A systematic review of the effectiveness of digital cognitive assessments of cognitive impairment in Parkinson's disease

Craig, S. N., Dempster, M., Curran, D., Cuddihy, A. M., & Lyttle, N. (2025). A systematic review of the effectiveness of digital cognitive assessments of cognitive impairment in Parkinson's disease. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2025.2454983

Published in:
Applied Neuropsychology: Adult

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Applied Neuropsychology: Adult

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hapn21

A systematic review of the effectiveness of digital cognitive assessments of cognitive impairment in Parkinson's disease

Saskia N. Craig, Martin Dempster, David Curran, Aoife. M. Cuddihy & Nigel Lyttle

To cite this article: Saskia N. Craig, Martin Dempster, David Curran, Aoife. M. Cuddihy & Nigel Lyttle (01 Feb 2025): A systematic review of the effectiveness of digital cognitive assessments of cognitive impairment in Parkinson's disease, Applied Neuropsychology: Adult, DOI: 10.1080/23279095.2025.2454983

To link to this article: https://doi.org/10.1080/23279095.2025.2454983

APPLIED NEUROPSYCHOLOGY: ADULT
https://doi.org/10.1080/23279095.2025.2454983

Routledge
Taylor & Francis Group

A systematic review of the effectiveness of digital cognitive assessments of cognitive impairment in Parkinson's disease

Saskia N. Craig[a], Martin Dempster[b], David Curran[a], Aoife. M. Cuddihy[a] and Nigel Lyttle[c]

[a]Department of Clinical Psychology, Queen's University, Belfast, Northern Ireland; [b]Department of Psychology Applied to Health & Illness, Queen's University, Belfast, Northern Ireland; [c]Department of Clinical Neuropsychology, Royal Victoria Hospital Belfast, Belfast, Northern Ireland

ABSTRACT

Background: Digitalization in healthcare has been extended to how we examine and manage Parkinson's Disease Mild Cognitive Impairment (PD-MCI). ***Methods***: Moyer Population (those with PD and in some cases control groups), Intervention (digital cognitive test) and Outcome (validity and reliability) (PIO) and Campbell et al. Synthesis Without Meta-analysis (SWiM) methods were employed. A literature search of MEDLINE, PsycINFO, CINAHL, OpenGrey, and ProQuest Theses and Dissertations Sources screened for articles. ***Results***: The digital trail-making test (dTMT) was the most used measure. There was strong validity between the dTMT and pencil-paper TMT, Mini-Mental State Examination (MMSE), and Montreal Cognitive Assessment (MoCA) scores (ranging from r = .55 to .90, p < .001). Validity between the TMT pencil-paper and digital versions were adequate (ranging from r = .51 to 90, p < .001). Reliability was demonstrated between PD and control groups' scores (ranging from r = .71 to .87). One study found excellent inter-rater reliability (ICC = .90 to .95). The dMoCA was the most used screen that assessed more than two cognitive domains. There was a range in the strength of agreement between digital and pencil-paper versions (ICC scores = .37 to .83) and only one study demonstrated adequate validity (r = .59, p < .001). Poor internal consistency (α = .54) and poor test re-test reliability (between PD and control groups' scores, p > .05) were found. ***Conclusion***: This review found that digitalized cognitive tests are valid and reliable methods to assess PD-MCI. Considerations for future research are discussed.

KEYWORDS

Digital methods; Mild Cognitive Impairment (MCI); neuropsychological cognitive testing; Parkinson's disease (PD)

Introduction

Idiopathic Parkinson's disease (PD) is the second most common neurodegenerative condition after Alzheimer's Disease (Pechstein et al., 2020). PD is characterized by changes within the brain, with a gradual depletion of dopamine-related neurons in the Substantia Nigra pars compacta (SNpc) (Hornykiewicz & Kish, 1987). PD has become more prevalent over the last 25 years, with cases rising from 2.5 million to 6.1 million (Rocca, 2018). The observed rise in confirmed cases might be explained by several reasons. Firstly, with an aging population worldwide (resulting from increasing life expectancy) there is a longer average disease duration. However, also, in clinical practice, there has been an improved awareness and recording of PD (Rocca, 2018).

While PD, for most, at least in the early stages, is considered to predominantly affect an individual's motor functioning in a fairly predictable and steady way, over time, the response to medications becomes less predictable. This results in significant fluctuations in the person's motor state. Early manifestations of the disease classically begin with symptoms of tremor, bradykinesia (slowness of movement), rigidity, and postural instability (Rawat & Pandey, 2022). In addition to the classic motor symptoms, PD in the early stages often includes cognitive alterations that may be overlooked, while psychological issues like low mood, anxiety, rapid eye movement, and sleep disorders can emerge as the disease progresses (Connolly & Fox, 2014; Jankovic & Tan, 2020). Moreover in the later stages, patients may experience more pronounced cognitive decline, significant motor impairment, and complex psychological difficulties (Rukavina et al., 2021). This highlights the need for comprehensive care and management tailored to the stages of the disease, including both pharmacological and non-pharmacological interventions (Calleo et al., 2012).

The full spectrum of cognitive impairment occurs in PD, ranging from subjective cognitive decline to mild cognitive impairment and Parkinson's disease dementia. To assist with the early identification of PD-Mild Cognitive Impairment (PD-MCI), The International Parkinson and Movement Disorder Society (IPMDS) has devised a working definition of PD-MCI: a gradual deterioration in cognitive functioning, reported by the patient, caregiver, or clinician, that is not caused by other comorbidities, as well as not interfering

CONTACT S. N. Craig scraig25@qub.ac.uk Department of Clinical Psychology, Queen's University Belfast, Belfast, Northern Ireland.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/23279095.2025.2454983.

with functional independence (Litvan et al., 2012). PD-MCI may occur early within the course/stage of the disease; however, this can vary from person to person. Furthermore, not everybody with PD-MCI will convert to PD Dementia at the same time/at all. The IPMDS further agreed on a two-level definition of assessing PD-MCI. Level 1 assessment involves clinicians either using a general cognitive screen, such as the Parkinson's Disease-Cognitive Rating Scale (PD-CRS), or a range of shorter tests to assess impairment in at least two tests of the same domain (Litvan et al., 2012). Level 2 assessment involves the inclusion of an extensive assessment of each domain (i.e., language, memory, attention, executive, and visuospatial functioning), identifying deficits in at least two tests within or across the five domains (Litvan et al., 2012).

In March 2020, the World Health Organization (WHO) declared Cornavirus-19 (COVID-19) to be a global pandemic, meaning typical face-to-face neurological care and psychological testing were compromised (Papa et al., 2020). PD patients were at an increased risk, due to the majority of those being within the elderly population and having other medical comorbidities (Papa et al., 2020). However, COVID-19 provided opportunities to demonstrate that standard pencil and paper cognitive testing could also be administered using digital platforms and accelerate efforts to offer accessible services to those who could not attend clinic appointments, for example, due to COVID-19 or living in remote locations (Libon et al., 2021). There is support for implementing videoconference (Brearly et al., 2017) and telephone methods (Grosch et al., 2011) within neuropsychological services for several neurological conditions. Focusing on PD, the validity of telemedicine has been well documented and is considered practical to use (Ben-Pazi et al., 2018). In the context of cognitive testing, test validity can be defined as the degree to which theory and findings support the interpretations of scores entailed by those that are proposed as cognitive tests (1999 Standards for Educational and Psychological Testing). At its most basic level, this means the degree to which a test measures what it is intended to measure (Schoenberg et al., 2011). Test reliability is also important to consider (and is defined as the degree to which results are accurate and would be the same if the data were replicated again, Cook et al., 2015). See Appendix A for definitions of these types of validity and reliability.

The IPMDS has developed a useful protocol for how clinicians can implement digital therapeutics (Papa et al., 2020). It could be argued that the need for using digital methods has never been greater, however, Artusi et al. (2020) have disputed that they tend to not be used in clinical practice. This may be related to it not being clear what cognitive capacity is required for using such technologies (Ellis & Earhart, 2021). Additionally, it remains unclear as to what mode of delivery (i.e. self-administered or not) and what cognitive tests are most valid, reliable, and sensitive to measure PD-MCI in clinical practice (Kokubo et al., 2018).

This review aims to explore the following questions by applying the Population Intervention Outcome (PIO) method (Moyer, 2008).

In PD patients (P), what digital psychometric assessments (I) have been used to measure mild cognitive impairment?

What is the validity, reliability, and test sensitivity of these measures to assess PD-MCI (O)?

To our knowledge, this is the first review that aims to identify and evaluate studies pertinent to the above research questions, using replicable, strict, and transparent procedures.

Materials and methods

Narrative synthesis

A Synthesis Without Meta-analysis (SWiM) (Campbell et al., 2020) design was employed due to the methodological heterogeneity (cross-sectional, case-control, and longitudinal studies) and clinical diversity (relative to the PIO; Population, Intervention, and Outcome method used) within the studies selected. SWiM methodology (Campbell et al., 2020) has been recommended by Cochrane guidelines, as a way to synthesize studies with high clinical and methodological heterogeneity (Higgins & Green, 2008). This review further adhered to the PRISMA Statement guidelines (Moher et al., 2008). The protocol for this paper can be accessed at www.crd.york.ac.uk/prospero (registration number, CRD42023429665).

Selection criteria

Information sources

To identify relevant literature, MEDLINE, PsychINFO, and CINAHL were searched. There were no restrictions in terms of language, publication period/timeframe, or sources. However, it was expected that most papers would be published within the last 10 years due to recent advances in technology and appreciation for utilizing digital cognitive assessments. *Opengrey* and *ProQuest theses and dissertations sources* were also searched from June to August 2023. Grey literature was included due to being recommended by Paez (2017). Five databases were searched, to decrease bias in searching, analyzing, and combining individual studies. No authors were contacted throughout this synthesis as all literature published was identifiable. The following key terms were used in various combinations and adapted according to each database: Parkinson* Disease, Mild Cognitive Impairment, Cognitive Assessment, Computer, Zoom, Smartphone, Telephone, and Laptop. This search strategy was completed in April 2023.

Eligibility and study Selection

All searches were screened for eligibility, using the following criteria:

Inclusion criteria

- Participants included in the study were adults who had a diagnosis of PD.
- Papers describe participants' PD diagnosis (either from a medical professional such as a Neurologist,

Movement Disorder Diagnostic Task Force Criteria, self-report, or how their diagnosis is not reported).
- Papers included people living with PD who have completed a test that measured their cognitive functioning.
- All or some aspects of the psychometric test were administered using online computerized technology methods (such as Zoom, Apps, PowerPoint, and laptop).
- Mixed designs, with a focus on the quantitative findings.

Exclusion criteria

- Studies that were only qualitative in design.
- Studies involving participants using only in-person assessments (such as being completed at a hospital clinic using pencil-paper methods).
- Telephone, 3D, or virtual reality assessments.

Two reviewers (SC and AC) completed the screening of titles and abstracts, using the eligibility criteria above. All remaining articles were then screened for eligibility (by SC and AC) using the full article text, being conducted independently (as recommended by Ouzzani et al., 2016). Inter-rater reliability, using weighted Cohen's Kappa statistic (measuring the agreement between two raters) was calculated (as recommended by Pérez et al., 2020) (see Appendix B). Any disagreements were addressed by a third reviewer (DC). SC performed citation chaining (forward and backward searching) on the final papers selected for inclusion, with no new articles identified. Therefore, 17 papers were eligible for inclusion (see Figure 1). Table 1 demonstrates the 4-level screening checklist that was used.

Quality assessment

Sirriyeh et al. (2012) Quality Assessment Tool for Studies with Diverse Designs (QATSDD) (see Appendix C) was selected as the quality appraisal instrument. The QATSDD tool combined the diverse studies, allowing each article to be assessed for quality assurance, being developed from existing quality assessment tools (such as CONSORT guidelines, STROBE, MOOSE, CASP approaches, and the Quality in Qualitative evaluation framework) (Sirriyeh et al., 2012). An exhaustive list of components that represented "good research design" was extracted from the above tools and synthesized to devise the QATSDD (as cited by Sirriyeh et al., 2012). Two reviewers (SC and DC) independently and subsequently compared notes to identify and resolve differences in interpretation, as suggested by Sirriyeh et al. (2012). This tool was shortened from 16 to 14 characteristics (two items were removed due to assessing qualitative studies). The two reviewers measured the severity of the quality of assessment using a 4-point classification system *(0= No mention, 1= Slightly, 2= Moderately, and 3= Complete*). The 4-point scale provided an accurate quality assessment (Sirriyeh et al., 2012). See Appendix F for the independent ratings given by the two raters. There were no guidelines available to interpret the overall quality. To assess reliability, SPSS statistical package version 29 calculated ICC estimates and their 95% confidence intervals between the two raters' scores, based on a mean-rating (k_x = 2), absolute-agreement, 2-way mixed-effects model.

Data extraction, data synthesis and heterogeneity

Data were extracted.

- Study information (authors, year published, and study location).
- Population Characteristics.
- Characteristics of the digital cognitive test.
- Validity of the digital cognitive test.
- Reliability of the digital cognitive test.

Two reviewers (SC and AC) independently screened all papers (based on their titles, abstracts, and full text) for potential inclusion. Their results were combined, and a third reviewer (DC) solved discrepancies. SC and MD performed data extraction independently, using Microsoft Excel, which recorded details about the study objectives, cognitive tests used, participants, study methods, and outcomes of significance to the review questions. The results were compared and synthesized by SC and MD.

Due to the high methodological heterogeneity (i.e. cross-sectional, case-control, and prospective study designs) and high clinical diversity within the studies selected, a meta-analysis was not appropriate as data were not sufficiently alike (Borenstein et al., 2017). Therefore, a narrative synthesis of results took place, which included descriptions of participants, study characteristics, and cognitive testing administered) and the validity and reliability of the cognitive tests. Textual descriptions and a tabulation approach aimed to summarize and explain the findings of the synthesis (Popay et al., 2006). Other graphical and tabular methods (such as forest and harvest plots) were not included within the studies, as only three out of the 17 studies had data reporting/available to complete these methods (Campbell et al., 2020). Additionally, grouping (organizing the studies and their findings, as suggested by Popay et al., 2006) allowed the studies to be collated into four domains: demographics, cognitive tests used, validity, and reliability. Studies were grouped using this system, aiming to answer the research questions, adhere to the PIO method, and provide insight into the authors that have advanced knowledge within this research area (Popay et al., 2006). Finally, data reported a common metric (p values) since the quantitative data within selected studies were presented in different numerical and statistical forms (Popay et al., 2006). Adhering to the SWiM guidance (Campbell et al., 2020), p values were chosen as they hold the advantage of combining a variety of statistical tests and were used due to the absence of effect sizes. Quantitative data refers to reporting the different types of validity (convergence, predictive, and external) and

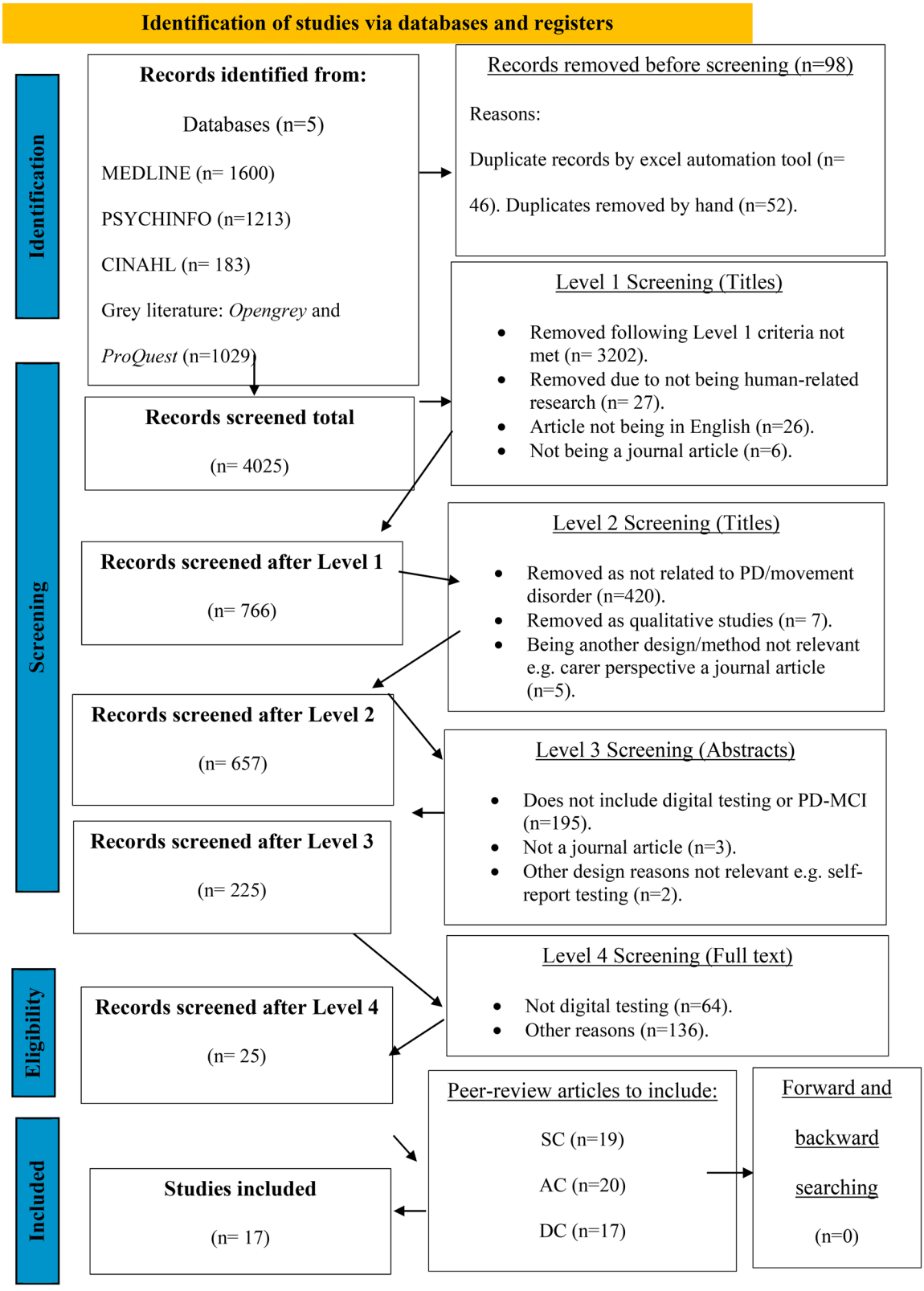

Figure 1. Prisma Flowchart visually depicting how selected studies were included.

reliability (inter-rater, parallel forms, internal consistency, and test re-test) that studies examined. The statistical tests utilized to examine validity included Pearson's correlation with tests of significance, and Bland Altman plots. The statistical tests utilized to examine reliability included Intra Class Correlation, Cohen's d, Pearson's correlation, Cronbach's alpha, and Kolmogorov–Smirnov statistic. Test sensitivity was reported independently under the reliability section, as it had been highlighted distinctly in the outcomes of seven studies. This was addressed separately because it was not initially included in the objectives outlined during the Prospero registration phase.

Table 1. PRISMA flowchart level checklist.

	Description	Checklist
Level 1	Article reviewed by Title	1. Was the study conducted in humans? Yes No Unclear 2. Does the paper title include Parkinson's Disease/ movement disorder? Yes No Unclear 3. Does the paper title include cognitive/ neuropsychometric testing? Yes No Unclear If yes or unclear to any, move to level 2.
Level 2	Article reviewed by Title	1. Does this study include digital cognitive assessment? (i.e. remotely such as via computer or telephone?) Yes No Unclear 2. Does the study appear to have a relevant study design (cannot be qualitative) Yes No Unclear If yes or unclear to any, move to level 3.
Level 3	Article reviewed by Abstract	1. Does this study include people living with PD completing cognitive assessment? Yes No Unclear 2. Does this study include delivery of cognitive assessment via digital methods? (i.e. computer or telephone?) Yes No Unclear 3. Does the study appear to have a relevant study design (cannot be qualitative) Yes No Unclear If yes or unclear to any, move to level 4.
Level 4	Article reviewed by Full text	1. Does this study include people living with PD completing cognitive assessment? Yes No Unclear 2. Does this study include delivery of cognitive assessment via digital methods? (i.e. computer or telephone?) Yes No Unclear 3. Does the study appear to have a relevant study design (cannot be qualitative) Yes No Unclear

Results

Study Selection

We identified 4,025 papers of potential relevance following the removal of duplicates (see Figure 1 for the PRISMA flow diagram). When titles and abstracts were screened, 4,000 papers were removed. From the remaining 25 papers, eight were excluded when screened at full text (level 4 screening). There was a statistically significant agreement between the two raters, $\kappa_{w=}$ .837, 95% *CI.* [.703–.970], p <.001. The strength of agreement was classified as very good according to Landis and Koch (1977) and excellent according to Fleiss et al. (2013). In total, 17 studies were included.

Quality assessment of included studies

Guidelines were not available to interpret the Sirriyeh et al. (2012) QATSDD tool. However, scores ranged from 25 to 39 out of a maximum of 42. The highest quality ratings were found among seven studies (scoring over 30), among the following criteria: explicit theoretical framework, clear description of the research setting, the procedure of data collection, detailed data recruitment, the fit between research question and analysis, and strengths/limitations discussed. The remaining studies' quality ratings scores ranged from 20 to 29, with the following criteria lowering their scores: sample size considered and service-user involvement. On the validity and reliability criterion, 11 studies scored highly (score of 2 or 3), with the remaining studies scored low (1 or 0 score).

To assess the reliability between the two raters for their quality of assessment scores, ICC estimates and their 95% confident intervals were calculated, with ICC = 0.99, p < .001, indicating excellent reliability between the two raters (Koo & Li, 2016, >.90= excellent reliability). Disagreements were resolved by the raters, who agreed on the highest number given when disagreements occurred (see Appendix C).

Primary aim

The digital cognitive tests that have been used to assess PD-MCI (see Table 2).

Author/s and publication year

All papers were published between 2010 and 2023. The studies took place in eight different countries: seven in the United States of America, three in Israel, two in Germany, one in New Zealand, one in Australia, one in Japan, one in Poland, and one in China.

Study design

All studies were quantitative, observational studies. There were four different types of research designs, highlighting the methodological and clinical diversity of studies included. Four studies were cohort cross-sectional; one was cohort longitudinal, nine were case-control cross-sectional and three were case-control longitudinal in their design.

Sample size

The sample sizes ranged from 15 (Lauraitis et al., 2020) to 230 (Dion et al., 2021).

Population characteristics

All studies included those with PD. Six studies (Abdolahi et al., 2016; Hanna-Pladdy et al., 2010; Rosenblum et al., 2021; Shao et al., 2023); Virmani et al., 2022; Weizenbaum et al., 2022) were cross-sectional, only recruiting those with PD. For case-control studies, six studies solely included healthy controls (Dion et al., 2021; Lee et al., 2018; Park & Schott, 2018; Ramos et al., 2022; Schejter-Margalit et al., 2021, 2022) and the remaining five studies included both healthy controls and a mix of neurological conditions such as Huntington's Disease, Amyotrophic Lateral Sclerosis, Spinocerebellar ataxia, MCI, and Dementia (Kokubo et al., 2018; Lauraitis et al., 2020; Mishra et al., 2022; Schmitz-Peiffer et al., 2022; Saban & Ivry, 2021).

Digital psychometric tests

Six studies used a smartphone, six were delivered via computer, one was delivered by smartphone or computer, three used a digital pen and one study was delivered by a tablet.

Table 2. Summary table of included studies.

Study	Design	Sample Characteristics	Cognitive test adapted digitally	Digital Interface	Designed for PD/MCI?	QATSDD (/42).
Park and Schott (2022), Germany	Case-control, cross-sectional.	32 HC aged 45–82 years old. 30 PD aged 45–80 years old. 17 men and 13 women in both samples.	dTMT	Touchscreen, keyboard using tablet.	No	39
Ramos et al. (2022), New Zealand	Case-control, longitudinal.	Time points 1: 38 HC and 38 PD. Time point 2: 28 PD and 31 HC.18 F and 20 M- HC .16 F and 22 M- PD. Time points 1: 54–83 HC, 43–70 PD. Time point 2: 54–78 HC, 43–79 PD.17 F and 14 M-HC. 12F and 16 M- PD.	dCBTt	Computer with touchscreen and mouse interactive function.	No	35
Virmani et al. (2022), USA	Cohort, cross-sectional.	50 PD. 30 men and 20 women.	Visuospatial subtests of dMoCA	Using Doxy.me®ctional, computer, which used interactive keyboard functions.	No	30
Weizenbaum et al. (2022), USA.	Cohort, cross-sectional.	27 PD aged (14 mean and 13 women).	dTMT and dBSST	Smartphone app, using touchscreen interactive functions.	No	33
Lee et al. (2018), Australia.	Case-control, longitudinal	99 PD and 34 HC 51 males and 48 females.	dVSt	Smartphone app, using touchscreen interactive functions.	Yes	29
Mishra et al. (2022), USA.	Cross-sectional, case-control	14 HC, 14 PDs, and 11 mild cognitive impaired (MCI). 23 males and 16 females.	dTMT	Sensor attached to the lower shin of the dominant leg using an elastic band- connected to computer screen.	Sensor, yes- but not TMT.	26
Rosenblum et al. (2021), Israel.	Cohort, longitudinal.	36 PD.	DailyCog. The two tasks we designed are: preparing a hot drink and preparing a shopping list (assessing executive functions abilities, visual–spatial abilities, and psychomotor speed)	Smartphone app developed for phones with an android operating system	Yes	34
Dion et al. (2021), USA.	Case-control, cross-sectional.	115 PD and 115 HC.	dCDT	Digital interactive pen, recorded via computer video (for marking purposes).	No	28
Kokubo et al. (2018), Japan.	Case-control, cross-sectional.	28 PD, 27 MCI&D and 29 HC	dTMT, Stroop, back-task and verbal memory subtests	Smartphone, interactive function.	Yes	29
Schmitz-Peiffer et al. (2022), Germany.	Case-control, cross-sectional.	62 PD, 41 HC and 43 ALS.	dTMT	Electronic gaze-controlled version, using a non-computer.	No	24
Hanna-Pladdy et al. (2010), USA.	Cohort, cross-sectional.	50 PD patients.37 males, 13 females.	NeuroTax Battery (assesses memory, executive, visuospatial, verbal, attention, processing speed, and motor skills)	Computer, with interactive mouse and keyboard function.	Yes	26
Schejter-Margalit et al. (2021), Israel.	Case-control, cross-sectional.	75 PD and 59 HC.45 males, 30 females.	dCDT	Digital, interactive pen ballpoint pen.	No	25
Schejter-Margalit et al. (2022), Israel.	Case-control, cross-sectional.	29 PD, 31 HC.	dCDT	Digital, interactive ballpoint pen.	No	32
Saban and Ivry (2021), USA.	Case-control, cross-sectional.	12 with PD, 12 with SCA and 12 HC.	DSP anddMoCA- which removed "Alternating Trail Making".	Computer with interactive keyboard function, Gorilla Experiment Builder.	Yes. No for TMT subtest.	31
Abdolahi et al. (2016), USA.	Case-control, longitudinal.	8 PD, 9HD.	dMoCA	Computer, with interactive mouse and keyboard function.	No	25
Lauraitis et al. (2020), Poland.	Case-control, cross-sectional.	1 with PD, 3 with HD, 1 with early D, 1 with NCI, 1S) and 8 HC10 male, 5 female.	Digital SAGE cognitive test (orientation, picture naming, similarities, construction, verbal fluency, problem solving, memory.	Smartphone, interactive touchscreen through Google Pay app.	Yes- also for Huntington's disease and Alzheimer's disease	28
Shao et al. (2023), China.	Cohort, cross-sectional.	41 PD20 male, 21 female.	Digital app (assesses visuospatial, memory, executive and processing speed abilities).	Smartphone, interactive touchscreen function, downloaded by Android App Market and Apple App Market.	Yes for combination of tests into the one battery. No for individual subtests.	31

Note. dTMT: digital Trail Making Test; dCBT: digital Corsi Block-Tapping test; dBSST: digital Backwards Spatial Span Task; dVSt: digital Victoria Stroop test; DailyCog: Daily Cognition test; dCDT: digital Clock Drawing Test; digital NEUROTAX: NeuroTrax Mindstreams; DSP: Discrete Sentence Production; digital Sage: self-administered cognitive testing methodology; Digital app: digital application.

Some of the digital tests were compared with paper-based test versions with a videophone component. Nine studies only used one subtest, with four using the digital Trail-Making Test (dTMT), one using a digital version of the Corsi-block Tapping test (dCBTt), one using a novel version of the digital Victoria Stroop test (dVSt), and three studies using a digital version of the Clock Drawing Test (CDT). Ten studies used two or more subtests, with one study using a dTMT and digital Backwards Spatial Span test (dBSSt), while three studies used the dMoCA. The remaining novel tests were based on commonly used subtests that assessed processing speed, attention, memory, executive, and visuospatial functioning.

Secondary aims

Validity and Reliability of Digital Cognitive Tests.

Validity

Nine of the 17 studies explored the validity of the digital cognitive tests (see Table 3). Regarding the dTMT, Kokubo et al. (2018) revealed that UX-TMT had good convergent validity with the pencil-paper Mini Mental State Examination (ppMMSE). Furthermore, the dTMT part-B score in the UX-TMT showed moderate correlations with the TMT part-B score in the ppMoCA (pencil paper MoCA). Schmitz-Peiffer et al. (2022) found good convergent validity, with significant moderate correlations between ppTMT part-A and dTMT part-A as well as between ppTMT part-B and dTMT for PD limited motor impairment group. Divergence was found, with no significant correlations between the pencil-paper and digital test versions for the PD-significant motor impairment group. Mishra et al. (2022) found good agreement between the dTMT and ppMoCA. Weizenbaum et al. (2022) also found divergence, with no significant relationships found between dTMT-B with ppMoCA, ppWMS-III, and ppTMT A & B Time. Finally, Park and Schott (2022) found moderate relationships between the completion times (correlation coefficient ranged between 0.82 and 0.90) and overall test scores (correlation coefficient ranged between 0.51 and 0.67) between pencil and digital methods of the TMT. Additionally, Bland-Altman plots

Table 3. Validity and reliability reported for included studies.

Study ID (Authors and Year)	Psychometric test	Validity	Reliability
Lee et al. (2018)	Digital Victoria Stoop subtest	Convergent validity found between the dVSt and ppVSt) (r = .48–.64, p < .0001). Small correlation between the ppMOCA score and CIT total time, time, and correct/s (r = .23–.32, $p \leq$ .02). ppMOCA-attention and executive scores had a moderate correlation with CIT total time (r = .23–47, $p \leq$ .03). Bland-Altman plots (94–98%) were within confidence interval.	Internal consistency- Intraclass correlation coefficient was moderate to strong (.46–.808, $p \leq$ .0002).
Weizenbaum et al. (2022)	Digital Backwards spatial task.	Convergent- between the dBSSt with ppMoCA (p < .0001) and ppTMT part B Time (p <.001). NO agreement with the ppWMS-III (p <.0001) and ppBRIEF-A (p= .712).	Internal consistency (.86) for within persons, and (.89) for between-persons.
Ramos et al. (2022)		None reported.	Test retest- 1 year after testing- medium-sized effect in impairment, comparing to the small effect sizes reported at baseline.
Viramni et al. (2022)	Digital MoCA	Convergent- Good agreement with ppMoCA (Blatman-Plot within 95% confidence interval).	Internal consistency- ICCs between ppMoCA and dMOCA (.825).
Saban & Ivry, 2021.		Not reported	Inter-rater- PD patients showed less improvement than HC in non-repetition condition (p = .017).
Abdolahi et al. (2016).		Convergent validity found for HD and PD samples (r= .59) between digital and ppMoCA. VWeak-moderate predictive value (r = .36).	Poor internal consistency (α=.54) and reliability (*ICC*=.37).
Park and Schott (2022)	Digital Trail-Making Test	Convergent- between time and test versions (r = .82 − .90, p <.001) Moderate relationships between digital and pencil paper scores (ranging from .51 to .67, p <.001). Bland-Altman plots (ranging from 92% to 98% were within the limits of agreement).	Inter-rater- the dTMT had excellent reliability, (ICC values between 0.90 and 0.95.
Weizenbaum et al. (2022)		Moderate predictive validity with ppMoCA (p =.637), but weak predicative validity with ppWMSI-III (p = .066), ppTMT Time (p = .186) and ppBRIEF-A (p = .148).	Parallel forms- Within-person means were significantly negatively correlated with *SD* (r= −.93, p< .001), Between-person reliability was .87. Between-person reliability was 0.87.
Mishra et al. (2022)		Convergent- dTMT time was significantly moderately correlated with ppMoCA scores (p = −.56, p < .01).	Not reported.
Kokubo et al. (2018)		Convergent- strong, significant correlation with the ppMMSE-J scores (r=.77, p<.001). There was a moderate, significant correlation with ppMoCA-J (r=.56, p<.001).	Internal consistency- Cronbach's alpha for the UX-TMT ranged from .71 to .83 in all three groups (PD, MCI-D and HC).
Schmitz-Peiffer et al. (2022)		External validity- For PD group with no motor impairments, there was a moderate, significant correlation with ppTMT-A (r = 0.55, p = 0.018) and ppTMT-B (r = .57, p = .010). Convergent- For the PD group with significant motor impairments, there were no significant correlations with ppTMT-A: p = .141) or with the ppTMT-B (p = .120).	Parallel forms- Pairwise comparisons were not significant between groups for dTMT-A. For patients with motor impairments, the made more errors in the dTMT, compared to HC (24.5; p = .048).
Hanna-Pladdy et al. (2010)	NeuroTax-Novel Screen	Not reported.	Internal consistency- with MMSE scores (Kκ = .291, p= .031).

Note. dTMT: digital Trail Making Test; dCBT: digital Corsi Block-Tapping test; dBSST: digital Backwards Spatial Span Task; dVSt: digital Victoria Stroop test; DailyCog: Daily Cognition test; dCDT: digital Clock Drawing Test; digital NEUROTAX: NeuroTrax Mindstreams; DSP: Discrete Sentence Production; digital Sage: self-administered cognitive testing methodology; Digital app: digital application.

between test variants (96%, 92%, and 94% of data points) were within the limits of agreement, demonstrating good convergence validity (Bland & Altman, 1997) between digital and paper versions.

For the dVSt, Lee et al. (2018) found moderate convergence validity between a pencil-paper version (r = .48–.64, as recommended by Dancy & Reidy, 2007). For the dSBBt, Weizenbaum et al. (2022) found good convergence, with a significant relationship found between the ppMoCA (p <.001) and the ppWMS-III (p <.001). Focusing on the dMoCA, with pencil-paper versions, Abdolahi et al. (2016) found moderate convergence validity (r= .59, Dancey & Reidy, 2007) for both HC and PD samples. For PD patients only, the total dMoCA score was suggested to have a moderate predictive value (r= .37). Conversely, Viramni et al. (2022) found convergence validity, with good agreement found between the digital and pencil-paper versions (within a 95% confidence interval, Bland & Altman, 1997).

Reliability

Eleven studies explored reliability (see Table 3), with three studies examining reliability for the dBSSt. Weizenbaum et al. (2022) for dBSSt and dTMT found that within-person means were significantly negatively correlated with standard deviation, meaning that high accuracy was associated with less intraindividual variability around the mean. Between-person reliability was good (indicated by Nunnally & Bernstein, 1994 threshold) for both tests, with statistical significance not found between study day/time with accuracy rates, nor with residual variance between persons. This indicated the absence of a practice effect. For the dTMT, Ramos et al. (2022) at a one-year follow-up found no significant differences between PD and HC (0.43, 95% *CI* [0.94, 0.09]). For PD patients, a medium-sized dysfunction effect was found (d ¼ 0.65, 95% *CI* [1.18, 0.13) that contrasts with the small-sized effect at baseline. Kokubo et al. (2018) used Cronbachs alpha for the dTMT cognitive test in all groups (PD, MCI-D, and HC) indicating good reliability (indicated by Sekaran & Bougie, 2010 reliability threshold). Abdolahi et al. (2016) found that for PD patients, the total dMoCA was suggested to have poor internal consistency. Finally, Park and Schott found the dTMT had excellent inter-rater reliability (indicated by Bland & Altman, 1997 reliability threshold). Lee et al. (2018) established good reliability for the dVSt (within a 95% confidence interval, Bland & Altman, 1997), and the intraclass correlation coefficient was overall moderate to strong, as indicated by Koo and Li (2016) reliability threshold. Two other studies (Schmitz-Peiffer et al., 2022; Saban & Ivry, 2021) also examined inter-rater reliability. Schmitz-Peiffer et al. (2022) found increased error rates for PD patients (with motor impairment) when compared to HC for the dTMT-B. Saban and Ivry (2021) further found good repeatability, with PD patients demonstrating less improvement than controls in non-repeated groups for the dMoCA cognitive test (p= .017). Focusing on novel tests, Hanna-Pladdy et al. (2010) digital NeuroTax test found fair agreement with the ppMMSE test and only slight agreement with the ppMoCA test (as recommended by Landis & Koch, 1977).

To further examine the reliability, this review identified that seven studies examined the test sensitivity.

1. Schejter-Margalit et al. (2021) Receiver Operating Curve (ROC) analysis demonstrated that the dCDT score had a higher Area Under Curve (AUC) in discriminating controls from PD patients (0.807, 95%*CI*, 0.726–0.880) than the ppMoCA (0.590, 95%*CI*, 0.496–0.685) or pencil paper Color Trials Test (ppCTT) (0.636, 95%*CI*, 0.535–0.738 and 0.717, 95%*CI*, 0.615–0.804 for the CTT-1 and CTT-2).
2. Hanna-Pladdy et al. (2010) NeuroTax categorized 40% with PD-MCI within average-borderline (while the ppMMSE did not categorize anyone as borderline and the ppMoCA only categorized 8% as borderline).
3. Kokubo et al. (2018) identified a cutoff score of 21 on the UX-TMT indicated PD-MCI, with high sensitivity (.82) found. Age improved test sensitivity (.97).
4. Schmitz-Peiffer et al. (2022) found errors made in the dTMT were a sensitive marker for PD-MCI (p = .048), indicative of executive functioning impairment.
5. Shao et al. (2023) found the following cutoff scores for a visuospatial disorder: 26 (Digit Span Test, DST), 17 (Visual Organization Test, VOT), and 19 (Facial Recognition Test, FRT). The AUCs were significantly different from 0.5 (DST, p < .001, VOT, p= .0002, and FRT, p= .0002).
6. Lee et al. (2018) demonstrated sensitivity; those with PD motor impairments and dVSTt time (r = .49, p < .0001), W time (r = .43, p < .0001), and C time (r = .35, p = .0003). A score greater than 19% (particularly for older adults and those that were categorized as having disease severity) warranted a further assessment.
7. Abdolahi et al. (2016) findings identified cognitive impairment for the PD sample between the in-person, and remote assessment, with a kappa statistic of 0.50 (95% *CI*: –.02, 1.00) and 75% agreement established.

Discussion

To the best of our knowledge, this systematic review is the first to consider what digital psychometric tests assess PD-MCI. Using a Level 1 assessment, the TMT was the most used subtest, while the dMoCA was the most used battery to determine a Level 2 assessment. The secondary aims examined the validity, reliability, and sensitivity of these measures.

One-cognitive domain subtests

The TMT assesses executive functioning skills, which play a key role in PD cognitive and motor functioning (Zhou et al., 2017). This review found that the dTMT was the most used subtest. Four studies found moderate-strong validity with comparator tests and between digital and pencil-paper versions, strengthening the validity of the dTMT (Brown et al., 2016; Schmand, 2019). Nonetheless, three other authors (Park

& Schott, 2022; Schmitz-Peiffer et al., 2022; Weizenbaum et al., 2022) did not find convergence between dTMT with ppMoCA, ppWMSI-III, and ppTMT time scores. However, ceiling effects may have contributed to these findings. Unfortunately the authors did not address why they selected specific psychometric tests as comparators, which would have strengthened the justification for these analyses. Schmitz-Peiffer et al. (2022) was the only study that distinguished between significant and non-significant motor impairment on dTMT performance. While the authors did not explain why the convergence indicators between paper-and-pencil and digital versions of assessments were low, it may relate to the difference in modalities of delivery (such as environmental test conditions, cognitive demand, and/or testing instructions). Additionally, individual differences among test-takers (such as the stage and severity of PD) may have influenced engagement and scores. However, there is a need for future studies to examine this further, as Schmitz-Peiffer et al's. (2022) study was vastly different from the other TMT methods, utilizing eye-tracking computer equipment rather than an interactive computer that required motor skills. The dTMT's internal consistency (Kokubo et al., 2018) and inter-rater reliability (Schmitz-Peiffer et al., 2022; Weizenbaum et al., 2022) were examined, supporting using the dTMT (DeVon et al., 2007; Gwet, 2014). However, these studies' methodological designs meant that other aspects of reliability (test re-test and parallel forms) were not examined, being important considerations for future research (Litvan et al., 2012). To further assess psychometric properties, two studies examined test sensitivity. However, only Kokubo et al. (2018) provided an optimal score, accounting for age norms, confirming previous research on how age affects PD cognition (Kim et al., 2009).

Only Lee et al. (2018) utilized the dVSt, which also measured executive functioning skills (Bayard et al., 2011). Convergence validity was not demonstrated with pencil-paper dBSSt and MoCA test versions which may be due to the MoCA assessing multiple cognitive domains. Reliability however was found (inter-rater and test re-test reliability), being like other PD-specific cognitive screens (Parkinson Disease-Cognitive Rating Scale, PDCS, Kulisevsky & Pagonabarraga, 2009), with a cutoff score calculated to indicate PD-MCI.

Only Weizenbaum et al. (2022) utilized the dBSSt (measures working memory abilities, Melin et al., 2023), finding validity with pencil-paper WMS-III Spatial Span and BRIEF-A screens. These findings can be explained by the strong relationship between executive functioning and working memory abilities (Baddeley & Hitch, 1974), which both the dBSST and comparator tests' assessed. Also supporting the dBSST's use, reliability (internal consistency and inter-rater) was found. However, future research should aim to develop a cutoff score that indicates PD-MCI.

Only Ramos et al. (2022) examined the reliability of the dCBTt, supporting its use with good test re-test reliability. However, there were subtle differences between PD patient's performances on subtests of the dCBT, with deficits found in the backward recall subtest of the dCBT. Nonetheless, the one-year follow-up may not have been long enough to detect subtle changes associated with PD cognition (Flannery et al., 2018).

The majority of studies that utilized the dCDT (measures executive functioning and visuospatial abilities, Pinto & Peters, 2009) did not examine the tests' validity or reliability. Unfortunately, However, Dion et al. (2021) found that the dCDT distinguished between PD and non-PD group scores, being parallel with prior research, demonstrating cognitive and motor slowing in PD (Muslimović et al., 2005). Schejter-Margalit et al. (2021) were able to take their findings further by examining PD subtypes of GBA-PD and LRRK2-PD when compared to HC's, with the dCDT being identifying subtle motor-cognitive impairments.

Screens assessing more than two cognitive domains

MoCA

Abdolahi et al. (2016) were the only authors who studied the examined validity of the dMoCA, by comparing scores for both HC and PD samples with pencil-paper versions. All studies however examined the reliability of the dMoCA. The difference between studies' internal consistency scores may be due to methodological differences in their study design. For example, Abdolahi et al. (2016) employed a longer follow-up study (leading to the potential for variation between face-to-face versus digital testing) and recruited a small sample. Only one study (Sban & Ivry, 2021) examined test-retest reliability demonstrating PD patients significantly revealed less improvement than controls in the non-repetition condition, being indicative of deficits in algorithm-based learning. However, studies did not identify a PD-MCI cutoff score, justifying future research.

Digital screens with novel designs

Four studies developed novel digital screens (assessing <2 cognitive domains). Only one study (Hanna-Pladdy et al., 2010) examined it's validity with the pencil-paper MMSE and MoCA tests. None of the studies examined the reliability and this again, warrants future research. One study (Shao et al., 2023) created a cutoff score for PD-MCI, which is helpful in understanding its potential in clinical practice.

Clinical implications

This review demonstrated mixed results as to how digitalized tests (that measure PD-MCI) compare with paper-pencil versions. This may be explained by the fact that the majority of studies did not define what standardized protocols were used for the administration and scoring. Saban and Ivry (2021) were the only authors who provided administration instructions, which minimizes experimenter bias and enables accurate measurement (Ding et al., 2022). However, the protocols were not specific to PD, with there a need for this due to the context of PD populations age and associated symptoms (such as experiencing vision, hearing and sensory-motor issues) (Goldman et al., 2018). Therefore, the following PD-specific digitalized cognitive testing procedures can be utilized in clinical practice.

- PD patients should complete digital assessments independently to monitor task completion- to assess potential executive functioning difficulties.
- Verify PD patients understand instructions or ask for question repetition, as noncompliance may signal MCI (e.g. inhibition or encoding issues).
- Check computer microphones to guarantee clear audibility and independent response, noting any deviations, potentially indicating PD-MCI.
- Recognize the variability in PD patients' performance due to "good and bad days", while considering other confounders (psychological state, fatigue, severity, discomfort, medication adherence and timing).
- Adjust for visual impairments, and use tools like the dCDT for motor-affected patients to assess cognitive performance.
- Requirements for administering in the home environment include: ensure participants have access to stable WIFI, are comfortable using digital methods, and complete testing under optimal conditions (distractions or using necessary hearing/visual aids).

This review identified promising findings for those with PD, including those with significant PD motor impairments and of a range of ages. This was important as research has found that these factors can influence cognitive performance (Wang et al., 2014; Wojtala et al., 2019). However, studies did not examine the impact medication, years and stage of disease had on cognitive performance between digital and pencil-paper versions.

The IPMDS recommends that cognitive deficits are evidenced on at least two cognitive tests (either in a specific cognitive domain or in two separate cognitive domains) (Goldman et al., 2018). If employing two cognitive tests to adhere to the IPMDS recommendation, the dTMT should be considered alongside an additional digitalized one-domain cognitive test. In terms of acceptability, this review highlights that the majority of studies used the dTMT by utilizing touchscreen and keyboard functions (using tablet, computer, or smartphone devices) as alternative methods to pencil-paper methods. Alternatively, leg-sensors may be an alternative option, particularly for those with significant motor impairments (Mishra et al., 2022; Schmitz-Peiffer et al., 2022).

Research implications

Firstly, future research should aim to utilize appropriate consideration to procedures that relate to participant recruitment and randomization of test order. Secondly, future research should be encouraged to explore the statistical significance of how demographic variables (motor impairment, education level, age and disease progression) influence cognitive performance for both PD and non-PD samples (with the same demographics recruited, allowing for comparisons to be made). Thirdly, the studies that recruited a small sample may be argued to have limited generalizability of findings. Focusing on the cognitive measures themselves, it would be helpful for research to evaluate the validity and reliability of digital versions of commonly used screens (such as MMSE), since this review only identified the MoCA being a common screen that has been adapted digitally to assess PD-MCI. In terms of the novel screens identified, future studies should address the methodological criticisms described, with more PD participants and controls to further examine their psychometric properties when administered digitally. Future studies should examine digital tests' effect sizes when comparing PD and control groups test performance to aid our understanding as to what difference (if any) there is between mode of test delivery (Funder & Ozer, 2019). Finally, assessing potential practice effects are of priority to examine, given that cognitive testing is commonly used to monitor disease progression in PD (Aarsland et al., 2017).

Strengths and limitations of the review

This review strengthens the evidence base, as to the best of our knowledge no studies have explored what cognitive tests have been adapted digitally to assess PD-MCI, with their psychometric properties examined. It was promising that traditional one and two domain cognitive tests employed the same metrics for digital and pencil-paper versions. Using established metrics ensures that the digital versions of tests maintain the same levels of validity and reliability as traditional methods. Additionally, maintaining the same metrics allows for the direct comparison of scores, facilitating consistency and accuracy during assessment procedures. However, the heterogeneity in the measures and designs, alongside utilizing non-representative recruitment and pseudo-randomization of PD patients are obvious limitations to the studies included. Therefore, this review could not discount baseline discrepancies and confounding variables. Other limitations included that studies used different statistical methods and sample sizes. However, this review attempted to address this by focusing on studies that assessed aspects of validity and reliability, alongside test sensitivity. A notable methodological limitation is the exclusion of the Scopus database from the search strategy. Despite using five databases to search for literature, incorporating Scopus may have enhanced the robustness of this review's findings. Finally, risk of bias was not assessed due to the studies' methodological designs. For example, the ROBINS-I tool was not used as studies did not measure two or more interventions (Sterne et al., 2016). The Newcastle-Ottawa (Wells et al., 2001) and QUADAS-2 (Whiting et al., 2011) scales could not be implemented due to the review's PIO method.

Conclusion

This review found that digitalized cognitive tests are valid and reliable methods to assess PD-MCI. Digitalized testing allows for the opportunity to collect more data to improve the accuracy of the classification of PD-MCI and monitor PD disease progression (Chan et al., 2021). By examining what digitalized tests have been used to assess PD-MCI, it is hoped that this can help researchers and clinicians in the future to develop standardized guidelines, as recommended by Litvan et al. (2012).

Acknowledgments

The research team acknowledges the work of Aoife Cuddihy for her role in the screening process. As such, she has been included as an author.

Author contribution

SC, MD, and DC developed the review concept. SC completed the systematic search, which was reviewed by AC. Extracted data were checked by MD and DC. The quality appraisal was completed by SC and DC. The synthesis of data was completed by SC under the supervision of MD. SC drafted the review, with MD, DC, and NL providing critical revisions. All authors have approved the version of the review for submission. SC acts as the guarantor for the article.

Ethical statement

The article does not contain any studies with human or animal subjects performed by any of the authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article.

References

Aarsland, D., Creese, B., Politis, M., Chaudhuri, K. R., Ffytche, D. H., Weintraub, D., & Ballard, C. (2017). Cognitive decline in Parkinson disease. *Nature Reviews. Neurology*, *13*(4), 217–231. https://doi.org/10.1038/s41572-021-00280-3

Abdolahi, A., Bull, M. T., Darwin, K. C., Venkataraman, V., Grana, M. J., Dorsey, E. R., & Biglan, K. M. (2016). A feasibility study of conducting the Montreal Cognitive Assessment remotely in individuals with movement disorders. *Health Informatics Journal*, *22*(2), 304–311. https://doi.org/10.1177/1460458214556373

Artusi, C. A., Imbalzano, G., Sturchio, A., Pilotto, A., Montanaro, E., Padovani, A., Lopiano, L., Maetzler, W., & Espay, A. J. (2020). Implementation of mobile health technologies in clinical trials of movement disorders: Underutilized potential. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, *17*(4), 1736–1746. https://doi.org/10.1007/s13311-020-00901-x

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of Learning and Motivation*. (pp. 47–89). Academic Press. https://doi.org/10.1016/s0079-7421(08)60452-1

Bayard, S., Erkes, J., & Moroni, C. (2011). Victoria Stroop Test: Normative data in a sample group of older people and the study of their clinical applications in the assessment of inhibition in Alzheimer's disease. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *26*(7), 653–661. https://doi.org/10.1093/arclin/acr053

Ben-Pazi, H., Browne, P., Chan, P., Cubo, E., Guttman, M., Hassan, A., Hatcher-Martin, J., Mari, Z., Moukheiber, E., Okubadejo, N. U., & Shalash, A. (2018). The promise of telemedicine for movement disorders: an interdisciplinary approach. *Current Neurology and Neuroscience Reports*, *18*(5), 26. https://doi.org/10.1007/s11910-018-0834-6

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ (Clinical Research ed.)*, *314*(7080), 572–572. https://doi.org/10.1136/bmj.314.7080.572

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

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

Brown, D. S., Bernstein, I. H., McClintock, S. M., Munro Cullum, C., Dewey, R. B., Jr, Husain, M., & Lacritz, L. H. (2016). Use of the Montreal Cognitive Assessment and Alzheimer's Disease-8 as cognitive screening measures in Parkinson's disease. *International Journal of Geriatric Psychiatry*, *31*(3), 264–272. https://doi.org/10.1002/gps.4320

Calleo, J., Burrows, C., Levin, H., Marsh, L., Lai, E., & York, M. K. (2012). Cognitive rehabilitation for executive dysfunction in Parkinson's disease: Application and current directions. *Parkinson's Disease*, *2012*, 512892. https://doi.org/10.1155/2012/512892

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ (Clinical Research ed.)*, *368*(1), l6890. https://doi.org/10.1136/bmj.16890

Chan, J. Y., Yau, S. T., Kwok, T. C., & Tsoi, K. K. (2021). Diagnostic performance of digital cognitive tests for the identification of MCI and dementia: A systematic review. *Ageing Research Reviews*, *72*, 101506. https://doi.org/10.1007/978-3-319-44775-9_7

Connolly, B., & Fox, S. H. (2014). Treatment of cognitive, psychiatric, and affective disorders associated with Parkinson's disease. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, *11*(1), 78–91. https://doi.org/10.1007/s13311-013-0238-x

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, *49*(6), 560–575. https://doi.org/10.1111/medu.12678

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Pearson Education.

De Frias, C. M., Dixon, R. A., Fisher, N., & Camicioli, R. (2007). Intraindividual variability in neurocognitive speed: A comparison of Parkinson's disease and normal older adults. *Neuropsychologia*, *45*(11), 2499–2507. https://doi.org/10.1016/j.neuropsychologia.2007.03.022

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing*, *39*(2), 155–164. https://doi.org/10.1111/j.1547-5069.2007.00161.x

Dion, C., Frank, B. E., Crowley, S. J., Hizel, L. P., Rodriguez, K., Tanner, J. J., Libon, D. J., & Price, C. C. (2021). Parkinson's disease cognitive phenotypes show unique clock drawing features when measured with digital technology. *Journal of Parkinson's Disease*, *11*(2), 779–791. https://doi.org/10.3233/JPD-202399

Ellis, T. D., & Earhart, G. M. (2021). Digital therapeutics in Parkinson's Disease: Practical applications and future potential. *Journal of Parkinson's Disease*, *11*(s1), S95–S101. https://doi.org/10.3233/JPD-202407

Flannery, S. L., Jowett, T., Garvey, A., Cutfield, N. J., & Machado, L. (2018). Computerized testing in Parkinson's disease: Performance deficits in relation to standard clinical measures. *Journal of Clinical and Experimental Neuropsychology*, *40*(10), 1062–1073. https://doi.org/10.1080/13803395.2018.1485880

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions* (3rd ed.). John Wiley & Sons.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515249209792 82

Goldman, J. G., Holden, S. K., Litvan, I., McKeith, I., Stebbins, G. T., & Taylor, J. P. (2018). Evolution of diagnostic criteria and assessments for Parkinson's disease mild cognitive impairment. *Movement Disorders: Official Journal of the Movement Disorder Society*, *33*(4), 503–510. https://doi.org/10.1002/mds.27323

Grosch, M. C., Gottlieb, M. C., & Cullum, C. M. (2011). Initial practice recommendations for teleneuropsychology. *The Clinical Neuropsychologist*, *25*(7), 1119–1133. https://doi.org/10.1080/13854046.2011.60940

Gwet, K. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Hanna-Pladdy, B., Enslein, A., Fray, M., Gajewski, B. J., Pahwa, R., & Lyons, K. E. (2010). Utility of the NeuroTrax computerized battery for cognitive screening in Parkinson's disease: Comparison with the MMSE and the MoCA. *The International Journal of Neuroscience*, *120*(8), 538–543. https://doi.org/10.3109/00207454.2010.496539

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. The Cochrane Collaboration and John Wiley & Sons.

Hirsch, L., Jette, N., Frolkis, A., Steeves, T., & Pringsheim, T. (2016). The incidence of Parkinson's disease: A systematic review and meta-analysis. *Neuroepidemiology*, *46*(4), 292–300. https://doi.org/10.1159/000445751

Hornykiewicz, O., & Kish, S. J. (1987). Biochemical pathophysiology of Parkinson's disease. *Advances in Neurology*, *45*, 19–34.

Jankovic, J., & Tan, E. K. (2020). Parkinson's disease: Etiopathogenesis and treatment. *Journal of Neurology, Neurosurgery, and Psychiatry*, *91*(8), 795–808. https://doi.org/10.1136/jnnp-2019-322338

Kim, J. W., Cheon, S. M., Park, M. J., Kim, S. Y., & Jo, H. Y. (2009). Cognitive impairment in Parkinson's disease without dementia: Subtypes and influences of age. *Journal of Clinical Neurology (Seoul, Korea)*, *5*(3), 133–138. https://doi.org/10.3988/jcn.2009.5.5.133

Klucken, J., Krüger, R., Schmidt, P., & Bloem, B. R. (2018). Management of Parkinson's disease 20 years from now: Towards digital health pathways. *Journal of Parkinson's Disease*, *8*(s1), S85–S94. https://doi.org/10.3233/JPD-181519

Kokubo, N., Yokoi, Y., Saitoh, Y., Murata, M., Maruo, K., Takebayashi, Y., Shinmei, I., Yoshimoto, S., & Horikoshi, M. (2018). A new device-aided cognitive function test, User eXperience-Trail Making Test (UX-TMT), sensitively detects neuropsychological performance in patients with dementia and Parkinson's disease. *BMC Psychiatry*, *18*(1), 220. https://doi.org/10.1186/s12888-018-1795-7

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kulisevsky, J., & Pagonabarraga, J. (2009). Cognitive impairment in Parkinson's disease: Tools for diagnosis and assessment. *Movement Disorders: Official Journal of the Movement Disorder Society*, *24*(8), 1103–1110. https://doi.org/10.1002/mds.22506

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://www.jstor.org/stable/2529310. https://doi.org/10.2307/2529310

Lauraitis, A., Maskeliūnas, R., Damaševičius, R., & Krilavičius, T. (2020). A mobile application for smart computer-aided self-administered testing of cognition, speech, and motor impairment. *Sensors (Basel, Switzerland)*, *20*(11), 3236. https://doi.org/10.3390/s20113236

Lee, W., Williams, D. R., & Evans, A. (2018). Feasibility of smartphone-based testing of interference in Parkinson's Disease. *Neuro-Degenerative Diseases*, *18*(2–3), 133–142. https://doi.org/10.1159/000488593

Libon, D. J., Baliga, G., Swenson, R., & Au, R. (2021). Digital Neuropsychological Assessment: New technology for measuring subtle neuropsychological behavior. *Journal of Alzheimer's Disease: JAD*, *82*(1), 1–4. https://doi.org/10.3233/JAD-210513

Litvan, I., Goldman, J. G., Tröster, A. I., Schmand, B. A., Weintraub, D., Petersen, R. C., Mollenhauer, B., Adler, C. H., Marder, K., Williams-Gray, C. H., Aarsland, D., Kulisevsky, J., Rodriguez-Oroz, M. C., Burn, D. J., Barker, R. A., & Emre, M. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Movement Disorders: Official Journal of the Movement Disorder Society*, *27*(3), 349–356. https://doi.org/10.1002/mds.24893

Macleod, A. D., & Counsell, C. E. (2016). Predictors of functional dependency in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *31*(10), 1482–1488. https://doi.org/10.1002/mds.26751

Melin, J., Göschel, L., Hagell, P., Westergren, A., Flöel, A., & Pendrill, L. (2023). Forward and backward recalling sequences in spatial and verbal memory tasks: What do we measure? *Entropy (Basel, Switzerland)*, *25*(5), 813. https://doi.org/10.3390/e25050813

Mishra, R. K., Park, C., Zhou, H., Najafi, B., & Thrasher, T. A. (2022). Evaluation of motor and cognitive performance in people with Parkinson's disease using instrumented trail-making test. *Gerontology*, *68*(2), 234–240. https://doi.org/10.1159/000515940

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ (Clinical Research ed.)*, *339*(4), b2535. https://doi.org/10.1136/bmj.b2535

Moyer, V. (2008). Weighing the evidence: PICO questions: What are they, and why bother? *AAP Grand Rounds*, *19*(1), 2–2. https://doi.org/10.1542/gr.19-1-2

Muslimovic, D., Post, B., Speelman, J. D., & Schmand, B. (2005). Cognitive profile of patients with newly diagnosed Parkinson disease. *Neurology*, *65*(8), 1239–1245. https://doi.org/10.1212/01.wnl.0000180516.69442.95

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.532-5415.2005.53221.x

Nunnally, J., & Bernstein, H. I. (1994). *Psychometric theory*. McGraw-Hill Companies.

Ou, Z., Pan, J., Tang, S., Duan, D., Yu, D., Nong, H., & Wang, Z. (2021). Global trends in the incidence, prevalence, and years lived with disability of Parkinson's disease in 204 countries/territories from 1990 to 2019. *Frontiers in Public Health*, *9*, 776847. https://doi.org/10.3389/fpubh.2021.776847

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Paez, A. (2017). Grey literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, *10*(3), 233–240. https://doi.org/10.1111/jebm.12266

Papa, S. M., Brundin, P., Fung, V. S., Kang, U. J., Burn, D. J., Colosimo, C., Chiang, H.-L., Alcalay, R. N., & Trenkwalder, C. (2020). Impact of the COVID-19 pandemic on Parkinson's disease and movement disorders. *Movement Disorders Clinical Practice*, *7*(4), 357–360. https://doi.org/10.1002/mdc3.12953

Park, S. Y., & Schott, N. (2022). Which motor-cognitive abilities underlie the digital Trail-Making Test? Decomposing various test scores to detect cognitive impairment in Parkinson's disease—Pilot study. *Applied Neuropsychology. Adult*, 1–15. https://doi.org/10.1080/23279095.2022.2147837

Pechstein, A. E., Gollie, J. M., & Guccione, A. A. (2020). Fatigability and cardiorespiratory impairments in Parkinson's disease: Potential non-motor barriers to activity performance. *Journal of Functional Morphology and Kinesiology*, *5*(4), 78. https://doi.org/10.3390/jfmk5040078

Pérez, J., Díaz, J., Garcia-Martin, J., & Tabuenca, B. (2020). Systematic literature reviews in software engineering—enhancement of the study selection process using Cohen's kappa statistic. *Journal of Systems and Software*, *168*, 110657. https://doi.org/10.1016/j.jss.2020.110657

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. https://doi.org/10.1159/000203344

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A Product from the ESRC Methods Programme Version*, *1*(1), b92. https://doi.org/10.13140/2.1.1018.4643

Ramos, A. A., Garvey, A., Cutfield, N. J., & Machado, L. (2022). Forward and backward spatial recall in Parkinson's disease and matched controls: A 1-year follow-up study. *Applied Neuropsychology. Adult*, *31*(4), 647–656. https://doi.org/10.1080/23279095.2022.2059372

Rawat, C. S., & Pandey, S. (2022). Parkinson's disease–an introduction. In Techniques *for assessment of Parkinsonism for diagnosis and rehabilitation* (pp. 1–24). https://doi.org/10.1007/978-981-16-3056-9_1

Rocca, W. A. (2018). The burden of Parkinson's disease: A worldwide perspective. *The Lancet. Neurology*, *17*(11), 928–929. https://doi.org/10.1016/S1474-4422(18)30355-7

Rosenblum, S., Richardson, A., Meyer, S., Nevo, T., Sinai, M., & Hassin-Baer, S. (2021). DailyCog: A real-world functional cognitive mobile application for evaluating Mild Cognitive Impairment (MCI) in Parkinson's Disease. *Sensors (Basel, Switzerland)*, *21*(5), 1788. https://doi.org/10.3390/s21051788

Rukavina, K., Batzu, L., Boogers, A., Abundes-Corona, A., Bruno, V., & Chaudhuri, K. R. (2021). Non-motor complications in late stage Parkinson's disease: Recognition, management and unmet needs. *Expert Review of Neurotherapeutics*, *21*(3), 335–352. https://doi.org/10.1080/14737175.2021.1883428

Saban, W., & Ivry, R. B. (2021). PONT: A protocol for online neuropsychological testing. *Journal of Cognitive Neuroscience*, *33*(11), 2413–2425. https://doi.org/10.1162/jocn_a_01767

Schejter-Margalit, T., Kizony, R., Ben-Binyamin, N., Hacham, R., Thaler, A., Maidan, I., & Mirelman, A. (2022). Neural activation in the prefrontal cortex during the digital clock drawing test measured with functional near-infrared spectroscopy in early-stage Parkinson's disease. *Parkinsonism & Related Disorders*, *105*, 9–14. https://doi.org/10.1016/j.parkreldis.2022.10.021

Schejter-Margalit, T., Kizony, R., Shirvan, J., Cedarbaum, J. M., Bregman, N., Thaler, A., Giladi, N., & Mirelman, A. (2021). Quantitative digital clock drawing test as a sensitive tool to detect subtle cognitive impairments in early-stage Parkinson's disease. *Parkinsonism & Related Disorders*, *90*, 84–89. https://doi.org/10.1016/j.parkreldis.2021.08.002

Schmand, B. (2019). Why are neuropsychologists so reluctant to embrace modern assessment techniques? *The Clinical Neuropsychologist*, *33*(2), 209–219. https://doi.org/10.1080/13854046.2018.1523468

Schmitz-Peiffer, H., Aust, E., Linse, K., Rueger, W., Joos, M., Löhle, M., Storch, A., & Hermann, A. (2022). Motor-independent cognitive testing in motor degenerative diseases. *Journal of Clinical Medicine*, *11*(3), 814. https://doi.org/10.3390/jcm11030814

Sekaran, U., & Bougie, R. (2010). *Research methods for business: A skill building approach*. John Wiley & Sons Ltd.

Shao, X., Wang, K., Zhang, Y., Zhen, X., Dong, F., Tian, H., & Yu, Y. (2023). Outcome of visuospatial dysfunction assessment in patients with Parkinson's disease using mobile application software. *Frontiers in Aging Neuroscience*, *15*, 1108166. https://doi.org/10.3389/frnagi.2023.1108166

Sirriyeh, R., Lawton, R., Gardner, P., & Armitage, G. (2012). Reviewing studies with diverse designs: The development and evaluation of a new tool. *Journal of Evaluation in Clinical Practice*, *18*(4), 746–752. https://doi.org/10.1111/j.1365-2753.2011.01662.x

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Clinical Research ed.)*, *355*, i4919. https://doi.org/10.1136/bmj.i4919

Virmani, T., Lotia, M., Glover, A., Pillai, L., Kemp, A. S., Iyer, A., Farmer, P., Syed, S., Larson-Prior, L. J., & Prior, F. W. (2022). Feasibility of telemedicine research visits in people with Parkinson's disease residing in medically underserved areas. *Journal of Clinical and Translational Science*, *6*(1), e133. https://doi.org/10.1017/cts.2022.459

Wang, Q., Zhang, Z., Li, L., Wen, H., & Xu, Q. (2014). Assessment of cognitive impairment in patients with Parkinson's disease: Prevalence and risk factors. *Clinical Interventions in Aging*, *9*, 275–281. https://doi.org/10.2147/CIA.S47367

Weizenbaum, E. L., Fulford, D., Torous, J., Pinsky, E., Kolachalama, V. B., & Cronin-Golomb, A. (2022). Smartphone-based neuropsychological assessment in Parkinson's disease: Feasibility, validity, and contextually driven variability in cognition. *Journal of the International Neuropsychological Society: JINS*, *28*(4), 401–413. https://doi.org/10.1017/S1355617721000503

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2001). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529–536. https://doi.org/10.7326/0003-4819-155-201110180-00009

Wojtala, J., Heber, I. A., Neuser, P., Heller, J., Kalbe, E., Rehberg, S. P., Storch, A., Linse, K., Schneider, C., Gräber, S., Berg, D., Dams, J., Balzer-Geldsetzer, M., Hilker-Roggendorf, R., Oberschmidt, C., Baudrexel, S., Witt, K., Schmidt, N., Deuschl, G., … Reetz, K. (2019). Cognitive decline in Parkinson's disease: The impact of the motor phenotype on cognition. *Journal of Neurology, Neurosurgery, and Psychiatry*, *90*(2), 171–179. https://doi.org/10.1007/s13311-020-00963-x

Zhou, H., Sabbagh, M., Wyman, R., Liebsack, C., Kunik, M. E., & Najafi, B. (2017). Instrumented trail-making task to differentiate persons with no cognitive impairment, amnestic mild cognitive impairment, and Alzheimer disease: A proof of concept study. *Gerontology*, *63*(2), 189–200. https://doi.org/10.1159/000452309