CogState computerised cognitive testing yields global cognitive measures that can detect lower cognitive performance associated with efavirenz in a low-income South African setting

Evan C Edmond[1, 2], Anna J Dreyer[1, 3], John Joska[1], and Sam Nightingale[1]

[1]HIV Mental Health Research Unit, Division of Neuropsychiatry, Department of Psychiatry and Mental Health, Neuroscience Institute, University of Cape Town, Cape Town, South Africa
[2]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
[3]University of Cambridge, Cambridge, UK

May 7, 2025

Contents

1 Abstract

2 Introduction

Cognitive testing is a key part of memory clinic assessment to identifying and managing low cognitive performance in people living with HIV. The gold standard approach of neuropsychological "pen & paper" testing is richly informative but time consuming and dependent on availability, training, and funding of skilled testers.

Expert availability per capita to diagnose and manage disorders of cognition is ∼100x greater in high income countries (2.2/100k) than in low-middle income countries (0.02/100k)[1]. This gap is not feasibly addressable by training new clinicians at current rates. Strategic goals 6 and 7 of the 2022 WHO dementia research blueprint[2] advocate for novel clinical diagnostics that are applicable to diverse settings and the entire disease spectrum including prodromal disease.

Computerised cognitive testing could address this need, with various tools applied to screen for HIV-associated low cognitive performance in sub-Saharan Africa[3–5]. Various validation concerns remain, with regards to case/cohort characteristics, lack of control groups, demographic data, and most importantly - construct validity ("does the measure behave as though it is measuring the (indirectly measured) property")[6,7].

This study investigated one of these tools (CogState brief battery) in the CONNECT study, a prospective cohort study of people with HIV in a low-income peri-urban area of Cape Town, South Africa, done in the context of a national switch from efavirenz to dolutegravir antiretroviral therapy (ART). CogState has been used extensively in people with HIV, with most published work originating from North America[8,9], Europe[10,11], and Australia[12], with one study in a low-income setting in Uganda[13]. CogState has significant advantages - it is a self-contained software package that delivers stimuli, records responses, and produces report data in a time-efficient, reliable, and reproducible manner.

This study investigated the feasibility of using CogState to measure cognitive performance in a low-income cohort of South African people with HIV, comparing performance with gold-standard neuropsychological "pen&paper" testing. This cohort is broadly representative of people with HIV in southern Africa, an under-represented group in neuro-HIV research. A demographically matched control cohort of people without HIV was also studied. Potential confounders were addressed through detailed socioeconomic and technology familiarity data collection. We compared overall cognitive performance as measured by GDS derived from CogState vs. "pen&paper" testing. Finally, we performed principal component analysis of both CogState and "pen&paper" data to investigate construct validity and form inferences on what underlying cognitive features were assessed.

3 Methods

3.1 Study design and participants

We recruited a prospective cohort of adults with and without HIV as part of the parent study: Cognition, Neuropsychiatric Symptoms and Neuroinflammation Switching to Dolutegravir in Cape Town (CONNECT), based at the Gugulethu Community Health Centre in a low-income periurban area of Cape Town, South Africa. Eligible participants had been receiving efavirenz-based ART for at least 1 year and were identified by the clinic to switch to dolutegravir- based ART as part of the national programmatic switch. The study was approved by the University of Cape Town Faculty of Health Sciences Human Research Ethics Committee (017/2019). Written informed consent was obtained in the language of participant preference (English or isiXhosa). Full inclusion and exclusion criteria are described in previous work[14].

3.2 Technology use questionnaire

To assess the feasibility of computerised testing in a low-middle income setting, data was collected on participants' computer and cellphone ownership and usage, as well as overall comfort with technology. Additional questions assessed difficulty with using the test laptop, as well as difficulty with each task.

3.3 CogState test battery

In this study, six CogState tasks were selected. The test battery was presented on a laptop computer. The participant was shown how to use the laptop keyboard to respond "Yes" or "No" for each task.

Detection The Detection task aims to measure processing speed. The on-screen instructions read: "Has the card turned face up?". A playing card is shown face down, and is shown to suddenly flip to face up at a random interval. The participant is instructed to press "Yes" as soon as the card flips over, working as quickly and accurately as possible. Button presses before the card is flipped are marked as errors. The test continues until 35 correct responses are made, or 3 minutes are up.

Identification The Identification task aims to measure attention using a choice reaction time task. The on-screen instructions read: "Is the card red?". A playing card is shown face down, and suddenly flips to face up. The participant is instructed to decide whether the card is red or not, and press "Yes" if it is red, and "No" if not, working as quickly and accurately as possible. The principal outcome measure of the task is the mean reaction time for correct responses in milliseconds, $\log_{10}$ transformed.

One-Back The One-Back task aims to measure working memory. The on-screen instructions read: "Is the previous card the same?". A sequence of playing cards is shown, one at a time, face-up, in the centre of the screen. The participant is instructed to decide whether the card is the same as the previous card, pressing "Yes" if it is the same, and "No" if not, working as quickly and accurately as possible. After the participant responds, the next card in the sequence is shown.

One Card Learning The One Card Learning task aims to measure visual learning. The on-screen instructions read: "Have you seen this card before in this test?". A sequence of playing cards is shown, one at a time, face-up, in the centre of the screen. The participant is instructed to decide whether they have previously seen this card at any point in this task, pressing "Yes" if so, and "No" if not, working as quickly and accurately as possible. After the participant responds, the next card in the sequence is shown.

Two-Back The Two-Back task aims to measure working memory. The on-screen instructions read: "Is the card the same as that shown two cards ago?". A sequence of playing cards is shown, one at a time, face-up, in the centre of the screen. The participant is instructed to decide whether the card is the same as the card shown two cards previously, pressing "Yes" if it is the same, and "No" if not, working as quickly and accurately as possible. After the participant responds, the next card in the sequence is shown.

Set Shifting The Set Shifting task aims to measure executive function. The on-screen instructions read: "Is this a target card?". A playing card is shown face up in the centre of the screen with the word "Number" or "Color" above it. If "Color" the participant is instructed to guess whether the target card is black or red. If the word is "Number" the participant must guess whether the current number displayed on the card is correct. At the start of the task, the participant guesses. Feedback on whether their answers are correct or not is provided. Subsequently, the participant may not proceed to the next trial until a correct response has been made. Part-way through the task the hidden rule changes. The participant is not told when these set-shifts occur, and they must learn the new target rule to proceed through the test. The participant is encouraged to work as quickly and accurately as possible.

The principal outcome measure of the Detection and Identification tasks is the mean reaction time for correct responses in milliseconds, $\log_{10}$ transformed. For the Two-Back and One Card Learning tasks the principal outcome is the overall accuracy (arcsine transformation of the square root of the proportion of correct responses).

CogState also reports the variability in responses ($\log_{10}$ transformed standard deviation of response times), and . The principal outcome measure is the total number of errors made during the test.

CogState also reports the variability in responses ($\log_{10}$ transformed standard deviation of response times), and the overall accuracy (arcsine transformation of the square root of the proportion of correct responses).

3.4 Pen&Paper cognitive testing

Comprehensive cognitive testing was done with a standard battery of validated tests that assess seven cognitive domains: executive functioning, verbal learning and memory, visuospatial learning and memory, verbal fluency, attention and working memory, information processing speed, and motor skills. This battery of tests takes approximately 2h. Tests were administered in either English or isiXhosa by a bilingual neuropsychology technician. A registered clinical neuropsychologist (AJD) supervised test administration and scoring protocols.

The cognitive domains, tests, and outcome variables were: (1) executive functioning, Color Trails Test 2 (CTT2) – completion time (in seconds); Wisconsin Card Sorting Test (WCST) – total score; (2) verbal learning and memory, Hopkins Verbal Learning Test-Revised (HVLT-R) – total across the three immediate recall trials, total on the delayed recall trial; (3) visuospatial learning and memory, Brief Visuospatial Memory Test-Revised (BVMT-R) – total across the

T score range	Deficit score
$T \geq 40$	0
$40 > T \geq 35$	1
$35 > T \geq 30$	2
$30 > T \geq 25$	3
$25 > T \geq 20$	4
$T < 20$	5

Table 1: Deficit score calculation thresholds

immediate recall trials, total on the delayed recall trial; (4) verbal fluency, category fluency test – total number of animals / total number of fruits and vegetables named in 1 minute; (5) attention/working memory, Wechsler Adult Intelligence Scale-Third Edition (WAIS-III) Digit Span subtest – total raw score; (6) processing speed, CTT1 – completion time (in seconds); WAIS-III Digit Symbol Coding subtest – total raw score; WAIS-III Symbol Search – total raw score; (7) motor skills, Grooved Pegboard Test (GPT) non-dominant hand (NDH) – completion time (in seconds); Finger Tapping Test NDH – completion time (in seconds).

3.5 Statistical analysis

3.5.1 Global deficit scores

Raw scores for each "pen&paper" task and the principal outcome measures for each CogState task were T-transformed (mean = 50, standard deviation = 10. Domain T-scores were calculated by taking the average of T-scores of the cognitive outcomes within each domain. Global T-scores were calculated by taking the average of the domain T-scores. Raw scores were adjusted if sex, years of education, and age were significantly associated with cognitive test performance in PWoH. We used the baseline data from PWoH to calculate demographically corrected z-scores, using standard regression-based norming processes[cysique2011, testa2009]. Details of these methods are described in Dreyer et al (2021)[dreyer2022]. The z-scores were then converted to T-scores (M = 50, SD = 10).

Deficit scores were calculated based on the T score thresholds listed in Table 1. Mean CogState deficit score greater than or equal to 0.5 were labelled "low cognitive performance". For "pen&paper" tests, mean T score for tests within each cognitive domain was calculated and used to calculate per-domain deficit scores in the same way. The mean deficit score across all domains was taken for each participants, with mean "pen&paper" deficit score greater than or equal to 0.5 labelled as "low cognitive performance".

3.5.2 Cross-correlation of individual CogState tests with pen&paper tests and domains

As a validation exercise, individual CogState tests were cross-correlated with individual pen&paper tests and also with pen&paper testing derived cognitive domains. The Pearson correlation coefficients are presented in heatmap form.

3.6 Principal component analysis

As only a subset of the CogState testing data (principal outcome measures) are used to calculate GDS in the conventional analytic approach, and the majority of data collected including inter-

trial variance is not used, we opted to additionally perform a data-driven analysis using all of the CogState data. Our aim was to explore the principal axes of variance in the raw CogState data, followed by cross-correlation with input features to understand redundancy/common themes in the dataset.

Finally, we tested each principal component's performance in classification of low/high cognitive performance against pen&paper by plotting ROC curves.

We also cross correlated the result with the gold-standard pen&paper data to address the question of whether these two approaches might be measuring correlated cognitive performance. Principal component analysis was performed using the scikit-learn python toolkit (v 1.4.2). Raw data for each input cognitive test was z-transformed (subtracting the mean and scaling to unit variance) before PCA was performed using the sklearn.decomposition.PCA function (separately for pen&paper and CogState test groups). The cumulative explained variance was plotted in a screeplot (Figure 9).

As similar proportions of explained variance were seen for both sets of inputs with 70% of variance explained by the first 6 components, these were selected for plotting. The loadings (eigenvectors) of input features (individual tasks) onto the principal components was plotted as a heatmap in Figure 6.

4 Results

91 people with HIV and 170 people without HIV were recruited between Aug 12, 2019, and Sept 16, 2022. 78.2% were female and mean age was 40.1. Demographic data are described in Table 1.

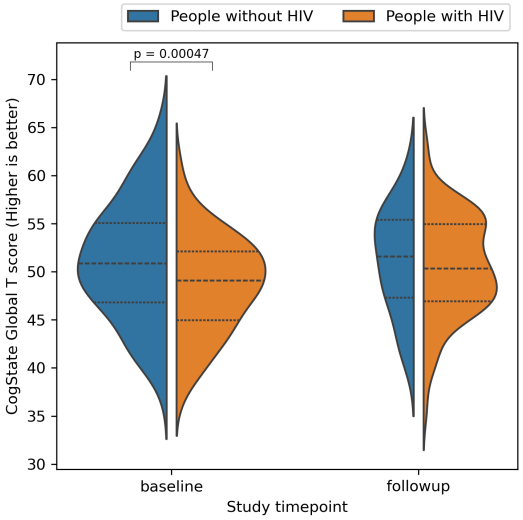
		Missing	Overall
n			261
Age at consent / years, mean (SD)		0	40.1 (8.2)
Ethnicity, n (%)	Black / Mnyama	1	260 (100.0)
Gender, n (%)	Female	0	204 (78.2)
	Male		57 (21.8)
First language, n (%)	Afrikaans	0	2 (0.8)
	English		7 (2.7)
	Setswane		1 (0.4)
	Shona / isiShona		5 (1.9)
	Sotho		4 (1.5)
	Xhosa / isiXhosa		241 (92.3)
	Zulu / IsiZulu		1 (0.4)
Testing language, n (%)	English	0	28 (10.7)
	Xhosa / isiXhosa		233 (89.3)
Income level, n (%)	R0-R499	10	40 (15.9)
	R1000-R1999		59 (23.5)
	R2000-R2999		30 (12.0)
	R3000-R3999		30 (12.0)
	R4000-R4999		27 (10.8)
	R500-R999		30 (12.0)
	R5000-R9999		35 (13.9)
HIV status, n (%)	Negative	0	91 (34.9)
	Positive		170 (65.1)

Figure 1: Demographics table (baseline combined) for CONNECT cohort.

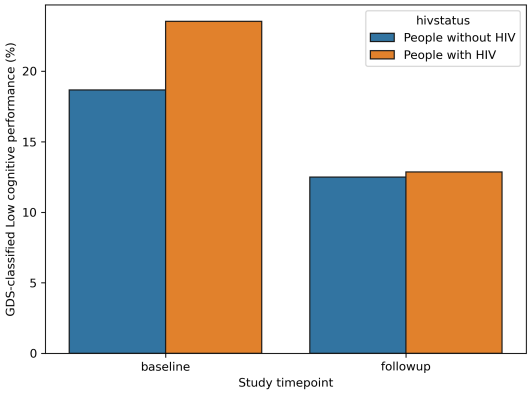
		Missing	Negative	Positive
n			91	170
Age at consent / years, mean (SD)		0	39.4 (9.3)	40.5 (7.6)
Ethnicity, n (%)	Black / Mnyama	1	90 (100.0)	170 (100.0)
Gender, n (%)	Female	0	68 (74.7)	136 (80.0)
	Male		23 (25.3)	34 (20.0)
First language, n (%)	English	0	5 (5.5)	2 (1.2)
	Sotho		2 (2.2)	2 (1.2)
	Xhosa / isiXhosa		84 (92.3)	157 (92.4)
	Afrikaans			2 (1.2)
	Setswane			1 (0.6)
	Shona / isiShona			5 (2.9)
	Zulu / IsiZulu			1 (0.6)
Testing language, n (%)	English	0	15 (16.5)	13 (7.6)
	Xhosa / isiXhosa		76 (83.5)	157 (92.4)
Income level, n (%)	R0-R499	10	13 (15.5)	27 (16.2)
	R1000-R1999		20 (23.8)	39 (23.4)
	R2000-R2999		13 (15.5)	17 (10.2)
	R3000-R3999		10 (11.9)	20 (12.0)
	R4000-R4999		5 (6.0)	22 (13.2)
	R500-R999		12 (14.3)	18 (10.8)
	R5000-R9999		11 (13.1)	24 (14.4)

4.1 Change over time - Does CogState have power to detect same longitudinal change (EFV switch)?

Change over time was analysed in the same way as done for Lancet paper[14] (calculation of global T, GDS-CI, and comparison between timepoints, split by HIV status).



(a) Violin plot of CogState global T score split by HIV status and timepoint. The area of each violin is proportional to the number of participants. Quartiles are marked.



(b) Percentage of individuals classified as "low cognitive performance" by GDS analysis of CogState data.

There was significantly lower cognitive performance in people with HIV at baseline compared to controls. At followup, the difference between the two groups was no longer significance. These results parallel the findings from pen&paper publication[14].

4.2 Global deficit scores

GDS generated using CogState and Pen&Paper are cross-tabulated in Table 3a. This showed only moderate agreement between GDS classification based on CogState versus pen&paper testing. The sensitivity, specificity, positive predictive value, and negative predictive value of CogState GDS is summarised below, taking the pen&paper testing as a gold standard. CogState GDS performed best as a predictor of normal/high cognitive performance, with negative predictive value of 0.84.

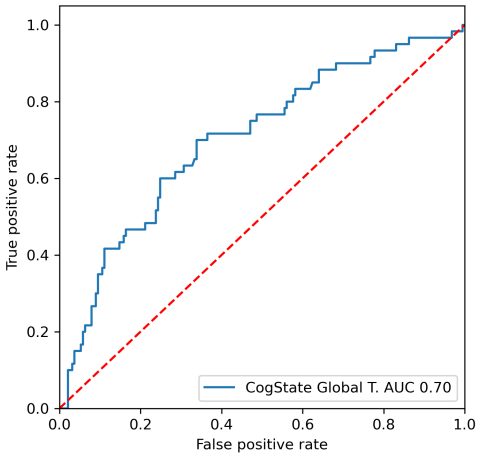
- Sensitivity: 0.52
- Specificity: 0.76
- PPV: 0.40
- NPV: 0.84

The performance of the global CogState T score against the pen&paper GDS is summarised in a receiver operating characteristic (ROC) curve in Figure 3b. Area-under-curve (AUC) of 0.69 suggests moderate performance of the global CogState T score in identifying low cognitive performance.

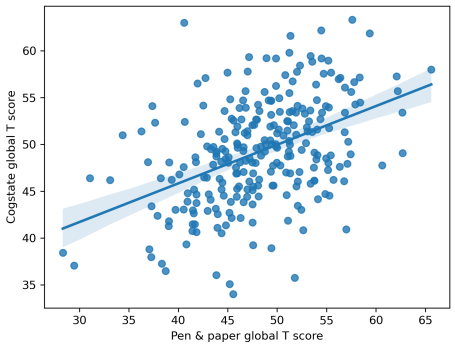
Plotting the CogState global T score (primary outcome measures) against the pen&paper global T score showed a weak positive correlation (Pearson R = 0.45, p = 3.6×10^{-14}) (Figure 3c).

		Pen & Paper cognitive performance	
		Low	Normal/High
CogState cognitive performance	Low	32	48
	Normal/High	30	152

(a) Table comparing classification as cognitively impaired (GDS) for Cogstate vs. gold standard.



(b) Receiver operating characteristic curve showing performance of CogState global T score from primary outcome measures against GDS classification of cognitive performance.



(c) Scatter plot of CogState vs. pen&paper global T scores, with linear regression trendline, and 95% confidence interval of the regression estimate.

4.3 Technology use questionnaire

Questionnaire data are presented in Table 4, with full detail on questions regarding the instructions and difficulty of individual tests presented in Supplementary Table 10. The majority (77%) of participants did not own a computer but mobile phone ownership was near universal (95%) - most of these (86%) with touchscreens. Most participants (79%) reported feeling somewhat comfortable, comfortable, or very comfortable using a computer. Most participants (85%) found it somewhat easy, easy, or very easy to use the computer during the testing session.

Participants who owned a computer had slightly better global CogState performance ($p = 0.014$, effect size $\delta T = 1.8$). Those who owned a touchscreen mobile phone performed slightly better than those who owned a mobile phone without a touchscreen ($p = 0.034$, effect size $\delta T = 1.8$).

		Missing	Overall	CogState Global T, mean	p
n			261		
Currently owns computer, n (%)	No	0	201 (77.0)	48.8	
	Yes		60 (23.0)	50.6	0.014
How often do you use a computer (desktop or laptop)?, n (%)	Never	0	129 (49.4)	48.4	
	Less than once a month		57 (21.8)	49.4	
	Once a month		19 (7.3)	50.3	
	Once a week		14 (5.4)	50.5	
	2-3 times a week		18 (6.9)	50.9	
	Most days		24 (9.2)	50.3	0.26
How comfortable do you feel using a computer?, n (%)	Very uncomfortable	0	14 (5.4)	47.3	
	Uncomfortable		14 (5.4)	49.2	
	Somewhat uncomfortable		27 (10.3)	46.6	
	Somewhat comfortable		60 (23.0)	49.5	
	Comfortable		88 (33.7)	49.4	
	Very comfortable		58 (22.2)	49.7	0.44
Do you currently own a mobile telephone?, n (%)	No	0	13 (5.0)	49.6	
	Yes		248 (95.0)	49.2	0.61
If yes, does it have a touchscreen?, n (%)	No	13	35 (14.1)	47.6	
	Yes		213 (85.9)	49.4	0.034
How did you find using the computer today?, n (%)	Very difficult	0	5 (1.9)	50.0	
	Difficult		3 (1.1)	48.7	
	Somewhat difficult		30 (11.5)	47.3	
	Somewhat easy		50 (19.2)	50.0	
	Easy		115 (44.1)	48.7	
	Very easy		58 (22.2)	50.4	0.127

Figure 4: Results of technology use questionnaire. p-values are derived from independent t-tests for Yes/No variables, and Kruskal-Wallis test for others.

4.4 Cross-correlation of CogState tests with pen&paper

Cross-correlation of selected CogState test data (those metrics associated best associated with a cognitive domain) with Pen&Paper derived cognitive domains is shown in Figure ¡ref new figure with trimmed rows¿. There was weak to moderate correlation between CogState derived

cognitive domain scores and P&P derived cognitive domain scores (maximum Pearson R = 0.53 between CogState One-back accuracy (Working Memory) and P&P Processing speed).

Full cross-correlation tables between raw CogState data (Speed, Variability, and Accuracy) vs. individual P&P tests (Figure 8) and P&P derived cognitive domains (Figure 5) are provided in Supplementary Materials.

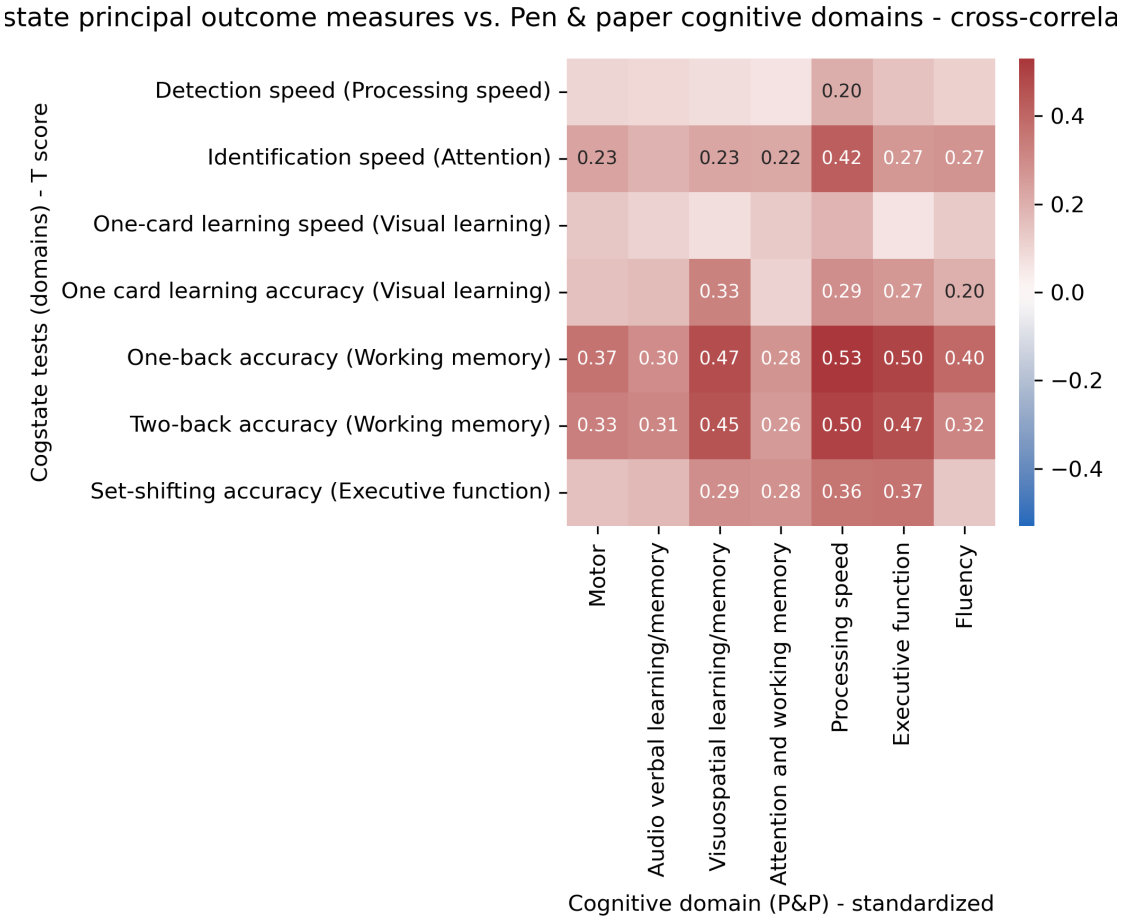


Figure 5: Principal CogState outcome measures and associated cognitive domains vs. P&P cognitive domains.

4.5 Principal component analysis

By inspecting the loadings (eigenvectors) of the components generated using all raw CogState data, we can determine which input data features contributed the greatest variance for each component. This is summarised in a heatmap in Figure 6.

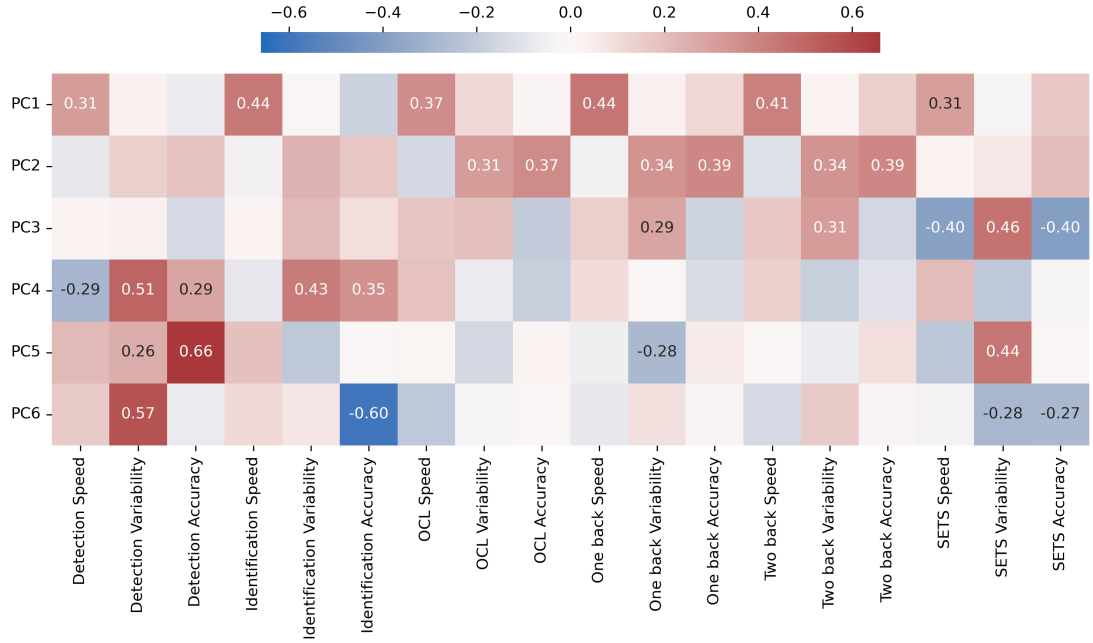


Figure 6: Heatmap of eigenvectors for each principal component.

PC1 - Speed The first principal component is loaded onto the "Speed" metrics from all of the CogState tasks - this component likely captures a global "Processing speed".

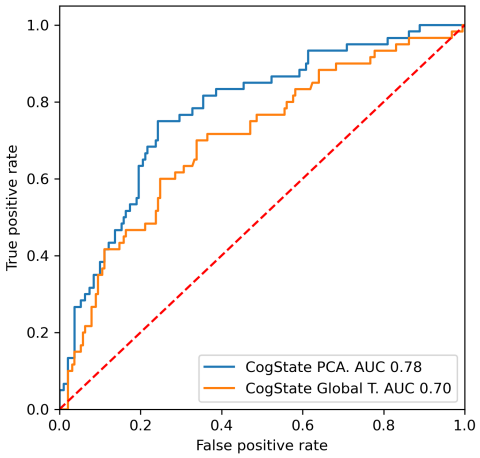
PC2 - Accuracy/Precision The second component is loaded onto the "Variability" and "Accuracy" metrics from several tests (One-card learning, One-back and Two-back), with loading in the opposite direction onto the "Speed" metric. This likely captures a common "Precision" across multiple tasks.

PC3 - Executive function / Choice The third component is most strongly loaded onto all three parameters from the Set shifting task, as well as "Variability" in the One-back and Two-back tasks. This likely represents a form of tradeoff between speed/accuracy and variability in these tests, and may reflect a common metric of "Executive function".

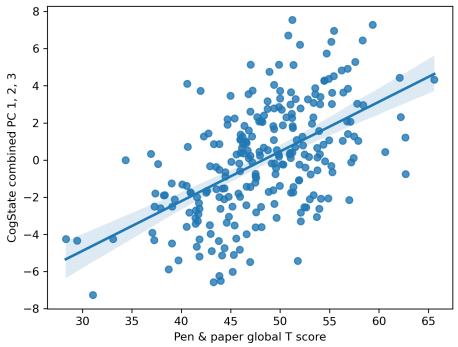
PC4 - Precision-Speed tradeoff The fourth component shows a similar pattern as PC2, and is most strongly loaded in opposite directions onto Detection task "Speed" vs. Detection and Identification "Variability" and "Accuracy". This may represent a tradeoff between speed and precision with these two, predominantly processing speed based, tasks.

PC5/PC6 - Other variability/accuracy tradeoffs? Noise? PC5 and PC6 show strong loading onto a smaller number of metrics from disparate tasks, involving opposite-direction loadings onto "Variability" and "Accuracy" metrics. These do not suggest a clear or intuitive explanation, and may reflect some other kind of tradeoff between variability and accuracy in performing these tasks. They represent a small proportion of explained variability in these tasks, and may reflect noise.

In comparison with pen&paper testing, selecting the first 3 principal components improved both the GDS classification performance and correlation shown in Figures 3b and 3c. The area under the ROC curve for classification of low cognitive performance improved to 0.78 (Figure 7a) and the strength of correlation with pen&paper global T score improved to 0.55 (Figure 7b).



(a) Receiver operating characteristic curve showing performance of CogState global T score from primary outcome measures against GDS classification of cognitive performance.



(b) Scatter plot of CogState vs. pen&paper global T scores, with linear regression trendline, and 95% confidence interval of the regression estimate.

4.6 Practice effects - How does CogState practice effect compare to pen&paper?

Aim to assess magnitude of practice effect (ie. People without HIV BL vs FU). Compare with PnP (which had quite a large practice effect). Anna suggested using reliable-change index.

No significant practice effects (paired one-sided t-test) with CogState tests (Table2). Several pen & paper tests showed significant improvement, likely a practice effect, at the followup timepoint (Table 3). Neither CogState nor pen & paper global rating showed significant practice effect. Reliable change index was calculated to count the number of individuals who could be classified as having improved. Unclear if this adds a lot to this analysis.

Task	Mean delta (Visit 2 - Visit 1)	n reliably improved	pvalue
Detection speed (Processing speed)	-3.11	5.00	0.95
Identification speed (Attention)	-0.14	6.00	0.54
One-card learning speed (Visual learning)	-0.64	4.00	0.66
One-card learning accuracy (Visual learning)	2.59	7.00	0.07
One-back accuracy (Working memory)	-0.98	2.00	0.76
Two-back accuracy (Working memory)	1.40	3.00	0.11
Set-shifting accuracy (Executive function)	-1.29	5.00	0.75

Table 2: Practice effects in CogState tests. Positive delta (red highlight) means better performance at followup. Negative delta (blue highlight) means worse performance at followup.

Task	Mean delta (Visit 2 - Visit 1)	n reliably improved	pvalue
Grooved Pegboard Test	4.33	3.00	**0.01**
Finger Tapping test	-1.36	2.00	0.80
HVLT-R - Total learning	5.02	9.00	**0.00**
HVLT-R - Delayed recall	2.76	10.00	**0.05**
BVMT-R - Total learning	4.16	9.00	**0.00**
BVMT-R - Delayed recall	1.77	4.00	0.09
Digit span	3.78	8.00	**0.01**
Digit symbol	2.85	6.00	**0.02**
Symbol search	0.14	4.00	0.47
CTT1 - Completion time	2.39	6.00	0.09
CTT2 - Completion time	2.19	4.00	0.08
WCST - total score	2.77	7.00	**0.05**
Fluency - animals	1.58	7.00	0.20
Fluency - fruit&veg	2.30	4.00	0.07
mean	2.48	5.93	0.14

Table 3: Practice effects in pen&paper tests. Positive delta (red highlight) means better performance at followup. Negative delta (blue highlight) means worse performance at followup.

5 Discussion

We applied CogState in a low-income South African setting demographically representative of the wider population of people living with HIV in southern Africa, with the aim of investigating its feasibility, confounding by technology familiarity, performance against P&P testing. We further performed principal component analysis to infer what underlying cognitive properties might be assessed by CogState.

We found that delivering CogState was feasible in this sociodemographic context. The majority of participants had little access to computers, but a large majority owned a mobile phone (most of these with touchscreens). We found slightly better global CogState performance in participants who owned a computer or a smartphone with a touchscreen.

CogState performed moderately in comparison to gold-standard pen&paper GDS classification (AUROC 0.7). It had good negative predictive value (0.84), but poor sensitivity (0.52) and positive predictive value (0.40). This cohort had relatively good cognitive performance and well-controlled HIV in relation to previous work.

Data-driven approaches may yield further insight into computerised cognitive testing data. Using PCA, we showed generalised "Speed" and "Accuracy" features that cut across multiple CogState tasks, as well as possible "Executive Function" or "Speed/Accuracy Tradeoff". More granular computerised characterisation of cognition may require a wider range of tasks that engage other cognitive domains. A simple Low/High cognitive performance dichotomy may not be the best way to validate computerised cognitive tests versus pen&paper tests - they are both indirect measures of common cognitive features that cannot be directly measured.

This study is limited by the relatively restricted set of CogState tasks in the brief battery. Similar data-driven analysis across larger datasets combining multiple sites could add robustness to these findings.

Therefore, a brief CogState battery might feasibly "rule-out" low cognitive performance in low-income settings, as well as characterising broad cognitive characteristics. This work suggests avenues for further development in computerised cognitive testing, aiming for more granular characterisation across more cognitive domains, and more sophisticated categorisation/characterisation of cognitive performance.

References

1. Seeher K, Cataldi R, Dua T, and Kestel D. Inequitable Access to Dementia Diagnosis and Care in Low-Resource Settings – A Global Perspective. Clinical Gerontologist 2023;46:133–7.
2. Cataldi R, Chowdhary N, Seeher K, Moorthy V, and Dua T. A Blueprint for the Worldwide Research Response to Dementia. The Lancet Neurology 2022;21:690–1.
3. Mwangala PN, Newton CR, Abas M, and Abubakar A. Screening Tools for HIV-associated Neurocognitive Disorders among Adults Living with HIV in Sub-Saharan Africa: A Scoping Review. AAS Open Research 2019;1:28.
4. Robbins RN, Brown H, Ehlers A, et al. A Smartphone App to Screen for HIV-Related Neurocognitive Impairment. Journal of mobile technology in medicine 2014;3:23–6.
5. Joska JA, Witten J, Thomas KG, et al. A Comparison of Five Brief Screening Tools for HIV-Associated Neurocognitive Disorders in the USA and South Africa. AIDS and behavior 2016;20:1621–31.
6. Kamminga J, Cysique LA, Lu G, Batchelor J, and Brew BJ. Validity of Cognitive Screens for HIV-Associated Neurocognitive Disorder: A Systematic Review and an Informed Screen Selection Guide. Current HIV/AIDS Reports 2013;10:342–55.
7. Wilson S, Milanini B, Javandel S, Nyamayaro P, and Valcour V. Validity of Digital Assessments in Screening for HIV-Related Cognitive Impairment: A Review. Current HIV/AIDS Reports 2021;18:581–92.
8. Penheiro R, Webber TA, Kiselica AM, and Woods SP. Executive Functions Are Independently Associated with Cognitive Dispersion in HIV Disease. Archives of Clinical Neuropsychology 2024:acae073.

9. Ridgely NC, Woods SP, Webber TA, Mustafa AI, and Evans D. Cognitive Intra-individual Variability in the Laboratory Is Associated With Greater Executive Dysfunction in the Daily Lives of Older Adults With HIV. Cognitive and Behavioral Neurology 2024;37:32.
10. De Francesco D, Underwood J, Anderson J, et al. Correlation between Computerised and Standard Cognitive Testing in People with HIV and HIV-negative Individuals. AIDS Care 2021;33:1296–307.
11. Underwood J, De Francesco D, Post F, et al. Associations between Cognitive Impairment and Patient-Reported Measures of Physical/Mental Functioning in Older People Living with HIV. HIV Medicine 2017;18:363–9.
12. Chaganti J, Gates TM, and Brew BJ. Reversible Large-Scale Network Disruption Correlates with Neurocognitive Improvement in HIV-associated Minor Neurocognitive Disorder with Combined Anti-Retroviral Therapy Intensification: A Prospective Longitudinal Resting-State Functional Magnetic Resonance Imaging Study. Neurological Sciences 2023;44:3261–9.
13. Yechoor N, Towe SL, Robertson KR, Westreich D, Nakasujja N, and Meade CS. Utility of a Brief Computerized Battery to Assess HIV-associated Neurocognitive Impairment in a Resource-Limited Setting. Journal of NeuroVirology 2016;22:808–15.
14. Nightingale S, Dreyer AJ, Thomas KGF, et al. Cognitive Performance, Neuropsychiatric Symptoms, and Cerebrospinal Fluid Viral Control Following Programmatic Switch from Efavirenz-Based to Dolutegravir-Based Antiretroviral Therapy in South Africa (CONNECT): A Prospective Cohort Study. The Lancet HIV 2024;11:e680–e689.

6 Supplementary materials

6.1 Cross-correlation of individual CogState and Pen&Paper tests

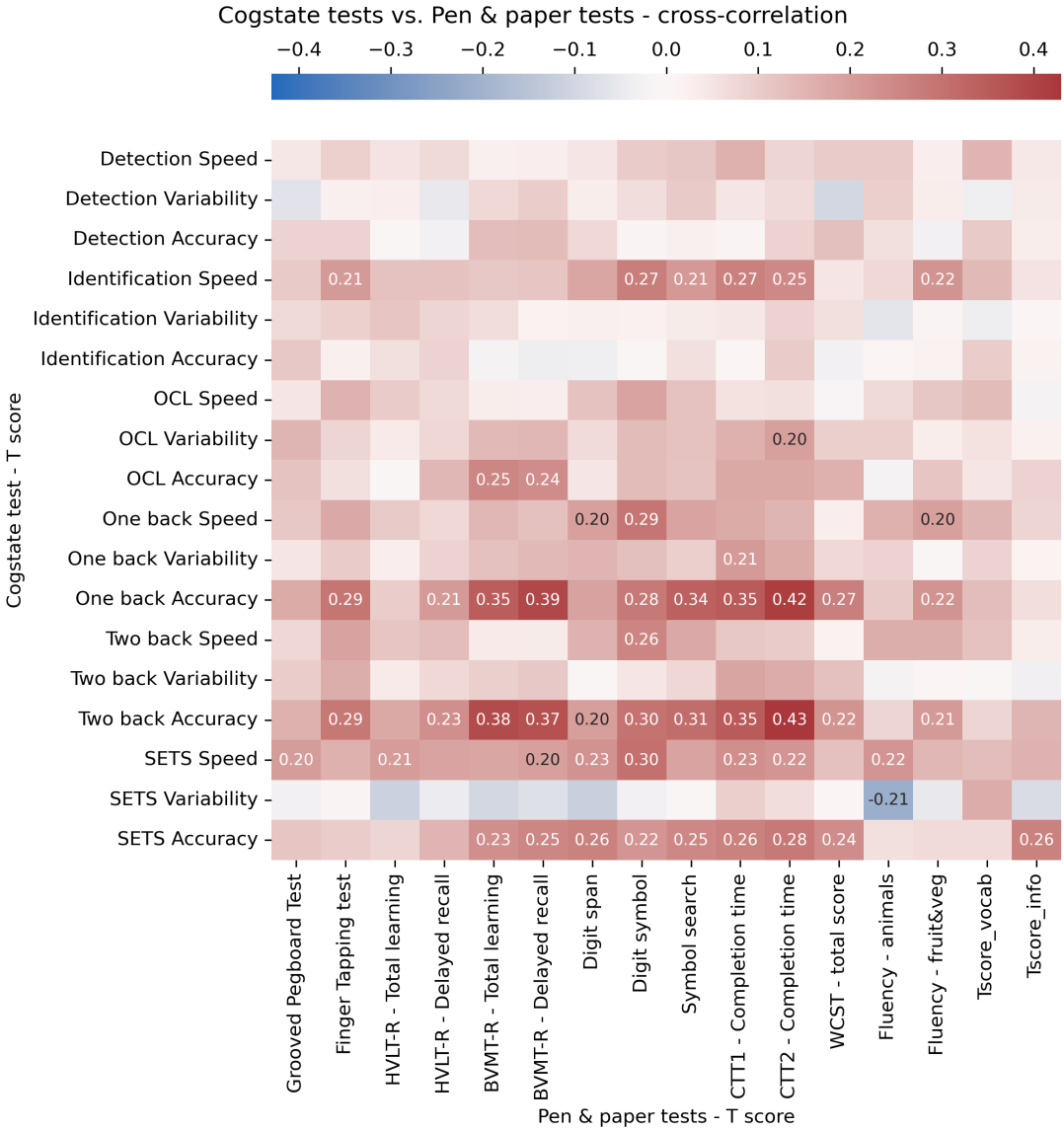


Figure 8: CogState tasks vs P&P tasks.

6.2 PCA screeplot

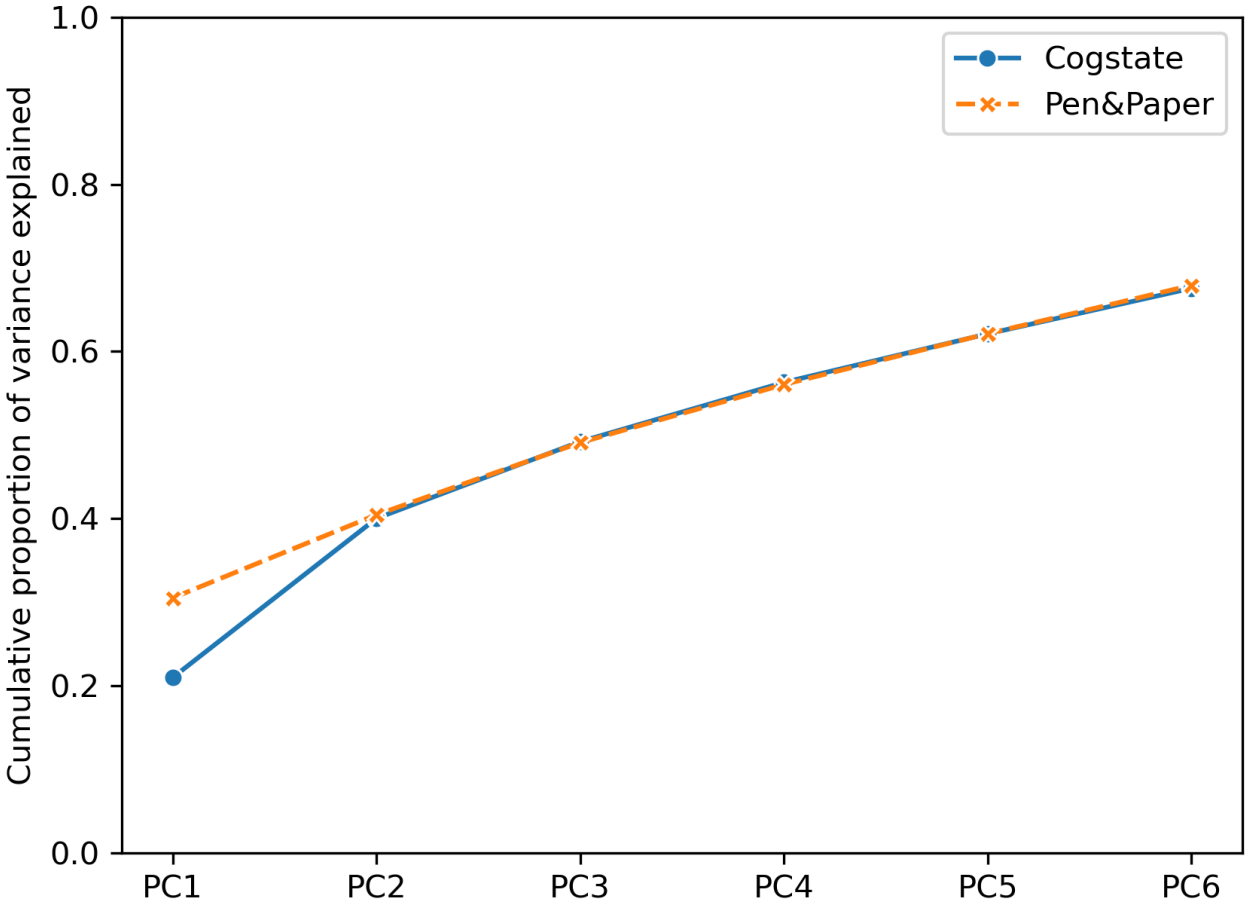


Figure 9: Plot of cumulatively explained variance by number of principal components for both Cogstate and P&P data.

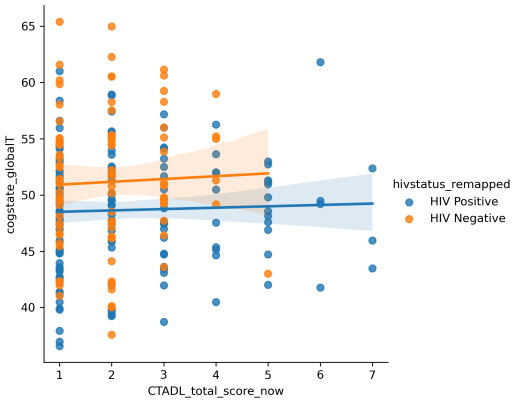
6.3 Difficulty of individual CogState tasks

		Missing	Overall	CogState Gl
			261	
find it to understand the instructions given to you in the tasks today?, n (%)	Difficult	0	7 (2.7)	
	Somewhat difficult		34 (13.0)	
	Somewhat easy		53 (20.3)	
	Easy		135 (51.7)	
	Very easy		32 (12.3)	
ection task, n (%)	No	0	256 (98.1)	
	Yes		5 (1.9)	
ntification task, n (%)	No	0	260 (99.6)	
	Yes		1 (0.4)	
e card learning task, n (%)	No	0	253 (96.9)	
	Yes		8 (3.1)	
e back task, n (%)	No	0	257 (98.5)	
	Yes		4 (1.5)	
o back task, n (%)	No	0	198 (75.9)	
	Yes		63 (24.1)	
-shifting task, n (%)	No	0	109 (41.8)	
	Yes		152 (58.2)	
tasks, n (%)	No	0	196 (75.1)	
	Yes		65 (24.9)	

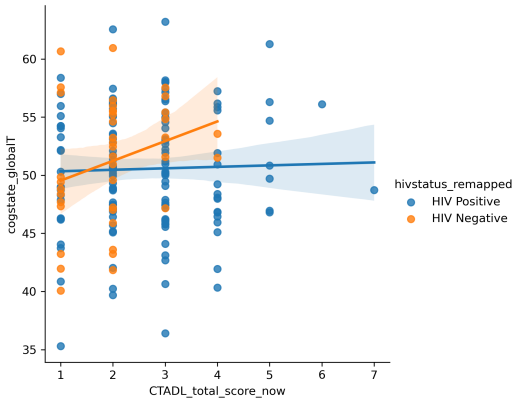
Figure 10: Survey data regarding instructions and difficulty of individual tasks.

6.4 Correlation to functional measures (PAOFI and CTADL)

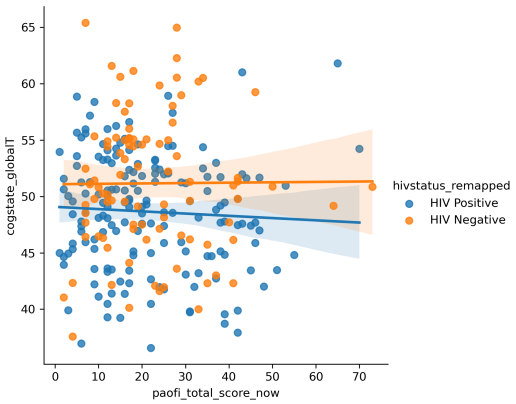
There was no clear correlation between baseline CTADL and PAOFI scores and CogState global T scores. At followup also, for people living with HIV there was no correlation. Numbers are limited at followup for people without HIV. Similar absence of effect to the previous paper. Could be included in supplementary materials for completeness.



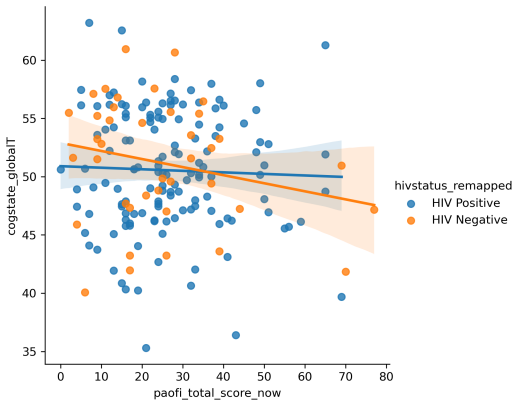
(a) Correlation between CTADL total score and CogState global T score, split by HIV status - baseline



(b) Correlation between CTADL total score and CogState global T score, split by HIV status - followup



(a) Correlation between PAOFI total score and CogState global T score, split by HIV status - baseline



(b) Correlation between PAOFI total score and CogState global T score, split by HIV status - followup