






Reliability, minimum detectable change and sociodemographic biases of selected neuropsychological tests among people living with HIV in south-eastern Nigeria



Authors:

Martins C. Nweke¹ 
 Nalini Govender² 
 Aderonke Akinpelu³ 
 Adesola Ogunniyi⁴ 
 Nombeko Mshunqane¹ 

Affiliations:

¹Department of Physiotherapy,
University of Pretoria, Pretoria,
South Africa

²Department of Basic Medical
Sciences, Durban University
of Technology, Durban,
South Africa

³Department of
Physiotherapy, University of
Ibadan, Ibadan, Nigeria

⁴Department of Medicine,
University of Ibadan,
Ibadan, Nigeria

Corresponding author:

Martins Nweke,
martins.nweke@gmail.com

Dates:

Received: 27 Aug. 2021

Accepted: 04 Jan. 2022

Published: 28 Apr. 2022

How to cite this article:

Nweke, M.C., Govender, N.,
Akinpelu, A., Ogunniyi, A., &
Mshunqane, N. (2022).

Reliability, minimum
detectable change and
sociodemographic biases of
selected neuropsychological
tests among people living
with HIV in south-eastern
Nigeria. *African Journal of
Psychological Assessment*,
4(0), a84. [https://doi.
org/10.4102/ajopa.v4i0.84](https://doi.org/10.4102/ajopa.v4i0.84)

Read online:



Scan this QR
code with your
smart phone or
mobile device
to read online.

Verification of the psychometric properties of neuropsychological (NP) tests in each society of people living with HIV (PLWHIV) will facilitate accurate classification of HIV-associated neurocognitive disorder. This study aimed to determine the reliability, minimum detectable change (MDC) and sociodemographic biases of selected NP tests among PLWHIV. The study took place at the HIV clinic of the University of Nigeria Teaching Hospital, Enugu. A total of 60 PLWHIV were randomly recruited into two groups of 30 each. The first group was evaluated by two independent examiners (inter-rater) and the other by a single rater (intra-rater). The NP tests utilised included the Hopkins Verbal Learning Test-Revised (HVLT-R), controlled oral word association test (COWAT), Trail Making Test-A (TMT-A) and -B (TMT-B), Digit Span Test-forward (DST-f) and -backward (DST-b). We examined agreement using intra-class correlation (ICC), standard error of measurement and MDC. We verified the influence of sociodemographic variables on test performance using Man-Whitney U-test and Kruskal-Wallis test. The HVLT-R- delay recall (DR), TMT-A, TMT-B and COWAT showed excellent inter-rater reliability with ICC values of 0.83, 0.86, 0.78 and 0.89, respectively. The HVLT-R- verbal learning (VL), DST-f and DST-b showed moderate inter-rater reliability with ICCs of 0.4.99, 0.52 and 0.60, respectively. The HVLT-R-DR, TMT-A, DST-b and COWAT showed excellent intra-rater reliability, with ICCs values of 0.76, 0.80, 0.84 and 0.97, respectively. The TMT-A, DST-f and DST-b were free from sociodemographic bias. The HVLT-R-DR, TMT-A, TMT-B, DST-f, DST-b and COWAT are reliable candidate NP tests for PLWHIV in our setting.

Keywords: HIV; neurocognitive disorder; neuropsychological assessment; reliability; Nigeria.

Introduction

Human immunodeficiency virus (HIV)-associated neurocognitive disorder (HAND) remains prevalent in the sub-Saharan part of Africa, especially in Nigeria, which has the world's second-highest HIV and AIDS incidence (Awofala & Ogundele, 2018; Nweke, Akinpelu, & Ezema, 2019; Yakasai et al., 2015). HIV targets the central nervous system (CNS) and compromises the blood-brain barrier, causing HIV-infected microglia to infiltrate the CNS and secrete neurotoxic viral proteins such as Tat and proinflammatory cytokines, as well as disrupting neurogenesis, dysregulation of CD4+ T cells and damaging synaptodendritic networks (Cody & Vance, 2016). These occurrences result in damage to specific brain structures and neural circuits, increasing the brain's predisposition for acquiring subsequent neuropsychological (NP) diseases, such as neurocognitive disorder (Watkins & Treisman, 2012). The HAND is a spectrum of disorders comprising HIV-associated dementia (HAD), mild neurocognitive disorder and asymptomatic neurocognitive impairment (Antinori et al., 2007). Memory, learning information processing speed, executive function, attention and/or concentration and verbal fluency are the most frequently impaired cognitive abilities in people living with HIV (PLWHIV) (Yakasai et al., 2015). The widespread use of antiretroviral therapy (ART) has reduced the burden of HAD, notwithstanding, the prevalence of mild but limiting phenotypes of HAND remains high (Yakasai et al., 2015). Therefore, proper screening tools are necessary to ensure effective evaluation and prompt initiation of treatment for HAND. Until now, in sub-Saharan Africa, brief measures such as the HIV Dementia Scale and the International HIV Dementia Scale, Montreal Cognitive

Copyright: © 2022. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Assessment Test, NeuroScreen and CogState Brief Battery dominate neurocognitive screening among PLWHIV (Mwangala, Newton, Abas, & Abubakar, 2019). Their simplicity of administration is fascinating but they are characterised with diagnostic weakness such as an inability to detect asymptomatic or mild cognitive impairment that makes them unacceptable when pursuing a definitive diagnosis of HAND (Singh et al., 2010). Towards the diagnosis of HAND, the comprehensive NP battery tests are the tools of choice (Robertson, Liner, & Heaton, 2009). They are the most useful instruments for identifying and classifying the impact of HIV or AIDS vis-a-vis the CNS (Robertson et al., 2009).

In a resource-limited environment where sophisticated laboratory and neuroimaging techniques are not accessible, the application of NP screening to characterise neurocognitive functioning among PLWHIV is essential to successful diagnosis and treatment of HAND (Robertson et al., 2009). Before now, the utility of the NP battery tests in SSA was narrow because they require instrumentations that are not common in the region. Also, their administration demands previous clinical or research experience among people living with HAND (Nweke, Mshunqane, Govender, & Akinpelu, 2021; Singh et al., 2010). In Nigeria, the use of NP tests in the diagnosis of HAND is gaining momentum in recent times (Jumar et al., 2017; Robertson et al., 2016). The NP tests are the gold standard for the diagnosis of HAND but they are cumbersome and culture-specific (Fernandez & Marcopulos, 2008) and this makes up one of the well-known challenges faced by neuro-HIV researchers or clinicians (Yeatesh & Taylor, 2011). They often require that two or more raters administer them to participants to reduce their burden (Andrews, Janzen, & Saklofske, 2001). This is especially true of clinical trials or normative studies where many participants undergo screening following the diagnostic criteria stipulated by the American Academy of Neurology's research nomenclature and diagnostic method (Antinori et al., 2007). Attrition is a common observation in clinical trials and results from premature discontinuation, missed or flawed assessments (Sridhara, Mandrekar, & Dodd, 2013). The cumbersomeness of NP testing may put the attrition rate above acceptable attrition margin (20%) (Nunan, Aronson, & Bankhead, 2018), thus undermining the feasibility of a clinical trial or observational study requiring large-scale testing. In the research setting where the same assessor or other assessors may follow up treatment outcomes, there is a need to examine the reliability of commonly used NP tests (McHugh, 2012). The significance of rater reliability stems from the fact that it indicates the degree to which the data obtained are accurate representations of the variables being examined (McHugh, 2012). In clinical research, reliability is most commonly verified using intra-class correlation (ICC) (Lee et al., 2013). However, the use of ICC in the estimation of reliability presents a weakness as it is a measure of relative reliability, thus underscoring the necessity of the estimation of minimum detectable change (MDC) (Lee et al., 2013). The MDC is a precise measure of measurement error and consistency (Lee et al., 2013). Based on the data available to us, no study has examined the psychometric properties of NP tests in the Nigerian context. Hence, this

study aimed at examining the reliability and MDC of selected NP tests and the effects of demographic characteristics.

Materials and methods

Participants

The study was a cross-sectional study conducted among HIV-positive adults aged between 18 and 64 years receiving ART at the HIV clinic of the University of Nigeria Teaching Hospital, Ituku-Ozalla, Enugu, Nigeria. The study took place between June and August 2020. The study constitutes one of the preliminary steps undertaken in pursuit of a clinical trial investigating the 'effects of an aerobic exercise programme on neurocognitive disorder in Nigeria'. Owing to the elaborateness of the NP evaluations, the study took place in two batches using different study samples to save patients' waiting time as many of the participants visited the clinic from distant locations while reducing contact time with patients, in the wake of coronavirus disease 2019 (COVID-19). A simple random sampling, balloting (3 'YES':1 'NO') was used to select study participants. Those who picked 'YES' from the ballot box were recruited into the study and consecutively assigned to raters who carried out the independent evaluation. The first batch involved a sample of 30 HIV-positive adults recruited to examine the inter-rater reliability between a neurological physiotherapist and a clinical psychologist, while the second batch, which involved a sample of 30 HIV-positive adults examined the intra-rater reliability.

The study population was PLWHIV on ART. In line with Viechtbauera et al. (2015), we estimated the sample size for this pilot study using a pilot sample size calculator available at <http://www.pilotsamplesize.com>. Using a HAND prevalence rate of 21.5% (Yusuf et al., 2014), an acceptable level of withdrawal or an incomplete assessment rate of 10%, we required a minimum sample size of 28 to detect the problem with 95% confidence. The inclusion criteria for the study include being HIV positive, adult between 18 and 65 years, being on ART for at least three months, formal (primary six) education with an ability to use English and capacity for consent. For the inter-rater examination, we used the consecutive assessment to assign participants to the raters. We excluded the following participants: individuals above 65 years of age, smokers, alcohol-dependents, substance abusers, individuals with cardiorespiratory disease (heart attack, asthma or chronic obstructive pulmonary disease [COPD]) disease; history of focal neurological deficit, traumatic brain injury with history of loss of consciousness, stroke, psychiatric illness including depression, opportunistic infection such as TB, candidiasis, hepatitis or record blood pressure over 140/90 mmHg, structured physical activity or cognitive enhancing medication were all excluded. Exclusion criteria were determined through the use of self-reports obtained during recruitment. Blood pressure, depression, alcohol abuse and substance abuse were measured using appropriate instruments.

Instruments

Neurocognitive evaluations were undertaken with the aid of NP tests. The NP tests constitute the gold standard instrument in screening and diagnosis of HAND (Yakasai et al., 2015). The

Frascati criteria stipulate that NP evaluation for PLWHIV should cover at least five ability domains commonly impaired among PLWHIV including verbal learning (VL), memory, working memory, attention, abstraction and/or executive function, information processing speed and verbal fluency (Antinori et al., 2007). In this study, we examined four NP tests covering seven ability domains. They include the Hopkins Verbal Learning Test-Revised (HVLTR), with test-retest reliability indices of 0.537–0.818 (O'Neil-Pirozzi, Goldstein, Strangman, & Glenn, 2012), Trail Making Test (TMT), with ICC of 0.7–0.98 (Salthouse, 2011), Digit Span Test (DST), with a test-retest reliability index of 0.7–0.78 (Groth-Marnat & Baker, 2003) and the Controlled oral word association Test (COWAT-F-A-S letter fluency), with an excellent inter-rater reliability index ≥ 0.9 (Ross et al., 2007). These tests have been used in African and Nigerian settings (Singh et al., 2010; Yakasai et al., 2015) and a recent clinical trial on HAND (Towe, Puja Patel, & Meade, 2017). The selection of these was mainly based on their utility in African setting and ease-of-administration.

Procedure

A clinical psychologist (evaluator 1) was trained by the lead investigator (evaluator 2), who is familiar with NP testing and an acceptable degree of agreement was established. Neuropsychological testing was carried out by the two independent evaluators in the first batch of the study, with each subject being assessed twice a day. Only evaluator 1 completed NP testing in the second batch, with each individual being assessed twice, one after the other. Data from the first evaluation by evaluator 1 and data from the second evaluation by evaluator 2 completed on the same day were utilised for the inter-rater reliability evaluation (Van Lummel et al., 2016). Data from the first and second measurements of evaluator 1 were used to assess intra-rater reliability. Every effort was made to ensure that participants who were being tested stayed away from the consulting bench to avoid contamination.

Data analysis

Exploratory statistics showed that some data sets were not normally distributed and log-transformation did not improve the distribution, hence we ignored log transformation as this could result from sociodemographic biases. Data were summarised using descriptive statistics. We used the ICC to verify the rater reliability. Specifically, inter- and intra-rater reliability was examined using the use $ICC_{2,1}$ and $ICC_{1,1}$, respectively. Both $ICC_{2,1}$ and $ICC_{1,1}$ were based on the absolute agreement in a two-way (mixed-effects) repeated-measures analysis of variance model. According to the literature, the ICC values were graded as follows: weak ($ICC < 0.40$), moderate (ICC between 0.40 and 0.75) and excellent ($ICC > 0.75$) (Sedrez et al., 2016). The following formula was used to calculate the standard error of measurement:

$$SEM = \text{standard deviation (SD)} \times (1 - [\text{square root of } ICC_{2,1}])$$

[Eqn 1]

We estimated the MDC using the formula:

$$1.96 \times [\text{Square root of } 2 \times SEM]$$

[Eqn 2]

based on a 95% confidence interval (CI). The MDC of $< 30\%$ and $< 10\%$ were deemed acceptable and excellent, respectively (Lee et al., 2013). After Bland-Altman procedures proportional bias was examined. The Mann-Whitney U test and the Kruskal-Wallis test were employed to test the putative impact of sociodemographic factors on NP test performance. The statistical package for social sciences version 22 (IBM Corp., 2012) was used, with an alpha set at 0.05.

Ethical considerations

This review was approved by the Research Ethics Committee of the Faculty of Health Sciences, University of Pretoria (Ethics Reference Number: 152/2020), which complies with the International Conference on Harmonisation-Good Clinical Practice (ICH-GCP) guidelines.

Results

A total of 60 PLWHIV (30 each for inter- and intra-rater reliability groups) with a mean age of 43.5 ± 8.2 years took part in the study. Participants in both groups were comparable concerning the sampled clinical and sociodemographic characteristics. In both groups, females were twice as many as men. Seventy per cent of the participants had secondary to tertiary education (Table 1).

Evaluation of the inter-rater reliability showed moderate to excellent reliability, with limited to an acceptable level of random measurement error. Specifically, the HVLTR-DR domain, TMT-A, TMT-B and COWAT showed excellent inter-rater reliability with ICC values of 0.83, 0.86, 0.78 and 0.89, respectively. Three tests, namely the HVLTR-VL, DST-f and DST-b showed moderate inter-rater reliability with ICCs of 0.4.99, 0.52 and 0.60, respectively. The result shows that all the tests except HVLTR possessed MDC percentage within the predefined limit of acceptance ($< 30\%$). There was no proportional bias throughout the inter-rater evaluations ($p > 0.05$) (Table 2).

Assessment of intra-rater reliability showed that all the tests except HVLTR-VL possessed moderate to excellent intra-rater reliability. Four tests, namely HVLTR-DR, TMT-A, DST-b and COWAT showed excellent intra-rater reliability, with ICC values of 0.76, 0.80, 0.84 and 0.97, respectively. The result shows that all the tests except HVLTR possessed MDC percentage within the predefined limit of acceptance ($< 30\%$). The DST-f, DST-b and COWAT had MDC percentage less than 10%. Three tests, namely the HVLTR-VL, DST-f and DST-b showed moderate intra-rater reliability with ICCs of 0.4.99, 0.52 and 0.60, respectively. Intra-rater proportional bias was observed in HVLTR-VL ($\beta = -0.475$; $p = 0.011$) and HVLTR-DR ($\beta = -0.371$; $p = 0.047$) (Table 3).

Examination of the sociodemographic biases of the NP test revealed that three of the seven tests, namely the TMT-A, DST-f

and DST-b were free from any form of sociodemographic bias, while four tests, namely the HVLTR-VL, HVLTR-DR, TMT-B and COWAT possessed at least one form of sociodemographic bias. Specifically, the inter-rater performance of the COWAT was significantly related to age ($Rho = 0.36; 0.048$). The intra-rater performance of both the HVLTR-VL ($p = 0.001$) and HVLTR-DR ($p = 0.010$) was affected by sex, with females showing better performance than their male counterparts. The inter-rater performance of the HVLTR-VL ($p = 0.046$) and the intra-rater performance of the TMT-b ($p = 0.003$) showed education bias, with individuals with tertiary and secondary education showing the better performance ($p = 0.007$) than those with primary education (Table 4).

Discussion

All the NP tests showed moderate to excellent inter-rater reliability, with acceptable levels of measurement error as determined by the MDC percentage. To facilitate the right interpretation of the results obtained during clinical or research follow-up, it is essential to note the variability

inherent in the measurement as defined by the SEM. The SEM shows that the measurement's precision varied from 0.2 to 5.3. We can consider these values clinically acceptable, showing that NP evaluation can be reliably conducted by two more raters using the selected tests. This simply proves the fact that NP assessments can be reliably conducted by two or more raters and yield similar results. By implication, NP evaluations using these tests can be undertaken by two or more raters with similar experience or training during a clinical or research follow-up. There is a paucity of literature on the inter-reliability of the selected NP tests among PLWHIV, notwithstanding they remain the gold standard instrument in the diagnosis of neurocognitive disorders among PLWHIV (Antinori et al., 2007). The finding of this study collaborates with that of Singh et al. (2010), in which an ICC of 0.89 was obtained between a psychiatrist and a psychologist. Similarly, Fals-Stewart (1992) reported excellent inter-rater reliability with the use of TMT-A and -B, although in an HIV-seronegative population. Using two or more evaluators in the assessment of neurocognitive performance is especially important in normative studies and clinical trials of many participants. The MDC percentage for the seven tests was within the predefined limit of acceptance, thus suggesting that the tests possess good sensitivity among PLWHIV in Nigeria. This provides clinicians and researchers with a baseline for the evaluation of the impact of a clinical intervention on NP performance of PLWHIV in Nigeria. There was no inter-rater proportional bias, indicating that measurement approaches agree evenly over the measurement range. This means that the boundaries of the agreement are unaffected by the measurement itself (Bland & Altman, 1999).

As touching intra-rater reliability, all but the HVLTR-VL showed good to excellent intra-rater reliability, with ICC values between 0.65 and 0.97. The MDC percentage values showed limited measurement error except for the HVLTR-VL, where the measurement error was greater than the pre-defined mark of 30%. The MDC values for the

TABLE 1: Clinical and sociodemographic variables.

Clinical/sociodemographic	Inter-rater data $N = 30$				Intra-rater data $N = 30$			
	Mean	\pm SD	n	%	Mean	\pm SD	n	%
Age	43	± 9	-	-	44	± 8	-	-
Systolic BP (mmHg)	121	± 13	-	-	119	± 11.9	-	-
Diastolic BP (mmHg)	77	± 9	-	-	79	± 7	-	-
Pulse rate (bpm)	74	± 10	-	-	79	± 12	-	-
Sex								
Male	-	-	10	33.3	-	-	10	33.3
Female	-	-	20	66.7	-	-	20	66.7
Education								
Primary	-	-	9	30.0	-	-	9	30
Secondary	-	-	13	43.3	-	-	14	46.7
Tertiary	-	-	8	26.7	-	-	7	23.3
Marital status								
Married	-	-	3	10.0	-	-	2	6.7
Single	-	-	23	76.7	-	-	21	70.0
Widow	-	-	4	13.3	-	-	7	23.3

Bpm: beats per minute; mmHg: millimetre of mercury; SD, standard deviation; BP, blood pressure.

TABLE 2: Result for inter-rater reliability showing intra-class correlation.

NP test	Rater	Mean	\pm SD	ICC _{2,1}	95% CI	p	SEM	MDC	MDC (%)	PB β	p
HVLTR-L	EV ₁	19.20	± 6.01	0.46	-0.12–0.74	0.048*	1.80	4.99	24.9	-0.167	0.254
	EV ₂	20.83	± 5.12	-	-	-	-	-	-	-	-
HVLTR-DR	EV ₁	5.55	± 2.83	0.83	0.65–0.92	< 0.001*	0.26	0.71	12.3	0.071	0.713
	EV ₂	5.90	± 2.97	-	-	-	-	-	-	-	-
TMT-A	EV ₁	61.03	± 33.32	0.86	0.71–0.93	< 0.001*	2.44	6.77	11.5	0.045	0.815
	EV ₂	56.67	± 34.31	-	-	-	-	-	-	-	-
TMT-B	EV ₁	149.07	± 42.82	0.78	0.52–0.90	< 0.001*	5.30	14.68	10.5	0.142	0.471
	EV ₂	133.86	± 47.59	-	-	-	-	-	-	-	-
DST-F	EV ₁	5.33	± 1.06	0.52	0.16–0.81	0.008*	0.37	1.03	17.9	0.333	0.072
	EV ₂	6.20	± 1.45	-	-	-	-	-	-	-	-
DST-B	EV ₁	3.10	± 0.96	0.60	0.15–0.81	0.009*	0.19	0.53	17.5	-0.276	0.139
	EV ₂	3.00	± 0.74	-	-	-	-	-	-	-	-
COWAT	EV ₁	26.87	± 10.47	0.89	0.77–0.95	< 0.001*	0.57	1.58	6.0	-0.106	0.579
	EV ₂	25.57	± 9.83	-	-	-	-	-	-	-	-

HVLTR-L: Hopkins Verbal Learning Test-Revised- Learning domain; HVLTR-L-DR: Hopkins Verbal Learning Test-Revised- Delayed Recall domain; TMT-A: Trail Making Test-A; TMT-B: Trail Making Test-B; DST-F: Digits Span Test-Forward; DST-B: Digits Span Test-Forward; COWAT: Verbal Fluency Test; SD, standard deviation; NP, neuropsychological; ICC, intra-class correlation; CI, confidence interval; SEM, Standard Error of Measurement; MDC, minimum detectable change; PB, proportional bias; EV, evaluator.

*, significant at $\alpha = 0.05$.

TABLE 3: Result for intra-rater reliability showing intra-class correlation.

NP test	M	Mean	± SD	ICC _{1,2}	95% CI	<i>p</i>	SEM	MDC	MDC%	PB β	<i>p</i>
HVLTR-L	M1	17.68	± 4.00	0.28	-0.18–0.53	0.143	2.73	7.57	38.5	-0.475	0.011*
	M2	21.93	± 6.62	-	-	-	-	-	-	-	-
HVLTR-DR	M1	5.21	± 2.57	0.76	0.34–0.90	< 0.001*	0.39	1.08	17.9	-0.371	0.047*
	M2	6.83	± 3.37	-	-	-	-	-	-	-	-
TMT-A	M1	56.67	± 24.31	0.80	0.56–0.91	< 0.001*	3.41	17.9	17.3	0.291	0.118
	M2	61.03	± 33.32	-	-	-	-	-	-	-	-
TMT-B	M1	133.37	± 44.27	0.65	0.28–0.83	0.002*	8.10	22.5	15.9	-0.168	0.374
	M2	149.43	± 38.24	-	-	-	-	-	-	-	-
DST-F	M1	4.62	± 0.76	0.67	0.32–0.85	0.002*	0.16	0.44	9.4	-0.281	0.140
	M2	4.86	± 0.99	-	-	-	-	-	-	-	-
DST-B	M1	3.00	± 1.0	0.84	0.65–0.92	< 0.001*	0.09	0.25	8.2	-0.141	0.465
	M2	3.07	± 1.10	-	-	-	-	-	-	-	-
COWAT	M1	19.63	± 10.64	0.97	0.93–0.98	< 0.001*	0.16	0.44	2.2	-0.035	0.855
	M2	20.07	± 10.78	-	-	-	-	-	-	-	-

HVLTR-L: Hopkins Verbal Learning Test-Revised- Learning; HVLTR-L-DR: Hopkins Verbal Learning Test-Revised- Delayed Recall; TMT-A: Trail Making Test-A; TMT-B: Trail Making Test-B; DST-F: Digits Span Test-Forward; DST-B: Digits Span Test-Forward; COWAT: Verbal Fluency Test; NP, neuropsychological; ICC, intra-class correlation; SEM, Standard Error of Measurement; MDC, minimum detectable change; PB, proportional bias; CI, confidence interval.

*, significant at $\alpha = 0.05$.

TABLE 4: Influence of sociodemographic characteristics on test score of selected neuropsychological tests.

NP test	Rater	HVLTR-L			HVLTR-DR			TMT-A			TMT-B			DST-F			DST-			COWAT		
		χ^2	MW-U test	<i>p</i>	χ^2	MW-U test	<i>p</i>	χ^2	MW-U test	<i>p</i>	χ^2	MW-U test	<i>p</i>	χ^2	MW-U test	<i>p</i>	χ^2	MW-U test	<i>p</i>	χ^2	MW-U test	<i>p</i>
Inter-rater data																						
Age	Ev ₁	0.43	-	0.806	1.89	-	0.389	4.56	-	0.102	3.03	-	0.220	1.930	-	0.381	4.77	-	0.092	6.960	-	0.031*
	Ev ₂	1.09	-	0.580	1.43	-	0.490	1.87	-	0.393	8.66	-	0.013*	1.540	-	0.463	0.50	-	0.779	7.330	-	0.026*
Sex	Ev ₁	-	73.50	0.242	-	95.00	0.825	-	77.00	0.311	-	94.00	0.962	-	79.000	0.346	-	71.00	0.154	-	80.500	0.390
	Ev ₂	-	73.00	0.234	-	58.50	0.091	-	82.50	0.441	-	81.50	0.505	-	88.000	0.579	-	80.50	0.340	-	78.000	0.332
Education	Ev ₁	0.38	-	0.828	0.27	-	0.872	2.66	-	0.264	8.04	-	0.018*	0.470	-	0.791	5.12	-	0.077	0.322	-	0.851
	Ev ₂	2.41	-	0.300	1.94	-	0.380	1.83	-	0.401	11.41	-	0.003*	1.500	-	0.472	0.80	-	0.669	0.757	-	0.685
Marital status	Ev ₁	2.98	-	0.226	0.84	-	0.658	2.55	-	0.280	1.31	-	0.520	1.620	-	0.446	0.08	-	0.961	1.850	-	0.396
	Ev ₂	1.62	-	0.446	1.03	-	0.597	1.29	-	0.524	1.27	-	0.531	2.480	-	0.289	5.89	-	0.052	3.750	-	0.153
Intra-rater data																						
Age	M ₁	-0.10	-	0.609	0.01	-	0.953	0.22	-	0.254	0.10	-	0.610	-0.030	-	0.880	0.06	-	0.770	-0.090	-	0.629
	M ₂	-0.17	-	0.379	-0.23	-	0.238	-0.24	-	0.211	0.16	-	0.394	0.120	-	0.577	0.03	-	0.884	0.150	-	0.416
Sex	M ₁	-	71.00	0.199	-	44.50	0.013*	-	73.50	0.248	-	67.00	0.243	-	95.000	0.806	-	94.50	0.797	-	72.500	0.226
	M ₂	-	20.50	0.001*	-	39.50	0.010*	-	57.50	0.061	-	82.00	0.420	-	68.500	0.290	-	88.50	0.939	-	70.500	0.194
Education	M ₁	0.80	-	0.670	5.14	-	0.076	4.55	-	0.103	3.28	-	0.194	0.481	-	0.786	3.15	-	0.207	3.500	-	0.174
	M ₂	6.16	-	0.046*	2.65	-	0.266	5.42	-	0.066	0.41	-	0.815	0.350	-	0.838	2.51	-	0.286	3.639	-	0.162
Marital status	M ₁	2.96	-	0.228	2.22	-	0.330	0.79	-	0.675	2.84	-	0.242	1.560	-	0.457	4.55	-	0.103	2.400	-	0.302
	M ₂	1.29	-	0.525	9.46	-	0.009*	0.27	-	0.873	1.15	-	0.562	1.430	-	0.490	5.32	-	0.070	1.370	-	0.503

HVLTR-L: Hopkins Verbal Learning Test-Revised- Learning; HVLTR-L-DR: Hopkins Verbal Learning Test-Revised- Delayed Recall; TMT-A: Trail Making Test-A; TMT-B: Trail Making Test-B; DST-F: Digits Span Test-Forward; DST-B: Digits Span Test-Forward; COWAT: Verbal Fluency Test; MW: Man-Whitney, χ^2 : Chi-Square value for the Kruskal Wallis test; NP, neuropsychological; EV, evaluator.

*, significant at $\alpha = 0.05$.

HVLTR-VL, HVLTR-DR, TMT-A, TMT-B, DST-f, DST-b and COWAT were approximately eight words, one word, 18 s, 23 s, one digit, one digit and one word, respectively. The poor intra-reliability of the HVLTR-VL reflects the considerable level of measurement error, which was the result of interruption experienced during evaluation. In this study, we assessed participants who were waiting to collect their drugs in an ART clinic. Occasionally, a few participants' attentions were drawn during NP evaluation and may this constitute a bias for the HVLTR-VL. For these participants, we ensured, most of the time, that they completed the test at hand before responding to the call. We also emphasised the need to pay attention during the test. However, the poor-intra-rater reliability of the HVLTR-VL suggests that the HVLTR-VL is sensitive to minute distractions.

Regarding the effects of sociodemographic characteristics on the selected NP tests, previous studies have been inconsistent in their support of age as a variable influencing verbal fluency performance, our finding supports the postulation that age affects verbal fluency performance. To nullify the potential age-related bias, the verbal fluency when used as a candidate NP test requires the use of aged-matched HIV negative control. This agrees with the findings of Barry, Bates and Labouvie (2008). However, contrary to Bates & colleagues, our study showed that older adults performed better than younger adults. Although verbal ability is a crystallised ability that does not deteriorate with age or minor brain dysfunction, phonemic fluency necessitates executive ability, specifically the ability to initiate and maintain effort, as well as the ability to organise information for retrieval, both of which are abilities that are sensitive to nuanced cerebral

dysfunction and ageing (Burke, Crowder, Hagan-Burke, & Zou, 2009; Henry & Beatty, 2006; Plumet, Gil, & Gaonac'h, 2005). Hence, the better COWAT performance of the older adults relative to younger adults found in this study could be because of the poor sample size, with skewed age distribution: older adults made up only 10% of the participants. Although education is a potential predictor of COWAT performance, in our study, educational levels and sex did not influence verbal fluency. This suggests that education- and sex-adjusted norms may not be necessary for this measure in our setting. The finding that age and sex affected the performance of the HVLT-R-DR is consistent with the result obtained by Vanderploeg et al. (2000), in which age and sex had a significant impact on the performance of the HVLT-R. However, unlike Vanderploeg and colleagues, our studies showed that educational level did not influence participants' performance on HVLT-R, suggesting that its election as a candidate NP tests for PLWHIV in our setting may not require educational consideration. Unlike Vanderploeg et al. (2000), it is likely that our sample size was insufficient to identify an influence of education on the HVLT-R or that educational differences were not correctly reflected by years of education. Although Feng et al. (2014) found that being single was linked to a higher risk of cognitive impairment than being married, our study is the first to report a significant effect of marital status on HVLT-R performance among PLWHIV. However, contrary to Feng et al. (2014), this study revealed better performance among single and widows compared with the ones who are married. The discrepancy could be because of the difference in the study population, with the former conducted amongst elderly Chinese.

In this study, the performance of TMT-B was influenced by sex. This agrees with the findings of Singh et al. (2010) in which gender was associated with TMT performance. It implies that the use of TMT-B in the diagnosis of HAND among PLWHIV must be based on sex-matched norms. However, the fact that education influenced the TMT-B in this study is consistent with Mitrushina, Boone, Razani and D'Elia (1999) and Tombaugh (2004) but contrary to Singh et al. (2010), which is a South African-based study among PLWHIV. This discrepancy may point to variation in sample size, and how many years of education is a true reflection of experience between the societies. Just like other NP tests, the effect of sociodemographic characteristics on DST performance is controversial (Choi et al., 2014). Some studies reported significantly higher DST score in females than in males (Ostrosky-Solis & Azucena, 2006; Singh et al., 2010), while others showed a minor or non-existent gender effect, implying that no gender-related modifications to the normative data are required (Anstey, Matters, Brown, & Lord, 2000; Pena-Casanova et al., 2009). Our studies showed that DST performance was not subject to any sex, age or educational bias, thus indicating that sociodemographic matched norms may not be necessary when it is used as a candidate NP test to aid HAND diagnosis. Overall, the imperfect agreement between the findings of this study and

the previous studies constitutes support to the postulation that the use of cross-cultural data for a NP test will lead to errors in HAND classification (Fernandez & Marcopulos, 2008).

The limitations of the study include the use of unequal sample categories, practice effect and the distractions encountered during NP evaluation; the healthcare professional drew the attention of some participants during NP assessment. The study draws its strength in being the first of its kind to report reliability and MDC for seven cognitive ability domains relevant to PLWHIV in Nigeria. The HVLT-R-DR, TMT-A, TMT-B, DST-f, DST-b and COWAT are reliable candidate NP tests for PLWHIV. In our setting, the TMT-A, DST-f and DST-b are free from demographic bias and hence may not require demographically adjusted norms, while TMT-B, HVLT-R and COWAT exhibit at least one form of sociodemographic biases. This may require that appropriate adjustments be made when they are selected as candidates for NP tests in pursuit of HAND diagnosis. Extra caution aiming at minimising distraction should be ensured when administering the HVLT-R. Further studies are recommended with large sample size to establish normative score of the selected NP tests and to examine the effect of socioeconomic status on NP test performance.

Acknowledgements

The authors would like to express their profound gratitude to the clinical psychologist, Mrs Jane Nwodo, and the research assistant, Maryjane Ukwuoma who converted the citations and list of references to APA format.

Competing interests

The authors declare that they have no financial or personal relationship that may have inappropriately influenced them in writing this article.

Authors' contributions

M.C.N. and N.M. conceived of the presented idea. M.N. carried out the research under the supervisions of N.M., N.G., A.A. and A.O. M.C.N. carried out statistical analysis and wrote the first draft of the manuscript. M.C.N., N.M., A.G., A.A. and A.O. revised the manuscript for intellectual content.

Funding information

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Student Financial Aid Scheme (NSFAS) via the University of Pretoria [Grant number: 1338].

Data availability

Raw data were generated at a Nigerian tertiary health institution. Derived data supporting the findings of this study are available upon special request from the corresponding author, M.N.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

References

- Andrews, J., Janzen, H.L., & Saklofske, D.H. (2001). Handbook of psychoeducational assessment ability, achievement, and behavior in children educational psychology (pp. 415–450). San Diego, CA: Academic Press.
- Anstey, K.J., Matters, B., Brown, A.K., & Lord, S.R. (2000). Normative data on neuropsychological tests for very old adults living in retirement villages and hostels. *The Clinical Neuropsychologist*, 14(3), 309–317. [https://doi.org/10.1076/1385-4046\(200008\)14:3;1-P;FT309](https://doi.org/10.1076/1385-4046(200008)14:3;1-P;FT309)
- Antinori, A., Arendt, G., Becker, J.T., Brew, B.J., Byrd, D.A., Cherner, M., Clifford, D.B., ... Wojna, V.E. (2007). Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*, 69(18), 1789–1799. <https://doi.org/10.1212/01.WNL.0000287431.88658.8b>
- Awofala, A.A., & Ogundele, O.E. (2018). HIV epidemiology in Nigeria. *Saudi Journal of Biological Sciences*, 25(4), 697–703. <https://doi.org/10.1016/j.sjbs.2016.03.006>
- Barry, D., Bates, M.E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology*, 15(2), 97–106. <https://doi.org/10.1080/09084280802083863>
- Bland, J.M., & Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135–160. <https://doi.org/10.1191/096228099673819272>
- Burke, M.D., Crowder, W., Hagan-Burke, S., & Zou, Y. (2009). A comparison of two path models for predicting reading fluency. *Remedial and Special Education*, 30(2), 84–95. <https://doi.org/10.1177/0741932508315047>
- Choi, H.J., Lee, D.Y., Seo, E.H., Jo, M.K., Sohn, B.K., Choe, Y.M., Byun, M.S., ... Woo, J.I. (2014). A normative study of the digit span in an educationally diverse elderly population. *Psychiatry Investigation*, 11(1), 39–43.
- Cody, S.L., & Vance, D.E. (2016). The neurobiology of HIV and its impact on cognitive reserve: A review of cognitive interventions for an aging population. *Neurobiology Of Disease*, 92 (Pt B), 144–156. <https://doi.org/10.1016/j.nbd.2016.01.011>
- Fals-Stewart, W. (1992). An interrater reliability study of the trail making test (parts A and B). *Perceptual & Motor Skills*, 74(1), 39–42. <https://doi.org/10.2466/pms.1992.74.1.39>
- Feng, L., Ng, X., Yap, P., Li, J., Lee, T., Håkansson, K., & Kua, E. (2014). Marital status and cognitive impairment among community-dwelling Chinese older adults: The role of gender and social engagement. *Dementia and Geriatric Cognitive Disorders Extra*, 4(3), 375–384. <https://doi.org/10.1159/000358584>
- Fernandez, A.L., & Marcopulos, B.A. (2008). A comparison of normative data for the trail making test from several countries: Equivalence of norms and considerations for interpretation 1. *Scandinavian Journal of Psychology*, 49(3), 239–246. <https://doi.org/10.1111/j.1467-9450.2008.00637.x>
- Groth-Marnat, G., & Baker, S. (2003). Digit span as a measure of everyday attention: A study of ecological validity. *Perceptual and Motor Skills*, 97(3 Pt 2), 1209–1218. <https://doi.org/10.2466/pms.2003.97.3f.1209>
- Henry, J.D., & Beatty, W.W. (2006). Verbal fluency deficits in multiple sclerosis. *Neuropsychologia*, 44(7), 1166–1174. <https://doi.org/10.1016/j.neuropsychologia.2005.10.006>
- IBM Corp. (2012). *IBM SPSS statistics for windows, version 21.0*. Armonk, NY: IBM Corp.
- Jumar, J., Sunshine, S., Ahmed, H., El-Kamary, S.S., Magder, L., Hungerford, L., Burdo, T., ... Royal 3rd, W. (2017). Peripheral blood lymphocyte HIV DNA levels correlate with HIV associated neurocognitive disorders in Nigeria. *Journal of NeuroVirology*, 23(3), 474–482. <https://doi.org/10.1007/s13365-017-0520-5>
- Lee, P., Liu, C., Fan, C., Lu, C., Lu, W., & Hsieh, C. (2013). The test-retest reliability and the minimal detectable change of the Purdue pegboard test in schizophrenia. *Journal of the Formosan Medical Association*, 112(6), 332–337. <https://doi.org/10.1016/j.jfma.2012.02.023>
- McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica (Zagreb)*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Mitrushina, M.N., Boone, K.L., Razani, J., & D'Elia, L. (1999). *Handbook of normative data for neuropsychological assessment* (2nd ed.). New York, NY: Oxford University Press.
- Mwangala, P.N., Newton, C.R., Abas, M., & Abubakar, A. (2019). Screening tools for HIV-associated neurocognitive disorders among adults living with HIV in sub-Saharan Africa: A scoping review. *AAS Open Research*, 1, 28. <https://doi.org/10.12688/aasopenres.12921.2>
- Nunan, D., Aronson, J., & Bankhead, C. (2018). Catalogue of bias: Attrition bias. *BMJ Evidence-Based Medicine*, 23(1), 21–22. <https://doi.org/10.1136/ebmed-2017-110883>
- Nweke, M.C., Akinpelu, A.O., & Ezema, C.I. (2019). Variation in spatio-temporal gait parameters among patients with HIV-related neurocognitive impairment. *Indian Journal of Physiotherapy and Occupational Therapy*, 13(4), 186–191. <https://doi.org/10.5958/0973-5674.2019.00158.8>
- Nweke, M.C., Mshunqane, N., Govender, N., & Akinpelu, O.A. (2021). Physiological effects of physical activity on neurocognitive function in people living with HIV: A systematic review of intervention and observational studies. *African Journal for Physical Activity and Health Sciences*, 27(1), 101–122. <https://doi.org/10.37597/ajphes.2021.27.1.8>
- O'Neil-Pirozzi, T.M., Goldstein, R., Strangman, G.E., & Glenn, M.B. (2012). Test-retest reliability of the Hopkins verbal learning test-revised in individuals with traumatic brain injury. *Brain Injury*, 26(12), 1425–1430. <https://doi.org/10.3109/02699052.2012.694561>
- Ostrosky-Solís, F., & Azucena, L. (2006). Digit span: Effect of education and culture. *International Journal of Psychology*, 41(5), 333–341. <https://doi.org/10.1080/00207590500345724>
- Pena-Casanova, J., Quinones-Ubeda, S., Quintana-Aparicio, M., Aguilar, M., Badenes, D., Molinuevo, J.L., Torner, L., ... Blesa, R. (2009). Spanish multicenter normative studies (NEURONORMA project): Norms for verbal span, visuospatial span, letter and number sequencing, trail making test, and symbol digit modalities test. *Archives of Clinical Neuropsychology*, 24(4), 321–341. <https://doi.org/10.1093/arclin/acp038>
- Plumet, J., Gil, R., & Gaonac'h, D. (2005). Neuropsychological assessment of executive functions in women: Effects of age and education. *Neuropsychology*, 19(5), 566–577. <https://psycnet.apa.org/doi/10.1037/0894-4105.19.5.566>
- Robertson, K., Jiang, H., Evans, S.R., Marra CM, Berzins B, Hakim J., Sacktor, N., ... Walawander, A. (2016). International neurocognitive normative study: Neurocognitive comparison data in diverse resource-limited settings: AIDS clinical trials group A5271. *Journal of NeuroVirology*, 22(4), 472–478. <https://doi.org/10.1007/s13365-015-0415-2>
- Robertson, K., Liner, J., & Heaton, R. (2009). Neuropsychological assessment of HIV-infected populations in international settings. *Neuropsychology Review*, 19, 232–249. <https://doi.org/10.1007/s11065-009-9096-z>
- Ross, T.P., Calhoun, E., Cox, T., Wenner, C., Kono, W., & Pleasant, M. (2007). The reliability and validity of qualitative scores for the controlled oral word association test. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 22(4), 475–488. <https://doi.org/10.1016/j.acn.2007.01.026>
- Salthouse, T. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence*, 39, 222–232. <https://doi.org/10.1016/j.intell.2011.03.001>
- Sedrez, J.A., Candotti, C.T., Rosa, M.I.Z., Medeiros, F.S., Marques, M.T., & Loss, J.F. (2016). Test-retest, inter-and intra-rater reliability of the flexicurve for evaluation of the spine in children. *Brazilian Journal of Physical Therapy*, 20(2), 142–147. <https://doi.org/10.1590/bjpt-rbf.2014.0139>
- Singh, D., Joska, J.A., Goodkin, K., Lopez, E., Myer, L., Paul, R.H., ... Sunpath, H. (2010). Normative scores for a brief neuropsychological battery for the detection of HIV-associated neurocognitive disorder (HAND) among South Africans. *BMC Research Notes*, 3, 28. <https://doi.org/10.1186/1756-0500-3-28>
- Sridhara, R., Mandrekar, S.J., & Dodd, L.E. (2013). Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clinical Cancer Research*, 19(10), 2613–2620. <https://doi.org/10.1158/1078-0432.CCR-12-2938>
- Tombaugh, T.N. (2004). Trail making test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2), 203–214. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8)
- Towe, S.L., Puja Patel, B.A., & Meade, C.S. (2017). The acceptability and potential utility of cognitive training to improve working memory in persons living with HIV: A preliminary randomized trial. *Journal of the Association of Nurses in AIDS Care*, 28(4), 633–643. <https://doi.org/10.1016/j.jana.2017.03.007>
- Van Lummel, R.C., Walgaard, S., Hobert, M.A., Maetzler, W., VanDieën, J.H., Galindo-Garre, F., & Terwee, C. (2016). Intra-rater, inter-rater and test-retest reliability of an instrumented timed up and go (iTUG) test in patients with Parkinson's disease. *PLoS One*, 11(3), e0151881. <https://doi.org/10.1371/journal.pone.0151881>
- Vanderploeg, R.D., Schinka, J.A., Jones, T., Small, B.J., Graves, A.B., & Mortimer, J.A. (2000). Elderly norms for the Hopkins verbal learning test-revised. *The Clinical Neuropsychologist*, 14(3), 318–324. [https://doi.org/10.1076/1385-4046\(200008\)14:3;1-P;FT318](https://doi.org/10.1076/1385-4046(200008)14:3;1-P;FT318)
- Viechtbauer, W., Smitsb, L., Kotz, D., Budee, L., Spigtc, M., Serroeyng, J., & Crutzen, R. (2015). A simple formula for the calculation of sample size in pilot studies. *Journal of Clinical Epidemiology*, 68(11), 1375–1379. <https://doi.org/10.1016/j.jclinepi.2015.04.014>
- Watkins, C.C., & Treisman, G.J. (2012). Neuropsychiatric complications of aging with HIV. *Journal of NeuroVirology*, 18(4), 277–290. <https://doi.org/10.1007/s13365-012-0108-z>
- Yakasai, A.M., Gudaji, M.I., Muhammad, H., Ibrahim, A., Owolabi, L.F., Ibrahim, D.A., Babashani, M., ... Habib, A.G. (2015). Prevalence and correlates of HIV-associated neurocognitive disorders (HAND) in Northwestern Nigeria. *Neurology Research International*, 2015, 486960. <https://doi.org/10.1155/2015/486960>
- Yeatesh, K.W., & Taylor, G. (2001). Neuropsychological assessment of children. In *Handbook of psychoeducational assessment, ability, achievement, and behavior in children educational psychology* (415–450). Retrieved from <https://doi.org/10.1016/B978-012058570-0/50016-1>
- Yusuf, A.J., Hassan, A., Maman, A.I., Muktar, H.M., Suleiman, A.M., & Baiyewu, O. (2017). Prevalence of HIV-associated neurocognitive disorder (HAND) among patients attending a tertiary health facility in Northern Nigeria. *Journal of the International Association of Providers of AIDS Care*, 16(1), 48–55. <https://doi.org/10.1177/2325957414553839>