Cognitive Assessment in Older Adults: Clinical Complexity, Performance Validity, and Lived Experience

Susan Lavin (B.A. Hons, MSc)

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

School of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

April 2026

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank all participants who contributed to this research. I am extremely grateful for your time, your contribution to this work, and the trust you placed in me.

I would like to extend my sincere thanks to my supervisors, Professor Jonathan Evans and Dr Clive Ferenbach, for their guidance and support throughout this process. Your expertise and feedback have been invaluable, and I am very grateful for your time, patience, and encouragement throughout this journey.

My thanks also go to the Psychological Therapies for Older People team - recruitment would not have been possible without your support. I am very grateful for the encouragement I received throughout this project and beyond, and for the opportunity to carry out this work within a team that genuinely values research in clinical practice.

I would also like to acknowledge the wider journey of this doctorate. Undertaking this training while balancing family life has been both challenging and deeply meaningful, and has shaped and strengthened me in ways that extend far beyond what can be put into words.

On a personal level, I am deeply grateful to my family and friends for their support throughout this time, and to those within this field who have walked a similar path - your perspective and encouragement in the midst of it all have meant more than you know.

Finally, to Conor, my husband - thank you for your love and patience, and for everything you have done to support this journey. You have been there through it all. As a father to Saorla, you are incredible. I feel very lucky to share this life with you, and I'm excited for everything still to come.

And to my daughter, Saorla, who joined along the way - you are my world. While you may not understand this yet, continue to believe in yourself, and all that you are capable of.

Chapter 1

Older Adults' Experiences of Cognitive Testing in Clinical Settings: A Qualitative Systematic Review

Prepared in accordance with the author requirements for the Journal of Applied Neuropsychology: Adult

Submit to Applied Neuropsychology: Adult

Abstract

Neuropsychological assessment is central to the identification of cognitive impairment and dementia in later life. Within this, cognitive assessment often involves structured testing procedures. However, qualitative research suggests that cognitive testing is not experienced as a neutral clinical procedure, but as a complex and meaningful encounter. Findings remain dispersed, and no qualitative systematic review has synthesised older adults' experiences of cognitive testing across clinical settings. This review aimed to identify, appraise, and synthesise qualitative research exploring these experiences.

A qualitative systematic review was conducted in accordance with PRISMA guidelines. PsycINFO, MEDLINE, and CINAHL were searched from inception to 24 March 2026. Eight studies met inclusion criteria. Data were analysed using thematic synthesis, and methodological quality was assessed using the CASP Qualitative Checklist.

Five analytical themes were generated: (1) managing uncertainty about cognitive change; (2) testing as an evaluative and identity-relevant encounter; (3) testing as a relational and emotionally meaningful experience; (4) trust and understanding shaping acceptance; and (5) the influence of context on testing and its alignment with lived experience. Across studies, testing was experienced as emotionally charged and shaped by clinician interaction, communication, and context, and was often difficult to relate to everyday functioning.

Overall, cognitive testing is experienced as a relational, context-dependent process rather than a purely objective assessment. These findings emphasise the need to interpret performance within a broader clinical framework that considers context, lived experience, and everyday functioning.

Keywords: dementia; older adults; cognitive testing; qualitative synthesis; lived experience

Introduction

Dementia is a leading cause of disability, dependency, and mortality among older adults worldwide and represents a significant and growing public health challenge, with prevalence expected to rise substantially in the coming decades as populations age (World Health Organization, 2023). It is caused by a range of progressive neurodegenerative conditions, with Alzheimer's disease being the most common (Prince et al., 2016). Advances in understanding the underlying pathological processes suggest that changes in the brain may begin many years before the onset of clinically detectable dementia, highlighting a prolonged preclinical phase of cognitive decline (Wilson et al., 2012). In response, there has been increasing emphasis on the early identification of cognitive impairment and its underlying causes, including neurodegenerative disease, as earlier detection may offer opportunities for intervention, support, and care planning (Dubois et al., 2015). However, dementia remains underdiagnosed in clinical practice, with early symptoms often subtle and difficult to distinguish from normal ageing, increasing reliance on clinical assessment processes to support identification and diagnosis.

Within this context, neuropsychological assessment has become a central component of clinical practice in the identification of cognitive impairment and the evaluation of potential underlying conditions. Within this broader process, cognitive assessment is routinely conducted across a range of healthcare settings, including memory clinics, primary care, acute hospitals, outpatient services, inpatient wards, and specialist older adult services. This often involves structured cognitive testing, including the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), and the Addenbrooke's Cognitive Examination III (ACE-III), as well as more comprehensive neuropsychological evaluations assessing multiple cognitive domains.

Despite its central role, the diagnostic pathway is often complex and fragmented. Individuals typically present across multiple services, with general practitioners frequently acting as the first point of contact and coordinating onward referral as part of the assessment process (Rolke et al., 2025). Structural and contextual barriers, including time constraints, resource limitations, variability in service provision, and stigma associated with cognitive decline can further complicate timely identification and diagnosis (Prince et al., 2016). As a result, individuals may experience uncertainty throughout the assessment process, both in relation to the purpose of cognitive testing and the meaning of outcomes. Cognitive testing is therefore not experienced in isolation, but as part of a broader and often ambiguous diagnostic journey.

Reflecting this, a growing body of qualitative research has explored how older adults experience cognitive testing. Across studies, individuals describe a range of emotional and psychological responses, including feelings of strain, performance pressure, and vulnerability during testing, sometimes accompanied by concerns about dignity and judgement (Krohne et al., 2011). Other work highlights uncertainty and emotional ambivalence surrounding neuropsychological assessment, with experiences shaped by clinician communication and the broader context of dementia screening (Bharath et al., 2023; Gruters et al., 2021).

In addition, individuals may approach cognitive testing with pre-existing expectations about ageing, cognitive decline, and the possibility of a dementia diagnosis, which can shape how the assessment process is experienced. For example, Aspö et al. (2024) found that anticipatory beliefs influenced engagement with the diagnostic workup process. Beyond immediate emotional responses, cognitive testing may intersect with identity and self-concept, with individuals engaging in meaning-making and face-preserving strategies following assessment (Lindeberg et al., 2022). The mode of assessment delivery may further influence experience, including telephone-based (Dumassais et al., 2024) and tablet-based approaches (Rolke et al., 2025). However, existing research on alternative delivery methods has largely focused on

feasibility and acceptability, with limited attention to the subjective and emotional experience of cognitive testing. Although these studies provide valuable insights, many have been conducted within specific clinical contexts and with relatively small samples, potentially limiting the transferability of findings across assessment settings.

This body of research suggests that cognitive testing is not a neutral clinical procedure, but a complex and potentially impactful experience for older adults. Experiences of assessment may shape how individuals understand their cognitive functioning, relate to healthcare professionals, and position themselves in relation to ageing and potential diagnosis. These experiences may also influence performance during testing, for example through anxiety, uncertainty, or perceived evaluation, highlighting the interactive and context-dependent nature of cognitive assessment.

Despite a growing body of qualitative research, findings remain dispersed across studies, settings, and modes of assessment delivery. Existing systematic reviews have primarily focused on attitudes toward dementia screening, help-seeking behaviours, or diagnostic disclosure, with limited integration of research exploring the lived experience of cognitive testing. To date, no focused qualitative systematic review has synthesised older adults' experiences of cognitive testing across clinical settings.

A qualitative systematic review is therefore warranted to integrate existing findings and generate deeper interpretive insights into how cognitive testing is experienced. Such synthesis may inform more person-centred approaches to assessment and enhance clinician–patient interactions within later-life services. The aim of this review is to systematically identify, appraise, and synthesise qualitative research exploring older adults' experiences of undergoing cognitive testing in clinical healthcare settings.

Methods

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). See Appendix A.

The study protocol was prospectively registered with PROSPERO (registration number: CRD420261339086) on 13 March 2026 and is available at: https://www.crd.york.ac.uk/PROSPERO/view/CRD420261339086

Search strategy

Electronic database searches were conducted in PsycINFO (via Ovid), MEDLINE (via Ovid), and CINAHL (via EBSCOhost) from inception to 24 March 2026. An initial search was undertaken on 19 February 2026, followed by an updated search on 24 March 2026 after consultation with an expert librarian at the University of Glasgow to refine the strategy and capture additional studies.

Database selection was informed by methodological guidance for qualitative evidence synthesis (Booth, 2016) and aimed to capture relevant research across psychological, nursing, and biomedical literature. PsycINFO and CINAHL were included to optimise retrieval of qualitative and dementia-related studies, while MEDLINE provided broader biomedical coverage.

The search strategy combined controlled vocabulary (e.g., MeSH terms in MEDLINE and equivalent subject headings in PsycINFO and CINAHL) with free-text terms to ensure comprehensive retrieval of relevant studies. Given recognised challenges in identifying qualitative research, including inconsistent indexing and variability in terminology (Booth, 2016), a sensitive search strategy was prioritised.

Search terms were structured using the SPIDER framework (Sample, Phenomenon of Interest, Design, Evaluation, Research type) and organised around three concepts: (1) older adults, (2)

cognitive or memory testing, and (3) qualitative research. Subject headings and free-text terms were combined using OR, and concept groups were combined using AND. Free-text terms were prioritised for the population concept to maximise sensitivity, with relevance determined during screening.

Terms and synonyms were developed for each concept, including older adults (e.g., older adult*, elderly, geriatric*, later life); cognitive or memory testing (e.g., neuropsychological tests, cognitive assessment, cognitive screening, MMSE, MoCA, ACE-III); and qualitative research, informed by established approaches to identifying qualitative studies (Booth, 2016), incorporating methodological terms (e.g., qualitative, thematic analysis, grounded theory) and data collection methods (e.g., interview*, focus group*).

Search strategies were adapted for each database to reflect differences in indexing terms and syntax. Full search strategies for each database are provided in Appendix B.

Study selection

All records were imported into EndNote (version 21) for storage and de-duplication, with duplicates removed prior to screening. Eligibility criteria used to guide screening and selection are outlined in Table 1.

Table 1. *Study Inclusion and Exclusion Criteria*

Inclusion	Exclusion
Adults undergoing cognitive testing in later-life or older adult clinical services	Younger or mid-life populations (e.g., mean age < 55 years)
Explored experiences of cognitive testing	Outside later-life clinical assessment contexts
Conducted in clinical settings (e.g., memory clinics, primary care, hospitals, outpatient/inpatient services), or community/research contexts with comparable structured assessments	Caregiver or clinician perspectives only
Used qualitative methodologies (e.g., interviews, focus groups, thematic analysis, grounded theory, phenomenology)	Quantitative-only studies or those examining general attitudes toward screening without direct testing experience
Published in peer-reviewed English-language journals	Diagnostic disclosure only (unless testing experience explored separately) or technology usability studies without qualitative experiential data

Study screening

Study selection was conducted in two stages: (1) title and abstract screening and (2) full-text screening. The first reviewer screened all records against the eligibility criteria, with a second reviewer (MB, Trainee Clinical Psychologist) independently screening a subset (10%) at the title and abstract stage to ensure consistency. Inter-rater agreement for this subset was 100%, with no discrepancies regarding application of the eligibility criteria.

Records meeting inclusion criteria were retrieved for full-text review. Reference lists of included studies were screened (backward citation searching) to identify additional papers; none met the inclusion criteria. No additional search methods (e.g., forward citation searching) were undertaken.

Data extraction

A standardised data extraction form was developed and piloted prior to use. Data extraction was conducted by the first reviewer, with a second reviewer independently checking a subset of studies (n = 3; 37.5%). Agreement between reviewers was high, with minor discrepancies resolved through discussion, supporting the reliability of the process for the remaining studies. The same subset was used for double quality appraisal.

The following data were extracted from each study: study characteristics (author(s), year of publication, country); study aims; participant characteristics; clinical setting; cognitive assessment used; qualitative methodology and analytic approach; and key findings relating to experiences of cognitive or memory testing.

Qualitative data, including participant quotations and authors' interpretations reported in the results or findings sections, were extracted and managed using Microsoft Excel. This supported systematic organisation, comparison across studies, and transparent tracking of coding decisions throughout the synthesis process.

Data synthesis

A thematic synthesis was conducted following the approach described by Thomas and Harden (2008). This involved three stages: (1) line-by-line coding of findings from included studies, (2) development of descriptive themes, and (3) generation of higher-order analytical themes.

Text labelled as results or findings in included studies was coded line-by-line, with attention to preserving participants' accounts, including direct quotations. Codes were developed

inductively and compared within and across studies, with multiple codes applied where appropriate to capture complexity.

Codes were organised into descriptive themes reflecting shared meanings across studies, with ongoing refinement as analysis progressed. Themes were grounded in data from multiple studies while retaining within-study nuance.

In the final stage, descriptive themes were interpreted in relation to the review question to generate higher-order analytical themes, providing a more conceptual understanding of participants' experiences while remaining grounded in the data. This involved moving beyond the original study findings to develop interpretive insights, while ensuring that themes were supported by the underlying data.

To enhance analytic rigour, strategies included maintaining a distinction between participant quotations and author interpretations, attending to disconfirming cases, and reviewing the contribution of each study to ensure balanced representation. Reflexive discussion supported ongoing interpretation, and a structured audit of theme coverage across studies enhanced transparency and consistency in the analysis.

Quality appraisal

Methodological quality of included studies was assessed using the Critical Appraisal Skills Programme (CASP) Qualitative Checklist (CASP, 2018). The checklist comprises ten items assessing key domains of qualitative research, including clarity of aims, appropriateness of methodology, research design, recruitment strategy, data collection, reflexivity, ethical considerations, rigour of analysis, and the value of the research.

Each study was appraised using ratings of "yes", "no", or "can't tell" to indicate whether each criterion was met. Quality appraisal was conducted by the first reviewer, with a second reviewer (MB, Trainee Clinical Psychologist) independently assessing a subset of studies (n =

3; 37.5%) to ensure consistency. Initial agreement was 90% (27/30 items), and discrepancies were resolved through discussion, resulting in full agreement. This level of agreement was considered sufficient to support the reliability of the appraisal process, and therefore the remaining studies were appraised by the first reviewer.

Studies were not excluded based on quality; rather, appraisal findings were used to inform interpretation of the synthesis.

Researcher reflexivity

Thematic synthesis is inherently interpretive and shaped by the researcher's perspectives and experiences. The primary reviewer was a Trainee Clinical Psychologist with experience working with individuals and families affected by cognitive decline, which may have influenced sensitivity to particular aspects of participants' experiences.

To support reflexivity, a reflective diary was maintained throughout the analytic process, alongside regular supervision to facilitate critical reflection on assumptions and interpretations. These processes supported ongoing examination of how interpretations were developed and refined, enhancing transparency and maintaining a reflexive approach to the analysis. See Appendix G for the reflexivity statement.

Results

A total of 9,242 records were identified through database searching across MEDLINE, PsycINFO, and CINAHL, including both the initial and updated searches. All records were imported into EndNote, and 2,590 duplicates were removed, resulting in 6,652 unique records for screening.

Titles and abstracts were screened against the predefined eligibility criteria, and 6,638 records were excluded, leaving 14 full-text articles for screening.

All 14 full-text articles were retrieved and assessed for eligibility. Of these, 6 were excluded (2 due to insufficient focus on the direct experience of cognitive testing, 2 due to context mismatch, and 2 due to population mismatch), resulting in 8 studies included in the final review. Reference lists of included studies were screened (backward citation searching); however, this did not identify any additional eligible studies. The study selection process is summarised in the PRISMA flow diagram (see Figure 1)

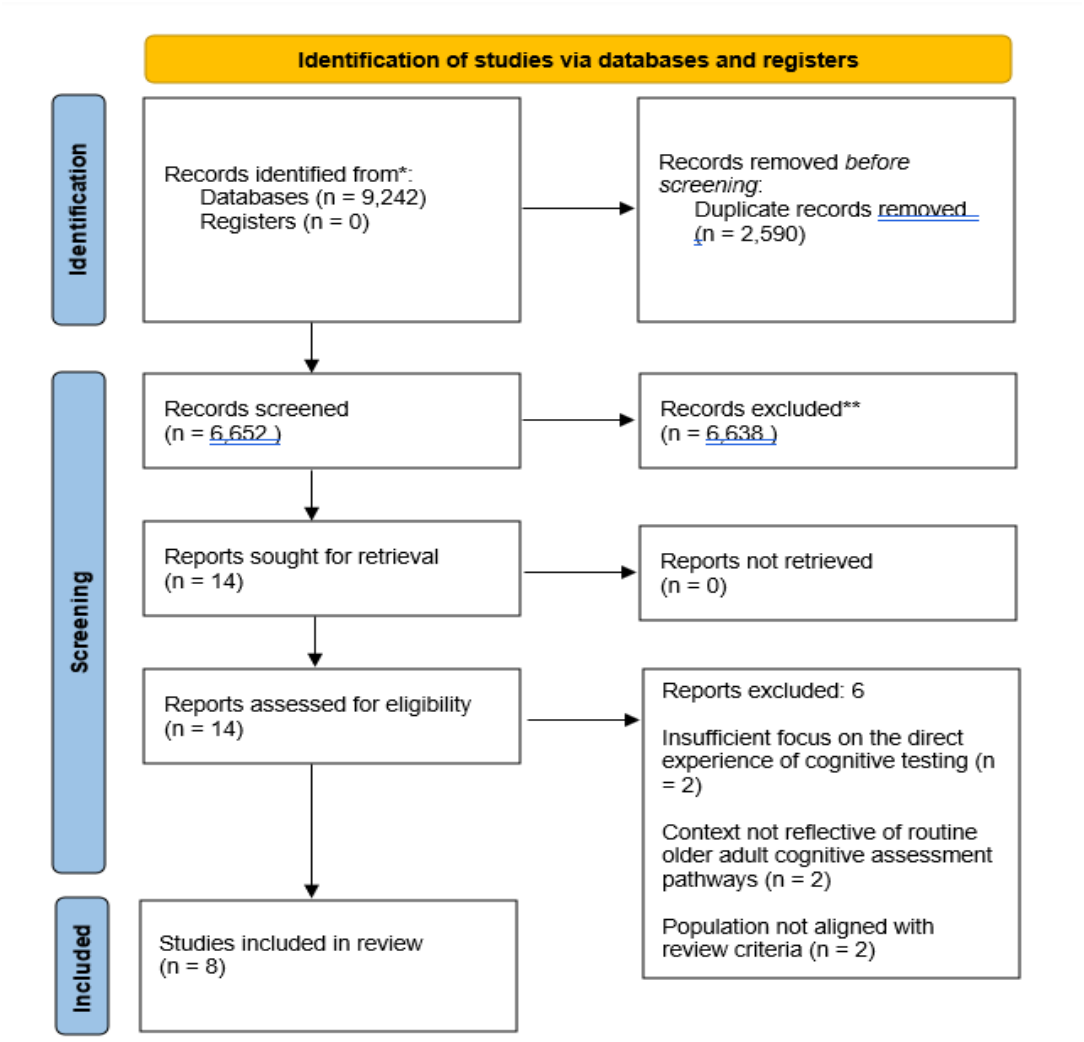

Figure 1. *PRISMA Flow Diagram*

Study characteristics

The included studies were conducted across multiple countries (India, UK, Ireland, Netherlands, Sweden, Australia, Canada, and Germany) and in diverse clinical settings, including memory clinics, primary care, neuropsychological services, and community-based screening. Participants were older adults undergoing cognitive or memory testing, consistent with the review inclusion criteria. Across the eight included studies, the synthesis reflected the experiences of 156 older adults undergoing cognitive testing in clinical healthcare settings.

A range of qualitative methodologies were employed, including phenomenological, longitudinal, descriptive, and mixed-methods designs with a qualitative component. Data were primarily collected via semi-structured interviews (n = 7), with one study using focus groups, and analysed using thematic (n = 5) and qualitative content analysis (n = 3).

Cognitive assessment approaches included commonly used screening tools (e.g., MMSE, MoCA), broader neuropsychological assessments, condition-specific tools (e.g., ADAS-Cog), and computerised platforms. Several studies used multiple assessment measures.

Further details are presented in Table 2.

Table 2. *Summary of Study Characteristics*

Author (Year)	Country	Aim	Participants	Setting	Assessment	Method	Collection	Analysis	Key Findings
Bharath et al. (2023)	India	Lived experience of neuropsych assessment	n=15	Community + screening	ADAS-Cog	Qualitative (phenomenological)	Interviews	Thematic	Mixed emotions (anxiety, curiosity, reassurance); cultural influences
Birt et al. (2020)	UK	Help-seeking and assessment experiences	n=41	NHS memory services	Cognitive assessment (not specified)	Qualitative (longitudinal)	Serial interviews	Thematic	Uncertainty, identity disruption; relational help-seeking
Cahill et al. (2008)	Ireland	Expectations and experiences of memory clinic	n=28	Memory clinic	Neuropsych testing	Mixed methods	Pre/post interviews	Thematic	Initial anxiety reduced; reassurance; clinician interaction important
Gruters et al. (2021)	Netherlands	Assessment and diagnostic disclosure experiences	n=27	Memory clinics	Neuropsych testing	Qualitative	Focus groups	Content	Diagnostic uncertainty; emotional ambivalence; “diagnostic paradox”
Lindeberg et al. (2021)	Sweden	Cognitive and language	n=8	Memory care	MMSE + language tests	Qualitative	Interviews	Content	Testing disconnected from real life;

Author (Year)	Country	Aim	Participants	Setting	Assessment	Method	Collection	Analysis	Key Findings
		assessment experiences							communication challenges
Rizkallah et al. (2024)	Australia	Sociocultural influences on assessment	n=10	Neuropsych setting	Neuropsych battery	Qualitative	Interviews	Thematic	Culture, language, education shape understanding; importance of rapport
Robillard et al. (2018)	Canada	Perspectives on computerised testing	n=19	Alzheimer's clinic	Cognigram	Qualitative	Interviews (pre/post)	Thematic	Technology anxiety; unfamiliarity affects performance; preference for human interaction
Rolke et al. (2025)	Germany	Acceptance of tablet-based testing	n=8	General practice	Tablet MoCA	Qualitative	Interviews	Content	Generally acceptable; barriers: time, resources

Quality Appraisal

The methodological quality of included studies was generally high, with 5 of the 8 studies rated as high quality, 2 rated as moderate-high quality, and 1 rated as moderate quality. All studies demonstrated clear research aims and appropriate qualitative methodologies, with designs largely suited to the research questions. Recruitment strategies were clearly described and generally appropriate.

Data collection methods were well described, with 7 studies using interviews and 1 using focus groups, all of which were appropriate for exploring participants' experiences. However, reflexivity was inconsistently addressed, with 3 of the 8 studies failing to consider the relationship between researcher and participants or the influence of researcher perspectives.

Ethical considerations were reported in 7 of the 8 studies although detail varied, and in some cases it was unclear whether ethical approval had been obtained. Data analysis procedures were reported across all studies, but depth and transparency varied, with some providing more detailed accounts of coding and theme development than others.

Overall, studies were considered of sufficient quality to contribute meaningfully to the synthesis. Identified limitations, particularly relating to reflexivity and reporting transparency, were considered during interpretation, with greater weight given to more detailed studies and more cautious interpretation of less well-reported findings.

Table 3. *Critical Appraisal of Included Studies Using the CASP Qualitative Checklist*

Study	Aims	Methodology	Design	Recruitment	Collection	Reflexivity	Ethics	Analysis	Findings	Value	Quality
Bharath et al. (2023)	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Moderate-High
Birt et al. (2020)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Cahill et al. (2008)	Yes	Yes	Yes	Yes	Yes	No	Can't tell	Yes	Yes	Yes	Moderate
Gruters et al. (2021)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Lindeberg et al. (2021)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Rizkallah et al. (2024)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Robillard et al. (2018)	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Moderate–High
Rolke et al. (2025)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High

Thematic Synthesis

Thematic synthesis of the eight included studies generated five analytical themes capturing the experiences of 156 older adults undergoing cognitive testing in clinical healthcare settings. These themes reflect interrelated emotional, relational, and contextual processes shaping how testing is experienced and understood. Rather than operating in isolation, they represent a dynamic and overlapping process through which individuals make sense of assessment in the context of cognitive change. The relationships between these themes are illustrated in Figure 2.

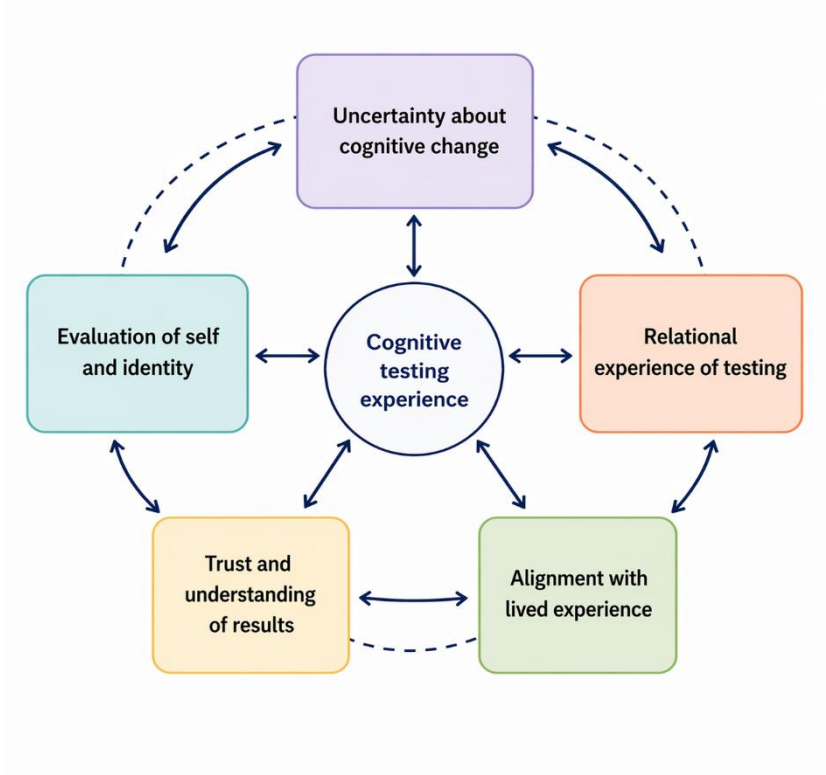

Figure 2. *Conceptual model of older adults' experiences of cognitive testing*

Across studies, cognitive testing was experienced as a complex and meaningful encounter rather than a neutral clinical procedure. Participants described entering assessment with uncertainty about cognitive change, often seeking reassurance about whether their experiences reflected normal ageing or the onset of illness. This was accompanied by feelings of evaluation, performance pressure, and concerns about competence and identity.

Experiences of testing were also shaped by relational factors, including clinician communication and perceived support, which influenced engagement and understanding. In turn, acceptance and interpretation of outcomes were closely linked to trust and the clarity of explanations provided.

The synthesis generated five analytical themes: (1) cognitive testing as managing uncertainty about cognitive change, (2) assessment as an evaluative and identity-relevant encounter, (3) assessment as a relational and emotionally meaningful experience, (4) trust and understanding shape acceptance of testing, and (5) cognitive testing is shaped by context and may not align with lived experience.

Theme 1: Cognitive testing as managing uncertainty about cognitive change

Across studies, cognitive testing was experienced within a broader context of uncertainty about cognitive change and ageing. Participants frequently described becoming aware of memory lapses or cognitive difficulties that were difficult to interpret, leading to questions about whether such changes reflected normal ageing or the onset of illness. As one participant reflected, *"I hope that isn't the start of dementia… I hope it's just being elderly"* (Birt et al., 2020), capturing the ambiguity that often preceded assessment.

This uncertainty formed the context within which testing was sought, experienced, and interpreted. For some, uncertainty emerged gradually through everyday lapses, while for others it was triggered by more acute or disorienting experiences. One participant described becoming lost in a familiar setting, prompting the decision to seek testing: *"I was on the bus and I thought, 'Where the devil am I going?'... So I went to the doctor"* (Birt et al., 2020). Similarly, participants in Rolke et al. (2025) described forgetfulness as *"scary,"* with one noting, *"sometimes I can't remember… what I was actually doing… That scares me a bit."* These

accounts highlight how uncertainty was grounded in lived experience and often emotionally charged.

Uncertainty was not limited to concerns about cognitive change, but also extended to the assessment process itself. Several participants described not understanding why testing had been arranged or what it would involve. For example, one participant stated, *"I was worried coming in because I don't really know why it was arranged"* (Cahill et al., 2008), while another reflected, *"I wasn't sure what you are going to do with the tests"* (Rizkallah et al., 2024). In this way, uncertainty was experienced both in relation to cognitive change and the purpose and interpretation of testing.

Testing was therefore often approached as a means of resolving uncertainty and gaining reassurance. Participants described using assessment to evaluate whether they were *"normal"* for their age, with one noting, *"I feel I am normal"* (Bharath et al., 2023), and another stating, *"it has given me more confidence that I am ok at this age."* Testing was also described as a way to *"know where we are… whether my memory is ok at this age"* (Bharath et al., 2023), indicating its role in benchmarking cognitive functioning. Some participants additionally framed testing in terms of early identification, suggesting it could *"put your mind at ease"* (Rolke et al., 2025) or support early detection.

However, reassurance was not consistently achieved. For some participants, uncertainty persisted during and after testing, particularly when they were unable to judge their own performance. One participant described, *"I don't know whether it's bad or good"* (Gruters et al., 2021). Others reported ongoing uncertainty following assessment, especially when explanations were limited or concerns were not taken seriously. As one participant noted, *"I'll have to wait until I am a little more obviously demented…"* (Birt et al., 2020), suggesting that uncertainty could remain unresolved.

Uncertainty also extended across the wider assessment pathway, with participants describing ongoing uncertainty and not knowing. One reflected, *"I wondered what was going on in my head and which direction it was going"* (Gruters et al., 2021). Even when a diagnosis was given, uncertainty was not always resolved, with some individuals continuing to express hope for improvement, such as wanting *"anything… as long as it gives me hope"* (Lindeberg et al., 2021).

Together, these findings suggest that cognitive testing takes place within a broader context of ongoing uncertainty about cognitive change. While it may provide reassurance for some, it does not consistently resolve uncertainty and may introduce further ambiguity. This highlights that testing is not a discrete event, but part of an ongoing process through which individuals make sense of cognitive change, underpinning subsequent themes of evaluation, relational experience, and trust.

This theme was represented across 7 of the 8 included studies and was supported predominantly by high-quality studies (4 high-quality, 2 moderate-high quality, and 1 moderate-quality), strengthening confidence in the findings. However, variation in the depth of participant accounts reported across studies may have influenced understanding of how uncertainty was experienced and navigated.

Theme 2: Assessment as an evaluative and identity-relevant encounter

Building on this uncertainty, participants often experienced cognitive testing as an evaluative encounter in which their abilities, competence, and identity were perceived to be judged. Performance was interpreted not simply as task completion, but as a reflection of personal worth, capability, and sense of self. Anticipation of testing often generated anxiety centred on the possibility of making mistakes and their implications. One participant asked, *"What if I am*

caught making some mistake?" (Bharath et al., 2023), reflecting fears of exposure and judgement.

During testing, this evaluative dimension became more pronounced, with difficulties often internalised as evidence of personal inadequacy rather than features of the assessment. For example, one participant reflected, *"I feel myself a dull person"* after struggling to answer questions (Bharath et al., 2023), while another stated, *"I felt really stupid when I couldn't do some of them"* (Cahill et al., 2008). These accounts suggest that errors were experienced as personally meaningful, contributing to negative self-evaluation and diminished self-worth. Similarly, participants described embarrassment when tasks appeared simple yet were difficult to complete, with one noting, *"I couldn't even draw the house… even a child could do that… it is embarrassing"* (Cahill et al., 2008), reinforcing the perception that testing exposed deficits in a publicly evaluative context. These accounts suggest that challenges to competence were not experienced as isolated task difficulties, but as threats to individuals' sense of self, linking cognitive performance to broader identity and self-concept.

For some participants, testing evoked a sense of failure, despite not being designed as a pass–fail exercise. One account described a *"very strong feeling of failing"* and the belief that *"I can't do it anymore"* (Gruters et al., 2021), illustrating how assessment could be experienced as a loss of competence, and more broadly, as a threat to their sense of self. This was further reflected in self-critical responses such as *"I could have done better. I should have done better"* (Robillard et al., 2018), suggesting that poor performance was often internalised as personal failure rather than understood as a feature of testing conditions or cognitive change.

The evaluative nature of testing was also shaped by broader social and cultural meanings attached to tests and performance. Participants sometimes drew on prior experiences of formal testing, where success and improvement were expected. As one participant explained, *"you're*

supposed to learn from every test and do better… [but] there's no way I can do better" (Robillard et al., 2018), highlighting the tension between familiar expectations of performance and the realities of cognitive decline. Structural and educational factors could also be internalised as personal deficits, with one participant stating, *"I am uneducated… I am not clever"* (Rizkallah et al., 2024), suggesting that testing may reinforce existing inequalities and self-perceptions of inadequacy.

Beyond self-evaluation, experiences of assessment were also shaped by interactions with clinicians and the need for concerns to be recognised as legitimate. Some participants described feeling dismissed when their difficulties were attributed to normal ageing, requiring them to justify their experiences to access assessment. One reflected, *"I knew he was wrong but you can't tell a consultant that he's wrong… I'm just told… don't trouble me"* (Birt et al., 2020), illustrating how professional authority could constrain individuals' ability to assert their concerns. In this way, the evaluative nature of assessment extended beyond task performance to include whether individuals were perceived as credible and deserving of investigation.

Importantly, the consequences of testing could further intensify its identity-relevant nature. Clinical interpretations and diagnostic labels were sometimes experienced as definitive judgements with significant personal implications. For example, one participant described receiving a diagnosis as *"like you get a stamp… Alzheimer's"* (Gruters et al., 2021), reflecting the perceived permanence and stigma associated with such labels. Similarly, when test results led to functional restrictions, such as losing the ability to drive, participants could struggle to understand the basis for these decisions, with one noting, *"I didn't know what those scores meant… but which quality was missing, I don't know"* (Lindeberg et al., 2021). This lack of clarity could reinforce the sense that assessment outcomes were imposed judgements rather than transparent or understandable conclusions.

Taken together, these findings suggest that cognitive testing is experienced as a deeply evaluative encounter in which individuals feel judged not only in terms of performance, but also in relation to competence and sense of self. Difficulties are often experienced as threats to identity and social value, with errors and outcomes internalised as shame, embarrassment, and diminished self-worth. At the same time, the need to be recognised and taken seriously highlights how assessment operates within broader dynamics of credibility and authority. This suggests that cognitive testing involves navigating not only cognitive demands, but also the emotional and identity-related implications of evaluation.

Experiences relating to evaluation and identity were evident across 7 of the 8 included studies and were supported by 4 high-quality, 2 moderate-high quality, and 1 moderate-quality study, strengthening confidence in this theme. Nevertheless, limited reflexive consideration in some studies may have influenced interpretation of participants' experiences of shame, competence, and threats to identity.

Theme 3: Assessment as a relational and emotionally meaningful experience

Extending beyond uncertainty and evaluation, participants described cognitive testing as a relational and emotionally meaningful experience shaped by interactions, the testing environment, and communication. Testing was experienced not only as a cognitive task, but as a relational encounter influencing emotional wellbeing, engagement, and meaning.

For some, relational aspects were experienced positively, providing opportunities for social interaction, emotional expression, and relief from loneliness. One participant described, *"I spent time with her [the clinician]… the loneliness goes"* (Bharath et al., 2023), suggesting that interaction with the clinician was experienced as familiar and emotionally meaningful.

More broadly, participants emphasised the importance of being listened to, supported, and given time during assessment. One participant described the clinic as a place where you could *"get things off your chest… everyone was very nice"* (Cahill et al., 2008), highlighting the role of emotional support in shaping a more positive experience. Communication style was also important, with one reflecting that difficulties with testing were not due to ability, but *"the way he put it over"* (Birt et al., 2020), indicating that how tasks were explained within the assessment could significantly shape engagement and perceived performance.

At the same time, relational dynamics were not uniformly positive and could shape experiences in more complex ways. Some participants described communication as insufficient or difficult to retain, with one noting, *"I forgot all the details… I would have liked to have received something on paper"* (Gruters et al., 2021), highlighting the need for ongoing and accessible communication. Others described tensions within support, particularly where adjustments were perceived as threatening autonomy, with one expressing discomfort at others slowing down communication (Lindeberg et al., 2021). This suggests that support within assessment is relationally negotiated and not always experienced as helpful.

Importantly, participants also described how cognition itself could be supported or disrupted within interaction. In some cases, skilled facilitation enabled participants to engage more fully, with one describing how staff were *"good at bringing out… asking questions and getting us to talk"* (Lindeberg et al., 2021), creating inclusive and supportive communicative spaces. Similarly, memory was sometimes described as being collaboratively supported, with participants using prompts such as *"have you heard about this…?"* to help retrieve information (Lindeberg et al., 2021), suggesting that cognitive performance was not entirely individual, but could be co-constructed in interaction.

Relational context also shaped emotional responses to testing. Informal conversation and rapport-building were described as reducing anxiety, with one participant noting, *"having a chat was entertaining… I got calmer somehow"* (Rizkallah et al., 2024). However, for others, interaction could increase pressure, particularly when difficulties in communication were exposed, leading one participant to suggest that a computer might feel less stressful than interacting with a person (Robillard et al., 2018). This illustrates how the interpersonal nature of testing can both support and complicate the experience.

Finally, the relationship between the participant and the person administering the test was significant. Spending more time with staff was associated with a more comfortable and engaging experience, allowing for greater closeness and reduced time pressure (Rolke et al., 2025). Shared language and cultural background could further enhance feelings of safety and understanding, with one participant noting that hearing their native language made them *"feel safe"* (Rizkallah et al., 2024), highlighting the role of relational familiarity in shaping the testing experience.

Overall, these findings suggest that cognitive testing is fundamentally relational. Experiences are shaped not only by tasks, but by interactions, communication, and social context. These dynamics can support engagement and reduce anxiety, but may also introduce complexity, particularly where issues of autonomy or communication arise, highlighting the relational and emotional demands of assessment.

The consistency with which relational and emotional aspects of cognitive testing were reported across all 8 included studies, including 5 high-quality, 2 moderate-high quality, and 1 moderate-quality study, strengthens confidence in this theme. However, differences in the reporting of interpersonal and contextual factors may have limited insight into how relationships and communication shaped participants' experiences of assessment.

Theme 4: Trust and understanding shape acceptance of testing

Across studies, participants' acceptance of cognitive testing was shaped by the extent to which the process felt trustworthy, transparent, and understandable. While the previous theme focused on relational and emotional experience, this theme highlights how participants evaluated and made sense of testing, including whether it was perceived as credible, fair, and meaningful. Trust in clinicians, clarity of communication, and perceived fairness influenced whether testing was experienced as acceptable or valid. Although closely related, trust and understanding operated as distinct processes, with understanding shaping the extent to which testing was trusted and accepted.

At the outset, trust in the clinician appeared to reduce anxiety and facilitate engagement with testing. One participant described entering assessment without fear because of confidence in the assessor: *"I did not have any worry or tension about the test… I knew that you would not show anything that makes me afraid or uncomfortable"* (Bharath et al., 2023). This suggests that trust functioned as a protective factor, shaping the emotional tone of the testing experience and enabling individuals to engage without anticipatory threat.

However, this sense of trust was fragile and could be undermined when the testing process appeared inconsistent, unclear, or unfair. Participants questioned the validity of assessment outcomes when results differed across contexts, with one noting: *"The tests I did at the clinic I got 85%, then… at my home I got 95%… if I had got 95% before I won't have been through all of this"* (Birt et al., 2020). Similarly, a lack of transparency during assessment contributed to perceptions of unfairness and loss of trust. One participant described: *"They kept firing questions at me… I don't feel that I got a fair assessment… all of a sudden I get landed with, 'You've got Alzheimer's!'… it's made me lose my confidence"* (Birt et al., 2020). This suggests

that unclear processes can lead to feelings of being judged without understanding, undermining both trust and confidence.

Understanding of the testing process and its outcomes also emerged as central to acceptance. Participants frequently described difficulty making sense of results and their implications, with one noting: *"I didn't know what those scores meant… 'you can't drive a car' – but which quality was missing, I don't know"* (Lindeberg et al., 2021). This illustrates how clinical conclusions can feel arbitrary when the link between performance and real-world consequences is not clearly explained, potentially undermining confidence in the assessment. Similarly, participants reported that explanations were often insufficient or difficult to retain: *"I forgot all the details of the information… I would have liked to have received something on paper"* (Gruters et al., 2021). Together, these accounts suggest that understanding depends not only on information provision, but on accessibility, timing, and the ability to revisit explanations, with difficulties in understanding shaping how far assessment outcomes are trusted and accepted.

Language and communication further shaped both trust and understanding. Even where shared language or interpretation was available, participants reported difficulties with clarity, with one noting: *"People use… words [that] we don't understand"* (Rizkallah et al., 2024). This suggests that linguistic concordance alone is insufficient; communication must also be culturally and educationally accessible.

Participants also highlighted the importance of human interpretation in making testing feel valid and acceptable. While computerised testing was perceived as objective, it was not seen as sufficient in isolation: *"[A computer test] is unbiased… but… a perceptive interviewer will read more than just the strict answer… the body language… the tone"* (Robillard et al., 2018). This suggests that trust is not grounded solely in objectivity, but in the integration of clinical expertise, relational understanding, and interpretation.

Finally, clarity regarding the purpose and outcomes of testing influenced whether it was experienced as worthwhile. Participants expressed a desire for clearer explanations of the process and next steps: *"That someone simply explains what the memory clinic is… and what you need to do"* (Rolke et al., 2025). This suggests that acceptance is closely tied to perceived usefulness and clarity in understanding one's situation.

Together, these findings suggest that cognitive testing is accepted not simply as a clinical procedure, but as a process that must feel trustworthy, fair, and understandable. Trust can facilitate engagement and reduce anxiety, whereas lack of transparency, inconsistency, and poor communication may undermine confidence in both the process and its outcomes. Understanding of tasks, results, and consequences is central to whether testing is experienced as meaningful or arbitrary, highlighting that assessment is shaped as much by how it is communicated as by what is measured.

Trust and understanding emerged as important influences on acceptance of cognitive testing across 7 of the 8 included studies and were supported by evidence from 5 high-quality and 2 moderate-high quality studies, strengthening confidence in this theme. However, variation in reporting depth across studies may have limited understanding of how participants developed trust in assessment processes and made sense of test outcomes.

Theme 5: Cognitive testing is shaped by context and may not align with lived experience

Across studies, experiences of cognitive testing were shaped not only by cognitive ability, but by the broader context in which assessment occurred, including task design, environment, and cultural and educational background. Participants frequently described a mismatch between test performance and everyday functioning, raising questions about the relevance of testing.

In some instances, participants identified ways in which testing aligned with everyday functioning. One reflected: *"Identifying things like date, day, year… while writing the cheque I remembered the date… it's useful that way"* (Bharath et al., 2023). However, such alignment was not consistently observed across studies.

More commonly, participants described difficulty understanding the purpose or relevance of tasks. One questioned: *"What I wrestled with? The things I had to draw… I don't know why I had to do that… what do you do with these results?"* (Gruters et al., 2021). This highlights how tasks could feel opaque and disconnected from real-world problem-solving, making performance difficult to interpret and contributing to uncertainty.

This sense of misalignment was further reinforced when tasks relied on unfamiliar formats or culturally specific conventions. For example, one participant noted: *"I have never done tests like this before… we did not do this in Greece when I was young"* (Rizkallah et al., 2024), while another explained: *"I haven't used a pencil… for 60 years… how will I write?"* (Rizkallah et al., 2024). These accounts suggest that performance was influenced by prior exposure and cultural experience, rather than cognitive ability alone.

The conditions under which testing was conducted shaped performance in several ways. Participants described the impact of time pressure, with one noting: *"If you're thinking and thinking and then you run out of time, you can feel a certain amount of time pressure"* (Rolke et al., 2025). Timed conditions may therefore introduce additional cognitive load, influencing performance through anxiety and disrupted retrieval. Performance was also affected by aspects of test delivery and interface design. One participant described difficulties with digital testing: *"You're really wanting to hit this one, but you're automatically hitting the other – wrong"* (Robillard et al., 2018), highlighting how motor demands and interface mechanics may shape performance independently of cognitive ability.

Physical and emotional context also played an important role in shaping engagement. Participants described how anxiety and unfamiliar environments affected performance, with one noting: *"Maybe I would have done better at home… if I wasn't so nervous"* (Cahill et al., 2008), and another reflecting: *"In the hospital… you meet a lot of unfamiliar people… At home… the context is more productive"* (Rizkallah et al., 2024). These accounts suggest that environmental and emotional factors may influence performance beyond cognitive ability. The way testing was conducted and explained further shaped how results were interpreted. One participant noted: *"I could do it quite easily once I knew what I was doing, it was the way he put it over"* (Birt et al., 2020), indicating that communication can shape how performance is interpreted, including whether difficulties are attributed to the task or to the individual.

Overall, these findings suggest that cognitive testing is shaped by contextual, cultural, and situational factors. Tasks may feel unfamiliar or lack relevance, while environmental conditions and delivery influence engagement and performance. This contributes to a perceived misalignment between test performance and everyday functioning, with participants questioning whether testing accurately reflects their real-world cognitive abilities.

Accounts of contextual, cultural, and situational influences on cognitive testing were evident across 7 of the 8 included studies and were supported by evidence from 4 high-quality, 2 moderate-high quality, and 1 moderate-quality study, strengthening confidence in this theme. Nevertheless, differences in the depth of contextual information reported may have influenced interpretation of the extent to which cognitive testing reflects everyday functioning and real-world cognitive abilities.

Discussion

Overview of findings

This review synthesised qualitative evidence from eight studies to explore older adults' experiences of undergoing cognitive or memory testing in clinical healthcare settings. The

findings suggest that cognitive testing is experienced not as a purely technical procedure, but as a complex process shaped by emotional, relational, and contextual factors.

Across studies, participants described entering assessment with uncertainty about cognitive change, reflecting concerns about ageing, memory, and the potential onset of illness. This is consistent with wider literature indicating that early symptoms of dementia can be difficult to distinguish from normal ageing (Prince et al., 2016). The present synthesis extends this by suggesting that uncertainty may persist throughout the assessment process rather than being fully resolved by testing alone, particularly where communication is limited or difficult to retain (Gruters et al., 2021). This suggests that cognitive testing may not always provide the clarity individuals seek and may, in some cases, contribute to ongoing uncertainty.

The identification of cognitive testing as an evaluative and identity-relevant encounter aligns with previous research highlighting its emotional impact. Older adults have been found to experience cognitive testing as effortful and pressurised, with performance linked to threats to dignity and self-worth (Krohne et al., 2011). Research also suggests that individuals actively manage how they account for their performance to protect their sense of competence (Lindeberg et al., 2022). The current synthesis extends this by showing how evaluation is experienced within the testing situation itself, with performance interpreted as indicative of identity, competence, and, in some cases, independence. These findings suggest that cognitive testing functions not only as an assessment of cognition, but also as a socially and personally meaningful evaluative event. Unlike many medical investigations, cognitive testing may be experienced as an evaluation of abilities closely linked to identity, autonomy, and future independence. For many participants, testing carried implications for identity, independence, future functioning, and how they were perceived by others.

The relational nature of cognitive testing identified in this review is consistent with qualitative findings emphasising the importance of clinician–patient interactions. Experiences of assessment depend, in part, on how individuals are received, listened to, and supported, with trust developing or diminishing across the diagnostic process (Birt et al., 2020). Clear explanation prior to and during assessment is also important, with inadequate preparation contributing to confusion and anxiety (Bharath et al., 2023). The present findings extend this literature by suggesting that relational factors shape not only emotional experience, but also how individuals make sense of and engage with assessment, indicating that cognitive testing is mediated through interaction rather than solely an individual cognitive task.

The findings relating to trust and understanding further support this interpretation. Participants valued clear explanations and professional conduct, but described difficulties retaining and making sense of information during and after assessment. Previous research similarly suggests that, although feedback is valued, retention may be limited where assessment is emotionally salient (Gruters et al., 2021). This highlights a gap between information provision and meaningful understanding, suggesting that effective communication depends not only on clarity, but also on timing, format, and individuals' capacity to process information in the moment. These findings are consistent with wider literature suggesting that effective consultations support patients to understand and engage with assessment outcomes, rather than simply receiving information (Birt et al., 2020). A key contribution of this review lies in identifying contextual influences on testing, particularly in relation to the perceived fit between assessment and everyday functioning. Participants described how performance may be shaped by factors such as task familiarity, cultural background, environmental context, and emotional state. They also questioned the extent to which testing reflects everyday cognitive demands, describing a perceived misalignment between test performance and real-world functioning.

This suggests that contextual factors shape both performance and how individuals interpret the relevance and meaning of assessment outcomes.

Recent work similarly highlights the importance of person-centred approaches within memory clinic assessment, noting that diagnostic work-ups do not always reflect or respond to individuals' expectations and needs (Aspö et al., 2024). The present findings extend this by suggesting that individuals may themselves perceive a misalignment between testing and lived experience, particularly where tasks feel unfamiliar, decontextualised, or difficult to interpret.

Importantly, while existing research has often examined individual aspects of cognitive testing, such as anxiety, communication, or diagnostic disclosure, the current synthesis suggests that these factors are interconnected within the lived experience of assessment. Emotional responses, relational dynamics, trust, and contextual influences appear to operate together, shaping how testing is experienced and understood. As such, cognitive testing may be better conceptualised as a dynamic process rather than a discrete clinical event.

Overall, these findings indicate that cognitive testing is embedded within a broader process through which individuals make sense of cognitive change, negotiate identity, and seek clarity about their future. While testing may provide useful clinical information, its meaning and impact appear to be shaped by emotional, relational, and contextual factors that extend beyond the test itself. This highlights the importance of understanding cognitive assessment not only as a measurement tool, but as an experience actively interpreted within the context of individuals' lives.

These findings raise questions about how such experiences may influence the assessment process. Difficulties in understanding, perceived misalignment with everyday functioning, and relational dynamics may shape how individuals engage with testing, interpret their performance, and respond to results. This suggests that the impact of cognitive testing may

extend beyond measurement alone, potentially influencing how outcomes are understood, accepted, and integrated into individuals' broader sense of self and future.

Strengths and Limitations

This review has several methodological and conceptual strengths. To our knowledge, it is one of the first qualitative syntheses focused on older adults' experiences of cognitive testing across clinical contexts. A systematic thematic synthesis approach enabled integration of findings across diverse studies, generating higher-order analytical themes that provide a nuanced understanding of how cognitive testing is experienced. Emphasis on preserving participant quotations enhanced the depth and credibility of the synthesis. A structured audit of theme representation further supported confidence that findings were grounded across the dataset rather than disproportionately shaped by individual studies.

However, several limitations should be considered. The number of included studies was relatively small (n = 8), and studies varied in setting, population, and assessment context, which may limit transferability. As with all qualitative synthesis, analysis was limited to published data and therefore constrained by the depth and quality of reporting. Some studies provided richer accounts than others, which may have influenced the prominence of certain themes.

Only studies published in English were included, which may have resulted in the exclusion of relevant research and introduced language bias. While efforts were made to distinguish between participant quotations and author interpretations, the synthesis remains an interpretive process shaped by the reviewer's perspective. Reflexive strategies were used to enhance transparency and analytic rigour; however, alternative interpretations remain possible. Although PRISMA was used to guide reporting, use of ENTREQ may have enhanced transparency regarding qualitative synthesis and reflexive processes.

Implications for Practice and Future Research

The findings of this review have important implications for clinical practice in the assessment of cognitive functioning in later life. Cognitive testing was experienced as more than a purely technical procedure, instead described as an emotionally and relationally meaningful encounter. Participants reported anxiety, vulnerability, and self-evaluation, suggesting that testing may evoke concerns about identity, competence, and future decline. While attending to emotional impact and building rapport are recognised components of good clinical practice, these findings reinforce their importance within cognitive assessment, where testing may carry particular personal significance.

The way in which testing is introduced, conducted, and explained shapes experience. Participants described uncertainty about the purpose of assessment and difficulty understanding tasks and results. While clear communication is standard practice, these findings suggest a need for greater emphasis on how assessment is explained and contextualised. Preparation for assessment, including discussion of its purpose and potential outcomes, may not be consistently experienced and may warrant more explicit attention. Clear explanations of what tasks assess and how results are interpreted may support engagement, while feedback in revisitable formats (e.g., written summaries) may enhance understanding, particularly during emotionally salient consultations.

The findings also highlight the importance of contextual and cultural factors. Participants described how task design, language, educational background, and testing environments influenced both performance and perceived relevance. Tasks relying on unfamiliar formats or culturally specific knowledge may disadvantage some individuals, while factors such as time pressure, anxiety, and unfamiliar settings may further shape performance. Clinicians should adopt a flexible and context-sensitive approach, recognising that performance may be influenced by non-cognitive factors and may not directly reflect everyday functioning. These

findings suggest a need for greater attention to how individuals understand and relate to testing, including exploring perceived misalignment between test performance and everyday functioning.

These findings also have implications for the interpretation of cognitive test performance. Participants' accounts suggest that results are shaped not only by cognitive ability, but also by contextual, emotional, and relational factors. This reinforces the importance of interpreting performance within a broader clinical framework, incorporating history-taking, informant perspectives, and everyday functioning. Such an approach may be particularly relevant in complex presentations, where outcomes may not straightforwardly reflect underlying ability or test engagement.

Future research should explore how cognitive assessment can better reflect individuals' everyday functioning and lived experience, alongside further examination of contextual, cultural, and relational influences on performance and interpretation. While tasks resembling everyday activities may enhance perceived relevance, this does not necessarily determine predictive validity; however, perceived misalignment may influence how individuals interpret and engage with assessment outcomes.

Conclusion

This review highlights that cognitive testing in later life is experienced not as a purely objective procedure, but as a complex, emotionally and socially meaningful process. Older adults engage with testing as a way of making sense of cognitive change, navigating uncertainty, and evaluating their abilities in relation to identity and future risk.

Experiences of testing are shaped by interacting emotional, relational, and contextual factors, influencing both how assessment is experienced and how performance is interpreted. The review also identifies a perceived misalignment between standardised cognitive testing and

individuals' lived experience of everyday functioning. Participants' accounts suggest that contextual and situational factors shape how test demands are experienced and interpreted, contributing to perceptions that testing may not fully reflect real-world functioning. While such perceptions do not necessarily indicate a lack of validity, they may influence how individuals understand and engage with their performance and its outcomes.

Overall, cognitive testing should be understood as part of a broader, context-sensitive clinical process. Integrating test results with clinical history, informant perspectives, and everyday functioning may support more meaningful and person-centred assessment in later life.

References

Aspö, M., Sundell, M., Protsiv, M., Wiggenraad, F., Ryden, M., Mangialasche, F., Kivipelto, M., & Visser, L. N. (2024). The expectations and experiences of patients regarding the diagnostic workup at a specialized memory clinic: An interview study. *Health Expectations*, *27*(2), e14021.

Bharath, D., Jones, S., Bailey, J., Chandra, M., Vaitheswaran, S., Shaji, K., Goswami, S. P., Spector, A., Fisher, E., & Satish, G. (2023). Older adults' experience of neuropsychological assessments for dementia screening in South India: a qualitative study. *Wellcome Open Research*, *8*(421), 421.

Birt, L., Poland, F., Charlesworth, G., Leung, P., & Higgs, P. (2020). Relational experiences of people seeking help and assessment for subjective cognitive concern and memory loss. *Aging & mental health*, *24*(8), 1356-1364.

Booth, A. (2016). Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. *Systematic reviews*, *5*(1), 74.

Cahill, S. M., Gibb, M., Bruce, I., Headon, M., & Drury, M. (2008). I was worried coming in because I don't really know why it was arranged' The subjective experience of new patients and their primary caregivers attending a memory clinic. *Dementia*, *7*(2), 175-189.

CASP, C. (2018). CASP qualitative checklist. *Critical appraisal skills programme*, *590*.

Dubois, B., Padovani, A., Scheltens, P., Rossi, A., & Dell'Agnello, G. (2015). Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges. *Journal of Alzheimer's disease*, *49*(3), 617-631.

Dumassais, S., Grewal, K. S., Aubin, G., O'Connell, M., Phillips, N. A., & Wittich, W. (2024). Exploring the qualitative experiences of administering and participating in remote

research via telephone using the Montreal Cognitive Assessment-Blind: cross-sectional study of older adults. *JMIR Formative Research*, *8*(1), e58537.

Gruters, A. A., Christie, H. L., Ramakers, I. H., Verhey, F. R., Kessels, R. P., & de Vugt, M. E. (2021). Neuropsychological assessment and diagnostic disclosure at a memory clinic: A qualitative study of the experiences of patients and their family members. *The Clinical Neuropsychologist*, *35*(8), 1398-1414.

Krohne, K., Slettebø, Å., & Bergland, A. (2011). Cognitive screening tests as experienced by older hospitalised patients: a qualitative study. *Scandinavian journal of caring sciences*, *25*(4), 679-687.

Lindeberg, S., Samuelsson, C., & Müller, N. (2021). Experiencing dementia: How does assessment of cognition and language relate to daily life? *Dementia*, *20*(4), 1408-1424.

Lindeberg, S., Samuelsson, C., Stål, H., Björkegren, J., & Müller, N. (2022). Conceptualizations of aging and cognitive competence: Evaluations and accounts after cognitive testing. *Journal of Aging Studies*, *63*, 101074.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*.

Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., & Karagiannidou, M. (2016). *World Alzheimer Report 2016. Improving healthcare for people living with dementia: Coverage, Quality and costs now and in the future* Alzheimer's Disease International].

Rizkallah, J., Staios, M., Analytis, P., Kosmidis, M. H., March, E., & Stolwyk, R. J. (2024). Exploring how sociocultural factors affect the experience of completing neuropsychological assessments within older Greek-Australians. *Archives of Clinical Neuropsychology*, *39*(1), 65-77.

Robillard, J. M., Lai, J.-A., Wu, J. M., Feng, T. L., & Hayden, S. (2018). Patient perspectives of the experience of a computerized cognitive assessment in a clinical setting. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*, 297-303.

Rolke, K., Rosendahl, C., Weckbecker, K., Hanke, A., Wagner, M., Nissen, L., Reimer, L. M., Jonas, S., Schaper, P., & Thyrian, J. R. (2025). How do patients, medical assistants and physicians accept and experience tablet-based cognitive testing by medical assistants in general practice?-A qualitative study. *BMC primary care*, *26*(1), 174.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC medical research methodology*, *8*(1), 45.

Wilson, R. S., Segawa, E., Boyle, P. A., Anagnos, S. E., Hizel, L. P., & Bennett, D. A. (2012). The natural history of cognitive decline in Alzheimer's disease. *Psychology and aging*, *27*(4), 1008.

World Health Organization. (2023). *Summary report on the Regional workshop on the implementation of the global action plan on the public health response to dementia (2017–2025), Doha, Qatar, 16-18 September 2023*.

Chapter 2

Understanding Medical Symptom Validity Test Performance in Older Adults: Clinical Complexity and Interpretation

Prepared in accordance with the author requirements for the Journal of Applied Neuropsychology: Adult

Submit to Applied Neuropsychology: Adult

Plain Language Summary

Title: When Memory Test Results May Not Tell the Full Story in Older Adults: Understanding Performance Validity Testing

Background: Cognitive assessments are commonly used to understand memory and thinking difficulties in older adults. As part of these assessments, clinicians use tools to check whether a person's performance reflects their true abilities. These are known as performance validity tests. One widely used example is the Medical Symptom Validity Test (MSVT), which helps identify when results may not be reliable.

Traditionally, poor performance on these measures has been interpreted as a lack of effort or engagement. However, this assumption may not always be accurate, particularly in older adults, where memory problems, physical health conditions, and mental health difficulties are common.

Aims and Questions: This study aimed to examine how older adults perform on the Medical Symptom Validity Test (MSVT) in a real-world NHS clinical setting. It explored whether test results differ between people with dementia, mental health difficulties, and those experiencing cognitive concerns without a clear diagnosis.

The study also asked whether factors such as physical health, medication use, and mental health are linked to poorer performance, and whether MSVT results in these groups reflect true ability or the complexity of underlying clinical difficulties.

Methods: The sample included individuals with dementia, mental health conditions (such as anxiety and depression), and people experiencing cognitive difficulties without a clear diagnosis. MSVT performance was compared across these groups, and additional clinical factors, including physical health, medication use, and mental health, were examined.

Main Findings and Conclusions: The findings showed that the MSVT worked well for people with mental health difficulties, with almost all showing valid performance. However, results were more complex for people with dementia and those with unclear diagnoses. In these groups, some individuals performed poorly despite having genuine cognitive difficulties, suggesting the test may sometimes incorrectly classify valid performance as invalid.

People with unclear or mixed presentations showed particularly varied results, reflecting the complexity of their difficulties. Overall, the findings suggest that MSVT results should not be interpreted in isolation. Clinicians should consider the wider clinical picture, including memory problems, mental health, and physical health.

This study highlights the importance of using a flexible, person-centred approach when interpreting cognitive assessment results in older adults.

References:

Green, P. (2004). *Green's Medical Symptom Validity Test for Microsoft Windows: User's manual*. Green's Publishing.

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist, 32*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

Roor, J. J., Peters, M. J., Dandachi-FitzGerald, B., & Ponds, R. W. (2024). Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates. *Neuropsychology Review, 34*(1), 299–319.

Abstract

Performance validity tests (PVTs) are routinely used in neuropsychological assessment to evaluate the credibility of cognitive test performance, but interpretation in older adults is complicated by cognitive impairment and multimorbidity.

This cross-sectional study examined Medical Symptom Validity Test (MSVT) outcomes in a clinically diverse sample of older adults and explored associations with diagnostic group and clinical factors. The sample combined retrospectively identified cases (n = 69) and newly recruited participants (n = 25), yielding 94 participants classified as dementia (n = 27), mental health (n = 25), and Concerns Raised regarding Cognition but Not diagnosed with Dementia (CRC-ND; n = 42). MSVT classifications (Pass, Fail, General Memory Impairment Profile [GMIP]) were compared across groups, and logistic regression within the CRC-ND subgroup examined associations with polypharmacy, anticholinergic burden, and physical health.

MSVT performance differed by diagnostic group ($\chi^2(4, N = 94) = 38.63$, $p < .001$). The mental health group showed high specificity, while the dementia group demonstrated predominantly GMIP classifications with some Fail outcomes; clinical record review indicated these were not interpreted as invalid responding. The CRC-ND group showed marked heterogeneity across MSVT outcomes, and none of the examined clinical variables significantly predicted MSVT failure, although wide confidence intervals suggest limited power to detect smaller effects. Exploratory subgroup analyses revealed distinct patterns across clinically derived subtypes.

These findings suggest that MSVT outcomes in older adults reflect broader clinical complexity rather than effort alone, and that PVT interpretation requires a context-sensitive, formulation-based approach.

Keywords: performance validity testing (PVT); Medical Symptom Validity Test (MSVT); older adults; dementia; clinical complexity

Introduction

Accurate interpretation of cognitive test performance is central to neuropsychological assessment, underpinning clinical decision-making across the lifespan. This may be particularly challenging in older adults, where diagnostic complexity is common. A key aspect is the assessment of performance validity, referring to the extent to which test performance provides a credible and interpretable estimate of an individual's functioning. Without establishing performance validity, test results may be misleading and cannot reliably inform diagnosis or treatment planning.

Performance validity tests (PVTs) are routinely incorporated into neuropsychological assessments to evaluate the credibility of test performance. These measures are designed to detect suboptimal or non-credible responding, defined as performance that does not accurately reflect an individual's true cognitive ability due to reduced engagement, misunderstanding, or other non-volitional factors (Lippa, 2018). PVTs are conceptually distinct from symptom validity tests (SVTs), which assess the accuracy of self-reported symptoms rather than cognitive performance (Lippa, 2018). In clinical practice, validity assessment may involve both standalone measures and embedded indicators derived from standard neuropsychological tests (McGuire et al., 2019).

Despite their widespread use, interpretation of PVT performance remains complex. Meta-analytic findings indicate that PVT failure occurs in approximately 15–16% of individuals undergoing neuropsychological assessment, with variability depending on clinical context and patient characteristics (Roor et al., 2024). This highlights the importance of interpreting PVT outcomes within the broader clinical context, as performance may be influenced by multiple factors rather than effort alone (Lippa, 2018; Roor et al., 2024). Failure to account for these influences may increase the risk of misclassification. This risk is further compounded by the relatively low base rate of non-credible responding in many clinical settings, whereby even

small false-positive rates may result in disproportionate misclassification, with implications for clinical decision-making (Roor et al., 2024). Taken together, this literature suggests that PVT performance, particularly in older adult populations, may reflect broader clinical complexity rather than a straightforward indicator of effort.

Emerging evidence challenges the assumption that PVT failure necessarily reflects intentional underperformance. Elevated rates of failure have been observed across neurological, psychiatric, and functional conditions, even in the absence of clear external incentives (McWhirter et al., 2020). Individuals with genuine clinical conditions may demonstrate suboptimal performance due to cognitive impairment, fatigue, emotional distress, or difficulties engaging with testing demands. In addition, the clinical significance of PVT failure appears to depend on the number and pattern of failures observed, with single failures often associated with performance comparable to valid cases, and multiple failures more consistently linked to invalid responding (Jennette et al., 2022). These findings indicate that rigid application of single-test cut-offs may be insufficient and highlight the need for a more nuanced, context-sensitive approach to interpretation.

Interpretation of a failed PVT requires careful clinical judgement because different conclusions carry distinct consequences. Labelling failure as intentional underperformance risks stigmatizing patients and withholding care, while treating it as evidence that all other test results are uninterpretable risks discarding genuine cognitive findings. Performance validity depends on assumptions about effort, understood as the extent to which a person can and will engage with testing, and this construct is complex. Therefore, a PVT failure should prompt targeted investigation rather than immediate attribution to intentional underperformance. Clinicians should consider the number and pattern of failures, severity of cognitive impairment, and factors such as fatigue, mood, sensory or language barriers, medication effects, testing conditions, and collateral information. Where uncertainty persists, additional validity measures

or repeat assessment may be warranted. This is consistent with Lippa's (2018) multi-model framework, which emphasises that performance validity findings should be interpreted within the broader clinical context rather than in isolation.

These challenges are particularly pronounced in older adults. Cognitive impairment and neurodegenerative processes can affect performance on validity measures, increasing the likelihood of false positive classifications. Research in dementia clinic populations indicates that PVT performance varies with severity of cognitive impairment, and that standard cut-offs may require adjustment to maintain specificity, as greater impairment can reduce performance even on relatively simple tasks, increasing the likelihood of false positive classifications (Martin et al., 2022). Age-related effects have also been observed, with higher failure rates reported in older individuals across both embedded and standalone measures (Rai et al., 2024). These findings suggest that commonly used thresholds may not generalise well to older adults, particularly when base rates and individual clinical characteristics are not adequately considered.

A widely used tool designed to address these challenges is the Medical Symptom Validity Test (MSVT), a standalone PVT used in neuropsychological assessment (Green, 2004). It is a brief, computerised verbal forced-choice memory task comprising several subtests, including Immediate Recognition, Delayed Recognition, and Consistency indices, designed to assess response validity. The MSVT was developed to be relatively easy, such that even individuals with significant neurological impairment are expected to perform well on primary effort indices (Green et al., 2011; Singhal et al., 2009). This assumption underpins its use as an indicator of suboptimal effort, often interpreted as evidence of invalid responding.

Initial validation studies supported this interpretation, demonstrating high specificity in distinguishing valid from non-credible performance, including in individuals with severe

cognitive impairment. Sensitivity is more variable, with Singhal et al. (2009) reporting approximately 80%, while Green et al. (2011) focused on specificity.

To address concerns regarding false positive classifications, profile-based interpretation methods have been developed, including the Genuine Memory Impairment Profile (GMIP), which aims to differentiate poor effort from genuine cognitive impairment based on patterns of performance (Howe & Loring, 2009; Resch et al., 2022). These approaches reflect an attempt to account for the impact of neurological impairment on test performance.

However, subsequent research has raised concerns about interpretive clarity in clinical samples. Individuals meeting GMIP criteria often show similar performance on other PVTs and neuropsychological measures to those classified as invalid, raising questions about whether GMIP reliably distinguishes genuine impairment from invalid responding (Axelrod & Schutte, 2010; Reslan & Axelrod, 2017). Alternative classification approaches have also yielded inconsistent findings, with overlapping profiles observed across groups (Morin et al., 2019). These findings suggest that distinguishing between suboptimal effort and genuine impairment using the MSVT is more complex in practice than initially assumed.

This complexity is further amplified in older adults, where clinical factors may influence performance independently of effort. Psychiatric conditions, including depression and anxiety, have been associated with increased rates of PVT failure (Kim et al., 2024), while affective symptoms may also impact cognitive functioning and overlap with early neurocognitive presentations (Moore, 2023). In this context, performance patterns influenced by these factors may resemble those typically interpreted as invalid responding on PVTs, challenging the assumption that performance validity can be conceptualised as a simple binary distinction between valid and invalid responding, particularly in clinically complex populations.

Medication-related factors also play a significant role. Anticholinergic medications have been associated with cognitive impairment and increased risk of dementia in older adults (Bishara et al., 2020), and these effects may be compounded by polypharmacy, which is highly prevalent and linked to adverse outcomes including cognitive decline and functional impairment (Masnoon et al., 2017). Together, these factors further highlight the complexity of cognitive performance in older adults.

Against this background, interpretation of MSVT performance in older adults becomes particularly challenging. Performance on validity measures may reflect not only effort, but also the combined effects of cognitive impairment, mental health difficulties, medication burden, and physical health conditions. These uncertainties raise important questions regarding the extent to which MSVT failure can be interpreted as evidence of invalid responding in clinically complex populations.

Despite widespread use and strong psychometric foundations, research examining how the MSVT functions in real-world older adult populations remains limited. Much of the existing literature has been conducted in controlled or simulation-based contexts, which may not adequately capture the complexity of routine clinical presentations. In particular, there is limited research examining MSVT performance in older adults with co-occurring cognitive, psychiatric, and physical health conditions. This gap is especially relevant within NHS mental health services, where clinical complexity is high and external incentives for invalid responding are typically minimal.

These uncertainties have important clinical implications, as misclassification may lead to inaccurate diagnosis and inappropriate treatment planning. Research is therefore needed to clarify whether MSVT outcomes in these contexts reflect performance validity or broader clinical complexity, consistent with Lippa's (2018) multi-model framework which emphasises

contextual interpretation of validity findings. This study addresses this gap by examining MSVT performance across diagnostically diverse older adult groups and investigating the extent to which outcomes reflect performance validity versus broader clinical complexity.

Study Significance and Aims

In response to the limitations identified in the existing literature, the present study aims to:

- Evaluate MSVT classifications (Pass, Fail, GMIP) across older adults with dementia, mental health conditions (e.g., depression/anxiety), and Concerns Raised regarding Cognition but Not diagnosed with Dementia (CRC-ND).
- Investigate clinical factors associated with MSVT failure, with particular focus on physical health comorbidity, polypharmacy, anticholinergic burden, and available psychiatric symptom data.
- Assess the interpretive validity of MSVT classifications within a complex, real-world clinical population.

Research Questions

1. What is the MSVT's specificity in people with dementia, and what proportion are classified as having a GMIP?
2. What is the MSVT's specificity in older adults with a mental health condition?
3. How do MSVT validity classifications (Pass, Fail, GMIP) differ across older adults with dementia, mental health conditions, and CRC-ND?
4. What clinical factors (e.g., depression, anxiety, polypharmacy, anticholinergic burden, physical health) are associated with MSVT failures in older adults with CRC-ND?
5. To what extent do individuals with CRC-ND and dementia display MSVT profiles indicative of valid or invalid responding?

Hypotheses

Based on existing literature, the following hypotheses are proposed:

- *Dementia group*: Individuals are expected to demonstrate GMIP profiles consistent with genuine memory impairment. However, some variability is anticipated, including Fail classifications, given evidence that performance validity tests may yield false positives in individuals with significant cognitive impairment.
- *Mental health group* (without significant cognitive concerns): Individuals are expected to pass the MSVT with high specificity, indicating valid performance.
- *CRC-ND group*: No specific directional hypotheses are proposed due to the clinical heterogeneity and diagnostic uncertainty within this group; analyses will therefore be exploratory.

Exploratory Analyses

Exploratory analyses will examine factors associated with MSVT failure within the CRC-ND group, focusing on polypharmacy, anticholinergic burden, physical health conditions, and broader clinical complexity (McWhirter et al., 2020; Roor et al., 2024).

Methods

Design

The present study employed a cross-sectional design using both retrospective and newly recruited data to examine performance on the MSVT across diagnostically diverse older adult populations within a real-world NHS clinical setting. The study aimed to explore whether MSVT performance varied according to diagnostic group and to examine the influence of clinical factors, including mental health, medication burden, and physical health comorbidity.

Participants

The dataset comprised 94 participants, including retrospectively identified cases (n = 69) and newly recruited participants (n = 25). Participants were categorised into three diagnostic groups: dementia (n = 27), mental health (n = 25), and CRC-ND (n = 42).

The retrospective sample was drawn from routine clinical records within an NHS Psychological Therapies for Older People (PTOP) service. Participants were referred for neuropsychological assessment due to suspected cognitive impairment in the context of diagnostic uncertainty (e.g., query of dementia) often alongside complex presentations. A small number of referrals also had comorbid neurological conditions noted in the record (e.g., acquired brain injury [ABI]), although these were not the primary reason for referral. Participants were classified into dementia and CRC-ND groups based on clinician-determined outcomes documented in referral information, clinical assessment reports, and outcome letters.

In some cases, individuals initially classified within the CRC-ND group were later diagnosed with dementia following further assessment. However, participants were classified according to diagnostic outcome at the time of MSVT assessment to preserve ecological validity and reflect clinical decision-making.

The newly recruited sample comprised older adults referred to community mental health and psychological therapy services for mental health difficulties (e.g., anxiety, depression, trauma-related presentations). Participants did not have a dementia diagnosis at the time of assessment and did not present with cognitive concerns beyond those typically expected in this population.

In the retrospective sample, the MSVT was administered as part of routine clinical practice and was not standardised across referrals, instead being used selectively based on clinician judgement, typically in cases of diagnostic uncertainty or concerns regarding the validity of cognitive presentation. MSVT performance may have contributed to diagnostic formulation,

introducing potential circularity whereby test outcomes influenced the clinical classifications used in this study. This reflects routine clinical practice and supports the study's ecological validity.

Inclusion criteria for the newly recruited sample included age ≥60 years, ability to engage in testing, and capacity to provide informed consent. Exclusion criteria included known learning disability, neurological conditions preventing meaningful participation, inability to communicate in English, or lack of capacity to consent.

Measures

Demographic and Clinical Variables

Demographic data collected included age, gender, ethnicity, education, and diagnostic information. Socioeconomic status was estimated using the Scottish Index of Multiple Deprivation (SIMD), obtained from participants' postcodes, which were pseudonymised following data entry.

Medical Symptom Validity Test (MSVT)

The MSVT is a computerised performance validity test designed to assess the credibility of cognitive test performance (Green, 2004). It includes indices of Immediate Recognition (IR), Delayed Recognition (DR), and Consistency (CNS), alongside more effortful memory measures, including Paired Associates (PA) and Free Recall (FR). The MSVT has demonstrated good sensitivity and specificity in distinguishing valid from non-credible performance across clinical populations (Green, 2004). However, individuals with severe cognitive impairment may fail even very easy recognition subtests despite genuine effort, introducing the potential for false positive classifications. Profile-based interpretation, based on discrepancies between easier (IR, DR, CNS) and more difficult (PA, FR) subtests, has been shown to improve specificity in such populations (Green et al., 2011).

MSVT outcome classifications (Pass, Fail, GMIP) were derived from raw scores using established scoring criteria. Failure was defined as performance at or below the standard cut-off (≤85%) on one or more primary validity indices (IR, DR, CNS), in combination with response patterns indicative of insufficient or inconsistent effort. These included a reduced easy-hard discrepancy (<20 percentage points) or atypical score patterns (e.g., PA ≤ FR).

GMIP classifications were assigned where low scores on primary validity indices occurred alongside a pattern consistent with clinically significant memory impairment. This was defined by a preserved easy-hard discrepancy (≥20 percentage points), with relatively higher performance on IR, DR, and PA compared to FR, consistent with MSVT interpretive guidelines and profiles observed in severe cognitive impairment (Green et al., 2011).

All classifications were derived manually due to the absence of software-generated outputs within the dataset. Interpretation was informed by MSVT manual guidelines (Green, 2004) and applied consistently across cases.

Mental Health

Mental health symptoms were assessed in the newly recruited sample only using the Hospital Anxiety and Depression Scale (HADS), as this information was not routinely available in the retrospective sample. The HADS, by Snaith and Zigmond (1986) is a 14-item self-report measure of anxiety and depressive symptoms. Total anxiety and depression scores were calculated, with higher scores indicating greater symptom severity. The HADS was selected because it is a routinely used and well-established clinical measure of anxiety and depression in older adults.

Medication and Anticholinergic Burden

Medication data at time of testing were extracted from clinical records (retrospective sample) and supplemented by self-report and records where available (newly recruited sample). Only

regularly prescribed medications were included, and acute or short-term medications (e.g., antibiotics) were excluded where appropriate. Self-medicated, over-the-counter, and complementary medications were not routinely recorded and were therefore not included in medication counts.

A total medication count was calculated for each participant, and polypharmacy was defined as the concurrent use of five or more medications (Masnoon et al., 2017).

Anticholinergic burden was assessed using the Medichec online tool (Bishara et al., 2020), which assigns each medication a score from 0 (none) to 3 (high) based on its anticholinergic effect on cognition (AEC). Scores are summed to produce a total AEC score for each participant, which can exceed 3 due to cumulative effects from multiple medications (e.g., amitriptyline [3] and tolterodine [2] = 5). Higher scores indicate greater potential impact on cognitive functioning. AEC scores were treated as a continuous variable and included to examine their relationship with MSVT performance.

Physical Health

Physical health conditions were identified through clinical record review (retrospective sample) and self-report in the newly recruited sample, cross-checked against records where possible. Chronic and clinically significant conditions were coded using predefined categories (diabetes, kidney disease, heart disease, autoimmune conditions, chronic obstructive pulmonary disease, osteoarthritis, COVID-19 history, cancer, and hypothyroidism), consistent with prior research (Moore, 2023).

Conditions were coded dichotomously (*present/absent*), with an additional “other” category capturing relevant conditions not represented. A total physical health count (range: 0–10) was calculated by summing conditions across categories, providing a continuous index of overall health burden (multimorbidity).

Pain-related conditions were captured within the broader physical health variables (e.g., osteoarthritis). A more detailed pain variable was available for the newly recruited sample only and, where recorded, was coded using IASP-informed definitions (Raja et al., 2020). The pain variable was included in recognition of the potential influence of chronic pain on cognitive functioning and test performance in older adults. However, this variable was not included in the present analyses due to limited availability across the full dataset.

Procedure

For the retrospective cohort, clinical data were extracted from electronic and paper-based PTOP records using a structured data extraction protocol. Clinical variables (including diagnosis, medication, and physical health conditions) were recorded to reflect participants' presentation at the time of MSVT administration. This approach was adopted to ensure that MSVT performance was interpreted within the clinical context at time of testing, rather than being influenced by diagnostic or treatment changes occurring after assessment.

For the newly recruited cohort, participants were recruited through clinical services and completed the MSVT as part of the study protocol, independent of their routine clinical care.

All data were anonymised prior to analysis. For retrospective data, identifiable information was accessed solely for data extraction and record linkage and was replaced with a unique study ID at the earliest opportunity. Identifiable data were stored separately on secure NHS systems, with only pseudonymised data used for analysis.

Ethical approval was obtained through NHS research governance procedures (IRAS ID: 363146; REC reference: 25/SC/0385), with a favourable ethical opinion granted by the Hampshire NHS Research Ethics Committee. Caldicott Guardian approval was obtained to permit access to patient-identifiable data (Caldicott reference: Z6723578), in line with

Caldicott principles. NHS Lanarkshire Research and Development approval was also granted (R&D ID: L25097).

All procedures adhered to the UK Policy Framework for Health and Social Care Research, NHS data governance requirements, and the UK General Data Protection Regulation (UK GDPR) and Data Protection Act 2018. Data were stored securely on NHS and University systems, with access restricted to authorised members of the research team.

Data Preparation and Diagnostic Classification

A structured data preparation process was undertaken to ensure consistency in clinical information extracted from records. Presenting concern data were originally recorded as free-text entries within the MSVT database and varied considerably in format and terminology.

To address this, a single variable (*Presenting_Concern*) was created through systematic data cleaning and coding. Free-text entries were reviewed, standardised, and corrected for minor typographical errors while preserving clinical meaning, then categorised into predefined higher-order groups (e.g., mental health, cognitive concern, dementia, neurological, medical, assessment, mixed/unclear). Where multiple concerns were recorded, a primary concern was identified, with unclear cases classified as mixed/unclear.

A second variable *(Diagnosis_Update_Clinical_Record_Description)* was derived from post-assessment clinical records, including PTOP documentation and correspondence to referrers. This reflected the clinical formulation at the time of assessment, incorporating diagnostic conclusions, comorbidities, and contributing factors (e.g., functional overlay). Where available, updated diagnoses were recorded to capture diagnostic progression.

These variables supported organisation and interpretation of the dataset but were not included in statistical analyses; instead, they informed classification into the primary diagnostic groups.

Participants were classified into three groups based on clinical formulation at the time of MSVT administration: (1) dementia, (2) mental health conditions, and (3) CRC-ND. This approach preserved ecological validity and avoided retrospective bias.

Where criteria for dementia were not met in the context of diagnostic uncertainty, cases were assigned to the CRC-ND group. Such presentations typically reflect mild cognitive impairment or require monitoring (e.g., 6–12 months) to assess progression.

Clinical records were reviewed to identify indications of invalid responding or reduced effort during assessment. Where available, clinician reports were examined for references to effort, validity of performance, and interpretation of MSVT findings.

Statistical Analysis

An a priori power analysis was conducted using G*Power to estimate the sample size required to detect group differences in MSVT classifications using chi-square tests. For a three-group comparison with a medium effect size (Cramér's V = 0.30), $\alpha = .05$, and power $(1–\beta) = .80$, the required sample size was n = 88.

All analyses were conducted in IBM SPSS Statistics. Data were screened for accuracy, completeness, and distributional assumptions prior to analysis. Descriptive statistics summarised participant characteristics (age, gender, education, polypharmacy, AEC, and physical health comorbidity) and MSVT outcomes (Pass, Fail, GMIP). Frequencies and percentages are reported for categorical variables, with means and standard deviations for continuous variables. Percentages are based on valid cases. Visual summaries were used to aid interpretation.

Primary analyses

Group differences in MSVT classifications across diagnostic groups (dementia, mental health, CRC-ND) were examined using chi-square tests of independence. When expected cell counts

were small, Fisher–Freeman–Halton exact tests were used. Effect sizes for chi-square tests are reported as Cramér's V.

MSVT specificity was estimated in the mental health group as the proportion classified as Pass. In the dementia group, MSVT classifications were examined descriptively, with particular attention to GMIP outcomes.

Exploratory analyses

Within the CRC-ND group, associations between MSVT failure (Pass = 0; Fail = 1) and clinical variables were examined using logistic regression. Predictor variables were polypharmacy status, AEC score, and number of physical health conditions. Univariable logistic regression models were run for each predictor, followed by a multivariable model including all predictors. Model fit was assessed with the Hosmer–Lemeshow test and by inspecting classification tables. Where events-per-variable (EPV) were low, results are treated as exploratory and interpreted with caution.

Given heterogeneity within the CRC-ND group, descriptive subgroup analyses were conducted (MCI only; MCI with functional overlay; MCI with later dementia diagnosis; functional neurological disorder; functional cognitive disorder; and ABI [static cognitive impairment]). Frequencies and percentages of MSVT outcome classifications, and mean (SD) MSVT subtest scores, are reported for each subtype.

All tests were two-tailed with $\alpha < .05$. Missing data were handled using pairwise exclusion unless otherwise stated. Key estimates (e.g., odds ratios) are reported with 95% confidence intervals to convey precision.

Results

The sample comprised 94 participants, including 69 retrospective and 25 newly recruited cases. Participants were classified into three diagnostic groups: dementia (n = 27, 28.7%), mental health (n = 25, 26.6%), and CRC-ND (n = 42, 44.7%).

The sample included 37 males (39.4%) and 57 females (60.6%), with a mean age of 69.12 years (SD = 7.64; range = 45–84), consistent with an older adult cohort within memory assessment services. Education data were available for 75 participants and indicated a broad range of attainment. A substantial proportion (41.3%) had low educational attainment (<11 years), with smaller proportions completing secondary or higher education, suggesting variability in cognitive reserve.

Socioeconomic status, indexed using the SIMD, indicated that 29.8% of participants were in the most deprived quintile, with the remainder distributed across other quintiles, suggesting notable socioeconomic disadvantage within the sample.

Indicators of clinical complexity were common within the sample. Polypharmacy (≥5 medications) was present in 69.1% of participants, reflecting a substantial medication burden. The mean number of physical health conditions was 2.11 (SD = 1.55; range = 0–7), indicating moderate multimorbidity. Anticholinergic burden (AEC score) had a mean of 1.78 (SD = 1.69; range = 0–7), reflecting variability in medication-related cognitive risk.

Across the full sample, MSVT outcomes were distributed across Pass (50.0%), Fail (19.1%), and GMIP (30.9%) classifications. Given the clinical heterogeneity of the sample, subsequent analyses focused on differences between diagnostic groups.

Mean (SD) MSVT subtest scores across diagnostic groups are presented in Table 4, providing descriptive context for performance patterns. Individuals with dementia showed consistently lower scores across all indices relative to the mental health group, consistent with the expected

impact of cognitive impairment. In contrast, the mental health group demonstrated uniformly high performance, indicative of intact effort and cognitive functioning. The CRC-ND group showed intermediate scores with greater variability, reflecting its clinical heterogeneity and mixed cognitive presentations.

Table 4. *Mean (SD) MSVT Subset Scores by Diagnostic Group*

Diagnostic Group	n	IR	DR	CNS	PA	FR
Dementia	27	79.63 (15.00)	75.56 (12.51)	71.48 (13.57)	58.52 (16.80)	21.85 (16.24)
Mental Health	25	97.60 (6.31)	96.40 (9.19)	96.00 (7.36)	94.80 (10.46)	70.20 (15.31)
CRC-ND	42	83.57 (15.90)	78.57 (18.85)	76.43 (18.98)	70.71 (24.83)	39.88 (22.05)

Note. *IR* = Immediate Recognition; *DR* = Delayed Recognition; *CNS* = Consistency; *PA* = Paired Associates; *FR* = Free Recall; *CRC-ND* = Concerns Raised regarding Cognition but Not diagnosed with Dementia; Values are presented as mean (SD).

MSVT Specificity in the Dementia Group

No concerns regarding invalid responding were documented within the dementia group, based on clinician comments regarding effort and performance validity.

Within the dementia group (n = 27), 6 participants (22.2%) achieved a Pass classification, 4 (14.8%) were classified as Fail, and the majority (63.0%) met criteria for a General Memory Impairment Profile (GMIP).

When considering Pass classifications alone as a conventional estimate of specificity, the MSVT showed a low pass rate (22.2%). However, this does not account for GMIP classifications, which reflect performance patterns consistent with genuine memory impairment. When Pass and GMIP classifications are considered together, the majority of

participants demonstrated performance consistent with valid responding, albeit affected by cognitive impairment.

The presence of Fail classifications (14.8%) indicates that some individuals with dementia were classified as demonstrating invalid performance. However, no concerns regarding invalid responding were documented in these cases, and performance was interpreted as consistent with the individual's cognitive profile. In the absence of documented concerns regarding effort, these findings may reflect reduced test specificity in this population rather than clear evidence of invalid responding.

MSVT Specificity in the Mental Health Group

Within the mental health group (n = 25), 24 participants (96.0%) achieved a Pass classification, 1 (4.0%) was classified as Fail, and no participants met criteria for a GMIP. No concerns regarding engagement or effort were observed in this case.

Using Pass classifications as an estimate of specificity, the MSVT demonstrated high specificity (96.0%) in this group. The predominance of Pass classifications, alongside the absence of GMIP and minimal Fail classifications, indicates that the MSVT performed as expected, with little evidence of misclassification.

MSVT Classification in the CRC-ND Group

Within the CRC-ND group (n = 42), 17 participants (40.5%) achieved a Pass classification, 13 (31.0%) were classified as Fail, and 12 (28.6%) met criteria for a GMIP.

This pattern reflects a heterogeneous distribution of MSVT outcomes within this group, with a substantial proportion of participants classified across all outcome categories.

Group Differences in MSVT Classifications

A chi-square test of independence was conducted to examine differences in MSVT classifications across diagnostic groups. The analysis revealed a significant association

between diagnostic group and MSVT outcome, $\chi^2(4, N = 94) = 38.63$, $p < .001$, with a large effect size (Cramér's V = .45).

Inspection of the crosstabulation (see Figure 3) indicated that individuals with dementia were most likely to be classified as GMIP, whereas those in the mental health group were predominantly classified as Pass. The CRC-ND group showed a more mixed profile across classifications, reflecting greater heterogeneity. These findings indicate that MSVT outcomes vary meaningfully across diagnostic groups.

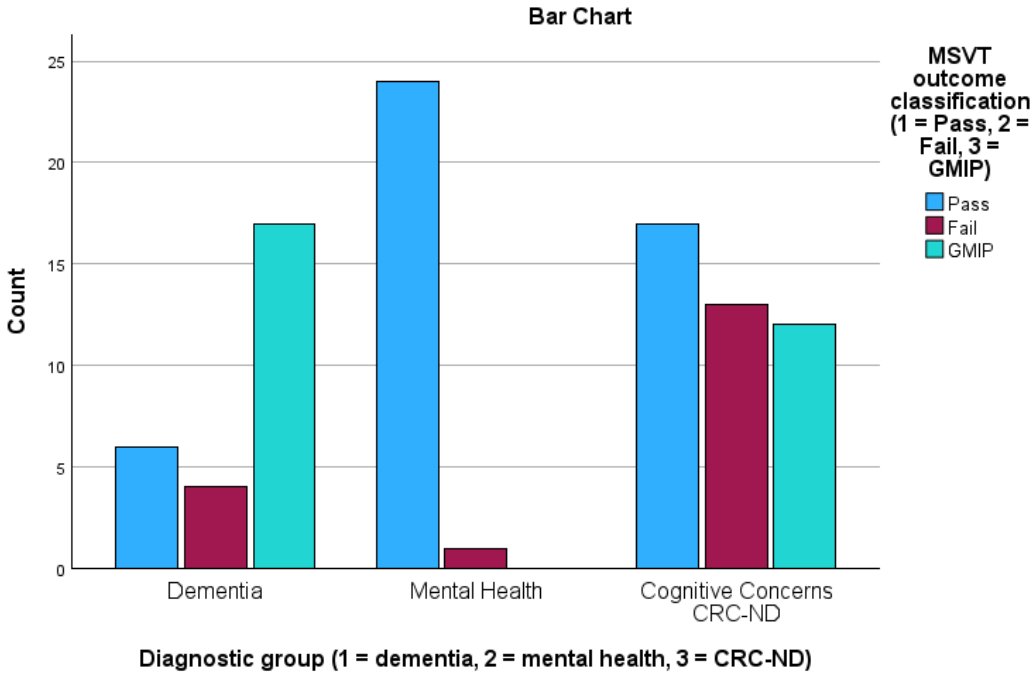

Figure 3. *Distribution of MSVT outcome classifications across diagnostic groups*

Clinical Factors Associated with MSVT Failure in CRC-ND

To explore factors associated with MSVT performance in a diagnostically uncertain group, logistic regression analyses were conducted within the CRC-ND subgroup. MSVT outcomes were recoded to distinguish between valid performance (Pass and GMIP = 0) and invalid performance (Fail = 1).

Univariable analyses indicated that anticholinergic burden (AEC score), polypharmacy status, and number of physical health conditions were not significantly associated with MSVT failure. Anticholinergic burden was not a significant predictor (B = -0.19, SE = 0.23, p = .428, OR = 0.83, 95% CI [0.53, 1.31]), nor were polypharmacy status (B = -0.68, SE = 0.72, p = .346, OR = 0.51, 95% CI [0.13, 2.07]) or number of physical health conditions (B = -0.46, SE = 0.32, p = .153, OR = 0.63, 95% CI [0.34, 1.19]).

A multivariable logistic regression model including AEC score, polypharmacy status, and number of physical health conditions was conducted to examine their combined effects, given their potential overlap and clinical relevance.

None of the predictors were significantly associated with MSVT failure: anticholinergic burden (B = -0.21, SE = 0.26, p = .402, OR = 0.81, 95% CI [0.49, 1.33]), polypharmacy status (B = 0.08, SE = 0.89, p = .929, OR = 1.08, 95% CI [0.19, 6.16]), and number of physical health conditions (B = -0.48, SE = 0.36, p = .181, OR = 0.62, 95% CI [0.31, 1.25]).

Model fit was acceptable based on the Hosmer–Lemeshow test ($\chi^2(7) = 8.69$, p = .276), indicating no evidence of poor fit. Wide confidence intervals across predictors likely reflect limited statistical power within the CRC-ND subgroup and warrant cautious interpretation.

Taken together, these findings indicate that MSVT failure within the CRC-ND group was not explained by the clinical variables examined.

To further explore heterogeneity within the CRC-ND group, exploratory subgroup analyses were conducted based on clinically derived diagnostic profiles. Mean (SD) MSVT subtest scores by CRC-ND subtype are presented in Table 5.

Table 5. *Mean (SD) MSVT Subset Scores by CRC-ND Subtype (n=42)*

Subtype	n	IR Mean (SD)	DR Mean (SD)	CNS Mean (SD)	PA Mean (SD)	FR Mean (SD)
ABI (Static Cognitive Impairment)	2	75.00 (28.28)	70.00 (21.21)	70.00 (14.14)	70.00 (28.28)	37.50 (17.68)
MCI + Functional Overlay	21	88.33 (14.86)	82.38 (16.10)	81.67 (16.07)	74.29 (24.61)	43.10 (23.64)
MCI + Later Dementia Diagnosis	6	85.00 (12.65)	77.50 (23.82)	74.17 (19.60)	75.00 (24.29)	31.67 (21.37)
Functional Neurological Disorder (FND)	2	75.00 (14.14)	67.50 (3.54)	67.50 (17.68)	65.00 (7.07)	60.00 (14.14)
MCI only	7	84.29 (13.67)	84.29 (23.17)	81.43 (20.15)	70.00 (30.00)	36.43 (25.77)
Functional Cognitive Disorder (FCD)	4	63.75 (14.93)	60.00 (13.54)	51.25 (19.74)	50.00 (24.50)	32.50 (8.66)

Note. *IR* = Immediate Recognition; *DR* = Delayed Recognition; *CNS* = Consistency; *PA* = Paired Associates; *FR* = Free Recall; *SD* = Standard Deviation; *ABI* = Acquired Brain Injury; *MCI* = Mild Cognitive Impairment; *FND* = Functional Neurological Disorder; *FCD* = Functional Cognitive Disorder.

As shown in Table 5, subgroups characterised by functional presentations, particularly functional cognitive disorder, demonstrated lower scores across MSVT indices, whereas MCI-

only and MCI with functional overlay groups showed relatively higher and more variable performance. These patterns should be interpreted cautiously given the small subgroup sizes.

As shown in Table 6, distinct patterns of MSVT classification were observed across clinically derived subtypes. Individuals with FCD, FND (excluding FCD), and ABI demonstrated high rates of MSVT failure, with all individuals in the functional neurological and ABI subgroups and the majority in the functional cognitive subgroup classified as Fail.

In contrast, those with MCI-only presentations were more likely to achieve a Pass classification (57.1%) and showed relatively low rates of GMIP (14.3%). Participants with a later dementia diagnosis exhibited a high proportion of GMIP outcomes (50.0%), consistent with emerging neurodegenerative impairment.

The MCI with functional overlay subgroup, which represented the largest proportion of the CRC-ND sample (n = 21), showed a heterogeneous distribution across Pass (52.4%), Fail (14.3%), and GMIP (33.3%) classifications, reflecting the clinical complexity of this group.

Table 6. *MSVT Outcomes by CRC-ND Subtype (n=42)*

Subtype	n	Pass, n (%)	Fail, n (%)	GMIP, n (%)
MCI + Functional Overlay	21	11 (52.4)	3 (14.3)	7 (33.3)
MCI + Later Dementia Diagnosis	6	2 (33.3)	1 (16.7)	3 (50.0)
MCI only	7	4 (57.1)	2 (28.6)	1 (14.3)
Functional Neurological Disorder	2	0 (0.0)	2 (100.0)	0 (0.0)
Functional Cognitive Disorder	4	0 (0.0)	3 (75.0)	1 (25.0)
ABI (Static Cognitive Impairment)	2	0 (0.0)	2 (100.0)	0 (0.0)
Total	**42**	**17 (40.5)**	**13 (31.0)**	**12 (28.6)**

Note. Percentages are calculated within subtype; *GMIP* = Genuine Memory Impairment Profile; *MCI* = Mild Cognitive Impairment; *ABI* = Acquired Brain Injury; *FND* = Functional Neurological Disorder; *FCD* = Functional Cognitive Disorder.

Review of available clinical records within the CRC-ND group indicated variability in the interpretation of MSVT failure. In some cases, clinicians described reduced or inconsistent engagement, whereas in others findings were considered ambiguous or no concerns were documented.

Together, Tables 5 and 6 illustrate the variability in both MSVT performance and classification outcomes across CRC-ND subtypes.

Summary of Findings

Overall, MSVT performance differed significantly across diagnostic groups, with a strong association between diagnostic category and test outcome. High specificity was observed within the mental health group, with most participants achieving a Pass classification (96.0%)

and minimal evidence of misclassification. In contrast, individuals within the dementia group were predominantly classified as exhibiting a General Memory Impairment Profile (GMIP; 63.0%), consistent with genuine cognitive impairment. Although a small proportion were classified as Fail, clinical record review indicated no concerns regarding invalid responding, suggesting these outcomes may reflect reduced MSVT specificity in this population.

The CRC-ND group demonstrated a heterogeneous pattern of MSVT outcomes, with participants distributed across Pass (40.5%), Fail (31.0%), and GMIP (28.6%) classifications. Logistic regression analyses indicated that MSVT failure within this group was not significantly associated with anticholinergic burden, polypharmacy, or number of physical health conditions, suggesting that performance was not explained by any single clinical factor.

Exploratory subgroup analyses provided further insight into this heterogeneity, revealing distinct patterns across clinically derived subtypes. Functional neurological, functional cognitive, and ABI subgroups demonstrated higher rates of MSVT failure, whereas MCI-only presentations were more likely to achieve a Pass classification and showed lower rates of GMIP. Participants with a later dementia diagnosis showed higher rates of GMIP, consistent with emerging neurodegenerative processes. In contrast, the MCI with functional overlay subgroup demonstrated a mixed profile across outcome categories, reflecting the complexity of this group. Clinical record review further highlighted variability in the interpretation of MSVT failure within the CRC-ND group.

Taken together, these findings highlight the complexity of interpreting MSVT performance in diagnostically uncertain populations and suggest that underlying clinical presentation, rather than any single factor, may meaningfully influence test outcomes.

Discussion

The present study examined MSVT performance in a clinically diverse sample of older adults, exploring associations with diagnostic group and broader clinical factors. The findings contribute to a growing body of literature indicating that PVT outcomes in older adult populations require careful, context-sensitive interpretation, rather than being assumed to reflect invalid responding alone.

MSVT performance differed significantly across diagnostic groups, with a strong association between diagnostic category and test outcome. The most notable finding was the marked heterogeneity observed within the CRC-ND group, which showed substantial variability across Pass, Fail, and GMIP classifications. This pattern was supported by subtest scores, with individuals with dementia demonstrating consistently lower performance, the mental health group performing near ceiling levels, and the CRC-ND group showing intermediate and more variable performance. Together, these findings suggest that MSVT outcomes in clinically complex populations are not readily captured by simple binary interpretations of validity.

More broadly, these findings contribute to ongoing debate regarding the interpretation of PVT outcomes. Although such measures have traditionally been conceptualised as indicators of effort, there is increasing recognition that cognitive performance in older adults is shaped by interacting neurological, psychological, and medical factors. Lippa's (2018) multi-model framework for performance validity assessment emphasises the integration of multiple sources of information, including test performance, behavioural observation, and clinical context, rather than reliance on single test outcomes (Bush et al., 2005). The present findings align with this perspective, highlighting the need for a more nuanced understanding of MSVT performance in diagnostically complex populations.

MSVT Performance in the Mental Health Group

The MSVT demonstrated high specificity within the mental health group, with most participants achieving a Pass classification. This is consistent with evidence that MSVT validity indices are generally robust to psychiatric presentations and can reliably distinguish valid from invalid performance (Cerny et al., 2022), supporting its use in mental health contexts where false positive classifications appear low.

This is particularly important, given that symptoms of anxiety and depression are often associated with reduced cognitive efficiency, including slowed processing speed, attentional disruption, and subjective cognitive difficulties. Meta-analytic evidence indicates that increasing severity of depression is associated with impairments in episodic memory, executive functioning, and processing speed (McDermott & Ebmeier, 2009).

Notably, the mental health group did not include individuals with significant cognitive concerns, representing a relatively "clean" psychiatric sample. Despite this, the findings suggest that mood-related difficulties alone do not necessarily result in elevated rates of MSVT failure. This may increase clinicians' confidence in interpreting MSVT performance in mental health populations where cognitive impairment is not a prominent feature, in contrast to the more heterogeneous patterns observed in groups with overlapping cognitive and functional presentations.

MSVT Performance in the Dementia Group

MSVT performance in the dementia group highlights important considerations for interpretation in clinically complex populations. Although a proportion of individuals were classified as Fail, clinical record review indicated that these outcomes were not interpreted as reflecting invalid responding. Instead, performance was understood within the context of each individual's cognitive profile and described as an accurate representation of their abilities. This

suggests that MSVT outcomes are not interpreted in isolation in clinical practice, but are integrated with broader neuropsychological findings and clinical judgement.

In this context, Fail classifications outside of the GMIP may not indicate invalid responding, but may instead reflect the impact of cognitive impairment on task performance. This raises important questions regarding MSVT specificity in dementia populations, suggesting that reduced performance may be misclassified as invalid when considered without clinical context. While reduced effort cannot be definitively excluded, there was no clinical evidence to support this interpretation in the current sample.

This pattern is consistent with concerns regarding false positive classifications in individuals with genuine cognitive impairment (Singhal et al., 2009). Individuals with severe memory impairment may fail validity indices despite providing genuine effort, particularly when standard cut-offs are applied. Base rate analyses further suggest that, in dementia evaluations where invalid performance is relatively uncommon, the likelihood that a failed PVT reflects true invalidity may be substantially reduced (Gaudet et al., 2022).

The present findings extend this literature by demonstrating that, even within a clinically defined dementia group, MSVT outcomes are not uniformly captured by GMIP classifications. This supports critiques that MSVT algorithms may not always differentiate between poor effort and genuine cognitive impairment (Axelrod & Schutte, 2010), and aligns with evidence that PVT specificity varies with level of impairment, with standard cut-offs potentially requiring adjustment in dementia populations (Martin et al., 2022).

Taken together, these findings highlight the need for cautious, context-sensitive interpretation of MSVT outcomes in dementia populations. Reliance on classification alone may risk misinterpretation, and a formulation-based approach to validity assessment may be more appropriate in older adults with complex presentations.

Heterogeneity and Diagnostic Uncertainty in the CRC-ND Group

A key contribution of the present study is the pattern of findings within the CRC-ND group. This group demonstrated substantial heterogeneity in MSVT outcomes, with participants distributed across Pass, Fail, and GMIP classifications. This variability reflects the diagnostic uncertainty and clinical complexity characteristic of this population and suggests that MSVT performance in such cases may not be adequately captured by binary interpretations of valid versus invalid responding.

A key consideration highlighted by these findings is the evolving nature of diagnosis in clinical practice. Participants were classified according to diagnostic status at the time of MSVT administration, reflecting the information available to clinicians during assessment. However, exploratory subgroup analyses identified individuals who later received a diagnosis of dementia, suggesting that MSVT performance may, in some cases, reflect emerging cognitive impairment not yet captured by formal diagnostic criteria. This is consistent with clinical practice within memory services, where presentations are often complex and evolving, and a "watch and wait" approach may be required while diagnostic clarity develops over time, raising the possibility that some Fail classifications may represent false positives rather than invalid responding.

These findings are consistent with broader literature indicating that performance on neuropsychological tests in clinical populations reflects the interaction of neurological, psychological, and medical factors (Lippa, 2018). They align with evidence suggesting that PVT failure is not solely attributable to malingering or intentional underperformance, but may be associated with psychiatric symptoms, cognitive inefficiency, and broader clinical burden (Kim et al., 2024). This is particularly relevant in older adult services, where overlapping contributors to performance are often the rule rather than the exception.

A particular strength of the present study was the use of exploratory subgroup analyses to further characterise this heterogeneity. These analyses were informed by rich clinical data and were intended to provide descriptive insight into patterns of performance rather than formal statistical comparison. However, subgroup sample sizes were small, limiting statistical power and the ability to draw firm conclusions, and findings should therefore be interpreted with caution.

Distinct patterns emerged across clinically derived subtypes. Individuals with functional neurological, functional cognitive, and ABI presentations demonstrated high rates of MSVT failure, suggesting that performance in these groups may be influenced by psychological, attentional, or functional factors rather than simply reflecting deliberate invalid responding. In contrast, participants with MCI-only presentations were more likely to achieve a Pass classification. This suggests that MSVT performance may align more closely with expected patterns in cognitive impairment where additional functional or psychological complexity is not prominent. Participants with later-confirmed dementia diagnoses showed a high proportion of GMIP classifications, consistent with emerging neurodegenerative processes influencing performance prior to formal diagnosis. Notably, individuals with MCI and functional overlay displayed a heterogeneous profile across all outcome categories, reflecting the interplay of cognitive and psychological factors.

Clinical record review further highlighted that MSVT failure within the CRC-ND group was not interpreted uniformly. In contrast to the dementia group, where there were no documented concerns regarding effort, clinician reports in the CRC-ND group reflected a more heterogeneous pattern. In some cases, MSVT failure was interpreted as indicating reduced or inconsistent effort, with reference to functional overlay or underperformance. In others, findings were described as ambiguous or requiring cautious interpretation, and in a subset of cases no concerns regarding effort were documented or reports were unavailable. This suggests

that MSVT failure within the CRC-ND group does not reflect a single process, but may arise from a range of interacting cognitive, psychological, and functional factors.

These findings have important implications for clinical practice. Attributing MSVT failure solely to deliberate underperformance may risk oversimplifying complex presentations and overlooking underlying psychological or functional processes. In particular, individuals presenting with functional neurological or cognitive symptoms may demonstrate patterns of performance that reflect difficulties with attention, engagement, or symptom experience, rather than intentional invalid responding. Misclassification in such cases may influence diagnostic formulation, reduce consideration of appropriate interventions, and potentially limit access to support or further assessment.

More broadly, these findings highlight the need for a formulation-based approach to performance validity assessment, consistent with Lippa's (2018) multi-model framework, in which MSVT outcomes are integrated with clinical history, behavioural observations, and neuropsychological findings. This may be particularly important in heterogeneous clinical populations such as CRC-ND, where the boundaries between cognitive impairment, functional symptoms, and effort are less clearly defined.

Clinical Factors and MSVT Performance

Clinical variables, including AEC, polypharmacy, and number of physical health conditions, were not significantly associated with MSVT failure within the CRC-ND group, indicating that performance variability was not explained by the factors examined in this study, although limited subgroup size may have reduced statistical power to detect smaller effects.

These findings suggest that no single clinical factor independently accounts for MSVT performance, and instead support a more multifactorial understanding of cognitive functioning in older adult populations. Although anticholinergic burden is associated with impairments in

memory, attention, and executive functioning, reflecting the cumulative cognitive effects of anticholinergic medications (Bishara et al., 2020), it did not independently predict MSVT failure in this diagnostically uncertain group.

It is also notable that mental health symptoms could not be examined across the full CRC-ND sample. Given evidence linking depression severity to cognitive impairment (McDermott & Ebmeier, 2009), it remains plausible that affective factors contributed to MSVT variability in ways not captured by the regression analyses.

Base Rates and Risk of Misclassification

The present findings should also be considered in light of base rate considerations, which are increasingly recognised as critical to the interpretation of PVT outcomes. Meta-analytic evidence suggests that the base rate of PVT failure in clinical populations is approximately 16% (Roor et al., 2024).

In populations with low base rates of invalid responding, such as older adults undergoing dementia assessment, even tests with high specificity may yield a proportion of false positive classifications. The current findings are consistent with evidence that PVT performance varies with the severity of cognitive impairment, and that standard cut-offs may require adjustment to maintain specificity in dementia populations (Martin et al., 2022).

Together, these findings highlight the importance of interpreting PVT outcomes within a framework that considers both base rates and clinical context, rather than relying on standard cut-offs alone.

Clinical Implications

The present findings have important implications for clinical practice, particularly in the interpretation of MSVT outcomes in older adult populations. PVT findings may require particular caution in older adult services, where cognitive impairment, diagnostic uncertainty,

physical health conditions, psychiatric symptoms, and functional difficulties frequently coexist. In these contexts, failed PVT outcomes may not necessarily reflect invalid responding and may instead be influenced by broader clinical complexity, increasing the risk of false positive classifications if interpreted in isolation. These findings therefore highlight the need for a cautious and context-sensitive approach, in which MSVT performance is not interpreted as a simple indicator of effort, but considered within the broader clinical context, including diagnostic uncertainty, functional overlay, and interacting cognitive and psychological factors.

This is consistent with guidance emphasising that validity assessment should draw on multiple methods and sources of information, rather than relying on a single test result (Bush et al., 2005). It also aligns with evidence that interpretation of PVT failure depends on the number and pattern of failures, as well as the clinical context, rather than isolated outcomes (Jennette et al., 2022). A formulation-based approach may therefore help to reduce misclassification and support more accurate, person-centred clinical decision-making.

These findings also highlight the potential value of incorporating normative data from real-world older adult clinical populations, including UK-based samples, into PVT tools to support more context-sensitive interpretation.

Recognition of this complexity is particularly important for clinical psychologists and neuropsychologists responsible for administering and interpreting PVTs, as well as multidisciplinary memory and older adult mental health teams involved in diagnostic formulation and clinical decision-making. Given the potential for cognitive impairment, psychiatric symptoms, physical health conditions, and functional difficulties to influence performance, PVT findings should be integrated with broader clinical information rather than interpreted in isolation.

Limitations

Several methodological limitations should be considered. First, the relatively modest sample size, particularly within subgroup and regression analyses, may have limited statistical power to detect associations between clinical variables and MSVT outcomes. This is particularly relevant given the heterogeneity of the CRC-ND group, where larger samples may be required to detect more subtle effects.

Second, the inclusion of both retrospective and newly recruited participants may have introduced variability in data quality and completeness. Retrospective variables were derived from routinely collected clinical data and may not fully capture the severity or functional impact of psychiatric, cognitive, and physical health conditions. At the same time, the use of routine clinical data enhances the ecological validity of the study, as findings reflect the complexity of real-world older adult mental health services. However, this strength comes at the expense of greater methodological control, with variability in data collection and availability of clinical information limiting standardisation across participants.

Third, the study relied on a single standalone performance validity test. Current guidance recommends the use of multiple validity indicators to improve classification accuracy and reduce the risk of misinterpretation, and findings should therefore be interpreted within the context of a single-measure approach.

Finally, the sample was drawn from a single NHS older adult mental health service, which may limit the generalisability of the findings to other clinical settings, particularly those with different patient characteristics or higher levels of external incentives.

Conclusion

Taken together, these findings suggest that MSVT performance in older adults is best understood as reflecting the interaction of cognitive impairment, psychological factors, and broader clinical complexity, rather than as a simple indicator of effort or response validity.

While the MSVT appears to perform robustly in mental health populations, its interpretation in dementia and diagnostically uncertain groups requires greater caution.

Patterns of MSVT performance across classification outcomes and subtest scores indicate that individuals with dementia show profiles consistent with genuine cognitive impairment, whereas those in the mental health group demonstrate largely intact performance. In contrast, the heterogeneity observed within the CRC-ND group, and the distinct patterns across clinical subtypes, highlight the limitations of binary interpretations of performance validity.

The present study contributes to ongoing efforts to refine the interpretation of performance validity in older adults by demonstrating that, in clinically complex populations, MSVT outcomes are shaped by underlying clinical presentation and cannot be fully understood in isolation. A context-sensitive, formulation-based approach is therefore essential to support accurate, clinically meaningful, and person-centred interpretation.

References

Axelrod, B. N., & Schutte, C. (2010). Analysis of the dementia profile on the Medical Symptom Validity Test. *The Clinical Neuropsychologist*, *24*(5), 873-881.

Bishara, D., Scott, C., Stewart, R., Taylor, D., Harwood, D., Codling, D., Banwell, C., & Sauer, J. (2020). Safe prescribing in cognitively vulnerable patients: the use of the anticholinergic effect on cognition (AEC) tool in older adult mental health services. *BJPsych bulletin*, *44*(1), 26-30.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, *20*(4), 419-426.

Cerny, B. M., Resch, Z. J., Rhoads, T., Jennette, K. J., Singh, P. G., Ovsiew, G. P., & Soble, J. R. (2022). Examining traditional and novel validity indicators from the medical symptom validity test across levels of verbal and visual memory impairment. *Archives of Clinical Neuropsychology*, *37*(1), 146-159.

Gaudet, C. E., Castelluccio, B., Gold, D., McLaughlin, N. C., & Correia, S. (2022). Limitations of performance validity tests in dementia evaluations: The role of base rates. *Psychological Assessment*, *34*(11), 1074.

Green, P. (2004). *Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's Manual*. Green's Publications. https://books.google.co.uk/books?id=wWWmMQEACAAJ

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology*, *18*(2), 86-94.

Howe, L. L., & Loring, D. W. (2009). Classification accuracy and predictive ability of the medical symptom validity test's dementia profile and general memory impairment profile. *The Clinical Neuropsychologist*, *23*(2), 329-342.

Jennette, K. J., Williams, C. P., Resch, Z. J., Ovsiew, G. P., Durkin, N. M., O'Rourke, J. J., Marceaux, J. C., Critchfield, E. A., & Soble, J. R. (2022). Assessment of differential neurocognitive performance based on the number of performance validity tests failures: A cross-validation study across multiple mixed clinical samples. *The Clinical Neuropsychologist*, *36*(7), 1915-1932.

Kim, S., Currao, A., Brown, E., Milberg, W. P., & Fortier, C. B. (2024). Importance of validity testing in psychiatric assessment: evidence from a sample of multimorbid post-9/11 veterans. *Journal of the International Neuropsychological Society*, *30*(4), 410-419.

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(3), 391-421.

Martin, P. K., Schroeder, R. W., & Olsen, D. H. (2022). Performance validity in the dementia clinic: Specificity of validity tests when used individually and in aggregate across levels of cognitive impairment severity. *The Clinical Neuropsychologist*, *36*(1), 165-188.

Masnoon, N., Shakib, S., Kalisch-Ellett, L., & Caughey, G. E. (2017). What is polypharmacy? A systematic review of definitions. *BMC geriatrics*, *17*(1), 230.

McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of affective disorders*, *119*(1-3), 1-8.

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, *34*(1), 114-131.

McWhirter, L., Ritchie, C. W., Stone, J., & Carson, A. (2020). Performance validity test failure in clinical populations—a systematic review. *Journal of Neurology, Neurosurgery & Psychiatry*, *91*(9), 945-952.

Moore, K. (2023). Exploring the influence of physical and mental health factors on older adults' performance on the Addenbrooke's Cognitive Examination III (ACE-III).

Morin, R. T., Axelrod, B. N., Reslan, S., & Schutte, C. (2019). Differentiating poor validity from probable impairment on the medical symptom validity test: a cross-validation study. *International Journal of Neuroscience*, *129*(3), 217-224.

Rai, J. K., Gervais, R. O., Rodrigues, N. B., & Erdodi, L. A. (2024). Base rates of failure on various performance validity tests as a function of age in adults referred for neuropsychological assessment. *Psychological Assessment*, *36*(8), 505.

Raja, S. N., Carr, D. B., Cohen, M., Finnerup, N. B., Flor, H., Gibson, S., Keefe, F. J., Mogil, J. S., Ringkamp, M., & Sluka, K. A. (2020). The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *Pain*, *161*(9), 1976-1982.

Resch, Z. J., Rhoads, T., Ovsiew, G. P., & Soble, J. R. (2022). A known-groups validation of the medical symptom validity test and analysis of the genuine memory impairment profile. *Assessment*, *29*(3), 455-466.

Reslan, S., & Axelrod, B. N. (2017). Evaluating the Medical Symptom Validity Test (MSVT) in a sample of veterans between the ages of 18 to 64. *Applied Neuropsychology: Adult*, *24*(2), 132-139.

Roor, J. J., Peters, M. J., Dandachi-FitzGerald, B., & Ponds, R. W. (2024). Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates. *Neuropsychology Review*, *34*(1), 299-319.

Singhal, A., Green, P., Ashaye, K., Shankar, K., & Gill, D. (2009). High specificity of the Medical Symptom Validity Test in patients with very severe memory impairment. *Archives of Clinical Neuropsychology*, *24*(8), 721-728.

Snaith, R., & Zigmond, A. (1986). The hospital anxiety and depression scale. *British medical journal (Clinical research ed.)*, *292*(6516), 344.

Chapter 3

Appendix A. PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review	p.7
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	p.8
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	pp. 9-11
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	p.11
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	p.14 (Table 1)
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	p.12
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	pp.12-13 + Appendix B (pp.90-92)
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	pp.14-15
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	p.15
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	p.15
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	p.15
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	pp.16-17

Section and Topic	Item #	Checklist item	Location where item is reported
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	pp.15-16
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	pp.15-16
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Table 2; pp.20-21
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	N/A
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure1 PRISMA; p.18
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	p.18
Study characteristics	17	Cite each included study and present its characteristics.	Table 2; pp.20-21
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Table 3; p.23
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	pp.24-36
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A

Section and Topic	Item #	Checklist item	Location where item is reported
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	pp.36-39
	23b	Discuss any limitations of the evidence included in the review.	pp.39-40
	23c	Discuss any limitations of the review processes used.	pp.39-40
	23d	Discuss implications of the results for practice, policy, and future research.	pp.40-42
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	p.12
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	p.12
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	N/A
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Appendix B: pp.90-92

Appendix B. Database Search Strategies

Appendix B1. Database Search Strategy (February 2026)

Database	Date	Search Strategy	Result(n)
MEDLINE (Ovid)	19 Feb 2026	1. (older adult* OR elderly OR geriatric* OR "late life" OR "later life").ti,ab. 2. (cognitive test* OR cognitive screen* OR cognitive assessment* OR memory test* OR memory assessment* OR neuropsychological test* OR neuropsychological assessment* OR MMSE OR MoCA OR ACE-III OR ACE-R).ti,ab. 3. (qualitative OR interview* OR "focus group*" OR thematic analys* OR grounded theory OR phenomenolog* OR "qualitative descriptive").ti,ab. 4. 1 AND 2 AND 3	1281
PsycINFO (Ovid)	19 Feb 2026	1. (older adult* OR elderly OR geriatric* OR "late life" OR "later life").ti,ab. 2. (cognitive test* OR cognitive screen* OR cognitive assessment* OR memory test* OR memory assessment* OR neuropsychological test* OR neuropsychological assessment* OR MMSE OR MoCA OR ACE-III OR ACE-R).ti,ab. 3. (qualitative OR interview* OR "focus group*" OR thematic analys* OR grounded theory OR phenomenolog* OR "qualitative descriptive").ti,ab. 4. 1 AND 2 AND 3	768
CINAHL (EBSCOhost)	19 Feb 2026	S1: (older adult* OR elderly OR geriatric* OR "late life" OR "later life") S2: (cognitive test* OR cognitive screen* OR cognitive assessment* OR memory test* OR memory assessment* OR neuropsychological test* OR neuropsychological assessment* OR MMSE OR MoCA OR ACE-III OR ACE-R) S3: (qualitative OR interview* OR "focus group*" OR thematic analys* OR grounded theory OR phenomenolog* OR "qualitative descriptive") S4: S1 AND S2 AND S3	1703

Appendix B2. Database Search Strategy (March 2026)

Database	Date	Search Strategy	Result(n)
MEDLINE (Ovid)	24 March 2026	1. (older adult* OR elderly OR geriatric* OR "late life" OR "later life").tw. 2. exp Neuropsychological Tests/ OR cognitive test*.tw. OR cognitive screen*.tw. OR cognitive assessment*.tw. OR memory test*.tw. OR memory assessment*.tw. OR neuropsychological test*.tw. OR neuropsychological assessment*.tw. OR MMSE.tw. OR MoCA.tw. OR ACE-III.tw. OR "ACE III".tw. OR ACE-R.tw. 3. Qualitative Research/ OR qualitative.tw. OR interview*.tw. OR "focus group*".tw. OR thematic analys*.tw. OR grounded theory.tw. OR phenomenolog*.tw. OR "qualitative descriptive".tw. 4. 1 AND 2 AND 3	2,498
PsycINFO (Ovid)	24 March 2026	1. (older adult* OR elderly OR geriatric* OR "late life" OR "later life").tw. 2. exp Neuropsychological Assessment/ OR cognitive test*.tw. OR cognitive screen*.tw. OR cognitive assessment*.tw. OR memory test*.tw. OR memory assessment*.tw. OR neuropsychological test*.tw. OR neuropsychological assessment*.tw. OR MMSE.tw. OR MoCA.tw. OR ACE-III.tw. OR "ACE III".tw. OR ACE-R.tw. 3. exp Qualitative Research/ OR qualitative.tw. OR interview*.tw. OR "focus group*".tw. OR thematic analys*.tw. OR grounded theory.tw. OR phenomenolog*.tw. OR "qualitative descriptive".tw. 4. 1 AND 2 AND 3	952
CINAHL (EBSCOhost)	24 March 2026	S1: TX (older adult* OR elderly OR geriatric* OR "late life" OR "later life") S2: (MH "Neuropsychological Tests+") S3: TX (cognitive test* OR cognitive screen* OR cognitive assessment* OR memory test* OR memory assessment* OR neuropsychological test* OR neuropsychological assessment* OR MMSE OR MoCA OR ACE-III OR "ACE III" OR ACE-R) S4: S2 OR S3	4,908

Database	Date	Search Strategy	Result(n)
		S5: (MH "Qualitative Studies+")	
		S6: TX (qualitative OR interview* OR "focus group*" OR thematic analys* OR grounded theory OR phenomenolog* OR "qualitative descriptive")	
		S7: S5 OR S6	
		S8: S1 AND S4 AND S7	

Appendix C: STROBE Checklist

STROBE Statement - Checklist of items that should be included in reports of *cross-sectional studies*

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	p.46
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	p.49
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	pp.50-54
Objectives	3	State specific objectives, including any prespecified hypotheses	pp.55-57
Methods			
Study design	4	Present key elements of study design early in the paper	p.56
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	pp.56-57
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	pp.57-58
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	pp.58-60
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	pp58-60
Bias	9	Describe any efforts to address potential sources of bias	p.112
Study size	10	Explain how the study size was arrived at	
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	pp.61-63
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	pp.63-64
		(*b*) Describe any methods used to examine subgroups and interactions	pp.63-64
		(*c*) Explain how missing data were addressed	pp.63-64
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	N/A

		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	pp.64-65
		(b) Indicate number of participants with missing data for each variable of interest	pp.64-65
Outcome data	15*	Report numbers of outcome events or summary measures	pp.66-67,71
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	pp.68-69
		(*b*) Report category boundaries when continuous variables were categorized	pp.64-65
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	pp.69-71
Discussion			
Key results	18	Summarise key results with reference to study objectives	pp.72-80
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	pp.80-81
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	pp.81-82
Generalisability	21	Discuss the generalisability (external validity) of the study results	p.81
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	N/A

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Appendix D: Ethical Approval Documents

D1. IRAS Approval Letter

NHS
Health Research Authority

South Central - Hampshire A Research Ethics Committee
Health Research Authority
2 Redman Place
Stratford
London
E20 1JQ

:

05 January 2026

Professor Jonathan Evans

Dear Professor Evans

Study title:	**Investigating Mental Health, Cognitive Function, and use of the Medical Symptom Validity Test in Older Adults: A Retrospective and Prospective Cohort Study**
REC reference:	**25/SC/0385**
Protocol number:	**0000**
IRAS project ID:	**363146**

Thank you for your letter of 19 December 2025, responding to the Proportionate Review Sub-Committee's request for changes to the documentation for the above study.

The revised documentation has been reviewed and approved on behalf of the PR sub-committee.

Confirmation of ethical opinion

On behalf of the Research Ethics Committee (REC), I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Good practice principles and responsibilities

The UK Policy Framework for Health and Social Care Research sets out principles of good practice in the management and conduct of health and social care research. It also outlines the responsibilities of individuals and organisations, including those related to the four elements of research transparency:

1. registering research studies

2. reporting results
3. informing participants
4. sharing study data and tissue

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All research should be registered in a publicly accessible database and we expect all researchers, research sponsors and others to meet this fundamental best practice standard.

It is a condition of the REC favourable opinion that **all clinical trials are registered** on a public registry before the first participant is recruited and no later than six weeks after. For this purpose, 'clinical trials' are defined as:

- clinical trial of an investigational medicinal product
- clinical investigation or other study of a medical device
- combined trial of an investigational medicinal product and an investigational medical device
- other clinical trial to study a novel intervention or randomised clinical trial to compare interventions in clinical practice.

A 'public registry' means any registry on the WHO list of primary registries or the ICMJE list of registries provided the registry facilitates public access to information about the UK trial.

Failure to register a clinical trial is a breach of these approval conditions, unless a deferral has been agreed by the HRA (for more information on registration and requesting a deferral see: Research registration and research project identifiers).

Where a deferral is agreed we expect the sponsor to publish a minimal record on a publicly accessible registry. When the deferral period ends, the sponsor should publish the full record on the same registry, to fulfil the condition of the REC favourable opinion.

If you have not already included registration details in your IRAS application form you should notify the REC of the registration details as soon as possible.

Publication of Your Research Summary

We will publish your research summary for the above study on the research summaries section of our website, together with your contact details, no earlier than three months from the date of this favourable opinion letter. Where a deferral is agreed, a minimum research summary will still be published in the research summaries database. At the end of the deferral period, we will publish the full research summary.

Should you wish to provide a substitute contact point, make a request to defer, or require further information, please visit: Research summaries - Health Research Authority (hra.nhs.uk)

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report
- Reporting results

The latest guidance on these topics can be found at https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/.

Ethical review of research sites

The favourable opinion applies to all NHS/HSC sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" above).

Approved documents

The documents reviewed and approved by the Committee are:

Document	*Version*	*Date*
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [UoG - Clinical Trials - 2025-2026 Client Information Letter]	1.0	28 July 2025
GP/consultant information sheets or letters [IRAS ID 363146 Letter to GP V1.1 19.12.25]	1.1	19 December 2025
GP/consultant information sheets or letters [Letter to GP V1.1 TRACK CHANGES]	1.1	19 December 2025
IRAS Application Form [IRAS_Form_31102025]		31 October 2025
IRAS Application Form XML file [IRAS_Form_31102025]		31 October 2025
IRAS Checklist XML [Checklist_19122025]		19 December 2025

Non-validated questionnaire [IRAS ID 363146 Demographic Information Form V1.0 28.10.25]	1.0	28 October 2025
Organisation Information Document [IRAS ID 363146 OID V1.0 28.10.25]	1.0	28 October 2025
Participant consent form [IRAS ID 363146 Consent Form V1.0 28.10.25]	1.0	28 October 2025
Participant information sheet (PIS) [IRAS ID 363146 Privacy Notice V1.0 28.10.25]	1.0	28 October 2025
Participant information sheet (PIS) [IRAS ID 363146 Results Opt In V1.0 28.10.25]	1.0	28 October 2025
Participant information sheet (PIS) [IRAS ID 363146 PIS V1.1 19.12.25]	1.1	19 December 2025
Participant information sheet (PIS) [IRAS 363146 Requested Information V1.0]	1.0	19 December 2025
Participant information sheet (PIS) [PIS - TRACK CHANGES]	1.1	19 December 2025
Participant information sheet (PIS) [Protocol V1.1 - TRACK CHANGES]	1.1	19 December 2025
Research protocol or project proposal [IRAS ID 363146 Protocol V1.1 19.12.25]	1.1	19 December 2025
Schedule of Events or SoECAT [IRAS ID 363146 SoE V1.0 28.10.25]	1.0	28 October 2025
Summary CV for Chief Investigator (CI) [JonEvansCV 2 Page Signed 220925]	1.0	22 September 2025
Summary CV for student [Susan Lavin CV 24.09.25]	1.0	24 September 2025
Summary CV for supervisor (student research) [Clive Ferenbach CV 30.09.25]	1.0	30 September 2025
Validated questionnaire [IRAS ID 363146 MSVT Description V1.0 31.10.25]	1.0	31 October 2025
Validated questionnaire [HADS]	1.0	30 October 2025

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at:
https://www.hra.nhs.uk/planning-and-improving-research/learning/

NHS
Health Research Authority

IRAS project ID: 363146	Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely

D2. NHS Ethics Approval Letter

CG Ref No:

NHS Lanarkshire
Application for Caldicott Guardian approval for use of patient identifiable data (PID)

You must address the Caldicott Principles (Appendix A) when submitting this request for data.
Applicants are requested to **type** their responses and email the completed document where possible.

1. Project Title

Investigating Mental Health, Cognitive Function, and Use of the Medical Symptom Validity Test in Older Adults: A Retrospective and Prospective Cohort Study

2. Details of individual / organisation requesting data

☒**NHS Lanarkshire** ☐**External**	Name / position / address: **Susan Lavin, NHS Lanarkshire, Trainee Clinical Psychologist, University of Glasgow**

3. Purpose for which data are to be used (Principle 1)

The data will be used for a doctoral research project.

As people get older, they may experience memory and thinking problems. These issues can be due to real memory loss or influenced by mental health conditions like depression or anxiety. To understand if memory problems are genuine or caused by other factors, psychologists use performance validity tests (PVTs). These tests check if someone is trying their best or if other issues, like mental/physical health or medication, are affecting their performance.

The PVT for this study was mainly designed for younger people or those without complex health problems. It is not clear how well it works for older adults who may have mental health conditions, physical health issues, or take multiple medications. This study explores how these factors relate to PVT results to help psychologists better understand and support older adults with memory concerns.

4. Outline why it is necessary to use identifiable data for this project. (Principle 2)

Identifiable data are required to:

- Identify eligible participants from NHS Lanarkshire clinical records.
- Link retrospective effort test results with demographic and clinical data.
- Avoid duplication and ensure data are accurately matched to the correct participant.

Identifiable data will be replaced with a unique study ID at the earliest opportunity following data extraction (retrospective data) or collection (prospective data). The key linking study IDs to identifiable information will be stored separately on secure NHS Lanarkshire systems, with access restricted to authorised personnel only. Precautions include secure NHS IT storage, password protection, and separation of identifiable data from analysis datasets. Only pseudonymised data will be used for analysis, and no identifiable information will be included in publications or shared outputs.

5. Which identifiable data items are required? Outline the reasons for using each individual item of PID. (Principle 3)

PID required		*Why necessary*
Chi Number	☒	**Required for safe and accurate record matching and de-duplication across datasets.**
Forename	☒	**Forename will be required to find the relevant patient data on Clinical Portal and ensure that this is the correct person i.e., used for checking purposes only – forename will not be part of the dataset.**
Surname	☒	**Surname will be required to find the relevant patient data on Clinical Portal and ensure that this is the correct person i.e., used for checking purposes only – surname will not be part of the dataset.**
Initials	☐	**Not required**
DoB	☒	**Required to calculate age and confirm eligibility**
Age	☐	**Age can be calculated from DOB obtained and is therefore not required.**
Gender	☒	**Needed for demographic analysis and description of sample.**
Address	☐	**Not required in full**
Post code	☒	**Used only to calculate Scottish Index of Multiple Deprivation (SIMD). Postcode will be deleted after coding and will not be included in the dataset.**
Clinical Information	☒	**Required to extract diagnosis, illness duration, comorbidities, medications, and test results for research purposes.**
Other (specify)	☐	
(Please provide any further details below)		

6. **Please provide full details of individuals who will have access to this identifiable data.** (**Principle 4**)

A Data Protection Impact Assessment (DPIA) has been completed for the project in line with University of Glasgow and NHS Lanarkshire requirements.

7. How will this identifiable data be kept secure from any further individuals? (Principle 4)

- Security measures are in place to protect participants personal data, including password protection, encryption, and secure NHS networks.
- Personal data (name, CHI number, contact details) will be stored securely on NHS Lanarkshire systems and any paper copies will be stored in locked cabinets within NHS premises.
- Research data will be pseudonymised. The data key linking your name to research code will be kept separately in secure NHS facilities, accessible only to the main researcher and chief investigator.
-
- No identifiable data will be transferred outside of the NHS secure network.
- Research data will be retained by the University for 10 years after the study is completed. After this time, data will be securely deleted.
- Paper copies containing identifiable information (e.g., consent forms, contact details) will be stored securely in a locked filing cabinet within NHS Lanarkshire premises. These will be retained for up to three months after data collection is complete to allow for data verification and audit checks, after which they will be shredded and confidentially disposed of in line with NHS information governance procedures.
- The main research database will be stored on secure NHS Lanarkshire IT systems and, where applicable (i.e., pseudonymised data for analysis), on the University of Glasgow secure network. All electronic files will be password protected, with access restricted to authorised members of the research team.

8. Outline action taken to ensure compliance with responsibilities and obligations to respect patient confidentiality e.g. individual policies or training. (Principle 5)

2025. These amendments have been completed and re-submitted to IRAS for review on 19th December 2025.

9. If you are not from NHS Lanarkshire outline who has statutory responsibility for information governance within your organisation? (Principle 6)

University of Glasgow Data Protection & FOI Office
University of Glasgow, Glasgow G12 8QQ

Email: **dpo@glasgow.ac.uk**

This application is being submitted on behalf of the University of Glasgow, which is the study sponsor, as the research is being conducted as part of a Doctorate in Clinical Psychology. While the study is not a service evaluation or audit required by NHS Lanarkshire, it is highly relevant to NHS clinical practice and is being conducted in collaboration with NHS Lanarkshire services.

NHS Lanarkshire is a participating organisation, and the study makes use of existing clinical data and recruits participants who are already receiving care within NHS Lanarkshire Older Adult services. Appropriate approvals are therefore being sought through NHS governance processes.

Information about the study has been shared with relevant clinical staff involved in the care of potential participants, including psychology and CMHT-OA teams. Clinicians will be responsible for identifying potentially suitable individuals from their caseloads and introducing the study to patients in a consistent and appropriate manner once approvals are in place.

10. Do any aspects of your project require a duty to share information? (Principle 7)

Yes - anonymised findings will be shared with NHS Lanarkshire services, published in a Doctorate thesis (University of Glasgow Enlighten repository), and may be submitted to peer-reviewed journals.

11. Have you considered informing patients and service users how you are using the confidential information? (Principle 8)

Yes - prospective participants will receive a Participant Information Sheet, Consent Form and Privacy Notice outlining how their data will be used and stored. Retrospective data will be pseudonymised, and outputs will contain no identifiable information.

Please list your organisation's Data Protection Registration Number
(if external to NHS Lanarkshire**)**

Note:

- Please provide copies of any other relevant supporting documentation (e.g. ethics approval, patient information leaflet etc.)

The release of data as described above is**:** **Approved** ☒ **Not approved** ☐

Date: 29/12/2025

Appendix A

Caldicott principles

Principle 1. **Justify the purpose(s)**
Every proposed use or transfer of patient-identifiable information within or from an organisation should be clearly defined and scrutinised, with continuing uses regularly reviewed, by an appropriate guardian.

Principle 2. **Don't use patient-identifiable information unless it is absolutely necessary**
Patient-identifiable information items should not be used unless there is no alternative.

Principle 3. **Use the minimum necessary patient-identifiable information**
Where use of patient-identifiable information is considered to be essential, each individual item of information should be justified with the aim of reducing identifiability.

Principle 4. Access to patient-identifiable information should be on a strict need-to-know basis
Only those individuals who need access to patient-identifiable information should have access to it, and they should only have access to the information items that they need to see.

Principle 5. **Everyone should be aware of their responsibilities**
Action should be taken to ensure that those handling patient-identifiable information - both clinical and non-clinical staff - are made fully aware of their responsibilities and obligations to respect patient confidentiality.

Principle 6. **Understand and comply with the law**
Every use of patient-identifiable information must be lawful. Someone in each organisation should be responsible for ensuring that the organisation complies with legal requirements.

Principle 7. **The duty to share information can be as important as the duty to protect patient confidentiality**
Health and social care professionals should have the confidence to share information in the best interests of their patients within the framework set out by these principles. They should be supported by the policies of their employers, regulators and professional bodies.

Principle 8. Inform patients and service users about how their confidential information is used
A range of steps should be taken to ensure no surprises for patients and service users, so they can have clear expectations about how and why their confidential information is used, and what choices they have about this. These steps will vary depending on the use: as a minimum, this should include providing accessible, relevant and appropriate information - in some cases, greater engagement will be required.

Caldicott Guardian for NHS Lanarkshire

D3. R&D Approval Letter

Professor Jonathan Evans

R&D Department

Date 21.01.2026

Enquiries to

Direct Line

Email

Dear Professor Evans

Project title: Investigating Mental Health, Cognitive Function, and use of the Medical Symptom Validity Test in Older Adults: A Retrospective and Prospective Cohort Study

R&D ID: L25097

I am writing to you as Chief Investigator of the above study to advise that R&D Management approval has been granted for the conduct of your study within NHS Lanarkshire as detailed below:

NAME	TITLE	ROLE	NHSL SITE TO WHICH APPROVAL APPLIES
Susan Lavin	Trainee Clinical Psychologist	Principal Investigator	Airbles Road Centre

For the study to be carried out you are subject to the following conditions:

Conditions

- You are required to comply with Good Clinical Practice, Ethics Guidelines, Health & Safety Act 1999 and relevant UK-GDPR and Data Protection 2018 legislation.
- The research is carried out in accordance with the Scottish Executive's Research Governance Framework for Health and Community Care (copy available via the Chief Scientist Office website: http://www.cso.scot.nhs.uk/

or the Research & Development Intranet site: http://firstport2/staff-support/research-and-development/default.aspx

- You must ensure that all confidential information is maintained in secure storage. You are further obligated under this agreement to report to the NHS Lanarkshire Data Protection Office and the Research & Development Office infringements, either by accident or otherwise, which constitutes a breach of confidentiality.
- Clinical trial agreements (if applicable), or any other agreements in relation to the study, have been signed off by all relevant signatories.
- You must contact the Lead Nation Coordinating Centre if/when the project is subject to any minor or substantial amendments so that these can be appropriately assessed, and approved, where necessary.
- You notify the R&D Department if any additional researchers become involved in the project within NHS Lanarkshire
- You notify the R&D Department when you have completed your research, or if you decide to terminate it prematurely.
- You must send brief annual reports followed by a final report and summary to the R&D office in hard copy and electronic formats as well as any publications.
- If the research involves any investigators who are not employed by NHS Lanarkshire, but who will be dealing with NHS Lanarkshire patients, there may be a requirement for an SCRO check and occupational health assessment. If this is the case, please contact the R&D Department to make arrangements for this to be undertaken and an honorary contract issued.

I trust these conditions are acceptable to you.

Yours sincerely,

Appendix E: Study Documentation (OSF Links)

The following study materials are available via the Open Science Framework (OSF):

- Participant Information Sheet: OSF | IRAS ID 363146 PIS V1.1 19.12.25.pdf
- Demographic Information Form: OSF | IRAS ID 363146 Demographic Information Form V1.0 28.10.25.pdf
- Consent Form: OSF | IRAS ID 363146 Consent Form V1.0 28.10.25.pdf
- Privacy Notice: OSF | IRAS ID 363146 Privacy Notice V1.0 28.10.25.pdf
- Letter to GP: OSF | IRAS ID 363146 Letter to GP V1.1 19.12.25.docx
- Results Opt In: OSF | IRAS ID 363146 Results Opt In V1.0 28.10.25.pdf

Appendix F: Data and Analysis Materials (OSF Links)

The following materials are available via the Open Science Framework (OSF):

- MRP Proposal: OSF | MRP Proposal .pdf
- MRP Protocol: OSF | IRAS ID 363146 Protocol V1.1 19.12.25.pdf
- Data Analysis Plan: OSF | Analysis Plan - MRP.docx
- Record of Data Analysis: OSF | MSVT_Analysis_Documentation.docx
- MRP SPSS Syntax: OSF | MSVT_Analysis_Syntax.sps

Appendix G: Reflexivity Statement

As a trainee clinical psychologist, my interest in examining the validity of the Medical Symptom Validity Test (MSVT) in older adults reflects a commitment to improving diagnostic accuracy and supporting equitable clinical decision-making. I occupy a position that is both "insider" and "outsider": as a clinician within NHS services, I am familiar with the systems and practices under study, yet I remain external to the lived experiences of the older adults whose data are represented.

I recognise that my clinical training and interest in formulation-based approaches may shape how I conceptualise performance validity, particularly in relation to assumptions about effort and cognitive functioning. Throughout the research process, I engaged in ongoing reflexive practice, including regular supervision and the use of a research log to document analytic decisions, uncertainties, and evolving interpretations. This supported critical reflection on how my perspectives may influence the framing and interpretation of findings.

Given the use of retrospective clinical data, I was mindful that participants were not directly involved in shaping the research process. I therefore sought to maintain a respectful and person-centred stance in analysis, recognising the complexity of the clinical contexts represented in the data and avoiding reductive interpretations of performance validity outcomes. Particular attention was given to the risk of reinforcing assumptions about "invalid" responding, especially in populations where cognitive, psychological, and functional factors overlap.

In presenting the findings, I aimed to balance clinical relevance with sensitivity to the potential implications of misclassification, reflecting on how interpretations may impact clinical understanding and care. Through ongoing reflexive engagement, I sought to ensure that the research remained grounded, ethically considered, and aligned with the values of clinical psychology practice.

Appendix H: Data Availability Statement

The data that support the findings of this study are not publicly available due to ethical and confidentiality restrictions. The dataset comprises sensitive clinical information collected within an NHS service, and the sharing of individual-level data would pose a risk of participant identification.

Anonymised data may be available from the author on reasonable request, subject to appropriate ethical approvals and data sharing agreements, and where this is consistent with NHS data governance and information governance policies.

Appendix I: Glossary of Neuropsychological and Study-Specific Terms

Term / Abbreviation	Definition
ABI (Acquired Brain Injury)	Damage to the brain occurring after birth due to injury, stroke, infection, or other neurological causes.
AEC (Anticholinergic Effect on Cognition) Score	A measure of the potential cognitive impact of medications with anticholinergic properties. Higher scores indicate a greater potential effect on cognition.
Alzheimer's Disease	The most common cause of dementia, characterised by progressive decline in memory, thinking, and everyday functioning associated with neurodegenerative changes in the brain.
Anticholinergic Burden	The cumulative effect of taking one or more medications with anticholinergic properties. Higher anticholinergic burden has been associated with difficulties in memory, attention, and other aspects of cognitive functioning, particularly in older adults.
Base Rate	The frequency with which a condition or outcome occurs within a particular population. Base rates influence the likelihood that a positive test result reflects a true finding.
Clinical Formulation	A structured understanding of an individual's difficulties that integrates biological, psychological, and social factors to guide assessment and intervention.

Term / Abbreviation	Definition
CNS (Consistency)	An MSVT validity index assessing consistency of responding across recognition memory trials.
CRC-ND (Concerns Raised regarding Cognition but Not diagnosed with Dementia)	A study-specific group comprising individuals referred for assessment of cognitive concerns who did not meet diagnostic criteria for dementia at the time of testing.
Dementia	A syndrome characterised by progressive decline in cognitive functioning that interferes with independence in everyday activities.
Diagnostic Complexity	The presence of multiple interacting factors that may influence assessment findings, such as cognitive impairment, mental health difficulties, physical health conditions, medication effects, and functional symptoms.
DR (Delayed Recognition)	An MSVT subtest assessing recognition memory following a delay.
Easy–Hard Discrepancy	A pattern of performance in which an individual performs unexpectedly poorly on easier tasks relative to more difficult tasks. Within performance validity testing, this pattern may suggest non-credible performance.
Effort	The degree to which an individual is able and willing to engage with cognitive testing. Contemporary frameworks recognise that effort may be influenced by cognitive, psychological, medical, and situational factors.

Term / Abbreviation	Definition
Executive Functioning	A group of higher-order cognitive abilities involved in planning, organisation, problem-solving, decision-making, self-monitoring, and cognitive flexibility.
False Negative Classification	When a test fails to identify a condition or characteristic that is actually present.
False Positive Classification	When a test incorrectly identifies a condition or characteristic that is not actually present. In the present study, this refers to valid test performance being incorrectly classified as invalid.
FCD (Functional Cognitive Disorder)	A condition characterised by persistent cognitive complaints that are not fully explained by neurodegenerative disease or other recognised neurological conditions.
FND (Functional Neurological Disorder)	A condition characterised by neurological symptoms that are not fully explained by recognised neurological disease and are associated with altered nervous system functioning.
FR (Free Recall)	An MSVT subtest assessing unaided recall of previously presented information.
Functional Overlay	The contribution of psychological, behavioural, or functional factors to the presentation of symptoms alongside any underlying medical or neurological condition.
GMIP (Genuine Memory Impairment Profile)	A classification on the MSVT indicating that poor performance may be attributable to genuine memory impairment rather than invalid responding.

Term / Abbreviation	Definition
HADS (Hospital Anxiety and Depression Scale)	A 14-item questionnaire used to assess symptoms of anxiety and depression.
IR (Immediate Recognition)	An MSVT subtest assessing recognition memory immediately after presentation of information.
Interpretive Validity	The extent to which conclusions drawn from test results accurately reflect the construct being assessed and are appropriate within the broader clinical context.
Malingering	The intentional exaggeration or fabrication of symptoms for an external benefit, such as financial gain, avoidance of responsibilities, or access to services.
MCI (Mild Cognitive Impairment)	A condition involving cognitive decline greater than expected for age, but not severe enough to meet diagnostic criteria for dementia.
Memory Clinic	A specialist service that assesses, diagnoses, and supports individuals experiencing memory difficulties or suspected dementia.
MSVT (Medical Symptom Validity Test)	A performance validity test designed to assess whether cognitive test performance provides a valid representation of an individual's abilities.
NHS (National Health Service)	The publicly funded healthcare system of the United Kingdom.
Neurodegenerative Disease	A condition characterised by progressive deterioration of the nervous system, such as Alzheimer's disease.

Term / Abbreviation	Definition
Neuropsychological Assessment	A structured assessment of cognitive functioning, including memory, attention, language, executive functioning, and other cognitive abilities.
Neuropsychologist	A psychologist with specialist training in the relationship between brain functioning, cognition, behaviour, and psychological wellbeing, who undertakes cognitive assessment and interpretation.
PA (Paired Associates)	An MSVT subtest assessing memory for associated word pairs.
Performance Validity	The extent to which cognitive test performance accurately reflects an individual's true cognitive abilities and level of engagement with assessment.
Polypharmacy	The concurrent use of multiple medications, commonly defined as five or more prescribed medications.
PTOP (Psychological Therapies for Older People)	A specialist NHS service providing psychological assessment and intervention for older adults experiencing mental health and related difficulties.
PVT (Performance Validity Test)	A measure designed to assess whether cognitive test performance is likely to provide a valid representation of an individual's functioning.
Sensitivity	The ability of a test to correctly identify individuals who do have a condition or characteristic. In the context of PVTs,

Term / Abbreviation	Definition
	high sensitivity reduces the likelihood of failing to identify genuinely invalid performance.
Specificity	The ability of a test to correctly identify individuals who do not have a condition or characteristic. In the context of PVTs, high specificity reduces the likelihood of valid performances being incorrectly classified as invalid.
Standalone PVT	A performance validity test specifically designed to assess the validity of cognitive test performance, rather than being embedded within a broader cognitive measure.
Tolterodine	A medication commonly prescribed for overactive bladder symptoms. It has anticholinergic properties and was relevant to the present study because medications such as tolterodine contribute to overall anticholinergic burden, which may affect cognitive functioning in older adults.
Validity Assessment	The process of evaluating whether cognitive test results provide an accurate and interpretable representation of an individual's functioning.
Validity Indicator	A measure or score used to evaluate whether cognitive test performance is likely to provide an accurate and interpretable representation of an individual's abilities.