Correlation between computerised and standard cognitive testing in people with HIV and HIV-negative individuals

Davide De Francesco (1), Jonathan Underwood (2, 3), Jane Anderson (4), Marta Boffito (5), Frank A. Post (6), Memory Sachikonye (7), Patrick W.G. Mallon (8), Lewis Haddow (1, 9), Jaime H. Vera (10), Ken M. Kunisaki (11, 12), Caroline A. Sabin (1) and Alan Winston (2) for the POPPY study group

(1) Institute for Global Health, University College London, London, UK;

(2) Division of Infectious Diseases, Imperial College London, London, UK;

(3) Division of Infection and Immunity, University of Cardiff, Cardiff, UK;

(4) Homerton University Hospital, London, UK;

(5) Chelsea and Westminster Healthcare NHS Foundation Trust, London, UK;

(6) King's College Hospital NHS Foundation Trust, London, UK;

(7) UK Community Advisory Board, London, UK;

(8) University College Dublin School of Medicine, Dublin, Ireland;

(9) Kingston Hospital NHS Foundation Trust, London, UK;

(10) Brighton and Sussex Medical School, Brighton, UK;

(11) Minneapolis Veterans Affairs Health Care System, Minneapolis, USA;

(12) University of Minnesota, Minneapolis, USA.

Corresponding author

Dr. Davide De Francesco

Institute for Global Health, UCL, Royal Free Campus, Rowland Hill Street, London, NW3 2PF.

Email: d.defrancesco@ucl.ac.uk

Telephone: +44 20 7794 0500 ext. 38827

Abstract

We investigated the correlations and agreement between cognitive assessments made using a computerised (CogState™, 6 domains) and a standard pen-and-paper battery (5 domains) in PWH and lifestyle-similar HIV-negative individuals. Demographically-adjusted domain and global T-scores were obtained and used to define cognitive impairment according to the multivariate normative comparison (MNC) criteria. Correlations between T-scores and the agreement between the classifications of cognitive impairment obtained from the two batteries were assessed using the Spearman's rank correlation and Cohen's κ, respectively. The correlation between global T-scores from the two batteries was 0.52 (95% CI 0.44-0.60) in PWH and 0.45 (0.29-0.59) in controls (p=0.38 for their difference). Correlations were generally stronger between domains within the same battery than between those from different batteries. The agreement between the two batteries in classifying individuals as cognitively impaired or not impaired was fair in PWH (κ=0.24) and poor in HIV-negative individuals (κ=-0.02). The moderate correlation between overall cognitive function and the modest agreement between binary classifications of cognitive impairment obtained from two different batteries indicate the two batteries may assess slightly different components of cognition.

Keywords: HIV; cognitive impairment; CogState; cognitive battery; computerised cognitive battery;

Running head: Correlation between cognitive batteries in PWH and HIV-negative individuals

Introduction

Despite effective and sustained antiretroviral treatment, high rates of cognitive disorders continue to be reported in people with HIV (PWH) (Alford & Vera, 2018; Schouten et al., 2011). Timely recognition of cognitive impairment is the first step towards effective management and treatment; therefore, cognitive assessment is an important clinical and research evaluation for diagnosing, managing and studying HIV-associated cognitive impairment in PWH. Current guidelines recommend the use of batteries of tests across several cognitive domains in PWH reporting complaints of cognitive problems without obvious confounding conditions (EACS, 2019). In research settings, assessment of cognitive function in PWH is instrumental to identify determinants of HIV-associated cognitive impairment and decline, and to understand the biological mechanisms underlying cognitive problems. Different cognitive batteries exist, including standard 'pen-and-paper' and computerised batteries. Compared to standard batteries, computerised assessment of cognitive function may offer a more uniform administration across participants, have automated scoring with a high degree of accuracy, may be a more rapid assessment and results may be less affected by language or cultural background (Wesnes, 2014). On the other hand, computerised testing could be more challenging for people unfamiliar with computers and the use of computer hardware (e.g. keyboard and mouse), potentially introducing systematic bias when comparing PWH from different cultural and socio-economic backgrounds with different propensities towards the use of computer-based technologies.

Both standard and computerised cognitive batteries have been extensively used to assess cognitive function in cohorts of PWH (Cysique et al., 2006; Garvey et al., 2011; Heaton et al., 2010; McDonnell et al., 2014; Schouten et al., 2016). A comparison of standard and computerised batteries is critical for our understanding of the degree of variability attributable to the battery used to assess cognitive function. Several studies have evaluated how performance on a computerised battery is related to performance on a traditional cognitive battery in the general population (de Jager et al., 2009; Kataja et al., 2017; Kuiper et al., 2017; Mielke et al., 2015) or in populations with specific neurodegenerative

disorders (Gagnon & Laforce Jr, 2016; Hammers et al., 2012; Maruff et al., 2009). To our knowledge, two studies have compared the prevalence of cognitive impairment obtained with a computerised battery to that obtained using a standard battery in PWH. Cysique et al reported the same prevalence of 62% among 60 PWH with advanced HIV infection (Cysique et al., 2006), regardless of the battery used, with, however, up to 23% (14/60) of PWH classified as 'impaired' according to one but not the other battery. In another study of 53 PWH, the prevalence of cognitive impairment was 51% when using a standard pen-and-paper battery and 43% when using a computerised battery, with 76% of PWH with cognitive impairment according to both batteries (Bloch et al., 2016). Moreover, Overton et al. reported only moderate correlations (ranging from 0.15 to 0.52) between the performance on tests of a computerised battery and an overall measure of cognitive function obtained from a standard battery in a sample of 46 PWH (Overton et al., 2011). These studies included high proportions of PWH with known cognitive disorders and had relatively small sample sizes. Data on the utility of computerised and standard cognitive batteries in larger contemporary cohorts of PWH with higher rates of viral suppression are lacking and little is known regarding any differences which may be presented when compared to lifestyle-similar HIV-negative individuals.

Our aims here were to evaluate the agreement between a computerised and a standard pen-and-paper battery in identifying PWH with cognitive impairment and to assess the correlation between cognitive function scores obtained from the two cognitive batteries in PWH and HIV-negative individuals with similar lifestyles.

Methods

Study participants

The Pharmacokinetic and Clinical Observations in People Over Fifty (POPPY) study is a prospective, observational cohort study of PWH and HIV-negative controls with similar lifestyles. Full details have been described previously (Bagkeris et al., 2018). Briefly, a group of PWH aged ≥50 years was recruited from eight HIV outpatient clinics in London and Brighton (United Kingdom), and in Dublin (Ireland).

Another group of PWH aged between 18 and 50 years was also recruited from the same HIV clinics and was frequency matched on gender, ethnicity, sexual orientation and location (in or out of London) to the group of PWH aged ≥50 years. Inclusion criteria for both groups of PWH were: documented presence of HIV infection, white or black-African ethnicity, likely route of HIV acquisition via sexual exposure and ability to comprehend the study information leaflet. In addition, a group of HIV-negative individuals aged ≥50 years was recruited from sexual health centres affiliated with the HIV clinics and was frequency matched to the group of PWH aged ≥50 years on gender, ethnicity, sexual orientation and location (in or out of London). Recruitment and first study visit took place between April 2013 and January 2016; a follow up visit was also conducted after approximately two years between June 2015 and May 2018. At both study visits, participants underwent detailed cognitive assessment using a computerised battery as reported in the following section.

As part of a sub-study to investigate sleep disorders in PWH, a subset of participants in each study group was recruited to undergo additional assessments (Kunisaki et al., 2020) including cognitive function using a standard pen-and-paper battery, as detailed in the following section. In order to participate in this sub-study, participants had to be able to wear a fingertip oximetry device and wrist actigraph for a week and to adhere to study procedures (according to the investigator's judgement). Study visits took place between March 2017 and July 2018. Here we present the analyses based on the subset of POPPY PWH and HIV-negative controls participating in this sub-study who completed the assessment of cognitive function for both the sleep sub-study and the main POPPY study (at the two-year follow up POPPY visit, closest to the sleep sub-study visit). The study was approved by the UK National Research Ethics Service (NRES; Fulham, London; UK number 12/LO/1409). All participants provided written informed consent.

Cognitive batteries

As part of the main POPPY study, participants underwent assessment of cognitive function using the CogState™ (Melbourne, Australia) computerised battery. The battery consisted of 10 different tests

covering six cognitive domains as reported in Table 1. Raw test scores were log-transformed or arcsine root-transformed as recommended by the CogState guidelines for analysis. Integrity and quality checks were applied to ensure that scores were generated from completed and fully understood tasks for each participant. Individual test scores not meeting integrity and quality checkswere excluded from the analysis. Individual test scores were converted into T-scores (with a mean of 50 and a standard deviation of 10) using CogState pediatric and adult normative data (Cogstate, 2017). Domain T-scores were obtained by averaging individual test T-scores as indicated in Table 1, and a global T-score of overall cognitive function was obtained by averaging T-scores across the six domains. For all T-scores, a higher value indicates better cognitive function.

Participants also completed a standard 'pen-and-paper' cognitive battery of 9 tests covering five domains (Table 1) administered by trained research staff. Tests and domains were purposely selected to cover major cognitive domains and to take approximately one hour to complete. Similarly to the scores of the computerised battery, individual test scores were converted into T-scores using appropriate normative data, with higher T-scores representing better cognitive function. Domain and global T-scores were obtained as average across test and domain T-scores, respectively.

Definition of cognitive impairment

The multivariate normative comparison (MNC) criterion (Huizenga et al., 2007) was applied to domain T-scores obtained from each battery, separately, to define cognitive impairment. The MNC simultaneously compares domain T-scores of each study participant to the average domain T-scores in the control group (in our case the HIV-negative group), taking into account the variances and covariance between T-scores. For each participant, a continuous measure of the deviation of the participant's cognitive profile from the average cognitive profile in the control group is then obtained. If this deviation exceeds a critical value associated with a 5% significance, the individual is classified as cognitively impaired (so that the false positive rate is approximately 5%). This definition of cognitive impairment was selected because the resulting prevalence of cognitive impairment does not depend

on the number of tests/domains tested, unlike other commonly used definitions such as the HIV-associated neurocognitive disorder classification, also known as Frascati criteria (Antinori et al., 2007), and the global deficit score (Carey et al., 2004) criteria (Underwood et al., 2018). As the computerised and standard batteries covered a different number of tests (i.e. 6 and 5, respectively), the agreement between the two batteries in terms of classification of individuals as impaired or not impaired is not artificially biased by the number of domains tested.

Statistical analysis

Continuous variables, including cognitive T-scores, were summarized using the median and the interquartile range (IQR); categorical variables were described using frequencies and percentages. Comparisons of sociodemographic characteristics, lifestyle, cognitive T-scores and prevalence of cognitive impairment between PWH and HIV-negative individuals were carried out using χ^2 and Wilcoxon rank-sum tests as appropriate. The agreement between the classification of cognitive impairment based on the computerised and standard battery was assessed using Cohen's κ statistics (Cohen, 1960) and interpreted following Landis and Koch (Landis & Koch, 1977) guidelines.

The correlation between cognitive T-scores, within and between domains from the two batteries, was assessed using the Spearman's correlation coefficient (r_s) and interpreted according to published guidelines as poor (absolute value of r_s between 0 and 0.2), fair (0.2-0.5), moderate (0.5-0.7) or very strong (0.7-1) (Chan, 2003). Correlations observed in PWH were compared to those seen in HIV-negative individuals using the z-test after Fisher's transformation (Fisher, 1992). Sensitivity analyses were conducted to verify whether correlations differed by time between cognitive assessments (≤3 months vs. >3 months) and by age (<50 years vs. ≥50 years) in PWH and HIV-negative controls combined. Due to the large number of statistical tests and to reduce the risk of false positive findings, p-values were adjusted for multiple testing using the Holm–Bonferroni method (Holm, 1979). In addition, the Bland–Altman plot (Bland & Altman, 1999) was used to assess the agreement between the two global T-scores obtained from the two batteries. Briefly, this graphically represent the

relationship between the two global T-scores by plotting the difference between the two global T-scores (computerised minus standard battery) against the average global T-score across the two batteries.

A cluster analysis of domain T-scores from both batteries was performed to identify groups of domains strongly related to each other and thus bringing similar information. A hierarchical clustering algorithm specifically developed to cluster variables (rather than observations) was used (Chavent et al., 2011) to find clusters in the 11 domain T-scores (the six domain T-scores obtained from the computerised battery plus the five domain T-scores obtained from the standard battery) plus the two global T-scores. The purpose of these analyses was to objectively assess similarities and differences between components of the two batteries using a data-driven approach that was 'agnostic' as to how they were performed. All analyses were performed using R v3.5.1 (the package 'ClustOfVar' was used to perform the cluster analysis (Chavent et al., 2011)), with p-values <0.05 considered as statistically significant.

Results

Participant characteristics

A total of 429 individuals (317 PWH and 112 HIV-negative) completed both the computerised and standard cognitive batteries with socio-demographic, lifestyle and HIV-related characteristics reported in Table 2. Compared to the HIV-negative individuals, PWH were younger [median (IQR) age 57 (51, 63) vs. 61 (57, 66) years, p<0.001] as expected due to the study design, more likely to be male (88.3% vs. 67.9%, p<0.001), men who have sex with men (82.0% vs. 45.5%, p<0.001) and to report current recreational drug (27.4% vs. 14.3%, p=0.005) and current/previous injection drug use (10.2% vs. 1.8%, p=0.005). A greater proportion of HIV-negative individuals (92.9%) reported current alcohol use compared to PWH (81.7%, p=0.005).

PWH had been diagnosed with HIV for a median (IQR) of 17.7 (11.0, 24.6) years previously, and 97.1% had a HIV RNA <40 copies/mL with a median (IQR) $CD4^+$ cell count of 624 (495, 850) cells/µL.

Domain and global T-scores in PWH and HIV-negative individuals

According to both batteries, global T-scores were approximately two points lower in PWH compared to HIV-negative individuals ($p<0.001$ for the computerised and $p=0.003$ for the standard battery, Table 2). In particular, PWH showed poorer performances than HIV-negative individuals in the psychomotor ($p<0.001$), visual attention ($p<0.001$), executive function ($p=0.03$) and verbal learning ($p=0.01$) domains tested using the computerised battery and the language ($p=0.005$), processing speed ($p<0.001$) and executive function ($p=0.003$) domains tested with the standard battery.

Classification of cognitive impairment: agreement between the two batteries

The prevalence of cognitive impairment was 15.8% in PWH vs. 9.8% in HIV-negative individuals according to the computerised battery ($p=0.16$) and 5.4% vs. 0.9% according to the standard battery ($p=0.05$). Among PWH, 10 were classified as cognitively impaired according to both batteries, 7 were classified as impaired based on T-scores of the standard battery but not on those of the computerised battery, whereas 40 were classified as impaired according only to the computerised battery. Cohen's κ statistic (95% confidence interval) was 0.24 (0.10, 0.38), indicating fair agreement between the two batteries in PWH.

In contrast, agreement between the two batteries in classifying HIV-negative individuals as cognitively impaired or not impaired was poor [Cohen's κ (95% confidence interval) = -0.02 (-0.05, 0.01)]. None of the HIV-negative individuals had cognitive impairment according to both batteries, with 12 participants being classified as cognitively impaired based on T-scores from one battery but not the other (11 classified by the computerised battery alone, 1 by the standard battery).

Global score of cognitive function: correlation between the two batteries

The correlation between the two global T-scores was fair in both PWH [r_s (95% CI) = 0.52 (0.44, 0.60)] and HIV-negative individuals [r_s (95% CI) = 0.45 (0.29, 0.59)] and did not differ between the two groups (p=0.38). Bland-Altman plots of the differences between the two global T-scores and their average, separately in PWH and HIV-negative individuals, are presented in Figure 1. Among PWH, the mean (95% CI) difference in global T-score was -2.2 (-3.0, -1.4) indicating, on average, lower T-scores with the computerised battery (p<0.001). In addition, the difference between global T-scores was negatively correlated with the average global T-score across batteries [r_s (95%) = -0.26 (-0.36, -0.16), p<0.001], suggesting larger differences (i.e. poorer agreement) in PWH with poorer cognitive scores. The lower and upper limits of agreement (95% CI) were -15.9 (-17.2, -14.6) and 11.5 (10.2, 12.8), respectively, indicating substantial discrepancies between the T-scores obtained from the two batteries.

A similar pattern was observed among HIV-negative individuals, with a mean (95% CI) difference between global T-scores of -1.9 (-3.0, -0.8) and a negative correlation between the difference and average global T-scores [r_s (95%) = -0.29 (-0.45, -0.11), p<0.001]. Limits of agreement (95% CI) were -13.5 (-15.4, -11.7) and 9.7 (7.8, 11.5) indicating substantial discrepancies between the T-scores obtained from the two batteries.

Correlations between scores within and between the two batteries

Correlations between cognitive T-scores within and between the two batteries are displayed in Figure 2, separately in PWH and HIV-negative individuals. P-values testing the hypothesis that the correlation between each pair of T-scores in PWH is different to that seen in HIV-negative individuals are reported in Table 3. Among PWH, all the six domain T-scores of the computerised battery were moderately or strongly correlated with the global T score, with r_s ranging from 0.57 (verbal learning) to 0.69 (visual attention). Correlations between domains tested with the computerised battery were generally fair to moderate ranging from 0.15 to 0.57; psychomotor with visual attention (r_s = 0.57) and visual

learning with executive function (r_s = 0.57) showed the strongest correlations. Correlations seen in PWH were generally similar to those observed in HIV-negative controls, however few correlations involving the psychomotor and visual attention domains appeared to differ between the two groups ($p<0.05$, Table 3).

Within the standard battery, correlations of the five domains with the global T-score were strong, ranging from r_s = 0.65 for motor function to r_s = 0.90 for processing speed, among PWH, and from r_s = 0.44 to r_s = 0.81 for the same domains among HIV-negative individuals (Figure 2). Among PWH, the correlations between the five domains were between fair and moderate, ranging from 0.21 to 0.65. Executive function and processing speed showed the strongest association (r_s = 0.65), followed by attention and processing speed, whereas motor function and language showed the weakest correlation (r_s = 0.21). Few correlations appeared to differ between PWH and HIV-negative individuals, especially those involving the attention and motor function domains (Table 3). In particular, the correlation between attention and executive function (r_s = 0.48 in PWH and r_s = 0.19 in HIV-negative individuals) and that between processing speed and motor function (r_s = 0.46 vs. 0.14) were significantly stronger in PWH compared to HIV-negative individuals (p=0.003 and p=0.006, respectively).

Among PWH, the computerised global T-score was fairly correlated with all domain T-scores assessed with the standard battery (r_s from 0.41 to 0.47) with the exception of motor function (r_s = 0.13), with no differences compared to the correlations observed among HIV-negative individuals. Correlations of the standard global T-score with domain T-scores from the computerised battery ranged from 0.32 to 0.42 in PWH, with the correlation of the visual attention domain being significantly stronger in PWH than in HIV-negative individuals (r_s = 0.42 vs. 0.13, p=0.004). Correlations between domain T-scores from different batteries were generally weaker than those between domain T-scores from the same battery in both PWH and HIV-negative individuals. Among PWH, verbal learning (computerised battery) and language (standard battery) showed the strongest correlation (r_s = 0.42) between

domains from different batteries. Also the processing speed domain (standard battery) showed relatively strong correlations with the visual attention (r_s = 0.37), working memory (r_s = 0.36), verbal learning (r_s = 0.35) and executive function (r_s = 0.33) domains assessed by the computerised battery. The correlation between the two T-scores of executive function from the two batteries was weak in both PWH and HIV-negative individuals (r_s = 0.27 vs. 0.25, p=0.82). Among both PWH and HIV-negative individuals, the cluster analysis suggested two distinct groups, one including the T-scores from the computerised battery and one including those from the standard battery (Figure 2).

Sensitivity analyses

Correlations between T-scores between and within the two cognitive batteries did not appear to differ significantly in participants who completed the two assessments within 3 months compared to participants who completed the assessments more than 3 months apart (all p's > 0.05, Tables S.1-S.3 in the Supplementary material). Correlations in younger participants aged <50 years did not differ significantly from those observed in older participants aged ≥50 years (all p's > 0.05, Tables S.4-S.6 in the Supplementary material).

Discussion

Computerised and standard pen-and-paper cognitive batteries yield moderately correlated overall cognitive scores; however, correlations of domains within the same battery were generally stronger than those of domains from different batteries, suggesting that the two batteries may assess slightly different aspects of cognition in treated PWH. Whilst correlations of domains within and between the two batteries were generally similar in PWH and lifestyle-similar HIV-negative individuals, correlations involving the motor/psychomotor and attention/visual attention, regardless of the battery used to assess them, appeared to be different.

Our results are in line with previous studies conducted in the general population (Kataja et al., 2017; Kuiper et al., 2017; Mielke et al., 2015) and PWH (Overton et al., 2011) reporting significant

correlations but with only weak to moderate strength between cognitive performances obtained from a computerised and a standard battery. Scores of overall cognitive function showed a stronger correlation than individual domain scores; however, concordance was somewhat lower than expected, with the computerised battery systematically estimating poorer cognitive function than the standard battery and the disagreement being inversely proportional to the average score. In addition, the two batteries appeared to assess two distinct cognitive profiles, with stronger correlations between domains within the same battery than between domains from different batteries, even for theoretically similar domains. Consistent with our findings, other studies have reported poor one-to-one correspondence between domains assessed with computerised and standard cognitive assessments (Hammers et al., 2012; Kataja et al., 2017; Maruff et al., 2009; Mielke et al., 2015). The two batteries appear to assess slightly distinct aspects of cognition, which is not unexpected given the cognitive tests for each cognitive domain differ between the two batteries. Furthermore, different task requirements and processing modalities, as well as genuine differences between batteries in the components of cognitive function being assessed, may have contributed to these results.

We believe our study is the first to assess the differences in the correlations between and within cognitive batteries observed in PWH to those observed in HIV-negative individuals with similar socio-demographic characteristics and lifestyles. We report that correlations seen in PWH are similar to those seen in HIV-negative individuals, with the only differences being in the correlations involving the motor/psychomotor and attention/visual attention domains, which appear to be stronger in PWH than in HIV-negative individuals. Previous studies suggest these domains are those that are most affected in PWH on antiretroviral treatment with cognitive disorders (May et al., 2020), and we also report poorer performances in PWH, compared to HIV-negative individuals, in the motor function and visual attention domains (according to the computerised battery). Such findings suggest that both cognitive batteries assessed in our study are sensitive batteries for assessing the specific cognitive profiles of PWH. Moreover, differences between PWH and HIV-negative individuals in the correlations involving these domains can be a result of the different range of performance/impairment between

the two groups, whereby a greater impairment in one domain can affect the performance in other domains, resulting in stronger correlations.

Interestingly, regardless of the battery, PWH showed, on average, 2 points lower global T-scores than controls. This difference is similar to the difference in overall cognitive function between PWH and appropriately chosen HIV-negative individuals reported in other studies (i.e. a difference of approximately 2 and 0.2 points in T- and Z-scores, respectively) (Cole et al., 2017; D De Francesco et al., 2019; Davide De Francesco et al., 2016). Whilst the difference reported in our study and these other studies is statistically significant, typically such a difference in itself would not be considered clinically meaningful (Nakasujja et al., 2013). However, this difference in T-score is likely to represent a reduction in cognitive reserve and therefore could have a lower threshold to become clinically relevant in the future, in addition to age-related changes in cognition compared to individuals with a higher cognitive T-score.

The agreement between the binary classifications of cognitive impairment in PWH was only fair, indicating substantial disagreement on whether or not an individual should be classified as cognitively impaired, depending on the battery used. The two batteries seem to assess different aspects of cognition, as also supported by the observed pattern of correlations between domains, and this may have resulted in a less than satisfactory agreement, even when applying the same criterion to detect cognitive impairment. In the absence of a gold standard to define cognitive impairment, it is difficult to ascertain the validity of both batteries in correctly identifying truly impaired PWH. Nevertheless, these results suggest that these binary classifications may be significantly affected by intra-individual variability that is an intrinsic characteristic of most cognitive tests (Schretlen & Sullivan, 2013). This finding adds to the literature that invites caution when using binary classifications of cognitive impairment (Davide De Francesco et al., 2016; Underwood et al., 2018) and when interpreting the overall results of a single cognitive battery in both clinical settings and research studies.

Our study has some limitations. One is the time passed between one cognitive assessment and the other, which varied between 1 and 23 months across study participants, with a median of approximately 10 and 7 months in PWH and HIV-negative individuals, respectively. Whilst there were differences between the two groups in this timespan, the correlations between domains did not seem to be affected by the time between the two cognitive assessments, and therefore it is unlikely that this has had a significant impact in our analysis. Similarly, PWH and HIV-negative individuals appeared to differ with regards to some socio-demographic and lifestyle characteristics, including age, which is an important factor associated with cognition. Since the T-scores used were adjusted for age, gender, ethnicity and education, as appropriate, this should not have biased the comparison of cognitive function and the rate of impairment between the two groups. Moreover, sensitivity analyses suggested age did not affect the pattern of correlations within and between batteries. Another limitation is the lack of clinical parameters, such as the presence of non-infectious comorbidities or other risk factors for cognitive impairment, which may have assisted in determining which cognitive battery is of more clinical relevance. Nonetheless, our main aim was to determine the relationship between cognitive test results when assessed by two distinct batteries, rather than comparing the ability of the two batteries to detect clinically relevant cognitive disorders. Finally, whilst the standard battery used covered all major cognitive domains known to be affected by HIV, some domains were assessed by a smaller number of tests than what is conventionally used. There is no agreement regarding the optimal number of tests to adequately assess domain function, however the use of a more comprehensive standard battery with more tests per domain, would have provided a more robust and reliable assessment of each domain and, consequently, of correlations with domains assessed with the computerised battery.

Conclusions

Whilst we report a moderate correlation between overall cognitive function when evaluated using a computerised and a standard battery, the two batteries seem to assess slightly distinct components

of cognition in both treated PWH and HIV-negative individuals with similar lifestyles. The agreement between batteries in classifying individuals with and without impairment was only modest, with only a small proportion of PWH being classified as impaired by both batteries. These results highlight the differences between cognitive batteries and cognitive tests which are commonly used to diagnose cognitive impairment and to study the development and consequences of cognitive problems in PWH.

Acknowledgements

POPPY-Sleep Management Team: Ken Kunisaki, Susan Redline, Alan Winston, Caroline Sabin, Paddy Mallon, Nicki Doyle, Amalia Ndoutoumou. POPPY-Sleep Sleep Reading Centre Team: Emily Kaplan, Dan Mobley, Michael Rueschman, Michelle Reid (Brigham and Women's Hospital, Boston/USA). POPPY-Sleep Scientific Steering Committee: Jane Anderson, David Asboe, Marta Boffito, Lucy Garvey, Paddy Mallon, Frank Post, Anton Pozniak, Caroline Sabin, Memory Sachikonye, Jaime Vera, Ian Williams, Alan Winston. POPPY-Sleep methodology/statistics/analysis: Caroline Sabin, Davide De Francesco, Mike Rueschman. POPPY-Sleep Sites and Trials Unit: Caldecot Centre, King's College Hospital (Frank Post, Beatriz Santana Suárez, Lucy Campbell); Department of Infection and Population Health, University College London (Lewis Haddow, Michelle Beynon, Abigail Severn, Andrea Cartier, Hinal Lukha); Elton John Centre, Brighton and Sussex University Hospital (Jaime Vera, Sarah Kirk); HIV Molecular Research Group, School of Medicine, PPY-Sleep Management Team: Ken Kunisaki, Susan Redline, Alan Winston, Caroline Sabin, Paddy Mallon, Nicki Doyle, Amalia Ndoutoumou. POPPY-Sleep Sleep Reading Centre Team: Emily Kaplan, Dan Mobley, Michael Rueschman, Michelle Reid (Brigham and Women's Hospital, Boston/USA). POPPY-Sleep Scientific Steering Committee: University College Dublin (Paddy Mallon, Alan Macken, Sumesh Babu, Aoife McDermott); Homerton Sexual Health Services, Homerton University Hospital (Jane Anderson, Sam Pelluri, Anna Price); Imperial Clinical Trials Unit, Imperial College London (Amalia Ndoutoumou, Daphne Babalis); St. Mary's Hospital London, Imperial College Healthcare NHS Trust (Alan Winston, Lucy Garvey, Jonathan Underwood, Lavender Tembo, Matthew Stott, Linda McDonald, Felix Dransfield); St Stephen's Centre, Chelsea and Westminster Hospital (Marta Boffito, Michelle Anthonipillai, Peter Fernando, Shane Hardwick, Chido Chiwome, Candida Fernandez, Ana Milinkovic). POPPY-Sleep Funders: The POPPY-Sleep sub study is funded by the US National Heart, Lung, and Blood Institute (R01 HL131049); the main POPPY study is funded from investigator initiated grants from BMS, Gilead Sciences, Janssen, MSD and ViiV Healthcare. We acknowledge the use of the NIHR/Wellcome Trust Clinical Research Facility at King's College Hospital.

The research is also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London.

Conflicts of interest

JU has received funding from Gilead Sciences for membership of advisory boards and for preparation of educational materials. CAS has received funding from Gilead Sciences, ViiV Healthcare, and Janssen-Cilag for membership of data safety and monitoring boards, advisory boards, and speaker panels and for preparation of educational materials. FAP reports grants and/or personal fees from Gilead Sciences, ViiV, Janssen, and MSD. JHV has received honoraria and research grants in trials sponsored by Merck, Janssen Cilag, Piramal and Gilead sciences.

The views presented here are those of the authors and do not reflect the views of the US Government, US Department of Veterans Affairs, or the authors' affiliated institutions.

Author contributions

DDF contributed to study concept and design, analysis and interpretation of data, drafting of manuscript. JU contributed to study concept and design, interpretation of data, critical revision of manuscript. JA, MB, MS, PWGM, contributed to the acquisition of data. FAP, LH, JHV contributed to the acquisition of data, critical revision of manuscript. KMK contributed to study concept and design, critical revision of manuscript. CAS and AW contributed to study concept and design, interpretation of data, critical revision of manuscript. All authors read and approved the final manuscript.

Tables/Figures

Table 1: Individual tests of the computerised and standard battery and how they map into cognitive domains.

Cognitive domain	Test	Scoring system
Computerised battery		
Visual Learning	CPAL	Total number of errors across the seven rounds
	GML – delayed recall	Total number of errors made after a delay
	One card learning task	Arcsine of the square root of the proportion of correct responses
Psychomotor	Detection task	Mean of the $\log_{10}$ transformed reaction times for correct responses
Visual Attention	Identification task	Mean of the $\log_{10}$ transformed reaction times for correct responses
Executive Function	GML test	Total number of errors made in five consecutive trials
	Set shifting task	Total number of errors across the five rounds
Verbal Learning	International Shopping list	Total number of correct responses made in three consecutive trials
		Total number of correct responses made after a delay
Working Memory	One back task	Mean of the $\log_{10}$ transformed reaction times for correct responses
		Arcsine of the square root of the proportion of correct responses
	Two back task	Mean of the $\log_{10}$ transformed reaction times for correct responses
		Arcsine of the square root of the proportion of correct responses
Standard battery		
Attention	PASAT 3	Total correct summations
Executive function	Trail Making Test-B	Total time to complete
Language	Category Fluency	Total number of animals in 1 minute
	Letter Fluency	Total number of words, 1 minute for each of 3 letters
Motor function	Grooved pegboard	Dominant hand: Time to complete
	Grooved pegboard	Non-dominant hand: Time to complete
Processing speed	Trail Making Test-A	Time to complete
	WAIS-III Digit Symbol	Total correct symbols
	WAIS-III Symbol Search	Total correct symbols
	Stroop colour-word test	Number of items completed

Note: CPAL: Continuous paired associate learning; GML: Groton maze learning; PASAT: Paced auditory serial addition test; WAIS: Wechsler adult intelligence scale

Table 2: Socio-demographic, lifestyle and HIV-related characteristics in PWH and HIV-negative individuals at the time of cognitive assessment with the computerised battery (earlier in time). Results of the cognitive assessments are also reported.

Median (IQR) or n (%)	**PWH (n=317)**	**HIV-negative (n=112)**	**p-value**
Male gender	280 (88.3%)	76 (67.9%)	<0.001
White ethncity	290 (91.5%)	106 (94.6%)	0.28
Age [years]	57 (51, 63)	61 (57, 66)	<0.001
BMI [Kg/cm^2]	25.2 (23.4, 28.6)	25.9 (23.9, 29.3)	0.10
MSM	260 (82.0%)	61 (45.5%)	<0.001
University degree or above	148 (46.7%)	58 (51.8%)	0.86
Years of education	16 (12, 18)	16 (13, 19)	0.64
Current alcohol use	259 (81.7%)	104 (92.9%)	0.005
Current recreational drugs	87 (27.4%)	16 (14.3%)	0.005
Ever injected drugs	32 (10.2%)	2 (1.8%)	0.005
Time between testing [months]	10.1 (4.1, 17.1)	6.9 (1.2, 16.1)	0.01
Current CD4+ count [cells/µL]	624 (495, 850)	N/A	N/A
Nadir CD4+ count [cells/µL]	197 (99, 290)	N/A	N/A
Years since HIV diagnosis	17.7 (11.0, 24.6)	N/A	N/A
On antiretroviral treatment	292 (92.1%)	N/A	N/A
HIV RNA <40 copies/mL	306 (97.1%)	N/A	N/A
Cognitive function [computerised battery]			
Cognitive impairment	50 (15.8%)	11 (9.8%)	0.16
Visual Learning T-score	52.4 (46.5, 56.7)	52.4 (48.9, 58.1)	0.17
Psychomotor T-score	44.5 (36.1, 50.4)	49.8 (43.8, 54.1)	<0.001
Visual Attention T-score	42.7 (35.3, 49.5)	46.7 (41.8, 51.1)	<0.001
Executive Function T-score	51.7 (47.0, 55.4)	53.6 (48.4, 57.0)	0.03
Verbal Learning T-score	53.3 (47.6, 59.7)	56.3 (50.1, 61.6)	0.01
Working Memory T-score	48.8 (44.6, 53.0)	49.7 (46.0, 54.0)	0.09
Global T-score	48.9 (44.5, 52.1)	51.1 (47.9, 54.0)	<0.001
Cognitive function [standard battery]			
Cognitive impairment	17 (5.4%)	1 (0.9%)	0.05
Language T-score	50.0 (43.3, 58.1)	52.8 (48.4, 59.8)	0.005
Attention T-score	45.5 (33.5, 52.7)	44.6 (28.6, 53.7)	0.47
Processing speed T-score	50.9 (44.8, 55.7)	53.5 (48.3, 57.3)	<0.001
Executive Function T-score	51.8 (43.3, 60.0)	55.2 (49.4, 62.5)	0.003
Motor function T-score	50.5 (42.6, 57.3)	51.9 (43.4, 58.1)	0.51
Global T-score	49.9 (44.0, 54.9)	52.0 (48.3, 55.6)	0.003

Table 3: Holm-Bonferroni adjusted p-values testing the difference in the correlation of domain T-scores between PWH and HIV-negative individuals. (C) indicates domains of the computerised battery, (S) domains of the standard battery.

	Visual Learning (C)	Psychomotor (C)	Visual Attention (C)	Executive Function (C)	Verbal Learning (C)	Working Memory (C)	Global T-score (C)	Language (S)	Attention (S)	Processing Speed (S)	Executive Function (S)	Motor Function (S)	Global T-score (S)
Visual Learning (C)		0.16	0.01	0.70	0.06	0.30	0.26	0.79	0.91	0.17	0.06	0.17	0.82
Psychomotor (C)			0.04	0.18	0.03	0.31	0.01	0.35	0.10	0.41	0.06	0.14	0.08
Visual Attention (C)				0.40	0.25	0.14	0.06	0.21	0.05	0.67	0.21	0.02	0.07
Executive Function (C)					0.09	0.97	1.00	0.86	0.83	0.53	0.82	0.77	0.99
Verbal Learning (C)						0.47	0.14	0.22	0.15	0.32	0.19	0.02	0.004
Working Memory (C)							0.58	0.74	0.69	0.63	0.72	0.98	0.74
Global T-score (C)								0.44	0.63	0.34	0.17	0.37	0.38
Language (S)									0.26	0.55	0.44	0.20	0.70
Attention (S)										0.01	0.003	0.51	0.14
Processing speed (S)											0.42	0.006	0.16
Executive Function (S)												0.01	0.24
Motor function (S)													0.02
Global T-score (S)													

Figure 1: Bland-Altman plot

Figure 2: Correlations (r_s) between domains and dendrogram resulting from the cluster analysis in PWH and HIV-negative individuals. (C) indicates domains of the computerised battery, (S) domains of the standard battery. A cross indicates that the correlation is not statistically significant (Holm-Bonferroni adjusted p-value >0.05).

Note: The order of domains is determined by the dendrogram and differ between PWH and HIV-negative individuals.

References

Alford, K., & Vera, J. (2018). Cognitive impairment in people living with HIV in the ART era: a review. *British medical bulletin*, *127*(1), 55-68.

Antinori, A., Arendt, G., Becker, J., Brew, B., Byrd, D., Cherner, M., Clifford, D., Cinque, P., Epstein, L., & Goodkin, K. (2007). Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*, *69*(18), 1789-1799.

Bagkeris, E., Burgess, L., Mallon, P. W., Post, F. A., Boffito, M., Sachikonye, M., Anderson, J., Asboe, D., Garvey, L., Vera, J., Williams, I., Johnson, M., Babalis, D., De Francesco, D., Winston, A., & Sabin, C. A. (2018). Cohort profile: The Pharmacokinetic and clinical Observations in PeoPle over fiftY (POPPY) study. *International journal of epidemiology*, dyy072-dyy072. https://doi.org/10.1093/ije/dyy072

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical methods in medical research*, *8*(2), 135-160.

Bloch, M., Kamminga, J., Jayewardene, A., Bailey, M., Carberry, A., Vincent, T., Quan, D., Maruff, P., Brew, B., & Cysique, L. A. (2016). A screening strategy for HIV-associated neurocognitive disorders that accurately identifies patients requiring neurological review. *Clinical Infectious Diseases*, *63*(5), 687-693.

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.

Chan, Y. (2003). Biostatistics 104: correlational analysis. *Singapore Med J*, *44*(12), 614-619.

Chavent, M., Kuentz, V., Liquet, B. ı., & Saracco, L. (2011). ClustOfVar: an R package for the clustering of variables. *arXiv preprint arXiv:1112.0295*.

Cogstate, L. (2017). Cogstate pediatric and adult normative data. Copyright.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Cole, J. H., Underwood, J., Caan, M. W. A., De Francesco, D., van Zoest, R. A., Leech, R., Wit, F., Portegies, P., Geurtsen, G. J., Schmand, B. A., Schim van der Loeff, M. F., Franceschi, C., Sabin, C. A., Majoie, C., Winston, A., Reiss, P., & Sharp, D. J. (2017). Increased brain-predicted ageing in treated HIV disease. *Neurology*. https://doi.org/10.1212/wnl.0000000000003790

Cysique, L. A., Maruff, P., Darby, D., & Brew, B. J. (2006). The assessment of cognitive function in advanced HIV-1 infection and AIDS dementia complex using a new computerised cognitive test battery. *Archives of Clinical Neuropsychology*, *21*(2), 185-194.

De Francesco, D., Underwood, J., Bagkeris, E., Boffito, M., Post, F., Mallon, P., Vera, J., Williams, I., Anderson, J., & Johnson, M. (2019). Depression, lifestyle factors and cognitive function in people living with HIV and comparable HIV-negative controls. *HIV medicine*, *20*(4), 274-285.

De Francesco, D., Underwood, J., Post, F. A., Vera, J. H., Williams, I., Boffito, M., Sachikonye, M., Anderson, J., Mallon, P. W., & Winston, A. (2016). Defining cognitive impairment in people-living-with-HIV: the POPPY study. *BMC infectious diseases*, *16*(1), 617.

de Jager, C. A., Schrijnemaekers, A.-C. M., Honey, T. E., & Budge, M. M. (2009). Detection of MCI in the clinic: evaluation of the sensitivity and specificity of a computerised test battery, the Hopkins Verbal Learning Test and the MMSE. *Age and ageing*, *38*(4), 455-460.

EACS. (2019). Guidelines Version 10.0 November 2019.

Fisher, R. A. (1992). Statistical methods for research workers (*Breakthroughs in statistics* (pp. 66-70). Springer.

Gagnon, M.-M., & Laforce Jr, R. (2016). Computerized vs. Paper-Pencil Assessment of Cognitive Change following Acute Ischemic Stroke. *Journal of neurological disorders*, *4*(8), 317.

Garvey, L., Surendrakumar, V., & Winston, A. (2011). Low rates of neurocognitive impairment are observed in neuro-asymptomatic HIV-infected subjects on effective antiretroviral therapy. *HIV clinical trials*, *12*(6), 333-338.

Hammers, D., Spurgeon, E., Ryan, K., Persad, C., Barbas, N., Heidebrink, J., Darby, D., & Giordani, B. (2012). Validity of a brief computerized cognitive screening test in dementia. *Journal of geriatric psychiatry and neurology*, *25*(2), 89-99.

Heaton, R. K., Clifford, D., Franklin, D., Woods, S., Ake, C., Vaida, F., Ellis, R., Letendre, S., Marcotte, T., & Atkinson, J. (2010). HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy CHARTER Study. *Neurology*, *75*(23), 2087-2096.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Huizenga, H. M., Smeding, H., Grasman, R. P. P. P., & Schmand, B. (2007). Multivariate normative comparisons. *Neuropsychologia*, *45*(11), 2534-2542. https://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2007.03.011

Kataja, E.-L., Karlsson, L., Tolvanen, M., Parsons, C., Schembri, A., Kiiski-Mäki, H., & Karlsson, H. (2017). Correlation between the Cogstate computerized measure and WAIS-IV among birth cohort mothers. *Archives of Clinical Neuropsychology*, *32*(2), 252-258.

Kuiper, J. S., Voshaar, R. C. O., Verhoeven, F. E., Zuidema, S. U., & Smidt, N. (2017). Comparison of cognitive functioning as measured by the Ruff Figural Fluency Test and the CogState computerized battery within the LifeLines Cohort Study. *BMC psychology*, *5*(1), 15.

Kunisaki, K. M., De Francesco, D., Sabin, C. A., Winston, A., Mallon, P. W., Anderson, J., Bagkeris, E., Boffito, M., Doyle, N., & Haddow, L. (2020). Sleep Disorders in HIV: A Substudy of the Pharmacokinetics and Clinical Observations in People Over Fifty (POPPY) Study. (Ed.),^(Eds.). Open Forum Infectious Diseases.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., & Pietrzak, R. H. (2009). Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Archives of Clinical Neuropsychology*, *24*(2), 165-178.

May, P. E., Heithoff, A. J., Wichman, C. S., Phatak, V. S., Moore, D. J., Heaton, R. K., & Fox, H. S. (2020). Assessing Cognitive Functioning in People Living with HIV (PLWH): Factor Analytic Results from CHARTER and NNTC Cohorts. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *Publish Ahead of Print*. https://doi.org/10.1097/qai.0000000000002252

McDonnell, J., Haddow, L., Daskalopoulou, M., Lampe, F., Speakman, A., Gilson, R., Phillips, A., Sherr, L., Wayal, S., & Harrison, J. (2014). Minimal cognitive impairment in UK HIV-positive men who have sex with men: effect of case definitions and comparison with the general population and HIV-negative men. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *67*(2), 120-127.

Mielke, M. M., Machulda, M. M., Hagen, C. E., Edwards, K. K., Roberts, R. O., Pankratz, V. S., Knopman, D. S., Jack Jr, C. R., & Petersen, R. C. (2015). Performance of the CogState computerized battery in the Mayo Clinic Study on Aging. *Alzheimer's & Dementia*, *11*(11), 1367-1376.

Nakasujja, N., Miyahara, S., Evans, S., Lee, A., Musisi, S., Katabira, E., Robertson, K., Ronald, A., Clifford, D. B., & Sacktor, N. (2013). Randomized trial of minocycline in the treatment of HIV-associated cognitive impairment. *Neurology*, *80*(2), 196-202.

Overton, E. T., Kauwe, J. S., Paul, R., Tashima, K., Tate, D. F., Patel, P., Carpenter, C. C., Patty, D., Brooks, J. T., & Clifford, D. B. (2011). Performances on the CogState and standard neuropsychological batteries among HIV patients without dementia. *AIDS and Behavior*, *15*(8), 1902-1909.

Schouten, J., Cinque, P., Gisslen, M., Reiss, P., & Portegies, P. (2011). HIV-1 infection and cognitive impairment in the cART era: a review. *Aids*, *25*(5), 561-575.

Schouten, J., Su, T., Wit, F. W., Kootstra, N. A., Caan, M. W., Geurtsen, G. J., Schmand, B. A., Stolte, I. G., Prins, M., & Majoie, C. B. (2016). Determinants of reduced cognitive performance in HIV-1-infected middle-aged men on combination antiretroviral therapy. *Aids*, *30*(7), 1027-1038.

Schretlen, D. J., & Sullivan, C. (2013). Intraindividual variability in cognitive test performance.

Underwood, J., Cole, J. H., Caan, M., De Francesco, D., Leech, R., van Zoest, R. A., Su, T., Geurtsen, G. J., Schmand, B. A., & Portegies, P. (2017). Gray and white matter abnormalities in treated human immunodeficiency virus disease and their relationship to cognitive function. *Clinical Infectious Diseases*, *65*(3), 422-432.

Underwood, J., De Francesco, D., Leech, R., Sabin, C. A., & Winston, A. (2018). Medicalising normality? Using a simulated dataset to assess the performance of different diagnostic criteria of HIV-associated cognitive impairment. *PloS one*, *13*(4), e0194760.

Wesnes, K. A. (2014). Moving beyond the pros and cons of automating cognitive testing in pathological aging and dementia: the case for equal opportunity. *Alzheimer's research & therapy*, *6*(5), 58.