} = 11.0$) and 23 HIV− ($M_{age} = 61.9$, $SD_{age} = 16.7$) individuals who were enrolled in NIH-funded studies at the University of California, San Diego (UCSD) HIV Neurobehavioral Research Program (HNRP) from 2005 to 2020, demonstrated capacity to consent, and provided written informed consent. All study procedures were approved by the UCSD Institutional Review Board and are in accordance with the Helsinki Declaration.

The current study is a secondary analysis of data from each participant's TNP evaluation, IPA1, and IPA2. Inclusion criteria were (1) age 18 years or older; (2) ability to provide informed consent; (3) negative urine toxicology for illicit drugs (excluding marijuana) or negative Breathalyzer test for alcohol on the day of the in-person study visits; (4) at least two completed IPAs; (5) greater than three months between the two prior IPAs; and (6) one remote TNP evaluation completed between March 2020 and December 2020. Exclusion criteria at in-person and TNP evaluations were consistent among all parent studies and included (1) psychotic disorder diagnosis; (2) history of a non-HIV related neurological condition known to impact neurocognitive functioning (e.g., stroke, head injury with neurological complications; epilepsy); and (3) non-HIV related medical conditions associated with neurocognitive disorders. To determine language of test administration, participants were asked to self-report how well they spoke English and Spanish using a Likert-type scale (0 = not well to 3 = very well). Participants reporting equal scores for each language were tested in their preferred language (Mungas et al., 2004). Ninety-six participants were evaluated in English (93.2%). Given that each participant is their own control, we considered it appropriate to include the Spanish speakers in analyses.

In-person psychiatric and neuromedical evaluation

The Composite International Diagnostic Interview (v2.1) was administered at the IPAs to assess current (i.e., past 12 months) and lifetime (i.e., >12 months ago) mood and substance use disorders (World Health Organization, 1997). The parent grants were funded before the publication of the DSM 5; therefore, diagnoses were made based on the DSM-IV criteria. HIV serostatus was determined using ELISA/Western blot by a CLIA-certified reference lab.

In-person neuropsychological assessment

Participants were administered a well-validated and comprehensive battery of neuropsychological assessments measuring seven cognitive domains: verbal fluency, executive function, processing speed, learning, delayed recall, working memory, and motor skills (Carey et al., 2004). This battery was designed in accordance with the international consensus conference recommendations (i.e., Frascati criteria) for HIV-associated neurocognitive disorder (Antinori et al., 2007; Heaton et al., 2010).

Teleneuropsychological evaluation setup

TNP evaluations were conducted using HIPAA-compliant Zoom with both the examiner and participants using their personal devices in their respective home environments. To the extent possible, examiner setups were standardized. Examiner standardizations included use of a virtual private network; collection of participant responses with an iPad and stylus; secluded setting to minimize interruptions; computers with a camera and microphone. Because examiner screens are shared with the participant, all computer notifications were disabled. To protect participant privacy, Zoom meeting rooms were password-protected and examiners

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

used headphones. All examiners operated from the same video-based platform; however, internet connection quality and computer hardware varied between examiners.

Because TNP evaluations were conducted in participants' naturalistic environments, participant standardizations were limited. HNRP schedulers recommended that participants find a private, quiet location in their homes, sit at a desk or table, wear headphones to improve audio quality and ensure confidentiality, and use a device with video capabilities for visual measures. Since June 2020, participants who did not have suitable home environments for TNP testing were provided the option of using a testing room (adhering to social distancing guidelines) at the HNRP for their TNP evaluation. Participants who connected by landline telephone received audio-only measures; participants who connected by tablet or personal computer received audio and visual measures; and participants who connected by smartphone received audio and visual measures, one of which needed to be adapted to conform to Zoom's non-adjustable mobile settings (i.e., Stroop Color and Word Test). Prior to testing, examiners administered a Remote Visit Questionnaire to assess participant testing environment (e.g., privacy, device used). After testing, examiners completed a second section of the Remote Visit Questionnaire to retrospectively capture the signal/connection quality during testing, audio quality, and interruptions during testing. Interruptions to testing were considered as any auditory or visual distraction that could influence neuropsychological performance (e.g., examinee's phone ringing, dog barking, family member speaking, garbage truck reversing).

Pre-testing sequence for the teleneuropsychological evaluation

Prior to beginning the TNP evaluation, participants received a brief introduction about the TNP procedure. To reduce the potential for distraction from self-view and video of the examiner, the video panel was minimized for participants only, leaving only test materials visible. Participants provided updated neurobehavioral and substance use histories for the interval between the prior IPA and the TNP visit. Examiners assessed substance use (e.g., alcohol, marijuana, methamphetamine) quantity since the previous IPA and lifetime quantity. Suspected intoxication at the TNP evaluation was indicated in behavioral notes.

Participants completed the Profile of Mood States (POMS), a 65-item self-report measure of mood (i.e., tension–anxiety, depression, anger–hostility, fatigue, confusion, vigor) over the previous seven days (McNair et al., 1981).

Teleneuropsychological assessment battery

A comparison of the in-person and TNP batteries is presented in Table 1. Inclusion criteria for neurocognitive measures in the TNP battery were (a) brevity, (b) common use among HNRP studies, and (c) suitability for video-administration and response recording. Measures included Hopkins Verbal Learning Test-Revised (HVLT-R Total and Delayed Recall) (Benedict et al., 1998; Diaz-Santos et al., 2021), Controlled Oral Word Association Test (COWAT; FAS, PMR) and Category (Animal) Fluency (Borkowski et al., 1967; Marquine, Morlett Paredes, et al., 2021), Action (Verb) Fluency (Woods et al., 2005), Wechsler Adult Intelligence Scale 3rd Edition (WAIS-III) Symbol Search and Letter Number Sequencing, Stroop Color and Word Test (Gooding et al., 2021; Rivera Mindt et al., 2021; Stroop, 1935), Paced Auditory Serial Addition Test (PASAT – Channel 1) (Diehr et al., 1998; Gooding et al., 2021), and the 60-item version of the Boston Naming Test (BNT; excluding item #48) (Kaplan et al., 1983).

Table 1. In-person battery versus teleneuropsychological battery

	In-person	Teleneuropsych
Verbal fluency		
Controlled oral word fluency test	X	X
Category fluency (animals)	X	X
Action (verb) fluency	X	X
Executive function		
Trail making test, part B	X	
Stroop color and word test	X	X
Wisconsin card sorting test (64-item)	X	
Language		
Boston naming test	X	X
Working memory		
WAIS-III letter number sequencing	X[a]	X
Paced auditory serial addition test (channel 1)	X	X
WMS-III spatial span	X[a]	
Processing speed		
Trail making test, part A	X	
WAIS-III digit symbol	X	
WAIS-III symbol search	X	X
Learning and memory		
Hopkins verbal learning test-revised	X	X
Brief visuospatial memory test-revised	X	
Motor skills		
Grooved pegboard test	X	

Note. WAIS-III = Wechsler Adult Intelligence Scale, Third Edition; WMS-III = Wechsler Memory Scale, Third Edition.
[a]Not administered to every participant in-person.

Individual test raw scores were converted into demographically adjusted (i.e., age, sex, education, race/ethnicity) T-scores ($M = 50$, $SD = 10$ in healthy subjects) (Antinori et al., 2007; Cherner et al., 2021; Heaton et al., 2004; Heaton et al., 2003). Individual neuropsychological tests were considered impaired when T-scores < 40 (Taylor & Heaton, 2001).

Given that neuropsychological assessments were designed to be administered in a face-to-face and in-person format, a few accommodations were made to facilitate TNP administration. First, considering internet and audio quality varies between participants, all TNP assessment instructions were presented both orally and visually. In the TNP format, verbal tasks (e.g., HVLT-R), and tasks that rely on verbal responses to visually presented stimuli (e.g., BNT; visual presentation of stimuli via shared screen) are administered similarly to in-person. Three tasks that require visual stimuli or physical interaction with stimuli were reformatted (i.e., stimuli presentation via screen share instead of booklet, verbal response instead of motor) to be included in the video format. Comparisons between IPA and TNP administration of WAIS-III Symbol Search, Stroop Color and Word Test, and PASAT – Channel 1 are described in Supplementary Table 1.

Statistical Analyses

HIV group differences on demographic characteristics were compared using independent *t*-tests and Chi-square statistics as appropriate. Raw scores were used for primary analyses. To examine test-retest reliability between IPA1 and IPA2, Pearson's

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

Table 2. Participant demographic characteristics at the teleneuropsychological assessment by HIV serostatus

	Total sample, $N = 103$	PWH, $N = 80$	HIV, $N = 23$	p-value
Age (years)	59.4 (12.5)	58.7 (11.0)	61.9 (16.7)	.278
Education (years)	14.2 (2.8)	13.9 (2.8)	15.3 (2.5)	**.037**
Sex (male)	84 (81.6%)	66 (83%)	18 (78%)	.761
Ethnicity (non-Hispanic White)	60 (58.2%)	45 (56%)	15 (65%)	.630
Testing language (English)	96 (93.2%)	73 (91%)	23 (100%)	.344
WRAT4 reading[a]	105.1 (12.9)	103.5 (11.3)	109.8 (16.0)	.093
Lifetime major depressive disorder	56 (56%)	45 (57.7%)	11 (50%)	.628
Lifetime substance use disorder	62 (62%)	54 (69.2%)	8 (36.4%)	**.007**
POMS total mood disturbance[b]	59.4 (39.8)	60.6 (38.7)	55.5 (43.9)	.601
HIV disease characteristics				
History of AIDS	–	51 (64.6%)	–	–
Detectable plasma viral load[c,d]	–	0 (0%)	–	–
Current CD4 count	–	643 [501, 850]	–	–
Nadir CD4 count	–	140 [49.3, 300.0]	–	–
Estimated years of HIV disease	–	24.0 [16.4, 30.7]	–	–
ARV status (on cART)	–	75 (96.2%)	–	–

Note. WRAT4 Reading = Wide Range Achievement Test; POMS = Profile of Mood States; Values are presented as *M* (*SD*) or Mdn [IQR]. Bolded values indicate $p < .05$; PWH = people with HIV
[a]$N = 62$, administered at the first in-person visit only, not administered to Spanish speakers.
[b]$N = 92$.
[c]Defined as >50 copies/mL in plasma.
[d]$N = 40$.

correlation coefficients (*r*) were used for normally distributed scores and Spearman's rho correlations for scores with skewed distributions. If raw scores on IPA1 and IPA2 were highly correlated (*r* or $\rho > .500$, $p < .05$), a mean in-person score was calculated for each neuropsychological test to represent average in-person performance (IPA-M) (Hemphill, 2003). Correlation coefficients were calculated for normally distributed scores and Spearman's rho correlations were used for scores with skewed distributions to examine test-retest reliability between IPA-M and TNP in the total sample and by HIV status. Paired *t*-tests were used to compare performance-level differences for normally distributed raw scores and Wilcoxon signed rank for skewed distributions between (1) IPA1 and IPA2; and (2) IPA-M and TNP in the total sample and by HIV status. Benjamini-Hochberg procedure was applied to correct for multiple comparisons using a false discovery rate of 0.05. Follow-up analyses using matched paired *t*-tests examined differences in T-scores between IPA2 and TNP to account for the effects of age, sex, education, and race/ethnicity.

Descriptive statistics from three of the Remote Visit Questionnaire items were calculated. Exploratory analyses were conducted to investigate the potential effects of technical aspects of the remote assessment environment on raw scores at the TNP evaluation. Linear regressions were used to examine the association between device type and change in raw scores from IPA2 to TNP (change score = TNP − IPA2). One-way analysis of variance (ANOVA) and Tukey's honest significant difference (HSD) tests were used to compare TNP raw scores between the four device types. A *t*-test was used to examine the effects of examiner reported interruptions to testing on TNP raw scores. Statistical analyses were performed using JMP Pro version 14.0.0 (JMP®, Version <*14.0.0*>. SAS Institute Inc., Cary, NC, 1989–2007).

Results

Demographic characteristics by HIV serostatus

Demographic characteristics at the TNP evaluation by HIV group are presented in Table 2. PWH had significantly fewer years of education and higher rates of lifetime substance use disorder ($ps < .05$) than HIV−. The groups did not differ on age, sex, ethnicity, Wide Range Achievement Test 4 Reading, lifetime diagnosis Major Depressive Disorder, and mood ($ps \geq .09$). All participants completed two IPAs (days apart: $M = 577$, $SD = 716$; Mdn = 365, IQR = 244–583; range = 108–3970) and one remote TNP evaluation (days apart from IPA2: $M = 414$, $SD = 238$; Mdn = 375, IQR = 277–465; range = 112–1655). 49.5% of participants had re-tests more than 1 year apart between IPA1 and IPA2. 55.3% of participants had re-tests more than 1 year apart between IPA2 and TNP.

Test-retest reliability of neuropsychological assessments

Results of correlation analyses are presented in Table 3. There were statistically significant correlations between IPA1 and IPA2 (*r* or $\rho = .603–.883$, mdn = .744, $ps < .001$) and between IPA-M and TNP (*r* or $\rho = .622–.958$, mdn = .801, $ps < .001$) across all neuropsychological assessment raw scores.

Correlations between IPA1 and IPA2 (Table 4) were statistically significant across neuropsychological assessment raw scores in PWH (*r* or $\rho = .596–.871$, mdn = .737, $ps < .001$) and HIV− groups (*r* or $\rho = .556–.943$, mdn = .826, $ps < .05$). Correlations between IPA-M and TNP (Table 5) were statistically significant across neuropsychological assessment raw scores in PWH (*r* or $\rho = .631–.960$, mdn = .820 $ps < .001$) except for the BNT ($\rho = .593$, $p = .122$), and in the HIV− group (*r* or $\rho = .593–.967$, mdn = .855, $ps < .05$). Correlations between IPA-M and TNP in the HIV− group for COWAT PMR and WAIS-III Letter Number Sequencing were not calculated due to small sample size ($n \leq 5$).

Performance-level differences between neuropsychological assessments

Results of matched paired *t*-tests or Wilcoxon Signed Rank test between IPA1 and IPA2 are presented in Table 6. There were no significant differences between raw scores on IPA1 and IPA2 ($ps \geq .012$; Benjamini-Hochberg procedure cut off of 0.05 required). Results examining performance-level differences between IPA-M and TNP are presented in Table 7. There were

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

Table 3. Correlation coefficients between the two most recent in-person assessments (IPA1 and IPA2) and the mean in-person assessment scores (IPA-M) and the teleneuropsychological scores (TNP)

	Correlations between IPA1 and IPA2		Correlations between IPA-M and TNP	
	Pearson's correlation coefficient or Spearman's rho	*p*-value	Pearson's correlation coefficient or Spearman's rho	*p*-value
Verbal fluency				
COWAT (FAS)	.862	**<.001**	.880	**<.001**
Category fluency – animals	.732	**<.001**	.801	**<.001**
Action (verb) fluency	.791	**<.001**	.622	**<.001**
COWAT (PMR)	.829	**.002**	.958	**<.001**
Executive function				
Stroop word	.883	**<.001**	.743	**<.001**
Stroop color	.876	**<.001**	.856	**<.001**
Stroop incongruent	.744	**<.001**	.824	**<.001**
Language				
Boston naming test	.669	**<.001**	.763	**<.001**
Working memory				
WAIS-III letter number sequencing	.603	**<.001**	.652	**<.001**
PASAT (channel 1)	.813	**<.001**	.805	**<.001**
Processing speed				
WAIS-III symbol search	.742	**<.001**	.762	**<.001**
Learning and memory				
HVLT-R total recall	.710	**<.001**	.696	**<.001**
HVLT-R delayed recall	.681	**<.001**	.838	**<.001**

Note. COWAT = Controlled Oral Word Fluency Test; WAIS-III = Wechsler Adult Intelligence Scale, Third Edition; PASAT = Paced Auditory Serial Addition Test (Channel 1); HVLT-R = Hopkins Verbal Learning Test-Revised; Bolded values indicate $p < .05$.

Table 4. Correlation coefficients between the two most recent in-person assessments (IPA1 and IPA2) by HIV status

	HIV+		HIV−	
	Pearson's correlation coefficient or Spearman's rho	*p*-value	Pearson's correlation coefficient or Spearman's rho	*p*-value
Verbal fluency				
COWAT (FAS)	.835	**<.001**	.943	**<.001**
Category fluency – animals	.659	**<.001**	.891	**<.001**
Action (verb) fluency	.733	**<.001**	.886	**<.001**
COWAT (PMR)	.813	**<.001**	$n = 0$	–
Executive function				
Stroop word	.871	**<.001**	.925	**<.001**
Stroop color	.858	**<.001**	.918	**<.001**
Stroop incongruent	.742	**<.001**	.789	**<.001**
Language				
Boston naming test	.678	**<.001**	.556	**.039**
Working memory				
WAIS-III letter number sequencing	.596	**<.001**	$n = 5$	–
PASAT (channel 1)	.788	**<.001**	.826	**<.001**
Processing speed				
WAIS-III symbol search	.737	**<.001**	.760	**<.001**
Learning and memory				
HVLT-R total recall	.678	**<.001**	.740	**<.001**
HVLT-R delayed recall	.676	**<.001**	.803	**<.001**

Note. COWAT = Controlled Oral Word Fluency Test; WAIS-III = Wechsler Adult Intelligence Scale, Third Edition; PASAT = Paced Auditory Serial Addition Test (Channel 1); HVLT-R = Hopkins Verbal Learning Test-Revised; Bolded values indicate $p < .05$.

significantly lower raw scores at TNP compared to IPA-M on COWAT (PMR) ($t(7) = -3.7$, $p = .007$), Stroop Word ($t(49) = -6.1$, $p < .001$), Stroop Color ($t(48) = -3.9$, $p < .001$), Stroop Incongruent ($t(49) = -2.8$, $p = .006$), WAIS-III Letter Number Sequencing ($t(27) = -2.8$, $p = .010$), and HVLT-R Total Recall ($t(81) = -3.6$, $p < .001$). Cohen's effect size values for Stroop Color ($d_z = .529$), Stroop Word ($d_z = .387$), WAIS-III Letter Number Sequencing ($d_z = .533$) and HVLT-R Total Recall ($d_z = .375$) suggest low to moderate practical significance. Effect size values for COWAT (PMR) ($d_z = 1.11$) and Stroop Word ($d_z = .867$) suggest high practical significance.

In PWH and the HIV− group, there were no significant differences between raw scores on IPA1 and IPA2, after correcting for multiple comparisons ($ps > .05$). In PWH, there were significantly lower raw scores on TNP assessments compared to IPA-M on COWAT (PMR) ($t(7) = -3.7$, $p = .007$), Stroop Word ($t(38) = -4.6$, $p < .001$), Stroop Color ($t(37) = -3.0$, $p = .004$), and HVLT-R Total Recall ($t(61) = -3.7$, $p < .001$) (Figure 1). In the HIV− group, there were lower raw scores on the TNP Stroop Word test compared to IPA-M ($t(10) = -5.8$, $p < .001$) (Figure 2).

Follow-up analyses using matched paired *t*-tests examining differences in T-scores between IPA2 and TNP showed significantly lower T-scores at the TNP assessment on Stroop Word (TNP− IPA2 = −5.72; $t(57) = -5.86$, $p < .001$), Stroop Color (TNP − IPA2 = −2.7; $t(56) = -3.26$, $p = .002$), and HVLT-R Total Recall (TNP − IPA2 = −3.65; $t(92) = -3.12$, $p = .002$). Neuropsychological test scores were considered impaired when T-score < 40. There were significantly more impaired scores in the TNP evaluation compared to IPA2 on Stroop Word (IPA2 = 18 (31%), TNP = 29 (50%); $p < .001$), Stroop Color (IPA2 = 19 (33%), TNP = 22 (39%); $p < .001$), and HVLT-R Total Recall (IPA2 = 39 (42%), TNP = 57 (61%); $p < .001$). On the Stroop Word test, 24% of participants went from an unimpaired to impaired score at the TNP evaluation and 5% went from impaired to unimpaired. On the Stroop Color test, 7% of participants went from an unimpaired to impaired score at the TNP evaluation and 2% went from impaired to unimpaired. On the HVLT-R Total Recall, 26% of participants went from an unimpaired to impaired score at the TNP evaluation and 6% went from impaired to unimpaired.

Additional follow-up analyses were conducted to examine the potential effects of administration language. Seven participants were excluded that were tested in Spanish. Results of matched paired *t*-tests were consistent with the total sample, with observed differences in raw scores on Stroop Word, Stroop Color, Stroop Incongruent, WAIS-III Letter Number Sequencing, and HVLT-R Total Recall ($ps \leq .01$).

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

Table 5. Correlation coefficients between the mean in-person assessment scores (IPA-M) and the teleneuropsychological scores (TNP) by HIV status

	HIV+		HIV−	
	Pearson's correlation coefficient or Spearman's rho	*p*-value	Pearson's correlation coefficient or Spearman's rho	*p*-value
Verbal fluency				
COWAT (FAS)	.885	**<.001**	.855	**<.001**
Category fluency – animals	.820	**<.001**	.784	**<.001**
Action (verb) fluency	.932	**<.001**	.967	**<.001**
COWAT (PMR)	.960	**<.001**	$n = 0$	-
Executive function				
Stroop word	.665	**<.001**	.926	**<.001**
Stroop color	.845	**<.001**	.883	**<.001**
Stroop incongruent	.792	**<.001**	.893	**<.001**
Language				
Boston naming test	.790	**<.001**	.593	.122
Working memory				
WAIS-III letter number sequencing[a]	.631	**<.001**	−.817	.184
PASAT (channel 1)	.830	**<.001**	.607	**.048**
Processing speed				
WAIS-III symbol search	.759	**<.001**	.836	**.001**
Learning and memory				
HVLT-R total recall	.685	**<.001**	.714	**<.001**
HVLT-R delayed recall	.824	**<.001**	.856	**<.001**

Note. COWAT = Controlled Oral Word Fluency Test; WAIS-III = Wechsler Adult Intelligence Scale, Third Edition; PASAT = Paced Auditory Serial Addition Test (Channel 1); HVLT-R = Hopkins Verbal Learning Test-Revised; Bolded values indicate $p < .05$.
[a]$n = 4$.

Remote visit questionnaire results

Device type, assessment type, and participant environment were evaluated from the Remote Visit Questionnaire. Of the total sample, 92 participants (89%) completed the Remote Visit Questionnaire. The most common device type used in the TNP evaluation was a smartphone (39%), followed by laptop/desktop (34%), tablet (14%), and traditional telephone (8%) (5% not documented). Audio-only assessments were not limited to only via traditional telephone. Over 75% of the TNP evaluations were conducted using both video and audio. 21% of participants were interrupted at least once during the TNP evaluation (ex. "cathedral bells caused some disruption to participant's attention span"). Results of a *t*-test comparing raw scores by participant interruption status revealed significant differences on the HVLT-R Total Recall such that participants performed worse when there were interruptions during testing ($M = 17.8$; $SD = 5.1$) compared to no interruptions ($M = 22.0$; $SD = 6.1$; $t(85) = -2.67$, $p = .009$). Additional follow-up analyses were conducted to examine the potential effects of interruptions to testing. Nineteen participants were excluded (16 PWH, 3 HIV−) that had interruptions to testing. Results of matched paired *t*-tests revealed differences in raw scores on Stroop Word, Stroop Color, Stroop Incongruent, and HVLT-R Total Recall ($ps < .05$).

In the exploratory analyses, results revealed no significant association between device used at the TNP evaluation and change in performance from IPA2 to TNP. An ANOVA showed a significant omnibus difference across device type groups in TNP raw scores on Category Fluency – Animals ($F(3,81) = 4.13$, $p = .008$). Follow-up pairwise comparisons showed poorer performance when administered via telephone compared to smartphone, tablet, and laptop/desktop ($ps < .05$).

Discussion

Results of this study add to a growing body of literature demonstrating that TNP assessments are reliable and valid across diverse populations and during the COVID-19 pandemic (Barcellos et al., 2021; Brearly et al., 2017; Cullum et al., 2014; Marra et al., 2020; Matchanova et al., 2020). Among our sample of PWH and HIV−, we established test-retest reliability between two IPAs that were approximately 1 year apart; and found significant and moderate to strong correlations between participants' IPA and TNP evaluations. Performance-level differences between IPA and TNP had variable effect sizes with small to moderate effect sizes for Stoop Color, Stroop Word, WAIS-III Letter Number sequencing, and HVLT-R Total Recall; and large effect sizes for COWAT (PMR) and Stroop Word. Importantly, there was only a small sample of participants completed the COWAT (PMR). There were lower raw scores on the Stroop Color and Word Test, COWAT (PMR), WAIS-III Letter Number Sequencing, and HVLT-R Total Recall at the TNP evaluation. Accounting for the effects of age, sex, education, and race/ethnicity, results indicated significant mean differences in T-scores on Stroop Word, Stroop Color, and HVLT-R Total Recall with the greatest T-score point difference on the Stroop Word test (TNP − IPA2 = −5.72). Across these three neuropsychological assessments, there were more participants that went from an unimpaired T-score at IPA2 to an impaired score at the TNP evaluation; however, several participants with impaired scores at IPA2 performed in the unimpaired range at the TNP evaluation. This could be attributable to practice effects, considering some participants had evaluations less than a year after their previous one; however, we might expect more participants to show improvement if there were significant practice effects (Dikmen et al., 1999).

Despite statistical significance, differences in raw scores and T-scores between IPA-M and TNP on the COWAT (PMR), WAIS-III Letter Number Sequencing, and HVLT-R Total Recall were minimal and possibly due to factors associated with COVID-19 (e.g., COVID-19 infection, stress, depression, social isolation), the TNP platform, or factors on the day of testing (e.g., pain, poor sleep, energy). (Hampshire et al., 2021; Suarez-Gonzalez et al., 2021). The marginal difference in raw scores from the in-person to TNP evaluation observed on the HVLT-R Total Recall (mean difference = −1.8) is consistent with another study which found poorer performance on the HVLT-R Total Recall in the video-based condition (mean difference = −2.11) among mildly impaired stroke patients (Chapman et al., 2020). The authors attribute this possibly to mishearing words in the TNP condition or participant anxiety with the TNP scenario. Results from our study indicate that interruptions to testing were significantly associated with worse performance on the HVLT-R Total Recall and significant differences remained even after excluding participants that experienced interruptions. Therefore, differences observed on TNP tests could be attributable to other technical aspects of the TNP environment. For example, audio glitches may affect participant's understanding of task instructions, ability to clearly hear verbal stimuli, and adequate response collection by examiners (Gardner et al., 2021). These testing environment

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

Table 6. Performance-level differences between the two most recent in-person assessments (IPA1 and IPA2)

	IPA1	IPA2	*p*-value	*Cohens d_z*
Verbal fluency				
COWAT (FAS) ($n = 69$)	43.1 (14.6)	42.4 (14.5)	.461	.092
Category fluency – animals ($n = 84$)	20.9 (6.4)	19.7 (6.0)	.012[a]	.263
Action (verb) fluency ($n = 79$)	15.5 (5.3)	15.7 (5.2)	.548	.059
COWAT (PMR) ($n = 7$)	47.8 (13.9)	45.4 (13.7)	.421	.297
Executive function				
Stroop word ($n = 78$)	87.7 (17.0)	87.1 (16.9)	.497	.073
Stroop color ($n = 77$)	63.6 (13.4)	61.8 (14.1)	.022[a]	.262
Stroop incongruent ($n = 77$)	36.8 (10.7)	35.5 (13.4)	.211	.145
Language				
Boston naming test ($n = 56$)	54.8 (5.8)	54.4 (8.6)	.864	.063
Working memory				
WAIS-III letter number sequencing ($n = 36$)	8.9 (2.8)	8.9 (2.2)	1.00	0
PASAT (channel 1) ($n = 76$)	32.2 (11.7)	32.8 (12.6)	.642	.080
Processing speed				
WAIS-III symbol search ($n = 86$)	30.1 (8.2)	29.9 (9.0)	.803	.032
Learning and memory				
HVLT-R total recall ($n = 87$)	22.8 (6.3)	22.7 (6.4)	.911	.015
HVLT-R delayed recall ($n = 83$)	7.2 (3.5)	7.3 (3.4)	.840	.036

Note. Values are presented as *M* (*SD*). COWAT = Controlled Oral Word Fluency Test; WAIS-III = Wechsler Adult Intelligence Scale, Third Edition; PASAT = Paced Auditory Serial Addition Test (Channel 1); HVLT-R = Hopkins Verbal Learning Test-Revised; Bolded values indicate $p < .05$.
[a]Results considered not significant after applying Benjamini-Hochberg procedure (false discovery rate of 0.05).

Table 7. Performance-level differences between the mean in-person assessment scores (IPA-M) and the teleneuropsychological scores (TNP)

	IPA-M	TNP	*p*-value	*Cohens d_z*
Verbal fluency				
COWAT (FAS) ($n = 68$)	42.9 (14.0)	41.1 (14.0)	.036[a]	.262
Category fluency – animals ($n = 83$)	20.3 (5.8)	19.6 (6.0)	.112	.188
Action (verb) fluency ($n = 77$)	15.6 (5.0)	14.8 (5.3)	.123	.178
COWAT (PMR) ($n = 7$)	46.4 (14.1)	41.9 (13.4)	**.007**	1.11
Executive function				
Stroop word ($n = 50$)	87.4 (16.3)	77.3 (16.2)	**<.001**	.867
Stroop color ($n = 49$)	63.8 (14.6)	59.8 (12.9)	**<.001**	.529
Stroop incongruent ($n = 49$)	36.6 (11.5)	33.8 (12.6)	**.006**	.387
Language				
Boston naming test ($n = 35$)	55.7 (4.2)	55.2 (5.0)	.295	.154
Working memory				
WAIS-III letter number sequencing ($n = 28$)	8.7 (2.0)	7.6 (2.7)	**.010**	.533
PASAT (channel 1) ($n = 52$)	34.1 (11.8)	32.5 (12.5)	.428	.210
Processing speed				
WAIS-III symbol search ($n = 58$)	31.2 (7.9)	29.9 (9.1)	.109	.218
Learning and memory				
HVLT-R total recall ($n = 82$)	22.9 (6.0)	21.1 (6.3)	**<.001**	.375
HVLT-R delayed recall ($n = 77$)	7.4 (3.2)	7.2 (3.2)	.591	.110

Note. Values are presented as *M* (*SD*). COWAT = Controlled Oral Word Fluency Test; WAIS-III = Wechsler Adult Intelligence Scale, Third Edition; PASAT = Paced Auditory Serial Addition Test (Channel 1); HVLT-R = Hopkins Verbal Learning Test-Revised; Bolded values indicate $p < .05$.
[a]Results considered not significant after applying Benjamini-Hochberg procedure (false discovery rate of 0.05).

characteristics may be important to capture in TNP practice to understand whether a poor performance may reflect change in neurocognitive function or limitations of videoconferencing.

Few studies have investigated the reliability and validity of the Stroop Color and Word Test, or a similar response inhibition measure, in the TNP setting. One study among a pediatric sample (aged 6–20) found no significant performance-level differences on the DKEFS Color Word Interference Test between in-person and home-based TNP assessment administered via tablet or laptop (Harder et al., 2020). Another study among middle-to-older adults (aged 40–86) suggests moderate correlations between in-person and TNP assessments via desktop on the Stroop Color and Word Test (Zeghari et al., 2022). In the current study, more pronounced differences in raw scores observed across the Stroop Color and Word Test may be attributable to limitations of administering a time-bound visual neuropsychological assessment via videoconferencing. Particularly, lag-time in communication between an incorrect response and examiner feedback could limit the remaining time in the task for correct responses. Additionally, to be administered in the TNP modality, presentation of the stimuli was reformatted to balance the number of words on each slide (i.e., 60 words/slide, 2 slides) with the number of slide changes. Communication between the participant and examiner about changing slides could have also limited opportunity for participant correct responses. Considering differences in T-scores between IPA2 and TNP were greatest for Stroop Word and progressively decreased for the remaining subtests, it is also possible participants could benefit from more practice administration to better acclimate to video-administration of this test. On average, participants T-score on Stroop Word dropped six points at the TNP evaluation

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

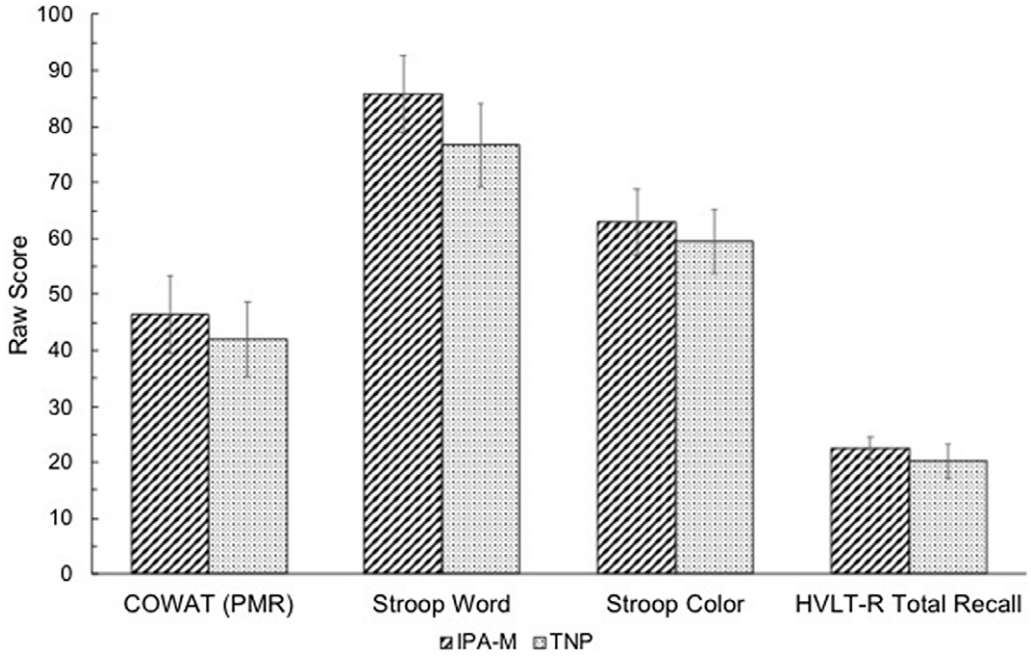

Figure 1. Statistically significant performance-level differences between the mean in-person assessment scores (IPA-M) and the teleneuropsychological scores (TNP) in the HIV + group.

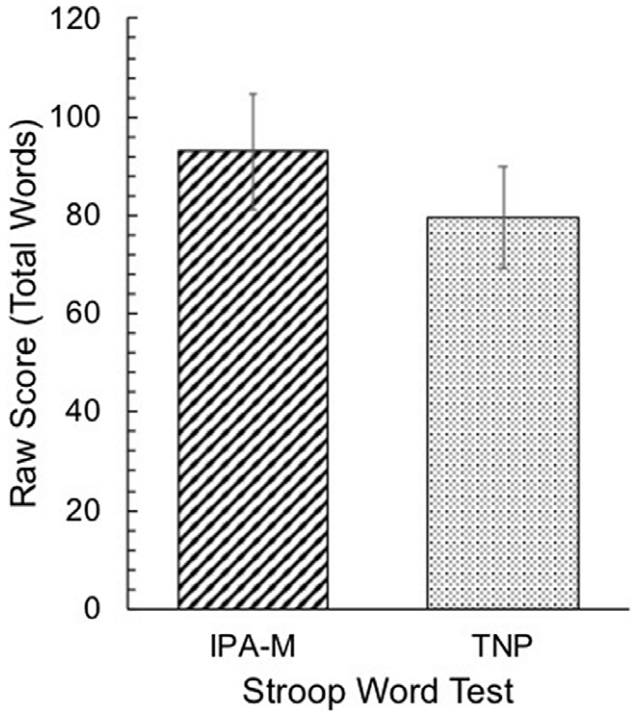

Figure 2. Statistically significant performance-level differences between the mean in-person assessment scores (IPA-M) and the teleneuropsychological scores (TNP) in the HIV− group.

and 24% of participants went from an unimpaired score at IPA2 to an impaired score at TNP. To the degree that this difference is constant across individuals, and unrelated to their last in-person score, it may be possible to create and apply a time constant correction for this neuropsychological test.

Results of an exploratory analysis revealed no significant association between device used at TNP and change in performance from IPA2 to TNP, and comparability between TNP raw scores across device types, which may be suspected to account for differences on the Stroop Color and Word Tests. Despite these findings, several challenges remain in digital neuropsychology with regard to device type and characteristics including (1) variability in the perceptual, motor, and cognitive abilities needed for response behaviors; (2) variability in hardware and software between devices that may affect stimulus presentation and response latency; and (3) rapid changes in hardware, software, and device ownership which may affect tests and test norms (Germine et al., 2019). These challenges are being rapidly investigated. Germine et al. (2019) found significant differences in response behavior on a digital trail making test with examinees (aged 18–35) taking less time to complete the task on an iPad compared to a personal computer, and more time on an iPhone. Thus, the question around device type and screen size warrants reevaluation with a greater sample size and power to detect meaningful differences in the future, which could inform TNP guidelines.

Among PWH, there were significant and strong correlations between IPA and TNP evaluations. Performance-level differences between IPA and TNP were minimal and consistent with findings from the total sample. Results suggest that TNP is a reliable alternative to IPA especially in the COVID-19 pandemic, but also more broadly when considering the health burden faced by this vulnerable population (Mirzaei et al., 2020). PWH are living longer and are more susceptible to age related neurodegenerative diseases and functional decline (Blackstone et al., 2012; Heaton et al., 2011; Wing, 2016). Neurocognitive and functional decline may limit the feasibility of attending in-person neuropsychological evaluations in aging PWH (Hearps et al., 2016). During the current COVID-19 pandemic, and especially among persons who are not vaccinated, PWH may be more fearful of going into the clinic for care or even COVID-19 testing services than the general population because of chronic immune impairment (Cooper et al., 2020; Fusco et al., 2020; Mirzaei et al., 2020). In addition to minimizing risk of additional infections, TNP evaluations have the

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

potential to better maintain consistent access to care for PWH and provide benefits such as decreased time commitment, transportation expenses, and increase overall convenience (Gajarawala & Pelkowski, 2021; Moffatt & Eley, 2010; Speedie et al., 2008).

This study is unique from other studies examining TNP in that evaluations were typically completed at home rather than in a clinical space. Although clinical spaces provide certain testing environment standardizations (e.g., adequate internet connectivity, distraction-free), evidence suggests that cognitive performance in a naturalistic environment may be more aligned with actual cognitive functioning compared to in a clinic setting (Bloem et al., 2020; Moore et al., 2021; Rentz, 2016). It is important to note that TNP may not be feasible for all PWH, particularly those of the most vulnerable backgrounds (e.g., those experiencing homelessness, lower socioeconomic status, less acculturated, limited access to technology) as it introduces other potential barriers like resources to ascertain the necessary technology, need for reliable internet access, and security concerns (Bilder et al., 2020; Mgbako et al., 2020). Although the HNRP does not currently provide technological devices (e.g., tablet) to participants for TNP testing, participants are provided the option of using a testing room at the HNRP for their evaluation.

There are several remaining questions that could not be addressed in this study: (1) Can published normative standards available for tests administered by IPA be used for those administered by TNP? Researchers may need to develop new normative data based on this modality or create adjustments for any differences derived from the TNP testing modality. (2) Can results of the same person using these two methods be compared to measure change or neurocognitive decline? Creating regression-based change scores would require a different set of visits than we used in the current study but can be helpful in determining significant change in neurocognitive performance. (3) Can results for different people in research studies be combined if some were administered the tests with IPA and some with TNP?

The current study is not without limitations. Our sample of PWH was relatively healthy which may have increased likelihood for test-retest reliability (Heaton et al., 2015). Due to the remote administration, motor tests (e.g., Grooved Pegboard) could not be administered. This may limit sensitivity in assessing some domains commonly impaired among PWH. Next, examiners did not explicitly ask participants not to write down information during the TNP testing. Although we detected significant differences between IPA-M and TNP on four neuropsychological assessments, there could be potential unmeasured confounders (e.g., distractibility, participant screen clarity, audio glitches) that may account for the differences. Audio glitches or disruptions when both the participant and examiner are speaking at the same time could have caused interference. While HNRP staff outline best practices for TNP evaluations, the lack of control over standardized testing environments in the TNP is a notable limitation. Furthermore, examinees were not asked to silence their device notifications during testing. While examiners noted this as an interruption to testing, we cannot fully rule out the potential impact of this distraction on test results. Results suggest that interruptions to testing were associated with worse performance on the HVLT-R Total Recall, despite no specific information about when the interruption occurred during testing. Thus, it may be beneficial to ask examinees to disable their device notifications during testing and for examiners to include standardized information about interruptions during testing. Since examiners do not directly assess substance use at the TNP and we were unable to conduct a urine toxicology test, we cannot confirm that examines were toxicology negative. Considering there was small sample of Spanish-speaking PWH that were administered the COWAT (PMR), results of significant differences between in-person and TNP may be interpreted with caution. Furthermore, there are other aspects of the Spanish-speaking sample that may make their neuropsychological evaluation particularly complex including socio-demographic, cultural, linguistic factors, and familiarity with telehealth and neuropsychological testing (Marquine, Rivera Mindt, et al., 2021). Future research may need to analyze the feasibility among Spanish-speaking PWH separately. The current study represents a secondary analysis of data from each participant's neuropsychological evaluations, not a randomized control trial to validate TNP assessment. The number of HNRP participants with a completed TNP evaluation is rapidly growing; therefore, follow-up analyses with a larger sample size and likely more statistical power will be beneficial. Future research may also investigate the role of emotional factors (e.g., depression, anxiety, financial instability, stress) on TNP performance during the COVID-19 pandemic, as well as whether performance on TNP assessments may adequately discriminate between impairment classifications.

Among our sample of PWH and HIV−, we provided evidence of test-retest reliability and performance-level comparability of our IPA and TNP. Considering the current COVID-19 pandemic, and possible additional pandemics in the future, there is an immediate need for reliable neuropsychological assessments that can be administered remotely, especially for PWH. TNP evaluation shows promise to improve access to neuropsychological services and maintain ongoing clinical research studies.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1355617722000777

Acknowledgements. This work was supported by NIMH-funded research programs including the HIV Neurobehavioral Research Center (HNRC), supported by award P30MH062512, the California NeuroAIDS Tissue Network (CNTN), supported by award U24MH100928, and the Multi-Dimensional Successful Aging Among HIV-Infected Adults, supported by award R01MH099987. Support for this study also includes a NIDA-funded research program that includes the Translational Methamphetamine AIDS Research Center (TMARC), supported by award P50DA026306; an NIH-funded research program that includes the CNS HIV Anti-Retroviral Therapy Effects Research (CHARTER), supported by award HHSN271201000036C; and the NIMHD-funded study Mechanisms of Disparities in Adverse Neurocognitive Outcomes among Hispanics Aging with HIV, supported by award R01MD013502. Stipend support to MK is funded by NIAAA award T32AA013525 and NIA award F31AG074838. Stipend support to NS is funded by NIDA award T32DA031098. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The San Diego HIV Neurobehavioral Research Center [HNRC] group is affiliated with the University of California, San Diego, the Naval Hospital, San Diego, and the Veterans Affairs San Diego Healthcare System, and includes: Director: Robert K. Heaton, Ph.D., Co-Director: Igor Grant, M.D.; Associate Directors: Ronald J. Ellis, M.D., Ph.D., and Scott Letendre, M.D.; Center Manager: Jennifer Iudicello, Ph.D.; Donald Franklin, Jr.; Melanie Sherman, Cheryl Kelley; NeuroMedical Core: Ronald J. Ellis, M.D., Ph.D. (Director/NeuroMedical Unit Head), Scott Letendre, M.D. (Co-I./Laboratory Unit Head), Christine Fennema-Notestine, Ph.D., (Co-I./Neuroimaging Unit Head); Debra Rosario, M.P.H., Neurobehavioral & Psychiatry Core: David J. Moore, Ph.D. (Co-Director/Neurobehavioral Unit Head), Murray B. Stein, M.D. (Co-Director/Psychiatry Unit Head), Erin E. Morgan, Ph.D. (Co-I./

Psychiatric Coordinator), Andrew H. Miller, Matthew Dawson, NeuroVirology & Biology Core: Sara Gianella Weibel, M.D. (Co-Director/NeuroVirology Unit Head), Sarah A. LaMere, D.V.M., Ph.D. (Associate Unit Head, NeuroVirology Unit), Cristian Achim, M.D., Ph.D. (Co-Director/Neurobiology Unit Head), Ana Sanchez, Ph.D. (Co-I./Neurobiology Unit), Adam Fields, Ph.D.(Associate Unit Head, Neurobiology Unit); Microbiome Core: Rob Knight, Ph.D. (Co-Director), Pieter Dorrestein, Ph.D. (Co-Director); Developmental Core: Scott Letendre, M.D. (Director), Ajay Bharti, M.D. (Co-I.), J. Allen McCutchan, M.D., Christine Fennema-Notestine, Ph.D.; Administrative Core: Robert K. Heaton, Ph.D. (Director/Coordinating Unit Head), Participant Accrual and Retention Unit: J. Hampton Atkinson, M.D. (Unit Head), Jennifer Marquie-Beck, M.P.H.; Data Management and Information Systems Unit: Ian Abramson, Ph.D. (Unit Head), Clint Cushman; Statistics Unit: Florin Vaida, Ph.D. (Unit Head), Anya Umlauf, M.S., Bin Tang, Ph.D.

Conflict of interest. The author(s) have nothing to disclose.

References

Antinori, A., Arendt, G., Becker, J., Brew, B., Byrd, D., Cherner, M., Clifford, D. B., Cinque, Q., Epstein, L. G., Goodkin, K., Gisslen, M., Grant, I., Heaton, R. K., Joseph, J., Marder, K., Marra, C. M., McArthur, J. C., Nunn, M., Price, R. W., . . . Wojna, V. E. (2007). Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*, *69*(18), 1789–1799.

Barcellos, L. F., Horton, M., Shao, X., Bellesis, K. H., Chinn, T., Waubant, E., & Schaefer, C. (2021). A validation study for remote testing of cognitive function in multiple sclerosis. *Multiple Sclerosis Journal*, *27*(5), 795–798.

Becker, J. T., Caldararo, R., Lopez, O. L., Dew, M. A., Dorst, S. K., & Banks, G. (1995). Qualitative features of the memory deficit associated with HIV infection and AIDS: cross-validation of a discriminant function classification scheme. *Journal of Clinical and Experimental Neuropsychology*, *17*(1), 134–142. https://doi.org/10.1080/13803399508406588

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins verbal learning test – revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist*, *12*(1), 43–55. https://doi.org/10.1076/clin.12.1.43.1726

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., Harder, L., Kanter, G., Lanca, M., Lechuga, D. M., Morgan, J. M., Most, R., Puente, A. E., Salinas, C. M., & Woodhouse, J. (2020). Inter organizational practice committee recommendations/Guidance for teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology*, *35*(6), 647–659.

Blackstone, K., Moore, D., Heaton, R., Franklin, D. R. Jr., Woods, S. P., Clifford, D. B., Collier, A. C., Marra, C. M., Gelman, B. B., McArthur, J. C., Morgello, S., Simpson, D. M., Rivera-Mindt, M., Deutsch, R., Ellis, R. J., Hampton Atkinson, J., & Grant, I. (2012). Diagnosing symptomatic HIV-associated neurocognitive disorders: Self-report versus performance-based assessment of everyday functioning. *Journal of the International Neuropsychological Society*, *18*(1), 79–88.

Bloem, B. R., Dorsey, E. R., & Okun, M. S. (2020). The coronavirus disease 2019 crisis as catalyst for telemedicine for chronic neurological disorders. *JAMA Neurology*, *77*(8), 927–928.

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*(2), 135–140. https://doi.org/10.1016/0028-3932(67)90015-2

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review*, *27*(2), 174–186. https://doi.org/10.1007/s11065-017-9349-1

Carey, C. L., Woods, S. P., Gonzalez, R., Conover, E., Marcotte, T. D., Grant, I., & Heaton, R. K. (2004). Predictive validity of global deficit scores in detecting neuropsychological impairment in HIV infection. *Journal of Clinical and Experimental Neuropsychology*, *26*(3), 307–319. https://doi.org/10.1080/13803390490510031

Chapman, J. E., Gardner, B., Ponsford, J., Cadilhac, D. A., & Stolwyk, R. J. (2020). Comparing performance across in-person and videoconference-based administrations of common neuropsychological measures in community-based survivors of stroke. *Journal of the International Neuropsychological Society*, *27*(7), 697–710.

Cherner, M., Marquine, M. J., Umlauf, A., Morlett Paredes, A., Rivera Mindt, M., Suárez, P., Yassai-Gonzalez, D., Kamalyan, L., Scott, T., Heaton, A., Diaz-Santos, M., Gooding, A., Artiola i Fortuny, L., & Heaton, R. K. (2021). Neuropsychological norms for the US-Mexico border region in Spanish (NP-NUMBRS) project: Methodology and sample characteristics. *The Clinical Neuropsychologist*, *35*(2), 253–268.

Cooper, T., Woodward, B., Alom, S., & Harky, A. (2020). Coronavirus disease 2019 (COVID-19) outcomes in HIV/AIDS patients: A systematic review. *HIV Medicine*, *21*(9), 567–577.

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

Díaz-Santos, M., Suárez, P. A., Marquine, M. J., Umlauf, A., Rivera Mindt, M., Artiola i Fortuny, L., Heaton, R. K., & Cherner, M. (2021). Updated demographically adjusted norms for the Brief Visuospatial Memory Test-revised and Hopkins Verbal Learning Test-revised in Spanish-speakers from the U.S.-Mexico border region: The NP-NUMBRS project. *Clinical Neuropsychologist*, *35*(2), 374–395. https://doi.org/10.1080/13854046.2020.1861329

Diehr, M. C., Heaton, R. K., Miller, W., & Grant, I. (1998). The Paced Auditory Serial Addition Task (PASAT): Norms for age, education, and ethnicity. *Assessment*, *5*(4), 375–387. https://doi.org/10.1177/107319119800500407

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, *5*(4), 346–356.

Fusco, F. M., Sangiovanni, V., Tiberio, C., Papa, N., Atripaldi, L., & Esposito, V. (2020). Persons living with HIV may be reluctant to access to COVID-19 testing services: Data from 'D. Cotugno' Hospital, Naples, Southern Italy. *Aids*, *34*(14), 2151–2152.

Gajarawala, S. N., & Pelkowski, J. N. (2021). Telehealth benefits and barriers. *The Journal for Nurse Practitioners*, *17*(2), 218–221.

Gardner, M. M., Aslanzadeh, F. J., Zarrella, G. V., Braun, S. E., Loughan, A. R., & Parsons, M. W. (2021). Cancer, cognition, and COVID: Delivering direct-to-home tele-neuropsychology services to neuro-oncology patients. *Neuro-Oncology Practice*, *8*(4), 485–496.

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *Clinical Neuropsychologist*, *33*(2), 271–286. https://doi.org/10.1080/13854046.2018.1535662

Gooding, A., Seider, T., Marquine, M., Suarez, P., Umlauf, A., Rivera Mindt, M., & Cherner, M. (2021). Demographically-adjusted norms for the paced auditory serial addition test and letter number sequencing test in Spanish-speaking adults: Results from the neuropsychological norms for the U.S.-Mexico border region in Spanish (NP-NUMBRS) Project. *Clinical Neuropsychologist*, *35*(2), 324–338. https://doi.org/10.1080/13854046.2019.1711199

Hampshire, A., Trender, W., Chamberlain, S. R., Jolly, A. E., Grant, J. E., Patrick, F., & Mehta, M. A. (2021). Cognitive deficits in people who have recovered from COVID-19. *EClinicalMedicine*, *39*, 101044. https://doi.org/10.1016/j.eclinm.2021.101044

Harder, L., Hernandez, A., Hague, C., Neumann, J., McCreary, M., Cullum, C. M., & Greenberg, B. (2020). Home-Based pediatric teleneuropsychology: A validation study. *Archives of Clinical Neuropsychology*, *35*(8), 1266–1275. https://doi.org/10.1093/arclin/acaa070

Hearps, A., Schafer, K., High, K., & Landay, A. (2016). HIV and aging: parallels and synergistic mechanisms leading to premature disease and functional decline. In *Advances in geroscience* (pp. 509–550). Springer.

Heaton, R. K., Clifford, D. B., Franklin, D. R., Woods, S. P., Ake, C., Vaida, F., Ellis, R. J., Letendre, S. L., Marcotte, T. D., Atkinson, J. H., Rivera-Mindt, M.,

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

Vigil, O. R., Taylor, M. J., Collier, A. C., Marra, C. M., Gelman, B. B., McArthur, J. C., Morgello, S., Simpson, D. M., ... Grant, I. (2010). HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER Study. *Neurology*, *75*(23), 2087–2096. https://doi.org/10.1212/WNL.0b013e318200d727

Heaton, R., Miller, S., Taylor, M., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Psychological Assessment Resources.

Heaton, R., Taylor, M., & Manly, J. (2003). Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III. In *Clinical interpretation of the WAIS-III and WMS-III* (pp. 181–210). Elsevier.

Heaton, R. K., Franklin, D. R., Ellis, R. J., McCutchan, J. A., Letendre, S. L., LeBlanc, S., Corkran, S. H., Duarte, N. A., Clifford, D. B., Woods, S. P., Collier, A. C., Marra, C. M., Morgello, S., Mindt, M. R., Taylor, M. J., Marcotte, T. D., Atkinson, J. H., Wolfson, T., Gelman, B. B., ... Grant, I. (2011). HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: differences in rates, nature, and predictors. *Journal of NeuroVirology*, *17*(1), 3–16. https://doi.org/10.1007/s13365-010-0006-1

Heaton, R. K., Franklin, D. R., Deutsch, R., Letendre, S., Ellis, R. J., Casaletto, K., Marquine, M. J., Woods, S. P., Vaida, F., Atkinson, J. H., Marcotte, T. D., McCutchan, J. A., Collier, A. C., Marra, C. M., Clifford, D. B., Gelman, B. B., Sacktor, N., Morgello, S., Simpson, D. M., ... Teshome, M. (2015). Neurocognitive change in the era of HIV combination antiretroviral therapy: The longitudinal CHARTER study. *Clinical Infectious Diseases*, *60*(3), 473–480. https://doi.org/10.1093/cid/ciu862

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, *58*(1), 78–79.

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Lea & Febiger.

Lillicrap, L., Hunter, C., & Goldswain, P. (2019). Improving geriatric care and reducing hospitalisations in regional and remote areas: The benefits of telehealth. *Journal of Telemedicine and Telecare*, *27*(7), 397–408. https://doi.org/10.1177/1357633X19881588

Marquine, M. J., Morlett Paredes, A., Madriaga, C., Blumstein, Y., Umlauf, A., Kamalyan, L., Rivera Mindt, M., Suarez, P., Artiola i Fortuni, L., Heaton, R. K., & Cherner, M. (2021). Demographically-adjusted norms for selected tests of verbal fluency: Results from the neuropsychological norms for the US-Mexico border region in Spanish (NP-NUMBRS) project. *Clinical Neuropsychologist*, *35*(2), 269–292. https://doi.org/10.1080/13854046.2020.1762931

Marquine, M. J., Rivera Mindt, M., Umlauf, A., Suarez, P., Kamalyan, L., Morlett Paredes, A., Yassai-Gonzalez, D., Scott, T. M., Heaton, A., Diaz-Santos, M., Gooding, A., Artiola i Fortuny, L., Heaton, R. K., & Cherner, M. (2021). Introduction to the neuropsychological norms for the US-Mexico border region in Spanish (NP-NUMBRS) project. *The Clinical Neuropsychologist*, *35*(2), 227–235. https://doi.org/10.1080/13854046.2020.1751882

Marra, D. E., Hamlet, K. M., Bauer, R. M., & Bowers, D. (2020). Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. *The Clinical Neuropsychologist*, *34*(7-8), 1411–1452.

Matchanova, A., Babicz, M. A., Medina, L. D., Rahman, S., Johnson, B., Thompson, J. L., Beltran-Najera, I., Brooks, J., Sullivan, K. L., Walker, R. L., Podell, K., & Woods, S. P. (2020). Latent structure of a brief clinical battery of neuropsychological tests administered in-home via telephone. *Archives of Clinical Neuropsychology*, *36*(6), 874–886.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1981). *Profile of mood states*. Educational and Industrial Testing Service.

Mgbako, O., Miller, E. H., Santoro, A. F., Remien, R. H., Shalev, N., Olender, S., Gordon, P., & Sobieszczyk, M. E. (2020). COVID-19, telemedicine, and patient empowerment in HIV care and research. *AIDS and Behavior*, *24*(7), 1990–1993.

Mirzaei, H., McFarland, W., Karamouzian, M., & Sharifi, H. (2020). COVID-19 among people living with HIV: A systematic review. *AIDS and Behavior*, *25*(1), 85–92.

Moffatt, J. J., & Eley, D. S. (2010). The reported benefits of telehealth for rural Australians. *Australian Health Review*, *34*(3), 276–281.

Moore, R. C., Paolillo, E. W., Sundermann, E. E., Campbell, L. M., Delgadillo, J., Heaton, A., Swendsen, J., & Depp, C. A. (2021). Validation of the mobile verbal learning test: Illustration of its use for age and disease-related cognitive deficits. *International Journal of Methods in Psychiatric Research*, *30*(1), e1859.

Mungas, D., Reed, B. R., Crane, P. K., Haan, M. N., & Gonzalez, H. (2004). Spanish and English Neuropsychological Assessment Scales (SENAS): Further development and psychometric characteristics. *Psychological Assessment*, *16*(4), 347–359. https://doi.org/10.1037/1040-3590.16.4.347

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.

Parsons, M. W., Gardner, M. M., Sherman, J. C., Pasquariello, K., Grieco, J. A., Kay, C. D., Pollak, L. E., Morgan, A. K., Carlson-Emerton, B., Seligsohn, K., Davidsdottir, S., Pulsifer, M. B., Zarrella, G. V., Burstein, S. M., & Mancuso, S. M. (2021). Feasibility and Acceptance of direct-to-home tele-neuropsychology services during the COVID-19 pandemic. *Journal of the International Neuropsychological Society*, *28*(2), 210–215.

Passell, E., Strong, R. W., Rutter, L. A., Kim, H., Scheuer, L., Martini, P., Grinspoon, L., & Germine, L. (2021). Cognitive test scores vary with choice of personal digital device. *Behavior Research Methods*, *53*(6), 2544–2557. https://doi.org/10.3758/s13428-021-01597-3

Rentz, D. M. (2016). Validating use of technology for cognitive test assessment. *EBioMedicine*, *11*, 23–24. https://doi.org/10.1016/j.ebiom.2016.08.002

Rivera Mindt, M., Marquine, M. J., Aghvinian, M., Scott, T. M., Cherner, M., Morlett Paredes, A., Taylor, M. J., Umlauf, A., Suarez, P., Diaz-Santos, M., Kamalyan, L., Heaton, A., Artiola i Fortuny, L., & Heaton, R. K. (2021). Demographically-adjusted norms for the processing speed subtests of the WAIS-III in a Spanish-speaking adult population: Results from the neuropsychological norms for the U.S.-Mexico border region in Spanish (NP-NUMBRS) project. *Clinical Neuropsychologist*, *35*(2), 293–307. https://doi.org/10.1080/13854046.2020.1723707

Shore, J. H. (2013). Telepsychiatry: Videoconferencing in the delivery of psychiatric care. *American Journal of Psychiatry*, *170*(3), 256–262. https://doi.org/10.1176/appi.ajp.2012.12081064

Speedie, S. M., Ferguson, A. S., Sanders, J., & Doarn, C. R. (2008). Telehealth: The promise of new care delivery models. *Telemedicine and e-Health*, *14*(9), 964–967.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Suarez-Gonzalez, A., Rajagopalan, J., Livingston, G., & Alladi, S. (2021). The effect of COVID-19 isolation measures on the cognition and mental health of people living with dementia: A rapid systematic review of one year of quantitative evidence. *EClinicalMedicine*, *39*, 101047. https://doi.org/10.1016/j.eclinm.2021.101047

Tailby, C., Collins, A. J., Vaughan, D. N., Abbott, D. F., O'Shea, M., Helmstaedter, C., & Jackson, G. D. (2020). Teleneuropsychology in the time of COVID-19: The experience of The Australian Epilepsy Project. *Seizure*, *83*, 89–97. https://doi.org/10.1016/j.seizure.2020.10.005

Taylor, M. J., & Heaton, R. K. (2001). Sensitivity and specificity of WAIS-III/WMS-III demographically corrected factor scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *7*(7), 867–874.

Wadsworth, H. E., Dhima, K., Womack, K. B., Hart, J., Jr., Weiner, M. F., Hynan, L. S., & Cullum, C. M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology*, *33*(8), 1040–1045. https://doi.org/10.1093/arclin/acx140

Wing, E. J. (2016). HIV and aging. *International Journal of Infectious Diseases*, *53*, 61–68. https://doi.org/10.1016/j.ijid.2016.10.004

Woods, S. P., Scott, J. C., Sires, D. A., Grant, I., Heaton, R. K., & Tröster, A. I. (2005). Action (verb) fluency: Test-retest reliability, normative standards,

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press

and construct validity. *Journal of the International Neuropsychological Society*, *11*(4), 408–415.

World Health Organization (1997). Composite International Diagnostic Interview (CIDI, version 2.1), Geneva.

Wosik, J., Fudim, M., Cameron, B., Gellad, Z. F., Cho, A., Phinney, D., Curtis, S., Roman, M., Poon, E. G., Ferranti, J., Katz, J. N., & Tcheng, J. (2020). Telehealth transformation: COVID-19 and the rise of virtual care. *Journal of the American Medical Informatics Association*, *27*(6), 957–962. https://doi.org/10.1093/jamia/ocaa067

Zeghari, R., Guerchouche, R., Tran-Duc, M., Bremond, F., Langel, K., Ramakers, I., Amiel, N., Lemoine, M. P., Bultingaire, V., Manera, V., Robert, P., & Konig, A. (2022). Feasibility study of an internet-based platform for tele-neuropsychological assessment of elderly in remote areas. *Diagnostics (Basel)*, *12*(4), 925. https://doi.org/10.3390/diagnostics12040925

https://doi.org/10.1017/S1355617722000777 Published online by Cambridge University Press