Bond University

Research Repository

# Montreal Cognitive Assessment (MoCA) for HIV-Associated Neurocognitive Disorders

Rosca, Elena Cecilia; Albarqouni, Loai; Simu, Mihaela

Link to publication in Bond University research repository.

# Montreal Cognitive Assessment (MoCA) for HIV-associated neurocognitive disorders

Rosca Elena Cecilia[1,2], Albarqouni Loai[3], Simu Mihaela[1,2]

[1]University of Medicine and Pharmacy "Victor Babes" Timisoara, Department of Neurology, Romania

[2]Clinical Emergency County Hospital Timisoara, Department of Neurology, Romania

[3]Faculty of Health Sciences and Medicine, Bond University, Australia

**Corresponding author:**

Rosca Elena Cecilia.

Address:  Department of Neurology, Clinical Emergency County Hospital, Bd. Iosif Bulbuca nr. 10, Timisoara 300736, Romania.

Email: roscacecilia@yahoo.com

Phone: +40 746173794

ORCID: 0000-0003-2412-1942

**Montreal Cognitive Assessment (MoCA) for HIV-associated neurocognitive disorders**

**ABSTRACT**

This study aims to systematically review the evidence on the accuracy of the Montreal Cognitive Assessment (MoCA) test for diagnosing HIV–associated neurocognitive disorders (HAND) and to outline the quality and quantity of research evidence available about the accuracy of MoCA in populations infected with HIV. We conducted a systematic literature review, searching five databases from inception until January 2019. We extracted dichotomized positive and negative test results at various thresholds and calculated the sensitivity and specificity of MoCA. Quality assessment was performed according to the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) criteria. Eight cross-sectional studies met the inclusion criteria for meta-analysis. Overall, 1014 patients were included but most studies recruited small samples. Recruitment period ranged from 2009 to 2015. We assessed most studies as being applicable to the review question though we had concerns about the selection of participants in three studies. The accuracy of MoCA for diagnosing HAND was reported at six cut-off points (scores 22-27). The MoCA test provides information about general cognitive functioning disturbances that contribute to a diagnosis of HAND. A lower threshold than the original cut-off of 26 is probably more useful for optimal screening of HAND, as it lowers false positive rates and improves diagnostic accuracy. Nonetheless, the choice of cut-off always comes with a sensitivity-specificity trade-off, the preferred cut point depending on whether sensitivity or specificity is more valuable in a given context.

HIV-associated neurocognitive disorders (HAND) in adults are estimated to be between 30% and 60%, despite the use of combined anti-retroviral therapy (ART; Simonini et al. 2010; Heaton et al. 2010; Heaton et al. 2011; Ciccarelli et al. 2013; Chan & Brew 2014). Epidemiological studies currently show that HIV-associated dementia (HAD) is rare (2-4%; McArthur 2004). Most patients present less severe forms of HAND, including asymptomatic neurocognitive impairment (ANI) and mild neurocognitive disorder (MND; Heaton et al. 2010; Munoz-Moreno et al. 2014; Sacktor et al. 2016).

At the beginning of the HIV epidemic, many patients developed severe neurological impairment in the final months of their illness. The clinical syndrome, comprising cognitive, behavioral and motor symptoms was termed AIDS dementia complex (ADC). The cognitive impairment predominantly consisted of mental slowing and attention and memory deficits. With the introduction of ART, ADC incidence decreased, but patients with treatment and long-term infection present milder cognitive symptoms. In addition, a shift has occurred in certain demographic variables and risk factors, like increased age and cardiovascular risk factors. Thereby, the phenotype of HAND has expanded, with broadening of neuropsychological profile (Cysique et al. 2004; Woods et al. 2009). Patients with HAND present a subcortical profile of cognitive impairment, with the core deficits being mental slowness, attention and memory deficits and impaired executive functions (Woods et al. 2009). The decreased speed of information processing is one of the most frequent cognitive abnormalities in HAND (Cysique et al. 2004; Woods et al. 2009; Schouten et al. 2011). Because mental speed facilitates most of the cognitive and motor processes, some authors even consider it the key deficit which in turn leads to impairments in other cognitive domains (Hardy & Hinkin 2002). Patients with HAND also present attention and working memory impairments. These cognitive functions are closely related, the ability to create a memory for temporary processing and to store the information being dependent on attention functions. In consequence, the two deficits occur simultaneously (Woods et al. 2009; Schouten et al. 2011). Also, patients with HAND have been reported to present impaired executive functions, with deficits in reasoning, planning, problem solving and shifting between tasks (Dawes et al. 2008; Schouten et al. 2011). In the memory domain, the patients present mainly learning new information and prospective episodic memory deficits, with impaired ability to execute a future intention or "remembering to remember" (Schouten et al. 2011). The most frequent language problem in HAND consists of impaired fluency, although this could be also due to mental slowness or executive dysfunction (Dawes et al. 2008; Schouten et al. 2011). Less frequently, patients with HAND

may present sensory-perceptual impairments, with disturbances in interpretation and integration of auditory, visual or sensorial stimuli (Schouten et al. 2011).

Currently, HAND can be classified according to the Frascati criteria, with different degrees of cognitive impairment that are separately diagnosed. In patients with ANI, the neuropsychological test performance is at least one standard deviation (SD) below the normative data in at least two domains with intact daily functioning. On the other hand, MND is characterized by similar neuropsychological test results, with impaired daily functioning. HAD is characterized by severe deficits in at least two cognitive domains, typically two SDs below normative data, and more severe daily functioning impairment (Antinori et al. 2007).

Recently, the validity of the Frascati criteria has been challenged (Gisslen et al. 2011; Nightingale et al. 2014). Researchers disagree over the validity of neuropsychological testing in characterizing ANI. Some researchers have argued against testing for ANI since there are no screening tools with high sensitivity and specificity that can be used in all clinical settings and there is no consensus on the therapeutic management of asymptomatic patients. In addition, screening can lead to unnecessary and expensive diagnostic procedures and a positive result might cause distress to some people living with HIV (PLWH; Nightingale et al. 2014). Furthermore, some observational studies have not shown an association between ART with estimated high central nervous system (CNS) effectiveness and neurocognitive function (Simonini et al. 2010; Giancola et al. 2006; Smurzynski et al. 2011).

On the other hand, there are some arguments that support screening for ANI. Several studies have shown that when the prevalence of ANI is high, patients may present poor adherence to medication, as well as high unemployment rates (Gorman et al. 2009). Additionally, ANI might be associated with an increased risk of progressive neurocognitive disease (Grant et al. 2014). Some studies have reported that ART with high CNS effectiveness is associated with improved cognitive function (Cysique et al. 2011) and after changes to ART on the basis of estimated CNS effectiveness, the levels of HIV RNA in the cerebrospinal fluid (CSF) declined, improving the cognitive functions (Smit et al. 2004). Furthermore, some ART have proven to be neurotoxic (Robertson et al. 2012). Although it has been argued that since ANI is "asymptomatic" it may have little clinical significance, recent research has demonstrated that ANI patients present grey and white matter abnormalities (Haziot et al. 2015) as well as abnormal blood plasma biomarkers (e.g., nadir CD4 count, neopterin, neurofilament light chains; Chan & Brew 2014).

In general, there is an agreement between international guidelines regarding the diagnosis of HAND (for a review see Underwood & Winston 2016). These guidelines which have a specific section regarding cognitive impairment, recommend a comprehensive assessment including a thorough medical history and examination, screening for depression, neuropsychological testing, magnetic resonance imaging (MRI) of the brain and lumbar puncture (European AIDS Clinical Society 2018, Mind Exchange Working Group 2013; HIV/AIDS Italian Expert Panel 2017). However, there is no clear consensus regarding the specific tests that should be used as part of the neuropsychological assessment. All guidelines refer to the Frascati criteria, which have recommendations on several preferred tests for each cognitive domain, and recommend a complex neuropsychological assessment and testing several cognitive domains (Antinori et al 2007). Furthermore, the Mind Exchange Working Group suggests that the tests selected should be validated in the language and culture of the population and evaluated according to appropriate normative data available to interpret the results (Mind Exchange Working Group 2013).

However, such tests are not available in many centers and require highly trained personnel (Antinori et al. 2007). Therefore, brief screening tests that are sensitive, easily accessible, and can be administered by clinical staff across a range of settings would be useful. Nonetheless, most HIV treatment guidelines do not make any specific recommendations about screening for neurocognitive impairment. Regarding the guidelines that propose recommendations, there is considerable variation in guidance reflecting the uncertainties in the literature (Underwood & Winston 2016). The European AIDS Clinical Society (EACS) guidelines (version 9.1, EACS 2018), recommend screening all HIV positive individuals without highly confounding conditions (such as severe psychiatric diseases, abuse of psychotropic drugs or alcohol, current CNS opportunistic infections or other neurological diseases, sequels of CNS disorders) at HIV diagnosis, before ART initiation and then later as indicated based on symptoms. The EACS screening method involves asking three questions: "Do you experience frequent memory loss?" "Do you feel that you are slower when reasoning, planning activities, or solving problems?" and "Do you have difficulties paying attention?". Answering "Yes", to at least one of these questions constitutes a positive screening test requiring further assessment. This approach differs from guidance given by the consensus report of the Mind Exchange Program which recommends screening within six months of diagnosis, before ART initiation, every 6–12 months if there is a high risk, every 12–24 months if there is low risk, and immediately if there is any clinical deterioration (Mind Exchange Working Group 2013). The recommended screening tool depends on the following aspects: availability of an appropriately trained clinician suitably trained to administer and interpret each

instrument; whether the clinician wants to screen for HAD only or for the milder forms of HAND; financial and time cost of testing; and the characteristics of the population in which the tool will be used. Because the neuropsychological resources are limited in many settings, a probable clinical diagnosis of HAND could be based on symptom questionnaires, screening tools, functional assessments, and limited neuropsychological testing. Patients with particular characteristics could then be targeted for full neuropsychological assessments (Mind Exchange Working Group 2013). However, some preferred screening tests are mentioned, such as the HIV Dementia Scale (HDS) and the International HIV Dementia Scale (IHDS). The British HIV Association (BHIVA) recommends that HIV-positive patients should have access to screening for cognitive difficulties within the first three months of receiving an HIV diagnosis and all HIV-positive patients should have access to repeated screening following events that are known to trigger or exacerbate cognitive difficulties, and otherwise on an annual basis (Angus et al. 2016). These recommendations are similar to those recommended in the guide published by the Infectious Diseases Society of America (Aberg et al. 2014). The World Health Organization (WHO) recommends that routine screening and management for mental health disorders should be provided for people from key populations living with HIV in order to optimize health outcomes and improve adherence to ART. However, the screening method and frequency have not been specified (World Health Organization 2016). The Italian Society for Infectious and Tropical Diseases recommends screening all people living with HIV if the patient presents cognitive complaints. Among the suggested tests, they recommend Montreal Cognitive Assessment (MoCA; HIV/AIDS Italian Expert Panel 2017).

The guidelines recommend a neurological examination, brain MRI, and CSF examination in order to exclude other pathologies, if the neuropsychological impairment detected on screening is confirmed by tests exploring multiple cognitive domains, including: verbal fluency, executive functions, speed of information processing, attention and working memory, verbal and visual learning, verbal and visual memory, motor skills, and assessment of daily functioning. In addition, these guidelines recommend an assessment of CSF HIV viral load level and, where appropriate, evidence for genotypic drug resistance (GDR) in a paired CSF and plasma sample (EACS 2018). After additional causes of cognitive impairment are excluded and a diagnosis of HAND is made, the clinician must take specific treatment and care measures (EACS 2018).

Few screening tools have been developed and validated, including the HDS and its derivative form, IHDS (Sacktor et al. 2005; Bottiggi et al. 2007). Both instruments are relatively insensitive to the milder cognitive symptoms that predominate in the combination ART era (Skinner et al. 2009). Although they are recommended as

screening tools by expert HIV guidelines (Mind Exchange Working Group 2013), a recent systematic review concluded that their accuracy is low. Summary estimates for the HDS as a test for HAND presented sensitivity and specificity of 42% and 91% respectively. On the other hand, when using IHDS as a test for all symptomatic HAND, the sensitivity and the specificity of this tool were 64% and 66% respectively (Haddow et al. 2013).

Other screening tests, such as Mini-Mental State Examination (MMSE) have been used in clinical practice to detect cognitive impairment regarding a variety of neurological disorders. Clinicians are familiar with its use and it is widely used as the first-choice tool for screening HAND. However, studies have indicated that this instrument is not very reliable in detecting HAND (Kami-Onaga et al. 2018; Milanini et al. 2016; Skinner et al. 2009).

The MoCA has been used in PLWH as another screening instrument with variable results. It was developed in 2005 for detecting mild cognitive impairment (MCI) and has been shown to be highly sensitive and specific in older adult population (Nasreddine et al. 2005). MoCA is a brief bedside test assessing short-term memory, attention and working memory and frontal-executive functions, which are commonly affected in patients with HIV infection. Scores on the MoCA range from zero to 30 points, with a score of 25 or lower indicating a cognitive impairment. This cut-point is now widely used as a threshold for detecting cognitive impairment and possible dementia. In order to minimize practice effects, three versions of MoCA have been developed in English, which test the same domains, but the content of the tasks is different. The alternative versions of MoCA present comparable reliability to the original test (Costa et al. 2012). Translations in multiple languages are also available and the administration time is typically 10 minutes.

The first item of MoCA, the modified trial making test, requires visuomotor, visuoperceptual skills and mental flexibility to shift between numbers and letters (Crowe, 1998; Sánchez-Cubillo et al. 2009). To perform the second sub-test of MoCA and copy a cube, an individual has to initially convert the two-dimensional contour to a tri-dimensional figure, ability that is enhanced by learning abilities (Sinha & Poggio 1996). After spatial planning, visuomotor coordination and integration of visual and fine motor sequences are also necessary. In Alzheimer's disease, a poor performance in drawing-to-command and copying conditions was reported in less educated, older age, female and depressed subjects (Gaestel et al. 2006). The clock drawing test has been extensively studied for detection of cognitive impairment. It evaluates the visuoconstructive skills. In addition, in order to draw the clock's face and to place the numbers correctly, subjects need to have intact planning, conceptualization and symbolic representation (Pinto & Peters 2009). When placing the hands of the clock to draw "ten past eleven", an inhibitory

response is also necessary. Although the scoring criteria for the clock drawing test in the MoCA has been simplified to decrease scoring complexity, scoring time, and minimize inter rater variability, suboptimal inter and intra-rater reliability for this item was reported (Price et al. 2011). Also, the task may be influenced by literacy status and education level (Nitrini et al. 2004). For the naming sub-test of MoCA, the patient has to name three animals that are presented visually. If the subject is unable to name it but can give contextual information on the animal, this is probably due to word finding difficulties or impaired semantic memory. If the patient cannot tell both, the name and the context, he probably presents impaired visuoperceptual skills or semantic memory. Cultural exposure and low education can also determine errors on this task. Attention is assessed by the digit span items. The digit span forward implies retention of auditory stimuli and articulatory rehearsal. The digit span backward necessitates executive processing, working memory, ability to transform the numbers in reverse order and language functions. For the letter A taping sub-test, the subjects need sustained and focused attention. In the MoCA validation study, MCI subjects and had comparable performance with normal controls, but the patients with Alzheimer's disease were significantly more impaired on this task (Nasreddine et al. 2005). The serial 7 subtractions sub-test evaluates calculation abilities. The sentence repetition tasks of MoCA evaluate language skills, attention and working memory (Small et al. 2000). The performance on this sub-test is also influenced by education. The letter fluency item of MoCA requires language abilities and intact executive functions with coordination of lexical and semantic knowledge, shifting from word to word, working memory, searching strategy and inhibition of irrelevant words (Troyer et al. 1998; Henry & Crawford 2004; Larsson et al. 2008). The abstraction subtest, where the patient has to find the similarities between objects, evaluates semantic knowledge and conceptual thinking. On the memory sub-test, the subjects have to recall 5 words, with 2 learning trials, and with 5 minutes between immediate recall and delayed recall. The category and multiple-choice cues can provide also information that helps distinguishing an encoding memory impairment which does not improve with cueing from a retrieval memory impairment that is improved with cueing. The last subtest assesses the patient's orientation in space and time. These items were demonstrated to have a low value for detecting MCI (Nasreddine et al. 2005), but temporal orientation was reported to have high sensitivity in detection of dementia (O'Keeffe et al. 2011). Furthermore, patients with temporal disorientation have been demonstrated to present also impaired verbal memory (Ryan et al. 2009).

The MoCA has a widespread international use, being recognized as one of the best screening tests (Ismail, Rajji, & Shulman, 2010; Jacova, et al. 2007) as several previous studies have consistently reported that it has good

overall psychometric properties and a good sensitivity in accurately identifying milder forms of cognitive

impairment in many clinical conditions. For example, in MCI, the internal consistency of MoCA was reported to be

excellent, with a Cronbach's alpha of 0.83 on the standardized items (Nasreddine et al. 2005). The test-retest

reliability was also good, wit a mean change in MoCA scores from the first to second evaluation of 0.9 points

(Nasreddine et al. 2005). In addition, in studies that applied Rasch analysis techniques, the researchers found that

scores on the MoCA can be used to quantify the amount of cognitive ability a person has and can be used to track

changes in cognitive ability over time (Koski et al. 2009). Also, in patients with a subcortical type of cognitive

impairment like Parkinson's disease, Cronbach's alpha was reported to be 0.66 to 0.79 (Ozdilek & Kenangil 2014;

Nie et al. 2012), with a good interrater reliability of 0.81 and a test-retest reliability of 0.79 (Gill et al. 2008).

Validation studies of the MoCA have been conducted concerning different types of neurological disorders,

such as MCI (Freitas et al. 2013), Alzheimer's disease (Freitas et al. 2013), Parkinson's disease (Hoops et al. 2009),

and Huntington's disease (Bezdicek et al. 2013). In a recent systematic review, MoCA as a screening test for

dementia and multidomain cognitive impairment in stroke patients, at the usual threshold of 26, presented a high

sensitivity (0.95) but at cost of specificity (0.45). An adjusted cutoff of 22 has improved its specificity (0.84)

without sacrificing sensitivity (0.78; Lees et al. 2014). In the diagnosis of dementia (including Alzheimer's disease,

vascular dementia, Lewy body dementia and frontotemporal dementia), the recommended threshold of 26 presented

a high sensitivity of 0.94 or above, but a low specificity of 0.60 or below. The systematic review pointed out that

cut-off scores lower than 26 were likely to be more useful for optimal diagnostic accuracy of MoCA concerning

dementia (Davis et al. 2015). A recent systematic review of the literature which evaluated the diagnostic accuracy of

the MoCA in differentiating healthy cognitive aging from possible MCI found that the optimal score, which

maximized true positives while minimizing false positives, was a cutoff of 23. Although sensitivity was lower at 23

(0.83) than at 26 (0.94), specificity was higher (0.88 vs. 0.66) and the balance between true positive and false

positive results was better (Carson et al. 2018).

To resume, early diagnosis and specific treatment and care of HAND is essential. Although all the

guidelines recommend for diagnosis the Frascati criteria, with an extensive battery of neuropsychological tests, this

is time consuming, expensive and necessitates trained personnel. Thus, screening for HAND and identifying the

patients that should be further investigated is essential, but the available guidelines on screening for HAND reflect

the uncertainties in the literature and clinicians are faced with a difficult choice: which screening test should they

use. MoCA fulfils very important feasibility criteria for use in clinical practice: it has a short administration time, it is freely available, with multiple translations and has minimal training requirements. Furthermore, online training and certification is available on the MoCA website. In addition, it has proven to have good psychometric properties in other populations and assesses a broad range of cognitive domains.

With regards to HIV-positive patients, several researchers have explored the utility of the MoCA to detect cognitive impairment, but the sensitivity and specificity values and the cut-off scores have differed across studies. Although the diagnostic assessment pathways may vary across different countries, often HAND is screened in specialized infectious diseases clinics during outpatient visits. The MoCA may help identify PLWH that require further assessments and specific care facilitating access to appropriate services. Nonetheless, being wrongly tested as positive, implies significant costs and harm due to further unnecessary investigations and psychological distress. Therefore, there is considerable value in determining the strength of the empirical evidence that supports the use of MoCA as a screening test for HAND. We aim to collate evidence from different studies, integrating the existing information and providing data for rational decision making, highlighting possible answers, that are easily accessible to clinicians, health care providers and policy makers.

The objective of this systematic review is to evaluate research regarding the accuracy of the MoCA test for diagnosing HAND against a concurrently applied reference standard and to highlight the quality and quantity of evidence available in this regard. Also, we aim to identify the gaps in the literature regarding this short neuropsychological test battery.

## Methods

This meta-analysis was performed following the recommendations described in the Cochrane Handbook for Diagnostic Test Accuracy Reviews (Handbook for Diagnostic Test Accuracy Review) and a Cochrane generic protocol for cross-sectional and delayed-verification studies (Davis et al. 2013). Results were reported according to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA; Moher et al. 2009).

### Search strategy and selection criteria

Supplemental Figure 1 shows the search strategy followed in the meta-analysis.

A computerized bibliographic search was performed from the beginning to January 2019 on the following databases: MEDLINE/PubMed, EMBASE, Scopus, Cochrane Library, Latin American and Caribbean Health Sciences Literature (LILACS) and PsychINFO. In addition, a complementary manual search was performed on the MoCA website as well as by checking reference lists of all relevant research papers in order to identify possible additional studies. We also searched grey literature through Google Scholar but only papers from peer-reviewed, non-predatory, and indexed journals were included.

The following key words were used: "Montreal cognitive assessment" or the acronym "MoCA", and "HIV infection" [MeSH] and "acquired immunodeficiency syndrome" [MeSH]. These search terms were for PubMed, the primary source of citations. Searches in other data sources used similar versions of these terms, appropriate for each database. We did not utilize search filters (collection of terms aimed at reducing the number needed to be screened) because our aim was to generate a comprehensive list of studies which would be suitable for answering the research question. Even the most sensitive filters have been found to miss relevant studies and perform inconsistently across subject areas and study designs, while at the same time have not significantly reduced the number of studies that need to be assessed for inclusion (Cochrane Handbook for DTA Reviews, Davis et al. 2013). In addition, we did not apply any restrictions on language.

Two authors reviewed the title, abstract and full text (when needed) of all retrieved research papers and assessed whether the study met the inclusion criteria. During the abstract review stage, in order not to miss any potentially eligible studies, we did not exclude the papers where we were not sure whether there was an appropriate reference standard or a full version of MoCA and if we were uncertain if the article was a diagnostic test accuracy study. We evaluated all these articles in full text. All initial reviews were rated equally, and participation of a third rater was not needed to address discrepancies.

Eligible studies were cross-sectional studies in which participants received the index test and the reference standard diagnostic assessment. Case-control studies were excluded owing to a high possibility of bias. We included studies reporting adults (over 18 years old) with confirmed HIV infection in which the association between MoCA score and HAND was assessed, MoCA being used as an index test. The index test was any full version of the MoCA. Although we expected to find the recommended cut-off score of 25or below to differentiate normal (26 and above) from impaired cognition (less than 26), we also included studies using other thresholds (22-27). The target condition was HAND, including ANI, MND, and HAD, as classified by the Frascati criteria (Antinori et al. 2007).

We used as a reference standard for HAND a complex neuropsychological assessment, evaluating at least five neurocognitive domains (including verbal and language, attention and working memory, abstraction and executive function, learning and recall, speed of information processing, and motor skills), with consensual recommendations on appropriate tests. In this study, as endorsed by international guidelines, a neurocognitive impairment was defined as an impairment in cognitive function on the above neuropsychological tests in which performance is considered clinically significant compared to appropriate controls matched by age and educational level (Antinori et al. 2007; European AIDS Clinical Society 2018, Mind Exchange Working Group 2013). We excluded studies of participants with confounding factors such as neurological disorders (e.g., recent traumatic brain injury, CNS infections, stroke, neurodegenerative disorders, and brain tumors), active psychosis, significant substance abuse, including alcohol and recreational drugs, and active infections.

Disagreements were resolved through discussion and a third rater was not needed to address differences. The methodological quality of the studies included was assessed by two authors independently (ECR, MS) according to the Cochrane Collaboration's tool for assessing the risk of bias (Handbook for Diagnostic Test Accuracy Reviews) using the unmodified Quality Assessment of Diagnostic Accuracy Studies 2 (QADAS-2) tool (Whiting et al. 2011).

### *Statistical analysis*

Pairwise meta-analysis was performed to estimate the sensitivities and specificities, with 95% confidence intervals. Regarding studies that reported more than one threshold, we extracted data on diagnostic accuracy for thresholds ranging from 22 to 27. We calculated for each cut-off score the following parameters: sensitivity (proportion of individuals diagnosed with HAND who tested positive on the MoCA); specificity (proportion of patients indicated as normal, who tested negative on the MoCA); the positive predictive value (PPV - proportion of individuals with a positive MoCA test who were diagnosed with HAND); and negative predictive value (NPV - proportion of patients who tested negative on the MoCA, without HAND). In addition, we calculated the likelihood ratios for positive results (LR+), representing the probability that a patient diagnosed with HAND could test positively on the MoCA as well as the likelihood ratios for negative results (LR-), referring to the probability that an individual diagnosed with HAND could present a negative MoCA result. LR+ may range from 0 to infinity. The larger LR+, the more informative is the test. Findings with LR+ greater than 1 indicate an increase in the odds of having a particular condition in a patient with a positive result. The bigger the number, the more convincingly the

finding suggests the presence of disease. If LR+ lies between 0 and 1, it argues against the diagnostic value of the test (McGee 2002). When diagnostic test accuracy results are around 1, it lacks of diagnostic value and we have learned nothing by ordering them (McGee 2002, Straus et al. 2019). A simple estimative method, independent of pretest probabilities and avoiding complex calculations, enables the clinician to make a bedside estimation: a LR+ of 2 causes a small change in disease likelihood, increasing the probability of the disease with 15%, one of 5 increases it 30%, determining a moderate change, and one of 10 increases it 45%, determining a large change in disease likelihood (McGee 2002). On the other hand, the LR- indicates the odds of having a disease in patients with negative test results. If the LR- is smaller than 1, then the post-test probability of the disease being present decreases. The smaller the LR-, the more informative is the test. A LR- of 1 means that the test is useless, because the odds the odds of having the condition have not changed after the test administration (McGee 2002, Straus et al. 2019).

Further calculations in our study included Phi, a measure of binominal effect size, and the Youden index, an estimated value of the optimal threshold at which sensitivity (true positive patients) is maximized and false positive results are minimized. The Youden's index, is a global measure of test performance, used for evaluating the overall discriminative power of a diagnostic procedure. It is calculated by deducting one point from the sum of a test's sensitivity and specificity expressed not as a percentage but as a part of a whole number: (sensitivity + specificity) – 1. Youden's index equals 0 in a test with poor diagnostic accuracy, and in a perfect test, Youden's index equals 1 (Šimundić 2009).

We used the bivariate model to perform meta-analysis of sensitivity and specificity and added cut-off values as a variate in the model (Reitsma et al. 2005). In addition, we compared the recommended threshold of 26 with the cutoff score of 27 and analyzed sensitivity and specificity for each cutoff using the hierarchical summary receiver-operating characteristic (HSROC) model (Rutter 2001) that allows for the possibility of variation in cutoffs between studies. We computed the accuracy of MoCA with 95% confidence and prediction intervals. Confidence intervals (CI) indicate the likely the true value within the population, whereas predictive intervals account for both the uncertainty in knowing the value of the population mean as well as the scatter of presented data (thus usually wider than CI). Whilst the confidence region depicts uncertainty in the overall average value caused by sampling variability, the prediction region depicts variation from between study heterogeneity.

The systematic assessment of study quality used unmodified QUADAS-2 to determine the overall risk of bias for each study. All calculations were performed using Review Manager, version 5.3 (Cochrane Collaboration,

Copenhagen, Denmark) and R software, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria), with the mada package.

## Results

### *Included studies*

From a total of 37 unique studies identified using the search strategy and assessed in full-text, we included eight studies whose characteristics are summarized in Supplemental Table 1. The PRISMA diagram describing the selection process of studies is detailed in Supplemental Figure 1. Twenty-nine studies were excluded for the following reasons: duplicated data (3), inadequate reference standard (11), insufficient data reported (3), MoCA test was modified (4) or the subtests were chosen specifically to optimize testing in individuals with limited educational level (1), the MoCA was not the index test (6) or the research paper was not a diagnostic test accuracy study (1).

### *Methodological quality of included studies*

The Qadas-2 scores for each domain are presented in Figures 1 and 2.

In the *Patient Selection* domain, the risk of bias was reduced by selecting only cross-sectional studies. A random or consecutive sample of patients was reported in four studies (Janssen et al. 2015; Nam Su Ku et al. 2014; Joska et al. 2016; Milanini et al. 2016). Two studies that included only patients aged 60 and above (Milanini et al. 2014) or adults aged 50 years and older (Fazeli et al. 2017), were considered to have a high risk of bias. Inappropriate exclusions were avoided in five studies (Janssen et al. 2015; Nam Su Ku et al. 2014; Joska et al. 2016; Milanini et al. 2016; Overton et al. 2013). On the other hand, one study excluded eight patients from the initial selection because of their low educational level (< grade 7) and one owing to incomplete testing (Koenig et al. 2016).

All patients were recruited in outpatient clinics from urban areas.

Regarding the *Index Test* domain seven studies were considered as presenting an unclear risk of bias (Janssen et al. 2015; Joska et al. 2016; Koenig et al. 2015; Ku et al. 2014; Milanini et al. 2014; Overton et al. 2013). In all studies except one (Milanini et al. 2016) it was unclear if the index test results were interpreted without knowledge of the results of the reference standard. In addition, the pre-specification of threshold was absent or unclear in four studies (Nam Su Ku et al. 2014; Fazeli et al. 2017; Koenig et al. 2016, Milanini et al. 2016), and this finding was also considered to present an unclear risk of bias.

With regards to the *Reference Standard* domain, all studies used a reference standard that would correctly diagnose HAND. However, only study specified that the reference standard was interpreted without knowledge of the index test results (Joska et al. 2016) and therefore all the other studies were classified as having unclear risk of bias.

Only three studies reported the *flow and timing* of the cognitive tests (Joska et al. 2016; Milanini et al. 2016; Overton et al. 2013) and one study spanned eight months. In this last case, we assumed that the interval between the index test and the reference standard might be inappropriate (Ku et al. 2014).

There were no exclusions from the analysis in four studies (Joska et al. 2016; Milanini et al. 2014; Milanini et al. 2016; Overton et al. 2013).

Generally, the studies had a low risk of bias and no study had more than one out of Quadas-2 items assessed as having a high risk of bias.

### *Findings*

In general, eight studies, that assessed 1014 patients were included. There was an overlap of participants because of the use of patients across several studies where multiple cut-off points were examined. Recruitment period was between 2009 and 2015. The study samples were selected from six different countries (USA, South Africa, South Korea, Canada, Netherlands and Italy). Samples ranged in size (from 67 to 200 participants), gender (37.2 % males to 96% males), median age (40 to 64 years), educational level, CD4 values, and viral load. All patients were on ART except in two studies where only 89.70% (Ku et al. 2014) and 98% (Fazeli et al. 2018) of patients were on antiretroviral medication. The reference standard used the Frascati criteria with extensive neuropsychological batteries measuring multiple cognitive domains in all the studies. The characteristics of the included studies are presented in Supplemental Table 1.

Table 1 and Supplemental Table 2 shows data related to the full meta-analysis. Supplemental Table 3 shows the detailed analysis of the Youden index and Phi for each threshold. In addition, the forest plots of MoCA at different thresholds are presented in Figure 3.

Data from four studies with 556 patients were pooled for analysis at a threshold of 27, revealing an overall sensitivity of 0.77, a specificity of 0.50 and a low LR+ of 1.54 (see Table 1 and Supplemental Table 2). For a cutoff score of 26, we extracted data from six studies with 784 patients; the overall sensitivity of MoCA was 0.73, with a specificity of 0.54 and a LR+ of 1.58 (see Table 1and Supplemental Table 2). We used the LR and found that

between the recommended cutoff scores of 26 and the 27 there were no statistically significant differences in the sensitivity and specificity of MoCA at both thresholds. The summary receiver operating characteristic (SROC) curve of both thresholds with 95% confidence and prediction interval is presented in Figure 4, where the 95% confidence region is a measure of within-study uncertainty (the precision of the test accuracy estimate) and the prediction region is a measure of between-study variability and defines the area in receiver operating characteristic (ROC) space where we are confident that a test performs within a stated degree of uncertainty.

Three studies, including 461 patients, contributed to the analyzed data using a threshold of 25; the sensitivity and specificity of MoCA were 0.61 and 0.72 respectively and the LR+ was 2.19 (see Table 1 and Supplemental Table 2). For a threshold of 24, 461 individuals were included, from three studies; the overall sensitivity of MoCA was 0.53, with a specificity of 0.82 and a LR+ of 2.87 (see Table 1and Supplemental Table 2).

Data from four studies, which included 586 participants was extracted for a cutoff score of 23. The current analysis indicated that a threshold of 23 offered the best diagnostic accuracy (see Table 1 and Supplemental Table 2). Although MoCA sensitivity was lower at 23 than at the recommended threshold of 26 (0.44 vs. 0.73), its specificity was higher (0.79 vs. 0.54), with a better balance between true positives and false positive results (Youden's index mean 0.384 vs. 0.293). However, the cut-off score of 23 presented a modest LR+ of 2.11.

For a cutoff score of 22, we could extract data from only two studies (see Table 1 and Supplemental Table 2), which included 287 HIV positive individuals. The sensitivity of MoCA was 0.23, with high specificity (0.90), but a lower Youden index than for the threshold of 23 (0.283 vs. 0.384). The LR+ was 2.27.

Owing to differences in the characteristics the studies and the small number of studies, we considered that performing a study of heterogeneity and advanced statistical analysis was not suitable.

### Discussion

The present meta-analysis allowed us to make several key observations.

Although the MoCA seemed to be a promising screening test for patients infected with HIV, our data revealed that it may not be the best discriminating tool for this specific population. At the original recommended cutoff score of 26, the test failed to adequately distinguish HIV-infected patients with cognitive impairment from those with normal cognition. The sensitivity decreased by changing the threshold, but the optimal cutoff score for diagnosing HAND was 23, offering the best balance between true positive and false positive results (see Table 1 and

Supplemental Table 2). Although the LR+ was of 2.11, with a modest increase in the odds of having HAND in a patient with positive result, the impact of a LR is very dependent on the baseline probability of having the condition (prevalence of the disease or pre-test probability). In our study, at a threshold of 23, the prevalence or the pre-test probability of HAND was 0.39 (39%) and the probability of finding cognitive impairment after MoCA administration was 0.57 (57%). The same LR would result in a post-test probability of 76% if the prevalence of HAND in the study population would be 60%. Therefore, LR should always be interpreted in the context of the pre-test probability or the prevalence of the outcome (Straus et al. 2019).

While clinicians may prefer a test with high sensitivity, this could increase the number of patients referred for formal cognitive testing and further assessment. On the other hand, a higher specificity may reduce unnecessary referrals, but many true cases could be missed. Alternatively, a lower cutoff score could provide a better balance between true positives and false positive results and could be used to identify individuals that should be repeatedly monitored (Overton et al. 2013).

In the literature, the original cutoff score of 26 demonstrated satisfactory sensitivity and specificity in general populations (Nasreddine et al. 2005). However, later systematic reviews revealed that MoCA, at the usual threshold presented a high sensitivity of 0.95 in stroke patients (Less et al. 2014), 0.94 or above in dementias, including Alzheimer's disease, vascular dementia, Lewy body dementia and frontotemporal dementia (Davis et al. 2015). Sensitivity regarding MCI was 0.94 (Carson et al. 2018), but at the cost of low specificity. The MoCA offered a better diagnostic accuracy for stroke patients at an adjusted cutoff score of 22 (Less et al. 2014). In MCI the best diagnostic accuracy was offered by the threshold of 23 (Carson et al. 2018). This is consistent with the results of the present study (see Table 1). However, in our study, the sensitivity and specificity of the test were lower, compared to other disorders: at the cut-off of 26, the sensitivity was 0.73, and the specificity was 0.54. The threshold of 23, although presented the best Youden index, offered a sensitivity of 0.44 and a specificity of 0.79. These differences could be due to several factors. A possible explanation could be that MoCA evaluates abstraction, object naming, clock drawing and language, which are domains that are not frequently related to HIV infection (Woods et al. 2009). Nonetheless, these items could be useful in older PLWH, which can present multiple comorbidities such as increased vulnerability to Alzheimer's disease, cardiovascular risk factors and cerebrovascular disease (Milanini et al. 2014; Fazeli et al. 2017; Devlin & Giovannetti 2017). Furthermore, the differences in sensitivity and specificity could be caused by the reference standard used. The current criteria for diagnosing MCI

include a subjective complaint and functional independence and there is no analog for ANI in DSM 5. The Frascati criteria identifies three severity levels for HAND, ANI being defined as neurocognitive impairment demonstrated by performance falling one standard deviation below the mean of demographically adjusted normative scores in two out of at least five measured domains; in this case, asymptomatic means that the patient has no clinically significant difficulties in everyday functioning (Antinori et al. 2007). Nevertheless, studies of neuropsychological batteries used for HAND in normal HIV-uninfected populations have suggested that, between 15 and 22% of individuals from an HIV-uninfected control group and 20% of a simulated normal population will score below the threshold for HAND, with false positive results. These errors are caused by two common practices to increase the sensitivity regarding milder neurocognitive abnormalities. First, extensive test batteries will have higher false-positive rates than individual tests because they involve multiple comparisons. The probability of an abnormal score increases as the number of tests performed per domain and the number of assessed domains increases (i.e., diagnosing a normal individual as impaired). Second, the high cutoff scores (z scores with a threshold of 1 SD) will increase the overlap between critical portions of test score distributions in individuals with and without disease (Gisslen et al. 2011, Meyer et al. 2013). The result of increased sensitivity is necessarily a reduction in specificity. Therefore, false-positive cases will lead to biased prevalence estimates and reductions in power for analytical estimates (Meyer et al. 203; Tierney et al. 2017). However, Frascati criteria are the most widely used criteria for diagnosing HAND in clinical settings and research. Direct validation of the criteria for ANI and MND rely on neuropsychological testing as there are no reliable longitudinal clinical-pathological correlation studies, nor a gold standard antemortem biomarker or imaging finding.

Our results confirm the main potential benefit of MoCA, as a test promising to significantly decrease the cognitive assessment time and costs. However, the optimal threshold for this tool should be lower than 26 (see Table 1 and Supplemental Table 2).

Overall, we recommend against approaches that use MoCA in isolation. A possible solution could consist of a short battery of tests, including the MoCA, requiring 10 to 30 minutes to complete, which could enhance both sensitivity and specificity and could be used in settings with limited resources (Joska et al. 2016). Future studies could compare multiple brief screening tests with a full neuropsychological battery in order to optimize a screening tool that can reliably detect HAND. Additionally, further cross-sectional studies are required to examine the optimum cut-off score for HAND. In this regard, different thresholds should be tested in individuals with multiple

cultural and educational backgrounds, and speaking different languages. Researchers should also consider the value of MoCA in a diagnostic workup so that clinicians understand how to use this screening test to attain relevant outcomes for patients, such as the benefits of earlier diagnostic and the harms of unnecessary testing.

The objective results from a screening test like MoCA are still likely to be more reliable than the information provided by patients or self-reports (De Francesco et al. 2016; Obermeit et al. 2017). Patients with abnormal screening results should be further assessed for the underlying causes of cognitive impairments such as mood disorders, cognition-impairing effects of ART, thyroid disease, syphilis and B12 deficiency. These abnormalities should be correctly identified before referring patients for a full neuropsychological assessment (Hakkers et al. 2017). A stepwise protocol including cognitive screening would be easy to implement in routine clinical practice showing physicians how to deal with this complex problem (Hakkers et al. 2017).

Our study has certain limitations. First, the relatively low number of the studies, where no more than 287 patients could be included in this analysis, requires particular caution when interpreting our results, especially in the case of the threshold score of 22. Second, there was significant heterogeneity among the studies with regards to demographic differences, language, cultural and educational background. Differences in cultural and educational experiences may result in lower performance on neuropsychological tests. Normative corrections (i.e., for age and gender, education, and ethnicity) are not always available for all populations of PLVH or they are based on a limited set of demographic factors. This can induce bias when evaluating cognitive impairment (Devlin & Giovannetti 2017). For example, although MoCA was demonstrated to be influenced by age, educational level and cultural background, its norms have been published and stratified by age and educational level only in the following languages: English, Quebec- French, Italian, Portuguese, Japanese and Czech (Carson et al. 2017). In addition, other factors may introduce unrecognized biases, such as the total central nervous system penetration-effectiveness (CPE) score, polypharmacy, medication side effects, CD4 count and viral loads (Koening et al. 2016). However, heterogeneity is reasonably assumed in diagnostic test accuracy studies and most approaches to pooling test accuracy data consider this aspect in the analysis.

In conclusion, despite limitations mentioned before, our meta-analysis represents the first systematic review of the literature published in this field and describes an accurate comparison between the MoCA thresholds in patients infected with HIV. The MoCA test appears to be a reasonable screening tool for HIV-infected patients, especially when our recommended threshold score of 23 is used, as it offers the best balance between true positive

and false positive results (with a sensitivity of 0.44 and a specificity of 0.79). Nonetheless, our findings indicate that optimal threshold for MoCA always comes with a sensitivity-specificity trade-off, the preferred cut point depending on whether sensitivity or specificity is more valuable in a given context.

**Conflict of interest:** The authors declare that they have no conflict of interest.

**References**

Aberg, J.A., Gallant, J.E., Ghanem, K.G., Emmanuel, P., Zingman, B.S., Horberg, M.A. (2014). Primary care guidelines for the management of persons infected with HIV: 2013 update by the HIV Medicine Association of the Infectious Diseases Society of America. *Clinical Infectious Diseases*, 58(1), e1-e34.

Angus, B., Awosusi, F., Barker,  G., Boffito, M., Brook, G.,Das, S., et al. (2016). BHIVA guidelines for the routine investigation and monitoring of adult HIV-1-positive individuals. https://www.bhiva.org/monitoring-guidelines. Accessed 12 January 2019.

Antinori, A., Arendt, G., Becker, J.T., Brew, B.J., Byrd, D.A., Cherner, M., et al. (2007). Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*, 69(18), 1789–1799.

Bezdicek, O., Majerova, V., Novak, M., Nikolai, T., Ruzicka, E., & Roth, J.  (2013). Validity of the Montreal Cognitive Assessment in the detection of cognitive dysfunction in Huntington's disease. *Applied Neuropsychology: Adult*, 20(1), 33-40.

Bottiggi, K.A., Chang, J.J., Schmitt, F.A., Avison, M.J., Mootoor, Y., Nath, A., et al. (2007). The HIV dementia scale: predictive power in mild dementia and HAART. *Journal of the Neurological Sciences*, 260(1–2), 11–15.

Carson, N., Leach, L., & Murphy, K.J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry*, 33(2), 379-388.

Ciccarelli, N., Fabbiani, M., Colafigli, M., Trecarichi, E.M., Silveri, M.C., Cauda, R., et al. (2013). Revised central nervous system neuropenetration- effectiveness score is associated with cognitive disorders in HIV-infected patients with controlled plasma viraemia. *Antiviral Therapy*, 18, 153–160.

Chan, P., & Brew, B.J. (2014). HIV associated neurocognitive disorders in the modern antiviral treatment era: prevalence, characteristics, biomarkers, and effects of treatment. *Current HIV/AIDS Reports*, 11(3), 317–324.

Costa, A.S., Fimm, B., Friesen, P., Soundjock, H., Rottschy, C., Gross, T., et al. (2012). Alternate-form reliability of the Montreal cognitive assessment screening test in a clinical setting. *Dementia and Geriatric Cognitive Disorders*, 33(6), 379-384.

Crowe, S.F. (1998). The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. *Journal of clinical psychology*, 54(5), 585–591.

Cysique, L.A., Maruff, P., & Brew, B.J. (2004). Prevalence and pattern of neuropsychological impairment in human immunodeficiency virus-infected/acquired immunodeficiency syndrome (HIV/AIDS) patients across pre and post highly active antiretroviral therapy eras: a combined study of two cohorts. *Journal of Neurovirology*, 10, 350–357.

Cysique, L.A., Waters, E.K., & Brew, B.J. (2011). Central nervous system antiretroviral efficacy in HIV infection: a qualitative and quantitative review and implications for future research. *BMC Neurology*, 11, 148.

Davis, D.H., Creavin, S.T., Noel-Storr, A., Quinn, T.J., Smailagic, N., Hyde, C., et al. (2013). Neuropsychological tests for the diagnosis of Alzheimer's disease dementia and other dementias: a generic protocol for cross-sectional and delayed-verification studies. *Cochrane Database Systematic Review*, 3, CD010460.

Davis, D.H., Creavin, S.T., Yip, J.L., Noel-Storr, A.H., Brayne, C., Cullum, S. (2015). Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane Database Systematic Review*, 10, CD010775.

Dawes, S., Suarez, P., Casey, C.Y., Cherner, M., Marcotte, T.D., Letendre, S., et al. (2008). Variable patterns of neuropsychological performance in HIV-1 infection. *Journal of Clinical and Experimental Neuropsychology*, 30, 613–626.

De Francesco, D., Underwood, J., Post, F.A., Vera, J.H., Williams, I., Boffito, M., et al. (2016). Defining cognitive impairment in people-living-with-HIV: the POPPY study. *BMC Infectious Diseases*, 16, 617.

Devlin, K.N. & Giovannetti, T. (2017). Heterogeneity of neuropsychological impairment in HIV infection : contributions from mild cognitive impairment. *Neuropsychology Review*, 27(2), 101-123.

European AIDS Clinical Society. EACS Guidelines 2018. www.eacsociety.org/guidelines/eacs-guidelines/eacs-guidelines.html. Accessed 12 January 2019.

Fazeli, P.L., Casaletto, K.B., Paolillo, E., Moore, R.C., Moore, D.J., The Hnrp Group. (2017). Screening for neurocognitive impairment in HIV-positive adults aged 50 years and older: Montreal Cognitive Assessment relates to self-reported and clinician-rated everyday functioning. *Journal of Clinical and Experimental Neuropsychology*, 39(9), 842-853.

Freitas, S., Simoes, M.R., Alves, L., & Santana, I. (2013). Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease. *Alzheimer's Disease & Associated Disorders*, 27(1), 37-43.

Gaestel, Y., Amieva, H., Letenneur, L., Dartigues, J.F., & Fabrigoule, C. (2006). Cube drawing performances in normal ageing and Alzheimer's disease: data from the PAQUID elderly population-based cohort. *Dementia and geriatric cognitive disorders*, 21(1), 22–32.

Giancola, M.L., Lorenzini, P., Balestra, P., Larussa, D., Baldini, F., Corpolongo, A., et al. (2006). Neuroactive antiretroviral drugs do not influence neurocognitive performance in less advanced HIV-infected patients responding to highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes*, 41(3), 332-337.

Gill, D.J., Freshman, A., Blender, J.A., & Ravina, B. (2008). The Montreal Cognitive Assessment as a screening tool for cognitive impairment in Parkinson's disease. *Movement Disorders*, 23, 1043-1046.

Gisslen, M., Price, R.W., & Nilsson S. (2011). The definition of HIV-associated neurocognitive disorders: Are we overestimating the real prevalence? *BMC Infectious Diseases*, 11, 356.

Gorman, A.A, Foley, J.M., Ettenhofer, M.L, Hinkin, C.H., & van Gorp, W.G. (2009). Functional consequences of HIV-associated neuropsychological impairment. *Neuropsychology Review*, 19, 186–203.

Grant, I., Franklin, D.R Jr., Deutsch, R., Woods, S.P., Vaida, F., Ellis, R.J., et al. (2014). Asymptomatic HIV-associated neurocognitive impairment increases risk for symptomatic decline. *Neurology*, 82, 2055-2062.

Haddow, L.J., Floyd, S., Copas, A., & Gilson, R. J. (2013). A systematic review of the screening accuracy of the HIV Dementia Scale and International HIV Dementia Scale. *PLoS One*, 8(4), e61826.

Hakkers, C.S., Kraaijenhof, J.M., Van Oers-Hazelzet, E.B., Visser-Meily, A.J.M.A., Hoepelman, A.I.M., Arends J.E., et al. (2017). HIV and cognitive impairment in clinical practice: the evaluation of a stepwise screening protocol in relation to clinical outcomes and management. *AIDS Patient Care and STDs*, 31(9), 363-369.

Handbook for Diagnostic Test Accuracy Reviews. http://methods.cochrane.org/sdt/handbook-dta-reviews . Accessed 25 January 2019.

Hardy, D.J. & Hinkin, C.H. (2002). Reaction time performance in adults with HIV/AIDS. *Journal of Clinical and Experimental Neuropsychology*, 24, 912–929.

Haziot, M.E.J., Barbosa Junior, S.P., Vidal, J.E., de Oliveira, F.T.M., de Oliveira, A.C.P. (2015). Neuroimaging of HIV-associated neurocognitive disorders. *Dementia & Neuropsychologia*, 9(4), 380-384.

Heaton, R.K., Clifford, D.B., Franklin, D.R. Jr., Woods, S.P., Ake, C., Vaida, F., et al. (2010). HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER Study. *Neurology*, 75, 2087–2096.

Heaton, R.K., Franklin, D.R., Ellis, R.J., McCutchan, J.A., Letendre, S.L., Leblanc, S., et al. (2011). HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: differences in rates, nature, and predictors. *Journal of Neurovirology*, 17, 3–16.

Henry, J.D., & Crawford, JR. (2004). Verbal fluency deficits in Parkinson's disease: a meta-analysis. *Journal of the International Neuropsychological Society*, 10(4), 608–622.

HIV/AIDS Italian Expert Panel. (2017). Linee Guida Italiane sull'utilizzo della Terapia Antiretrovirale e la gestione diagnostico-clinica delle persone con infezione da HIV-1. https://penta-id.org/news/new-italian-guidelines-on-the-use-of-antiretroviral-therapy-and-the-clinical-diagnostic-management-of-hiv-1-affected-patients/. Accessed 13 January 2019.

Hoops, S., Nazem, S., Siderowf, A.D., Duda, J.E., Xie, S.X., Stern, M.B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, 73(21), 1738–1745.

Ismail, Z., Rajji, T.K. & Shulman, K.I. (2010). Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry,* 25(2), 111-120.

Jacova, C., Kertesz, A., Blair, M., Fisk, J.D., & Feldman, H.H. (2007). Neuropsychological testing and assessment for dementia. *Alzheimer's & Dementia,* 3(4), 299-317.

Janssen, M.A.M., Bosch, M., Koopmans, P.P., & Kessels, R.P.C. (2015). Validity of the Montreal Cognitive Assessment and the HIV Dementia Scale in the assessment of cognitive impairment in HIV-1 infected patients. *Journal of Neurovirology*, 21(4), 383–390.

Joska, J.A., Witten, J., Thomas, K.G., Robertson, C., Casson-Crook, M., Roosa, H., et al. (2016). A comparison of five brief screening tools for HIV-associated neurocognitive disorders in the USA and South Africa. *AIDS and Behavior*, 20(8), 1621–1631.

Lees, R., Selvarajah, J., Fenton, C., Pendlebury, S.T., Langhorne, P., Stott, D.J., & Quinn, T.J. (2014). Test accuracy of cognitive screening tests for diagnosis of dementia and multidomain cognitive impairment in stroke. *Stroke*, 45(10), 3008-3018.

Kami-Onaga, K., Tateyama, M., Kinjo, T., Parrott, G., Tominaga, D., Takahashi-Nakazato, A., et al. (2018). Comparison of two screening tests for HIV-Associated Neurocognitive Disorder suspected Japanese patients with respect to cART usage. *PLOS ONE*, 13(6), e0199106.

Koenig, N., Fujiwara, E., Gill, M.J., & Power, C. (2016). Montreal Cognitive Assessment Performance in HIV/AIDS: impact of systemic factors. *Canadian Journal of Neurological Sciences*, 43(1), 157-162.

Koski, L., Xie, H. & Finch, L. (2009). Measuring cognition in a geriatric outpatient clinic: Rasch analysis of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology*, 22, 151-160.

Ku, N.S., Lee, Y., Ahn, J.Y., Song, J.E., Kim, M.H., Kim, S.B., et al. (2014). HIV-associated neurocognitive disorder in HIV-infected Koreans: Korean NeuroAIDS Project. *HIV Medicine*, 15(8), 470–477.

Larsson, M.U., Almkvist, O., Luszcz, M.A., & Wahlin, T.B.R. (2008). Phonemic fluency deficits in asymptomatic gene carriers for Huntington's disease. *Neuropsychology*, 22(5), 596–605.

McArthur, J.C. (2004). HIV dementia: an evolving disease. *Journal of Neuroimmunology*, 157, 3-10.

McGee, S. (2002). Simplifying Likelihood Ratios. *Journal of General Internal Medicine*, 17(8), 647–650.

Meyer, A. C., Boscardin, W. J., Kwasa, J. K., & Price, R. W. (2013). Is it time to rethink how neuropsychological tests are used to diagnose mild forms of HIV-associated neurocognitive disorders? Impact of false-positive rates on prevalence and power. *Neuroepidemiology*, 41(3-4), 208-216.

Milanini, B., Wendelken, L.A., Esmaeili-Firidouni, P., Chartier, M., Crouch, P.C., Valcour, V. (2014). The Montreal cognitive assessment to screen for cognitive impairment in HIV patients older than 60 years. *Journal of Acquired Immune Deficiency Syndromes*, 67(1), 67–70.

Milanini, B., Ciccarelli, N., Fabbiani, M., Baldonero, E., Limiti, S., Gagliardini, R., et al. (2016). Neuropsychological screening tools in Italian HIV+ patients: a comparison of Montreal Cognitive Assessment (MoCA) and Mini Mental State Examination (MMSE). The Clinical Neuropsychologist, 30(sup1), 1457-1468.

Mind Exchange Working Group. (2013). Assessment, diagnosis, and treatment of HIV-associated neurocognitive disorder: a consensus report of the mind exchange program. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 56(7), 1004-1017.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G.; and the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269.

Munoz-Moreno, J.A., Perez-Alvarez, N., Munoz-Murillo, A., Prats, A., Garolera, M., Jurado, M.À., et al. (2014). Classification models for neurocognitive impairment in HIV infection based on demographic and clinical variables. *PLoS One*, 9(9), e107625.

Nasreddine, Z.S., Phillips, N.A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699.

Nie, K., Zhang, Y., Wang, L., Zhao, J., Huang, Z., Gan, R., et al. (2012). A pilot study of psychometric properties of the Beijing version of Montreal Cognitive Assessment in patients with idiopathic Parkinson's disease in China. *Journal of Clinical Neurosciences*, 19, 1497-1500.

Nightingale, S., Winston, A., Letendre, S., Michael, B.D., McArthur, J.C., Khoo, S., et al. (2014). Controversies in HIV-associated neurocognitive disorders. *Lancet Neurology*, 13(11), 1139-1151.

Nitrini, R., Caramelli, P., Herrera, E., Porto, C.S., Charchat- Fichman, H., Carthery, M.T., et al. (2004). Performance of illiterate and literate nondemented elderly subjects in two tests of long-term memory. *Journal of the International Neuropsychological Society*, 10(4), 634–638.

Obermeit, L.C., Beltran, J., Casaletto, K.B., Franklin, D.R., Letendre, S., Ellis, R., et al. (2017). Evaluating the accuracy of self-report for diagnosis of HIV-associated neurocognitive disorder (HAND): defining "symptomatic" versus "asymptomatic" HAND. *Journal of Neurovirology*, 23, 67-78.

O'Keeffe, E., Mukhtar, O., O'Keeffe, S.T. (2011). Orientation to time as a guide to the presence and severity of cognitive impairment in older hospital patients. *Journal of neurology, neurosurgery, and psychiatry*, 82(5): 500–504.

Overton. E.T., Azad, T., Parker, N., Demarco Shaw, D., Frain, J., Spitz, T., et al. (2013). The Alzheimer Disease-8 and Montreal Cognitive Assessment as Screening Tools for Neurocognitive Impairment in HIV-Infected Persons. *Journal of Neurovirology*, 19(1), 109–116.

Ozdilek, B., & Kenangil, G. (2014). Validation of the Turkish version of the Montreal Cognitive Assessment Scale (MoCA-TR) in patients with Parkinson's disease. *Clinical Neuropsychology*, 28, 333-343.

Pinto, E., & Peters, R. (2009). Literature review of the Clock Drawing Test as a tool for cognitive screening. *Dementia and geriatric cognitive disorders*, 27(3): 201–213.

Price, C.C., Cunningham, H., Coronado, N., Freedland, A., Cosentino, S., Penney, D.L., et al. (2011). Clock drawing in the Montreal Cognitive Assessment: recommendations for dementia assessment. *Dementia and geriatric cognitive disorders*, 31(3): 179–187.

Reitsma, J.B., Glas, A.S., Rutjes, A.W., Scholten, R.J., Bossuyt, P.M., & Zwinderman, A.H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10), 982-990.

Robertson, K., Liner, J., & Meeker, R.B. (2012). Antiretroviral neurotoxicity. *Journal of Neurovirology*, 18(5), 388–399.

Rutter, C.M., & Gatsonis, C.A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, 20(19), 2865-2884.

Ryan, J.J., Glass, L.A., Bartels, J.M., Bergner, C.M., & Paolo, A.M. (2009). Predicting neuropsychological test performance on the basis of temporal orientation. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition*, 16(3), 330–337.

Sacktor, N.C., Wong, M., Nakasujja, N., Skolasky, R.L., Selnes, O.A., Musisi, S., et al. (2005). The international HIV dementia scale: a new rapid screening test for HIV dementia. *AIDS*, 19(13), 1367–1374.

Sacktor, N., Skolasky, R.L., Seaberg, E., Munro, C., Becker, J.T., Martin, E., et al. (2016). Prevalence of HIV-associated neurocognitive disorders in the Multicenter AIDS Cohort Study. *Neurology*, 86(4), 334-3410.

Sánchez-Cubillo, I., Periáñez, J.A., Adrover-Roig, D., Rodríguez-Sánchez, J.M., Ríos-Lago, M., Tirapu, J., et al. (2009). Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15(3), 438–450.

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.

Simioni, S., Cavassini, M., Annoni, J.M., Rimbault, Abraham, A., Bourquin, I., Schiffer, V., et al. (2010). Cognitive dysfunction in HIV patients despite long-standing suppression of viremia. *AIDS*, 24, 1243–1250.

Šimundić, A.M. (2009). Measures of diagnostic accuracy: basic definitions. *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine EJIFCC*, 19(4): 203–211.

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature*, 384:460–463.

Skinner, S., Adewale, A.J., DeBlock, L., Gill, M.J., & Power, C. (2009). Neurocognitive screening tools in HIV/AIDS: comparative performance among patients exposed to antiretroviral therapy. *HIV Medicine*, 10, 246–252.

Small, J.A., Kemper, S., & Lyons, K. Sentence repetition and processing resources in Alzheimer's disease. *Brain and language*, 75(2): 232–258.

Smit, T.K., Brew, B.J., Tourtellotte, W., Morgello, S., Gelman, B.B., Saksena, N.K. (2004). Independent evolution of human immunodeficiency virus (HIV) drug resistance mutations in diverse areas of the brain in HIV-infected patients, with and without dementia, on antiretroviral treatment. *Journal of Virology*, 78(18), 10133-10148.

Smurzynski, M., Wu, K., Letendre, S., Robertson, K., Bosch, R.J., Clifford, D.B., et al. (2011). Effects of central nervous system antiretroviral penetration on cognitive functioning in the ALLRT cohort. *AIDS*, 25, 357-365.

Straus, S.E., Glasizou, P., Richardson, W. S., & Haynes, R. B. (2019). Evidence-based medicine: How to practice and teach EBM (5th ed.). Elsevier Limited.

Tierney, S. M., Sheppard, D. P., Kordovski, V. M., Faytell, M. P., Avci, G., & Woods, S. P. (2017). A comparison of the sensitivity, stability, and reliability of three diagnostic schemes for HIV-associated neurocognitive disorders. *Journal of Neurovirology*, 23(3), 404-421.

Troyer, A.K., Moscovitch, M., Winocur, G., Alexander, M.P., & Stuss, D. (1998). Clustering and switching on verbal fluency: the effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, 36(6): 499–504.

Underwood, J., & Winston, A. (2016). Guidelines for evaluation and management of cognitive disorders in HIV-Positive individuals. *Current HIV/AIDS Reports*, 13(5), 235–240.

Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., et al. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155, 529-536.

Woods, S. P., Moore, D. J., Weber, E., & Grant, I. (2009). Cognitive neuropsychology of HIV-associated neurocognitive disorders. *Neuropsychology Review*, 19(2), 152-68.

World Health Organization. (2016). Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations – 2016 update. Geneva: World Health Organization.

**LEGENDS:**

**Figure captions:**

Figure 1 Risk of bias and applicability concerns graph: review author's judgements about each domain presented as percentages across included studies

Figure 2 Risk of bias and applicability concerns summary: review author's judgements about each domain for each included study

Figure 3 Forest plots of MoCA at different thresholds

Figure 4 SROC curve for the cut-off scores of 27 and 26

**Tables:**

Table 1 Evaluation of MoCA thresholds

**Electronic Supplementary Material:**

Supplemental Figure 1 Study selection flow chart

Supplemental Table 1 Characteristics of included studies.

Supplemental Table 2 Detailed evaluation of MoCA thresholds.

Supplemental Table 3 Detailed analysis of Youden index and Phi for each threshold.

| | Risk of Bias | | | | Applicability Concerns | |
|---|---|---|---|---|---|---|
| Patient Selection | 3 | 1 | 4 | | 8 | |
| Index Test | 7 | | 1 | | 8 | |
| Reference Standard | 7 | | 1 | | 8 | |
| Flow and Timing | 5 | | 3 | | | |

**Risk of Bias**   **Applicability Concerns**

■ High   ■ Unclear   ■ Low
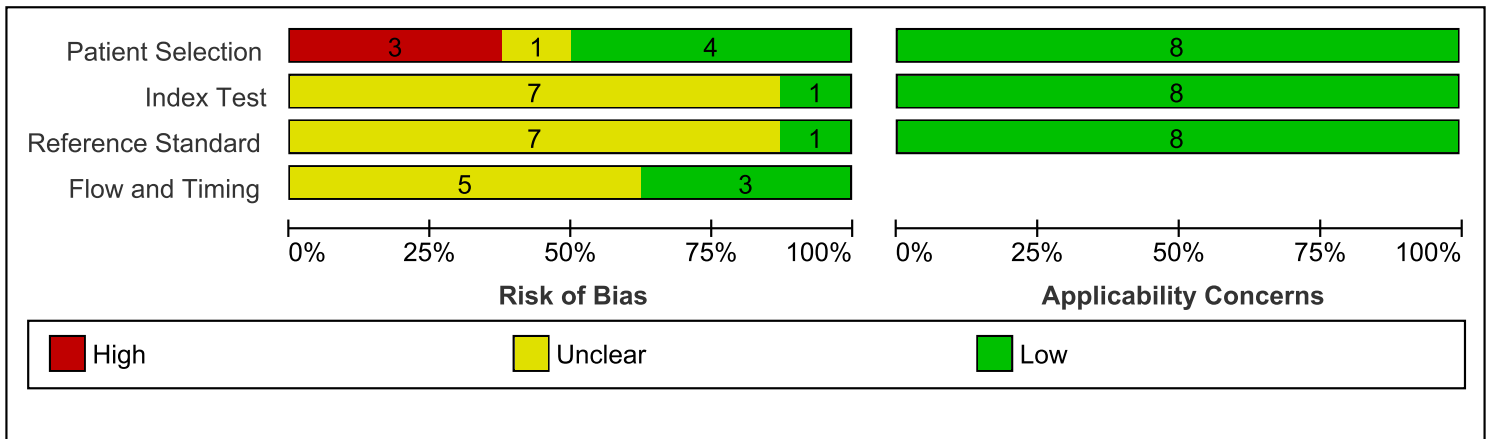
|  | Risk of Bias | | | | | Applicability Concerns | | |
|---|---|---|---|---|---|---|---|---|
|  | Patient Selection | Index Test | Reference Standard | Flow and Timing | | Patient Selection | Index Test | Reference Standard |
| Fazeli 2017 | − | ? | ? | ? | | + | + | + |
| Janssen 2015 | + | ? | ? | ? | | + | + | + |
| Joska 2016 | + | ? | + | + | | + | + | + |
| Koenig 2016 | − | ? | ? | ? | | + | + | + |
| Ku 2014 | + | ? | ? | ? | | + | + | + |
| Milanini 2014 | − | ? | ? | ? | | + | + | + |
| Milanini 2016 | + | + | ? | + | | + | + | + |
| Overton 2013 | ? | ? | ? | + | | + | + | + |

**−** High     **?** Unclear     **+** Low

## MoCA threshold 27

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Janssen 2015 | 22 | 21 | 17 | 35 | 0.56 [0.40, 0.72] | 0.63 [0.49, 0.75] |
| Ku 2014 | 42 | 86 | 9 | 57 | 0.82 [0.69, 0.92] | 0.40 [0.32, 0.48] |
| Milanini 2014 | 9 | 6 | 18 | 34 | 0.33 [0.17, 0.54] | 0.85 [0.70, 0.94] |
| Overton 2013 | 114 | 42 | 13 | 31 | 0.90 [0.83, 0.94] | 0.42 [0.31, 0.55] |

## MoCA threshold 26

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Fazeli 2017 | 38 | 17 | 6 | 23 | 0.86 [0.73, 0.95] | 0.57 [0.41, 0.73] |
| Janssen 2015 | 20 | 13 | 19 | 43 | 0.51 [0.35, 0.68] | 0.77 [0.64, 0.87] |
| Joska 2016 | 58 | 68 | 7 | 20 | 0.89 [0.79, 0.96] | 0.23 [0.14, 0.33] |
| Ku 2014 | 34 | 63 | 17 | 80 | 0.67 [0.52, 0.79] | 0.56 [0.47, 0.64] |
| Milanini 2014 | 13 | 6 | 14 | 34 | 0.48 [0.29, 0.68] | 0.85 [0.70, 0.94] |
| Overton 2013 | 96 | 33 | 31 | 31 | 0.76 [0.67, 0.83] | 0.48 [0.36, 0.61] |

## MoCA threshold 25

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Ku 2014 | 27 | 38 | 24 | 105 | 0.53 [0.38, 0.67] | 0.73 [0.65, 0.80] |
| Milanini 2014 | 19 | 13 | 8 | 27 | 0.70 [0.50, 0.86] | 0.68 [0.51, 0.81] |
| Overton 2013 | 80 | 21 | 47 | 52 | 0.63 [0.54, 0.71] | 0.71 [0.59, 0.81] |

## MoCA threshold 24

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Ku 2014 | 19 | 18 | 32 | 125 | 0.37 [0.24, 0.52] | 0.87 [0.81, 0.92] |
| Milanini 2014 | 22 | 16 | 5 | 24 | 0.81 [0.62, 0.94] | 0.60 [0.43, 0.75] |
| Overton 2013 | 67 | 13 | 60 | 60 | 0.53 [0.44, 0.62] | 0.82 [0.71, 0.90] |

## MoCA threshold 23

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Koenig 2016 | 18 | 38 | 6 | 63 | 0.75 [0.53, 0.90] | 0.62 [0.52, 0.72] |
| Ku 2014 | 10 | 10 | 41 | 133 | 0.20 [0.10, 0.33] | 0.93 [0.88, 0.97] |
| Milanini 2014 | 24 | 22 | 3 | 18 | 0.89 [0.71, 0.98] | 0.45 [0.29, 0.62] |
| Overton 2013 | 48 | 4 | 79 | 69 | 0.38 [0.29, 0.47] | 0.95 [0.87, 0.98] |

## MoCA treshold 22

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Ku 2014 | 7 | 4 | 44 | 139 | 0.14 [0.06, 0.26] | 0.97 [0.93, 0.99] |
| Milanini 2016 | 8 | 19 | 5 | 61 | 0.62 [0.32, 0.86] | 0.76 [0.65, 0.85] |

Table 1. Evaluation of MoCA at different thresholds.

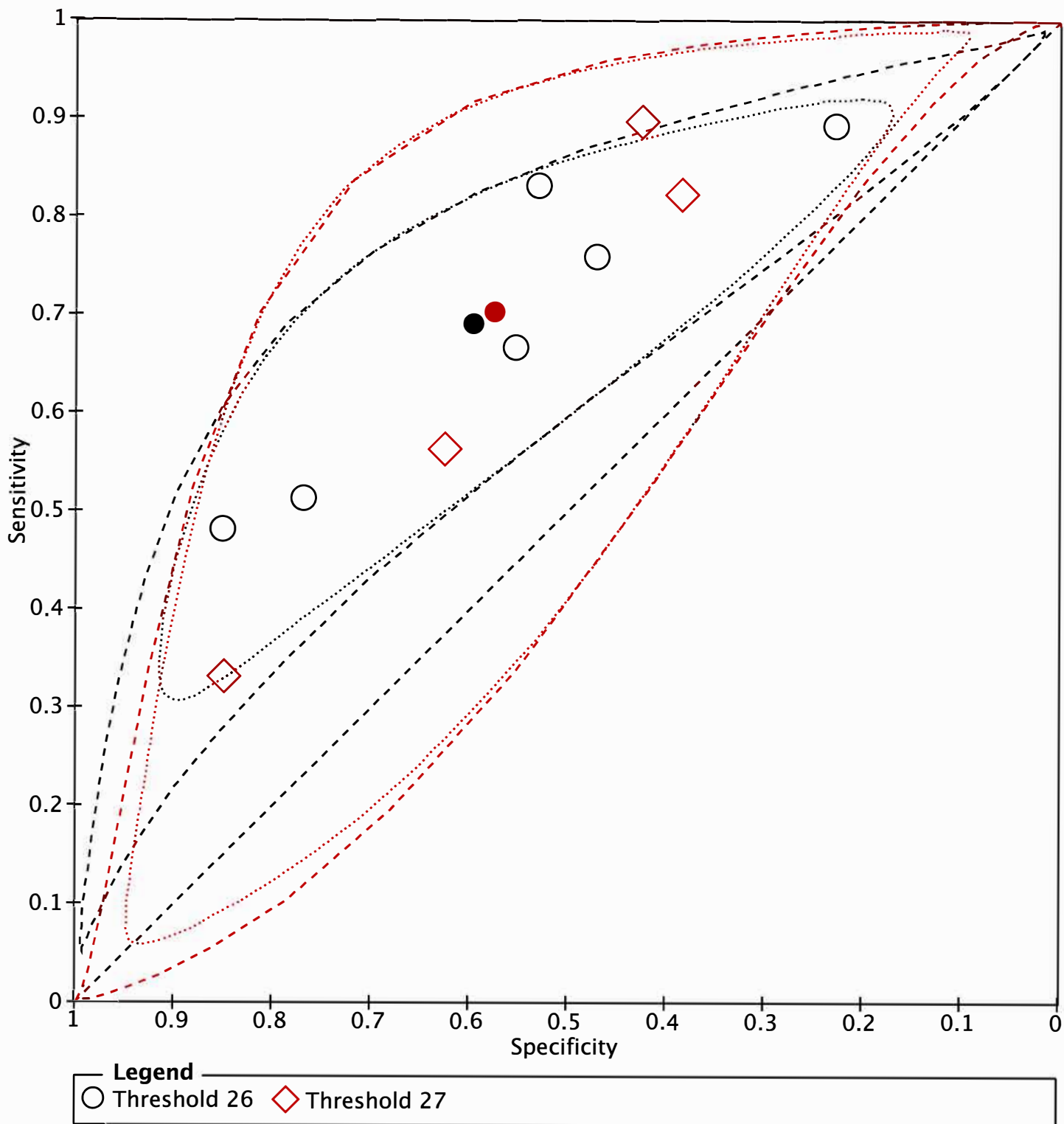| Studies | Threshold | Number of patients | Sensitivity (95%CI) | Specificity (95% CI) | TP | FP | FN | TN | LR+ (95%CI) | LR- (95%CI) | Youden index (mean) | Phi (mean) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2, 4, 5, 6 | 27 | 556 | 0.77 (0.71, 0.82) | 0.50 (0.45, 0.56) | 187 | 155 | 57 | 157 | 1.54 (1.35, 1.76) | 0.46 (0.36, 0.60) | 0.280 | 0.428 |
| 1, 2 4, 5, 6,7 | 26 | 784 | 0.73 (0.68, 0.78) | 0.54 (0.49, 0.58) | 259 | 200 | 94 | 231 | 1.58 (1.40, 1.78) | 0.50 (0.41, 0.60) | 0.293 | 0.385 |
| 4, 5, 6 | 25 | 461 | 0.61 (0.54, 0.68) | 0.72 (0.66, 0.77) | 126 | 72 | 79 | 184 | 2,18 (1.75, 2.73) | 0.54 (0.44, 0.65) | 0.332 | 0.390 |
| 4, 5, 6 | 24 | 461 | 0.53 (0.46, 0.60) | 0.82 (0.76, 0.86) | 108 | 47 | 97 | 209 | 2.87 (2.15, 3.83) | 0.58 (0.50, 0.68) | 0.375 | 0.313 |
| 3, 4, 5, 6 | 23 | 586 | 0.44 (0.37, 0.50) | 0.79 (0.75, 0.83) | 100 | 74 | 129 | 283 | 2.11 (1.64, 2.71) | 0.71 (0.63, 0.81) | 0.384 | 0.292 |
| 5, 8 | 22 | 287 | 0.23 (0.14, 0.36) | 0.90 (0.85, 0.93) | 15 | 23 | 49 | 200 | 2.27 (1.26, 4.09) | 0.85 (0.74, 0.98) | 0.283 | 0.426 |

Note: Studies are represented by numbers. 1 = Joska et al. 2016; 2 = Janssen et al. 2015; 3 = Koenig et al. 2016 ;  4 = Milanini et al. 2014; 5 =  Ku et al. 2014; 6 = Overton et al. 2013 ; 7 = Fazeli et al. 2017 ; 8 = Milanini et al. 2016. TP = true positives ; FP = false positives ; FN = false negatives ; TN = true negatives ; LR+ = likelihood ratio for positive result ; LR- = likelihood ratio for negative result.

Supplemental Table 1. Characteristics of included studies.

| Study | Setting | Number of patients | Gender | Reference test | Age (years) | Education | CD4 Cells/ml | CD4 nadir Cells/ml | Viral load | Percent of patients on cART |
|---|---|---|---|---|---|---|---|---|---|---|
| Fazeli et al. 2017 | USA[1] | 84 | 88% males | Frascati criteria (6 cognitive domains) | 58.2 (SD 6.5) | 14.3 years (SD 2.6) | 597.0 (365-776) | 135.5 (39.5-300) | 92% undetectable | 98% |
| Joska et al. 2016 | South Africa[2] and USA[3] | 156 | 37.2% males | Frascati criteria (9 tests) | 40 (median) | 11 years (median) | 460 (median) | - | 95% suppressed | 100% |
| Milanini et al. 2016 | Italy[4] | 93 | 75.3% males | Frascati criteria (6 cognitive domains) | 47 (median) | 13 years (median) | 644 (481-789) | 193 (67-298) | 90% suppressed | 100% |
| Janssen et al, 2015 | Netherlands[5] | 95 | 87.4% males | Frascati criteria (9 major cognitive domains) | 48.2 (SD 10.1) | 7 – 17 years | - | 213 (IQR 100-305) | < 50 copies/ml (100% of patients) | 100% |

| Koenig et al, 2016 | Canada[6] | 125 | 83.2% males | Frascati criteria (10 tests) | 46.6 (SD 10.3) | < Grade 12 (30.4%) | 485.9 (SD 292.6) | 144.8 (SD 108.7) | - | 100% |
|---|---|---|---|---|---|---|---|---|---|---|
| Ku et al, 2014 | South Korea[7] | 194 | 93.8% males | Frascati criteria (10 tests) | 45.12 (range 21-72) | 13.4 (SD 3.3) | 481.4 (SD 236.0) | 187.0 (SD 138) | - | 89.70% |
| Milanini et al, 2014 | USA[8] | 67 | 96% males | Frascati criteria | 64 (range 60-84) | 16.3 (SD 2.32) | 549 (SD 246) | 240 (SD 162) | < 70 copies/ml in 79% of patients | 100% |
| Overton et al, 2013 | USA[9] | 200 | 71.5% males | Frascati criteria (8 tests) | 43.3 (SD 10.7) | 13.1 (SD 2.6) | 538 (IQR 361-695) | 191 (IQR 70-300) | < 400 copies/ml in 100% of patients | 100% |

cART = combination anti-retroviral therapy

SD = standard deviation

IQR = interquartile range

[1] University of California, San Diego, (patients from Successfully Aging Seniors with HIV study, aged 50 years and above)

[2] Cape Town (the participants were previously recruited into larger parent cohort studies)

[3] Baltimore (the participants were previously recruited into larger parent cohort studies)

[4] Outpatient clinics, Institute of Clinical Infectious Diseases, Catholic University of Sacred Heart, Rome

[5] Outpatient clinics for infectious diseases at the Department of Internal Medicine in the Radboud University Medical Center in Nijmegen and at the same department in the Rijnstate Hospital

[6] Southern Alberta Clinic, Calgary

[7] Urban teaching hospitals in Seoul

[8] University of California, San Francisco (UCSF HIV over 60 cohort)

[9] Outpatient infectious diseases clinic at Washington University in St. Louis

Supplemental Table 2. Detailed evaluation of MoCA thresholds.

| Studies | Threshold | Number of patients | Sensitivity | Specificity | TP | FP | FN | TN | LR+ | LR- | PPV | NPV | Pre-test probability (Prevalence) | Pre-test odds | Post-test odds | Post-test probability | Youden index (mean) | Phi (mean) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2, 4, 5, 6 | 27 | 556 | 0.77 | 0.50 | 187 | 155 | 57 | 157 | 1.54 | 0.46 | 0.55 | 0.73 | 0.44 | 0.79 | 1.22 | 0.55 | 0.280 | 0.428 |
| 1, 2 4, 5, 6,7 | 26 | 784 | 0.73 | 0.54 | 259 | 200 | 94 | 231 | 1.58 | 0.50 | 0.56 | 0.71 | 0.45 | 0.82 | 1.30 | 0.56 | 0.293 | 0.385 |
| 4, 5, 6 | 25 | 461 | 0.61 | 0.72 | 126 | 72 | 79 | 184 | 2,18 | 0.54 | 0.64 | 0.70 | 0.44 | 0.79 | 1.73 | 0.64 | 0.332 | 0.390 |
| 4, 5, 6 | 24 | 461 | 0.53 | 0.82 | 108 | 47 | 97 | 209 | 2.87 | 0.58 | 0.70 | 0.82 | 0.44 | 0.79 | 2.27 | 0.70 | 0.375 | 0.313 |
| 3, 4, 5, 6 | 23 | 586 | 0.44 | 0.79 | 100 | 74 | 129 | 283 | 2.11 | 0.71 | 0.57 | 0.69 | 0.39 | 0.64 | 1.35 | 0.57 | 0.384 | 0.292 |
| 5, 8 | 22 | 287 | 0.23 | 0.90 | 15 | 23 | 49 | 200 | 2.27 | 0.85 | 0.39 | 0.80 | 0.22 | 0.28 | 0.64 | 0.39 | 0.283 | 0.426 |

Note: Studies are represented by numbers. 1 = Joska et al. 2016; 2 = Janssen et al. 2015; 3 = Koenig et al. 2016 ; 4 = Milanini et al. 2014; 5 = Ku et al. 2014; 6 = Overton et al. 2013 ; 7 = Fazeli et al. 2017 ; 8 = Milanini et al. 2016. TP = true positives ; FP = false positives ; FN = false negatives ; TN = true negatives ; LR+ = likelihood ratio for positive result ; LR- = likelihood ratio for negative result; PPV = positive predictive value ; NPV = negative predictive value.

Supplemental Table 3. Detailed analysis of Youden index and Phi for each threshold.

| Threshold | Youden index | | | | Phi | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Median | 2.50 percentile | 97.5 percentile | Mean | Median | 2.50 percentile | 97.5 percentile |
| 27 | 0.280 | 0.281 | 0.077 | 0.472 | 0.428 | 0.423 | 0.226 | 0.669 |
| 26 | 0.293 | 0.292 | 0.198 | 0.383 | 0.385 | 0.382 | 0.297 | 0.493 |
| 25 | 0.332 | 0.331 | 0.238 | 0.426 | 0.390 | 0.385 | 0.286 | 0.514 |
| 24 | 0.375 | 0.381 | 0.201 | 0.513 | 0.313 | 0.302 | 0.151 | 0.521 |
| 23 | 0.384 | 0.395 | 0.191 | 0.536 | 0.292 | 0.285 | 0.164 | 0.446 |
| 22 | 0.283 | 0.284 | 0.099 | 0.470 | 0.426 | 0.420 | 0.210 | 0.675 |