International Psychogeriatrics (2022), 34:5, 453–465
doi:10.1017/S1041610222000175

REVIEW

Detecting depression in persons living with mild cognitive impairment: a systematic review

Britney Wong,[1,2] Zahinoor Ismail,[2,3,4,5] and Zahra Goodarzi[1,2,3,6,7]

[1]*Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada*
[2]*Cumming School of Medicine, University of Calgary, Calgary, AB, Canada*
[3]*Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada*
[4]*Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada*
[5]*Department of Psychiatry, University of Calgary, Calgary, AB, Canada*
[6]*Department of Medicine, University of Calgary and Alberta Health Services, Calgary, AB, Canada*
[7]*O'Brien Institute of Public Health, University of Calgary, Calgary, AB, Canada*

ABSTRACT

Objective : Depression is common in persons experiencing mild cognitive impairment (MCI), with 32% (95% CI 27, 37) overall experiencing depression. Persons with MCI who have depression have more cognitive changes compared to those without depression. To understand how we can detect depressive symptoms in persons with MCI, we undertook a systematic review to identify tools that were validated compared with a reference standard.

Design: We searched MEDLINE, EMBASE, PsycINFO, and Cochrane from inception to April 25, 2021, and conducted a gray literature search. Title/abstract and full-text screening were completed in duplicate. Demographic information, reference standards, prevalence, and diagnostic accuracy measures were then extracted from included articles (PROSPERO CRD: CRD42016052120).

Results : Across databases, 8,748 abstracts were generated after removing duplicates. Six hundred and sixty-five records underwent full-text screening, with six articles included for data extraction. Nine tools were identified compared to a reference standard, with multiple demonstrating a sensitivity of 100% (Brief Assessment Schedule Depression Cards, Beck Depression Inventory-II, Cornell Scale for Depression in Dementia, Zung Self-Rated Depression Scale, and the Neuropsychiatric Inventory). The second highest sensitivity reported was 89% (Patient Health Questionnaire-9). Too few studies were available for a meta-analysis.

Conclusions Multiple depression detection tools have been examined amongst MCI outpatients, with several showing high sensitivity. However, this evidence is only present in single studies, with little demonstration of how differing MCI types affect accuracy. More research is needed to confirm the accuracy of these tools amongst persons with MCI. At this time, several tools could be suitable for use in cognitive clinics.

Keywords mild cognitive impairment, depression, diagnostic accuracy

Introduction

Mild cognitive impairment (MCI) is characterized by memory or cognitive deficits that do not affect daily function (Langa and Levine, 2015). There are two primary subtypes of MCI, amnestic and non-amnestic MCI (Mariani *et al.*, 2007). Amnestic MCI is predominantly a decline in memory, while non-amnestic MCI is impairment in other non-memory cognitive domains, such as attention, visuospatial functioning, language, or executive functioning (Csukly *et al.*, 2016; Ganguli *et al.*, 2010).

Prevalence estimates of MCI range between 16% and 20% worldwide, primarily affecting older adults (Roberts and Knopman, 2013). Persons with MCI and depression have a higher rate of progression to Alzheimer's disease than MCI patients without depression, with progression rates of 31% and 13.5%, respectively (Ma, 2020).

Ultimately, comorbid depression with MCI affects 32% (95% CI 27, 37) of MCI patients (Ismail *et al.*, 2017), and individuals experiencing both have difficulty with immediate and delayed memory tasks

Correspondence should be addressed to: Zahra Goodarzi, Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada. E-mail: Zahra.goodarzi@albertahealthservices.ca Received 08 Nov 2021; revision requested 11 Dec 2021; revised version received 27 Jan 2022; accepted 04 Feb 2022. First published online 31 March 2022.

in comparison with non-depressed persons with MCI (Ismail *et al.*, 2017). Additionally, persons with both MCI and depression typically have lower processing speeds and show a decrease in executive function, flexibility, and lexico-semantic function than MCI patients without depression (Ma, 2020). Unfortunately, these patients more frequently experience a poorer quality of life than MCI patients without depression (Ismail *et al.*, 2017). Furthermore, depressive symptoms in MCI are associated with greater amyloid-β burden (Krell-Roesch *et al.*, 2019; Miao *et al.*, 2021).

Given the substantial burden that MCI patients experience when facing cooccurring depressive symptoms, it is crucial that healthcare practitioners be able to detect depressive symptoms accurately. Depressive symptoms are currently assessed clinically using various different rating scales. However, the accuracy of these tools, in persons living with MCI, is not fully elucidated. The objective of our systematic review is to determine which tools for detecting depressive symptoms are the most accurate and feasible among outpatients with MCI, compared with a reference standard clinical diagnosis of depression.

Current publications in the IPG examine the prevalence of depressive symptoms among persons with MCI or elaborate on the prevalence of MCI amongst persons with geriatric depression. However, there is a lack of studies currently characterizing how efficacious depression detecting tools are in the context of MCI. Many of these articles feature common depression detection tools yet do not explore how accurate they are. Our study serves as a benchmark for future research endeavors, to develop novel tools specifically for patients with MCI, or conduct diagnostic accuracy studies on existing tools.

Methods

The protocol has been registered with PROSPERO (Making the Case for Investing in Mental Health in Canada, 2013) (CRD42016052120). The study is reported as per the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and guidelines (Moher *et al.*, 2015).

Search strategy

A literature search strategy was developed in conjunction with an experienced librarian. The databases MEDLINE, EMBASE, PsycINFO, and Cochrane Database for Systematic Reviews were searched from inception until April 25, 2021. They were searched using clusters of terms (key words and search terms specific to each database). The main search clusters were MCI, depression, and diagnostic accuracy terms. Within each cluster, keywords and specific search terms were combined using "or," and clusters were then subsequently combined using "and." The specific MEDLINE search used is shown in Appendix 4 (see Appendix 4 published as supplementary material online). All relevant terms describing depressive symptoms were included in the search, in addition to related derivatives of MCI. A gray literature search was also conducted from inception to January 10, 2021 (Appendix 3) (see Appendix 3 published as supplementary material online). All languages were included in this search. The organizations searched include mental health organizations, cognitive sites, general gray databases, search engines, international databases, and theses (Moher *et al.*, 2015).

Selection and eligibility

All abstracts were assessed for eligibility, in duplicate, by two independent authors (B.W and Z.G). We have defined MCI based on adult outpatients using Petersen criteria (Petersen *et al.*, 1997) or an NIA-AA diagnosis of MCI (Albert *et al.*, 2011), or a tool designed to assess MCI. Studies must use any depression detection tool (i.e. Geriatric Depression Scale (Yesavage and Sheikh, 1986), Neuropsychiatric Inventory Scale (Cummings *et al.*, 1994), and so forth), or depressive symptoms assessment as a way to detect depressive symptoms, compared to a reference standard. Reference standards included were a clinician's diagnosis, any diagnosis of depression from any version of the Diagnostic and Statistical Manual of Mental Disorders (DSM) (American Psychiatric Association, 2013) (i.e. DSM IV, V, and so forth) or the International Classification of Diseases (ICD) (The ICD-10 Classification of Mental and Behavioural Disorders, 1992) (i.e. ICD 9, 10, and so forth).

No language or age restrictions were used. At abstract screening, an article was included if it discussed MCI and a depression detection tool. Any abstract included by either author was included for full text. At full-text screening stage, articles were included if they looked at a tool compared to a reference standard and reporting of diagnostic accuracy outcomes. The full texts were reviewed in duplicate by two independent authors. All non-English abstracts were translated using the online translation software Google Translate and similarly assessed at the full-text stage using this software.

https://doi.org/10.1017/S1041610222000175 Published online by Cambridge University Press

Assessment of risk of bias

A risk of bias assessment was completed in duplicate by two independent authors (Z.G and B.W) using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool (Whiting *et al.*, 2011). The QUADAS-2 is an improved version of the QUADAS tool and is designed specifically for systematic reviews of diagnostic accuracy studies (Whiting *et al.*, 2011). Therefore, assessment of the quality of included studies is appropriate using the QUADAS-2. The QUADAS-2 has four domains, including Patient Selection, Index Test, Reference Standard, and its Flow of Timing (Whiting *et al.*, 2011).

Data extraction and synthesis of evidence

The data extraction criteria were created together by two authors (Z.G and B.W). The data were then extracted by one author (B.W) and verified by another (Z.G). Specific information extracted is outlined in Table 1. These elements include demographics such as total sample size and the percentage of females in the total sample or MCI subgroup. In addition, the prevalence of depression based on the reference standard, specific index tool used and its cutoff value, and the sensitivity, specificity, positive and negative likelihood ratios (positive likelihood ratio [PLR] and negative likelihood ratio [NLR], respectively), and positive predictive value and negative predictive value were extracted.

Results

Database searches

The database search generated 11,190 abstracts, and upon removing duplicates, 8,748 remained. 2,542 records were identified through the gray literature search. 10,625 records from all sources were excluded at level 1 abstract screening, because articles lacked an MCI population or subpopulation, or did not report a depression assessment tool. 665 records were then included at full-text screening. At this stage, articles were excluded because they did not report the diagnostic accuracy of their index tools (sensitivity, specificity, or likelihood ratios) (n = 73), did not have a reference standard diagnosis of depression (n = 275), or the article found was a conference abstract (n = 184). The latter represents abstracts where the full-text version was searched for but could not be found. Authors of included studies that did not report their sensitivity and specificity measures or prevalence of depression in the MCI subgroup based on the reference standard were emailed for verification or were calculated by the authors if enough information was given. Two out of four contacted authors responded. The exclusion criteria are listed in the PRISMA diagram (Figure 1). From full-text screening, six articles were included for qualitative analysis.

Summary of included studies

Six studies were included in the final qualitative synthesis, with dates ranging from 2006 to 2011 (Artero *et al.*, 2008; Boyle *et al.*, 2011; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011). All included articles were written in English. The total percent female ranged from 56.50% to 66.70% (Artero *et al.*, 2008; Boyle *et al.*, 2011; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011), whilst the sample size ranged from 113 to 6,892 individuals (Artero *et al.*, 2008; Boyle *et al.*, 2011; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011). The studies were conducted in France, Italy, Belgium, Australia, and Spain (Artero *et al.*, 2008; Boyle *et al.*, 2011; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011). The mean age of participants ranged from 70.1 ± 6.5 to 86.64 ± 6.59 years of age (Artero *et al.*, 2008; Boyle *et al.*, 2011; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011).

Five articles reported on a strict MCI diagnosis (Artero *et al.*, 2008; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011), whilst 1 other article (Boyle *et al.*, 2011) used a broader subgroup of patients experiencing cognitive impairment. The specific type of MCI was generally not mentioned among the articles; however, Di Iulio *et al.* (2010) reported on multidomain and amnestic MCI (Di Iulio *et al.*, 2010). MCI was assessed by two studies using Petersen's criteria of MCI (Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007). Other studies (n = 5) evaluated MCI using varying tools. Specifically, one study assessed MCI with cognitive tests such as the Benton Visual Retention Test, the Trail making Test, the Isaacs' Set Test, and a word recall test with both delayed free recall and recall with semantic prompts (Artero *et al.*, 2008). Furthermore, one study used the SMMSE to assess MCI (McCabe *et al.*, 2006). The other tools used to assess MCI were the MEC (n = 1) (Ros *et al.*, 2011) and the Six-Item Screen (Boyle *et al.*, 2011).

The depression tools evaluated in the included studies are the Geriatric Depression Scale (GDS-15) (n = 2) (Dierckx *et al.*, 2007; McCabe *et al.*, 2006), Center for Epidemiological Studies-Depression (CES-D) (n = 2) (Artero *et al.*, 2008; Ros *et al.*, 2011), Brief Assessment Schedule Depression Cards (BASDEC) (n = 1) (McCabe *et al.*, 2006), Beck Depression Inventory-II (BDI-II) (n = 1)

https://doi.org/10.1017/S1041610222000175 Published online by Cambridge University Press

Table 1. Data extraction information obtained from each included article

Author	Location	Year	Mean age of participants	Total % female	Total sample size	Type of cognitive impairment subgroup	% and number of females in MCI/CI subgroup	Method used to diagnose MCI/CI	Score on MCI diagnostic test
Artero et al.	France	2008	74.6 ± 5.7	60%	6,892	MCI	66.70% 1,626	Benton Visual Retention Test, Trail Making Test, Isaacs' Set Test, and word recall test	NR
di Lulio et al.	Italy	2010	Multidomain MCI: 71.0 ± 6.5 Amnestic MCI: 70.1 ± 6.5	56.50%	352	Multidomain and Amnestic MCI	41.26% 126	Petersen criteria by trained clinical neurologists and psychologists	Only criteria
Dierckx et al.	Belgium	2007	75 ± 6.0	72%	151	MCI	47.50% 40	Based on Petersen criteria, diagnosed by clinicians	Only criteria
McCabe et al.	Australia	2006	86.64 ± 6.59	73.45%	113	MCI	NR 61	SMMSE	20–23
Ros et al.	Spain	2011	72.74 ± 7.70	60%	623	MCI	61.73% 162	MEC	≥ 23
Boyle et al.	USA	2011	77	68%	236	CI	54% 37	Six-Item Screen	≥ 2

Author	Type of GS	Prevalence of depression by GS	Tool(s) used	Optimal cutoff	SN	SP	PLR	NLR	PPV	NPV
Artero et al.	DSM-IV (MINI)	3.1% (50/1626)	CES-D	>16	84%	81%	4.42	0.20	NR	NR
di Lulio et al.	Structured interview with clinician	37.3% (47/126)	NPI (depression domain)	>4	100%	56.96%	2.32	0	58.02%	100%
Dierckx et al.	Clinical diagnosis based on the DSM-IV criteria	NR	GDS	8	58%	85%	3.866	0.494	NR	NR
McCabe et al.	Semi-structured Clinical Diagnostic Interview for DSM-IV (SCID-I): Administered by a Clinical Psychologist and reviewed in consultation with a Geropsychiatrist	11.48% (7/61)	BASDEC BDI-2 CSDD GDS-15 SDS	BASDEC: 5 BDI-2: 10 CSDD: 10 GDS-15: 6 SDS: 40	BASDEC: 1.00 (0.59–1.00) BDI-2: 1.00 (0.54–1.00) CSDD: 1.00 (0.59–1.00) GDS-15: 1.00 (0.54–1.00) SDS: 1.00 (0.54–1.00)	BASDEC: 0.86 (0.73–0.94) BDI-2: 0.92 (0.80–0.98) CSDD: 0.94 (0.84–0.99) GDS-15: 0.92 (0.81–0.98) SDS: 0.82 (0.69–0.91)	BASDEC: 7.14 BDI-2: 12.25 CSDD: 17.00 GDS-15: 12.50 SDS: 5.56	BASDEC: 0 BDI-2: 0 CSDD: 0 GDS-15: 0 SDS: 0	NR	NR

Ros et al.	Composite International Diagnostic Interview: according to the DSM-IV and ICD-10	51% (82/162)	CES-D	13	86.25%	72.37%	3.123	0.19	76.67%	83.33%	
Boyle et al.	Structured Clinical Interview for DSM-IV Psychiatric Disorders (SCID)	24.32% (9/37)	PHQ-2 PHQ-9	PHQ-2: ≥ 3 PHQ-9: ≥ 10	PHQ-2 = 78% PHQ-9 = 89%	PHQ-2: 74% PHQ-9: 71%	PHQ-2: 2.72 PHQ-9: 3.11	PHQ-2: 0.31 PHQ-9: 0.16	PHQ-2: 46% PHQ-9: 50%	PHQ-2: 76% PHQ-9: 95%	

GS, gold standard; NLR, negative likelihood ratio; NPV, negative predictive value; NR, not reported; PLR, positive likelihood ratio; PPV, positive predictive value; SN, sensitivity; SP, specificity.

(McCabe *et al.*, 2006), Cornell Scale for Depression in Dementia (CSDD) (n = 1) (McCabe *et al.*, 2006), Zung Self-Rating Depression Scale (SDS) (n = 1) (McCabe *et al.*, 2006), the depression domain of the Neuropsychiatric Inventory (NPI) (n = 1) (Di Iulio *et al.*, 2010), and the Patient Health Questionnaire-2 and 9 (n = 1) (Boyle *et al.*, 2011). Descriptions of each tool are shown in Table 2.

Risk of bias assessment

The risk of bias assessment is shown in Table 3. Most studies (n = 5) had a low risk that the included patients did not match the review question (Artero *et al.*, 2008; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011). Multiple studies (n = 3) did not report if administrators of the index test were blind to the reference standard (Artero *et al.*, 2008; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007). It was unclear in several studies (n = 4) if the reference standard was interpreted without knowledge of the index test (Artero *et al.*, 2008; Boyle *et al.*, 2011; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007). As a result, it was unclear whether administrators were blinded in either direction. Furthermore, multiple studies (n = 4) did not report a time interval between the index test and the reference standard (Artero *et al.*, 2008; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; Ros *et al.*, 2011). The flow and timing among studies were generally unclear. Overall included articles did not report blinding of reference standard and index test, and timing between tools clearly. This indicates that some of the included articles had some risk of bias which could affect the accuracy of the tools.

Diagnostic accuracy of tools in included studies

Studies with MCI diagnosis as per Petersen or NIA-AA criteria

Prevalence. Each study had a unique prevalence of depression among MCI participants. Sample sizes in the MCI subgroups ranged from 37 to 1,626 MCI participants, with depression prevalence ranging from 3.1% to 51% as defined by the reference standard (Artero *et al.*, 2008; Di Iulio *et al.*, 2010; McCabe *et al.*, 2006; Ros *et al.*, 2011). Moreover, one included study did not report their depression prevalence rates as defined by the reference standard (Dierckx *et al.*, 2007).

GDS-15. There were two studies evaluating the diagnostic accuracy of the GDS-15 tool (Dierckx *et al.*, 2007; McCabe *et al.*, 2006). Sensitivity measures ranged from 58% to 100% whilst specificity ranged from 85% to 92%. The highest PLR reported was 12.50, while the lowest NLR was 0. All three studies used a reference standard consisting of a

https://doi.org/10.1017/S1041610222000175 Published online by Cambridge University Press

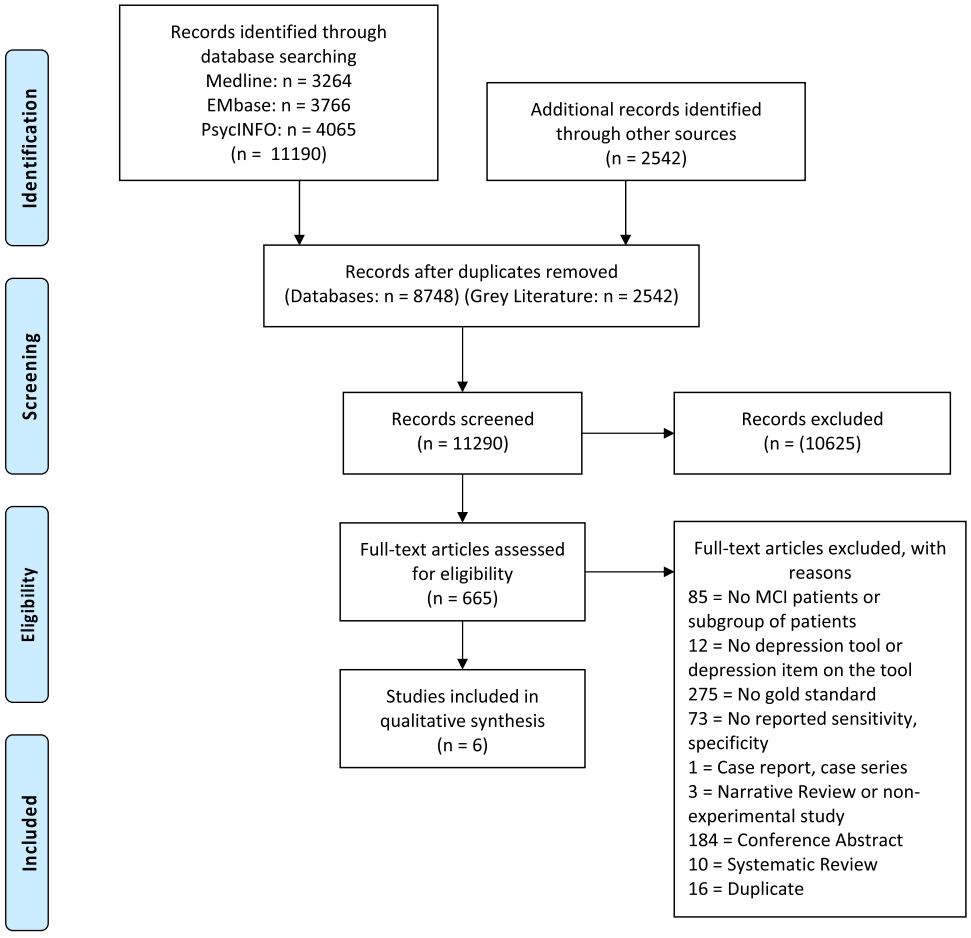

Figure 1 PRISMA diagram depicting the review flow (Moher *et al.*, 2009). 11,290 articles were screened in level 1 screening, from which 665 remained. After level 2, full-text screening, six articles were chosen for qualitative analysis.

clinical diagnosis based on DSM-IV criteria, administered by a physician.

BASDEC, BDI-II, CSDD, SDS. One study evaluated the diagnostic accuracy of the BASDEC, BDI-II, CSDD, and SDS tools (McCabe *et al.*, 2006). The reported sensitivity was 100% across all tools, while the corresponding specificities were 86% (BASDEC), 92% (BDI-II), 94% (CSDD), and 82% (SDS) (McCabe *et al.*, 2006). The reference standard used was a semistructured clinical diagnostic interview for DSM-IV (SCID-I) administered by a clinical psychologist and reviewed in consultation with a geropsychiatrist.

CES-D. Two studies reported on the diagnostic accuracy of the CES-D tool (Artero *et al.*, 2008; Ros *et al.*, 2011). Sensitivity values ranged from 86.25% to 84% while specificity ranged from 81% to 72.37%. Structured interviews with a physician using DSM-IV and ICD-10 criteria were used as reference standards.

NPI. One study evaluated the diagnostic accuracy of the NPI tool (Di Iulio *et al.*, 2010). A sensitivity of 100% and a specificity of 56.96% were reported for the depression domain of the NPI, compared to a structured interview with a clinician as a reference standard. PLR and NLR values were 2.32 and 0, respectively.

Studies discussing "mild cognitive impairment" but not diagnosed by criteria

One included article stated they examined individuals with MCI but did not specify whether diagnoses were made based on specific criteria (Boyle *et al.*, 2011).

Boyle *et al.* (2011) measured the PHQ-2 and PHQ-9 for diagnostic accuracy in a population of persons living with general cognitive impairment (Boyle *et al.*, 2011). The reference standard comparison was the Structured Clinical Interview for DSM-IV Psychiatric Disorders (SCID). An optimal cutoff of ≥ 3 and ≥ 10 were found for the PHQ-2 and PHQ-9, respectively (Boyle *et al.*, 2011). For the PHQ-2, the sensitivity and specificity were reported as 78% and 74%, respectively (Boyle *et al.*, 2011). The sensitivity of the PHQ-9 was 89%, while the specificity was 71% (Boyle *et al.*, 2011). The values

https://doi.org/10.1017/S1041610222000175 Published online by Cambridge University Press

Table 2. Descriptions of each tool assessed in the included studies

Tool name	Abbreviation	Rater	Permission or copyright restrictions	Items (n)	Time to complete	Description of tool
Geriatric Depression Scale-15	GDS-15	Patient (Albert *et al.*, 2011)	No	15	2–5 min (Smarr and Keefer, 2011)	The GDS-15 is a yes/no-based self-reported questionnaire and is the short-form version of the original GDS-30 scale. It consists of 15 items, with 10 items reaching a positive score and 5 items displaying a negative score when indicating the presence of depression. A score from 0 to 4 is considered normal and nondepressive, while 5–8, 9–11, and 12–15 indicate mild, moderate, and severe depression, respectively (Yesavage and Sheikh, 1986).
Center for Epidemiological Studies-Depression	CES-D	Patient (Burns *et al.*, 2002)	No	20	<10 min (Smarr and Keefer, 2011)	The CES-D is a brief self-report scale comprised of 20 items. It comprises six scales which reflect depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance (Radloff, 1977).
Brief Assessment Schedule Depression Cards	BASDEC	Patient response prompted by Interviewer (Burns *et al.*, 2002)	Yes	19	2–8 min (Burns *et al.*, 2002)	The BASDEC is comprised of 19 cards with statements derived from the Brief Assessment Schedule depression scale. These statements are answered with either "True" or "False" by the participant, according to their feelings towards the questions (Adshead *et al.*, 1992).
Beck Depression Inventory-II	BDI-II	Patient (Davidson *et al.*, 2019)	Yes	21	5–10 min (Smarr and Keefer, 2011)	The BDI-II questionnaire is a self-reported test that measures the intensity, severity, and depth of depression symptoms among patients ranging from 13 to 80 years old. It is the most updated and current version of the Beck Depression Inventory scale. Each of the 21 questions is comprised of 4 possible responses, where responses are assigned a score from zero to three to indicate symptom severity that the patient experienced over the past 2 weeks (Davidson *et al.*, 2019).
Cornell Scale for Depression in Dementia	CSDD	Informant and Patient (Alexopoulos *et al.*, 1988)	No	19	30 min (Alexopoulos *et al.*, 1988)	The CSDD is a 19-item instrument that specifically rates depression symptomology among persons living with dementia. The severity of each item is rate according to three grades: absent, mild or intermittent, and severe. It is administered in two steps. First, clinician interviews with the patient's caregiver are conducted using all 19 items, then interviews between the clinician and the patient are administered. A score greater than 10 defines a probable major depressive disorder episode, while a score greater than 18 defines major depressive disorder episode (Alexopoulos *et al.*, 1988).

Table 2. Continued

Tool name	Abbreviation	Rater	Permission or copyright restrictions	Items (n)	Time to complete	Description of tool
Zung Self-Rating Depression Scale	SDS	Patient (Jokelainen *et al.*, 2019)	No	20	10 min (Zung, 1965)	The SDS is a short self-rated questionnaire assessing affective, psychological, and somatic depression symptoms (Jokelainen *et al.*, 2019). It provides a way to monitor longitudinal changes in depression severity. Each item is scored on a Likert scale from 1 to 4, with the individual items summing to a total score (Zung, 1965). This score ranges from 20 to 80, with most individuals with depression scoring between 50 and 69 (Zung, 1965). A raw score above 40 indicates the presence of depression (Dunstan and Scott, 2019).
Neuropsychiatric Inventory Scale	NPI	Informant (By Clinician in Interview with Carer) (Burns *et al.*, 2002)	Yes	12	10–20 min (Cummings *et al.*, 1994)	The NPI assesses 10 behavioral and 2 neurovegetative areas, including delusions, hallucinations, agitation and/or aggression, depression and/or dysphoria, anxiety, elation and/or euphoria, apathy and/or indifference, disinhibition, irritation and/or lability, aberrant motor behavior, sleep and nighttime behavior disorders, and appetite and/or eating disorders. The interview is conducted with the caregiver, rather than the patient, to facilitate an open discussion of the patient's behaviors. The depression item consists of firstly a series of yes/no questions regarding the patient's behaviors. There are also four additional ratings present on the scale: frequency, severity, total (frequency × severity), and caregiver distress. Frequency ranges from rarely to very often, severity from mild to severe, and distress from not at all to very severely or extremely (Cummings *et al.*, 1994).
Patient Health Questionnaire-2	PHQ-2	Patient (Kroenke *et al.*, 2003)	No	2	<5 min (Kroenke *et al.*, 2003)	The PHQ-2 is a self-rated, two-item questionnaire that evaluates the frequency of depressed mood and anhedonia over the past 2 weeks prior to assessment (Kroenke *et al.*, 2003). The scale is made up of the first two items of the PHQ-9 and is presented in a yes/no style format (Li *et al.*, 2007). If the participant answers affirmatively to either of the questions, and thus scores positively on the scale, then the participant should subsequently be screened with the rest of the items on the PHQ-9 to determine if they have met the criteria for depressive disorder (Li *et al.*, 2007).

Table 2. Continued

Tool name	Abbreviation	Rater	Permission or copyright restrictions	Items (N)	Time to complete	Description of tool
Patient Health Questionnaire-9	PHQ-9	Patient (Kroenke *et al.*, 2003)	No	9	<5 min (Kroenke *et al.*, 2001)	The PHQ-9 is a self-rated, 9-item questionnaire that evaluates the frequency of depressed mood and anhedonia over the past 2 weeks prior to assessment (Kroenke *et al.*, 2003). Each item is scored from 0 to 3, making a highest attainable score of 27 (Kroenke *et al.*, 2001). An overall score above 9 indicates the likelihood of moderate to severe depression (Kroenke *et al.*, 2001).

of other diagnostic accuracy measures obtained are recorded in Table 1.

Discussion

We were able to identify five articles reporting the diagnostic accuracy of various depression tools within a defined MCI population (Artero *et al.*, 2008; Di Iulio *et al.*, 2010; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011), while one study by Boyle *et al.* (2011) validated two tools using a broader definition of MCI (Boyle *et al.*, 2011).

Seven tools were self-rated by patients (GDS-15 (Yesavage and Sheikh, 1986), CES-D (Radloff, 1977), BASDEC (Burns *et al.*, 2002), BDI-II, SDS (Jokelainen *et al.*, 2019), PHQ-2 (Kroenke *et al.*, 2003), PHQ-9 (Kroenke *et al.*, 2003), one was a caregiver and/or informant-rated scale (NPI) (Burns *et al.*, 2002), and the CSDD was both informant and patient-rated (Conradsson *et al.*, 2013) (Table 2). Moreover, we focused on the NPI as a tool to identify depressive symptoms, via its depression domain.

Currently, there is insufficient evidence to conclude which depression tool is most accurate within the context of MCI. Moreover, certain tools are not yet validated in the MCI population, such as the Hamilton Depression Rating Scale (HAM-D). At this time, the tool reported with the best balance and highest sensitivity and specificity was the CSDD by McCabe *et al.* (2006), with a sensitivity of 100% and a specificity of 94% (McCabe *et al.*, 2006). A sensitivity of 100% was also reported for the BASDEC, BDI-II, SDS, and the depression domain of the NPI (Di Iulio *et al.*, 2010; McCabe *et al.*, 2006), but specificity varied.

The CSDD was designed for the detection of depressive symptoms in persons with dementia and incorporates mood, physical symptoms, and collateral history (Alexopoulos *et al.*, 1988). The CSDD has previously been reported as a useful tool in the cognitively impaired population with high accuracy in persons with dementia (Goodarzi *et al.*, 2016). The CSDD does take more time to complete; however, in practice, this has been reasonable to implement, given that prior implementation has shown to be effective (Goodarzi and Watt, 2020). As such, the CSDD may be reasonable to use in clinics with a high proportion of persons with MCI; however, given the limited amount of literature, additional testing must be done to support this inference and determine whether the same accuracy is seen in the context of MCI as in dementia.

The GDS-15, which is commonly used in older adults, is based solely on the patients' responses (Yesavage and Sheikh, 1986). There were two

https://doi.org/10.1017/S1041610222000175 Published online by Cambridge University Press

Table 3. The QUADAS-2 risk of bias assessment for all included studies (n = 6)

	Artero *et al.*	Boyle *et al.*	di Lulio *et al.*	Dierckx *et al.*	McCabe *et al.*	Ros *et al.*
Was a consecutive or random sample of participants enrolled? (Yes/No/Unclear)	Yes	Yes	Yes	Yes	Yes	Yes
Was a case–control design avoided? (Yes/No/Unclear)	Yes	Yes	No	No	Yes	Yes
Did the study avoid inappropriate exclusions? (Yes/No/Unclear)	Yes	No	Yes	Yes	Yes	Yes
Could the selection of participants have introduced bias? (Low/High/Unclear)	Low	Unclear	Unclear	Unclear	Low	Low
Is there concern that the included patients do not match the review question? (Low/High/Unclear)	Low	Unclear	Low	Low	Low	Low
Were the index test results interpreted without knowledge of the results of the reference standard? (Yes/No/Unclear)	Unclear	Yes	Unclear	Unclear	Yes	Yes
If a threshold was used, was it prespecified? (Yes/No/Unclear)	Yes	No	Yes	Yes	Yes	Yes
Could the conduct or interpretation of the index test have introduced bias? (Low/High/Unclear)	Unclear	Low	Unclear	Unclear	Unclear	Unclear
Is there concern that the index test, its conduct, or interpretation differ from the review question? (Low/High/Unclear)	Unclear	Low	Unclear	Unclear	Unclear	Unclear
Is the reference standard likely to correctly classify the target condition? (Yes/No/Unclear)	Yes	Yes	Yes	Yes	Yes	Yes
Were the reference standard results interpreted without knowledge of the results of the index test? (Yes/No/Unclear)	Unclear	Unclear	Unclear	Unclear	Yes	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias? (Low/High/Unclear)	Unclear	Unclear	Unclear	Unclear	Low	Low
Is there concern that the target condition as defined by the reference standard does not match the review question? (Low/High/Unclear)	Low	Low	Low	Low	Low	Low
Was there an appropriate time interval between index test and reference standard? (Yes/No/Unclear)	Unclear	Yes	Unclear	Unclear	Yes	Unclear
Did all patients receive a reference standard? (Yes/No/Unclear)	Unclear	Yes	Yes	Yes	Yes	Yes
Did all patients receive the same reference standard? (Yes/No/Unclear)	Yes	Yes	Yes	Yes	Yes	Yes
Were all patients included in the analysis? (Yes/No/Unclear)	Unclear	No	Unclear	No	Yes	Yes
Could the patient flow have introduced bias? (Low/High/Unclear)	Unclear	Low	Unclear	Unclear	Low	Unclear

studies reporting on the accuracy of the GDS-15 tool (Dierckx *et al.*, 2007; McCabe *et al.*, 2006); however, there was a significant range in the reported sensitivity from 58% to 100%. McCabe *et al.* (2006) reported the highest sensitivity (100%) when administered with a cutoff of 6 (Dierckx *et al.*, 2007). More testing should be done to determine which diagnostic accuracy measures and cutoff point of the GDS-15 are most clinically relevant to the MCI population.

Limitations

Despite an exhaustive search, there were very few diagnostic accuracy studies identified that assess depression tools among persons living with MCI. Many studies used a depression tool but did not validate them, thus (n = 275) were excluded due to a lack of a reference standard or missing diagnostic accuracy outcomes (n = 73). Many studies evaluated a depression tool among the general geriatric population but did not include MCI as a subgroup.

The included studies varied in how they defined MCI, and most did not specify the MCI subtype (e.g. amnestic or non-amnestic impairment) (Artero *et al.*, 2008; Boyle *et al.*, 2011; Dierckx *et al.*, 2007; McCabe *et al.*, 2006; Ros *et al.*, 2011). As such, we lack evidence for validation of depression tools among different MCI subtypes or pathologies.

The risk of bias of included studies varied. Studies poorly described blinding of the index tests from reference standards, and similarly it was not clear what the interval was between measures. Both of these aspects can impact the study's precision at determining the index test's diagnostic accuracy. Accordingly, these unreported measures could have impacted the findings of our study.

The prevalence of depression was reported across studies. Despite having similar reference standards, the prevalence ranged from 3.1% to 51% (Artero *et al.*, 2008; Di Iulio *et al.*, 2010; McCabe *et al.*, 2006; Ros *et al.*, 2011). This range could be reflective of how studies examine different clinical care settings, variation in patient recruitment, or potential differences in MCI subtypes. For instance, Ismail *et al.* (2017) reported that the prevalence of depression in patients with MCI was 25% in community-based samples, compared to 42% in clinic-based samples (Ismail *et al.*, 2017). One proposed reason for this difference is that persons with depression are more likely to present to clinicians compared with non-depressed individuals, resulting in a higher reported prevalence of depression in a clinical population (Ismail *et al.*, 2017).

The included reference standards were a clinician's diagnosis, or any diagnosis of depression from any version of the DSM or ICD (The ICD-10 Classification of Mental and Behavioural Disorders, 1992). Amongst all studies, no collateral component was featured as part of the reference standard. As such, the reference standards are seen as patient-reported approaches to diagnosis. However, certain scales (i.e. CSDD (Radloff, 1977) and NPI (Cummings *et al.*, 1994)) were completely or partially informant-rated scales (Table 2). There thus might be a discrepancy when using the reference standard to assess the diagnostic accuracy of informant-rated scales.

Our study has several strengths. We followed all PRISMA checklist and Cochrane methods for systematic reviews (Moher *et al.*, 2015). A gray literature search was conducted alongside the database search to exhaust the search further. However, despite conducting an exhaustive search, it is possible that we could have missed literature that could impacted the results of our study.

Future Directions

Given the current lack of literature, more research must be done to evaluate the diagnostic accuracy of common depression tools in the context of MCI. The HAM-D is a commonly used tool used among persons with neurodegenerative disorders and is often considered the gold standard of observer-rated depression rating scales (Burke *et al.*, 2019). However, there is a lack of literature characterizing its accuracy amongst persons with MCI. Therefore, future research may target the HAM-D amid other important tools (i.e. CSDD, GDS, and so forth) for more rigorous evaluation of diagnostic accuracy.

Future studies examining accuracy of depression tools in persons with MCI should ensure clear criteria are used to identify the diagnosis of MCI, as well as the specific subtype of MCI symptoms, and report the blinding of the index tool to reference standard. Given the findings of our review, future work should examine the use of tools such as the CSDD in the MCI population to corroborate the above findings.

Conclusion

There are few tools validated to evaluate depressive symptoms in individuals with MCI. The CSDD was reported to have a high sensitivity and specificity, whilst several other tools (i.e. BASDEC, BDI-II, SDS, and the depression domain of the NPI) reported high sensitivity but variable specificity. Additional research is needed to make a more rigorous conclusion on which tools are the most accurate at depression detection.

https://doi.org/10.1017/S1041610222000175 Published online by Cambridge University Press

Acknowledgements

This project was funded by the Strategic Clinical Network (SCN) Summer Studentship Award as part of Alberta Health Services (AHS). We would also like to give a special thanks to librarian Lorraine Toews at the University of Calgary for assisting us with developing the search strategy.

Conflict of interest

BW has received funding via the Strategic Clinical Network (SCN) Summer Studentship Award with Alberta Health Services. ZG holds independent peer-reviewed project funding from the Canadian Institutes of Health Research (CIHR), Brenda Strafford Foundation, Hotchkiss Brain Institute (HBI), and O'Brien Institute of Public Health at the University of Calgary. ZI holds voluntary positions as Chair of the Canadian Conference on Dementia, and the Canadian Consensus Conference on the Diagnosis and Treatment of Dementia, but no conflict of interests are associated with either position.

Author contributions

Britney Wong was responsible for developing the search strategy and was a main abstract and full-text screener. She also wrote the manuscript. Zahinoor Ismail was responsible for abstract and full-text screening in duplicate with Britney Wong and provided edits to the manuscript. Zahra Goodarzi was responsible for developing the search strategy, screened abstracts and full texts in duplicate with Britney Wong, and provided edits to the manuscript.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1041610222000175

References

Adshead, F., Cody, D. D. and Pitt, B. (1992). BASDEC: a novel screening instrument for depression in elderly medical inpatients. *BMJ (Clinical Research ed.)*, 305, 397. DOI 10.1136/bmj.305.6850.397.

Albert, M. S. *et al.* (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 7, 270–279. DOI 10.1016/j.jalz.2011.03.008.

Alexopoulos, G. S. *et al.* (1988). Cornell scale for depression in Dementia. *Biological Psychiatry*, 23, 271–284. DOI 10.1016/0006-3223(88)90038-8.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: Diagnostic and Statistical Manual of Mental Disorders*. 5th edition, Arlington, VA: American Psychiatric Association.

Artero, S. *et al.* (2008). Risk profiles for mild cognitive impairment and progression to dementia are gender specific. *Journal of Neurology, Neurosurgery, and Psychiatry*, 79, 979–984. DOI 10.1136/jnnp.2007.136903.

Boyle, L. L. *et al.* (2011). How do the PHQ-2, the PHQ-9 perform in aging services clients with cognitive impairment? *International Journal of Geriatric Psychiatry*, 26, 952–960. DOI 10.1002/gps.2632.

Burke, A. D. *et al.* (2019). Diagnosing and treating depression in patients with Alzheimer's disease. *Neurology and Therapy*, 8, 325–350. DOI 10.1007/s40120-019-00148-5.

Burns, A., Lawlor, B. and Craig, S. (2002). Rating scales in old age psychiatry. *The British Journal of Psychiatry: The Journal of Mental Science*, 180, 161–167. DOI 10.1192/bjp.180.2.161.

Conradsson, M. *et al.* (2013). Usefulness of the Geriatric Depression Scale 15-item version among very old people with and without cognitive impairment. *Aging & Mental Health*, 17, 638–645. DOI 10.1080/13607863.2012.758231.

Csukly, G. *et al.* (2016). The differentiation of amnestic type MCI from the non-amnestic types by structural MRI. *Frontiers in Aging Neuroscience*, 8, 52. DOI 10.3389/fnagi.2016.00052.

Cummings, J. L. *et al.* (1994). The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in Dementia. *Neurology*, 44, 2308–2314. DOI 10.1212/wnl.44.12.2308.

Davidson, T. *et al.* (2019). Beck depression inventory. *The Gale Encyclopedia of Mental Health: A-C*, 188–190.

Di Iulio, F. *et al.* (2010). Occurrence of neuropsychiatric symptoms and psychiatric disorders in mild Alzheimer's disease and mild cognitive impairment subtypes. *International Psychogeriatrics*, 22, 629–640. DOI 10.1017/S1041610210000281.

Dierckx, E. *et al.* (2007). Differentiation between mild cognitive impairment, Alzheimer's disease and depression by means of cued recall. *Psychological Medicine*, 37, 747–755. DOI 10.1017/S003329170600955X.

Dunstan, D. A. and Scott, N. (2019). Clarification of the cut-off score for Zung's self-rating depression scale. *BMC Psychiatry*, 19, 177. DOI 10.1186/s12888-019-2161-0.

Ganguli, M. *et al.* (2010). Prevalence of mild cognitive impairment by multiple classifications: the Monongahela-Youghiogheny Healthy Aging Team (MYHAT) project. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 18, 674–683. DOI 10.1097/JGP.0b013e3181cdee4f.

Goodarzi, Z. *et al.* (2016). Guidelines for dementia or Parkinson's disease with depression or anxiety: a systematic review. *BMC Neurology*, 16, 244. DOI 10.1186/s12883-016-0754-5.

Goodarzi, Z. and Watt, J. (2020). Bridging knowledge gaps in the detection and treatment of neuropsychiatric symptoms in dementia: a review of current evidence'.

In: CAGP-CCSMH Annual Scientific Meeting; 20/20: Clear Vision of the Future of Seniors' Mental Health in Canada. Calgary, Canada, http://www.cagp.ca/resources/Documents/Book%20of%20Abstracts%20-%20FINAL.pdf.

Ismail, Z. *et al.* (2017). Prevalence of depression in patients with mild cognitive impairment: a systematic review and meta-analysis. *JAMA Psychiatry*, 74, 58–67. DOI 10.1001/jamapsychiatry.2016.3162.

Jokelainen, J. *et al.* (2019). Validation of the Zung self-rating depression scale (SDS) in older adults. *Scandinavian Journal of Primary Health Care*, 37, 353–357. DOI 10.1080/02813432.2019.1639923.

Krell-Roesch, J. *et al.* (2019). Cortical β-amyloid burden, neuropsychiatric symptoms, and cognitive status: the Mayo Clinic Study of Aging. *Translational Psychiatry*, 9, 123. DOI 10.1038/s41398-019-0456-z.

Kroenke, K., Spitzer, R. L. and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. DOI 10.1046/j.1525-1497.2001.016009606.x.

Kroenke, K., Spitzer, R. L. and Williams, J. B. W. (2003). The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care*, 41, 1284–1292. DOI 10.1097/01.MLR.0000093487.78664.3C.

Langa, K. and Levine, D. (2015). The diagnosis and management of mild cognitive impairment: a clinical review. *JAMA*, 312, 2551–2561. DOI 10.1001/jama.2014.13806.

Li, C. *et al.* (2007). Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *Journal of the American Geriatrics Society*, 55, 596–602. DOI 10.1111/j.1532-5415.2007.01103.x.

Making the Case for Investing in Mental Health in Canada (2013). Mental Health Commission of Canada. *Available at:* https://www.mentalhealthcommission.ca/wp-content/uploads/drupal/Investing_in_Mental_Health_FINAL_Version_ENG_0.pdf.

Ma, L. (2020). Depression, anxiety, and apathy in mild cognitive impairment: current perspectives. *Frontiers in Aging Neuroscience*, 12, 9. DOI 10.3389/fnagi.2020.00009.

Mariani, E., Monastero, R. and Mecocci, P. (2007). Mild cognitive impairment: a systematic review. *Journal of Alzheimer's Disease*, 12, 23–35. DOI 10.3233/jad-2007-12104.

McCabe, M. P. *et al.* (2006). Depression among older people with cognitive impairment: prevalence and detection. *International Journal of Geriatric Psychiatry*, 21, 633–644. DOI 10.1002/gps.1538.

Miao, R. *et al.* (2021). Plasma β-amyloid in mild behavioural impairment - neuropsychiatric symptoms on the Alzheimer's Continuum. *Journal of Geriatric Psychiatry and Neurology*, 6, 8919887211016068. DOI 10.1177/08919887211016068.

Moher, D. *et al.* (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLOS Medicine*, 6, 1–6. DOI 10.1371/journal.pmed.1000097.

Moher, D. *et al.* (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4, 1. DOI 10.1186/2046-4053-4-1.

Petersen, R. C. *et al.* (1997). Aging, memory, and mild cognitive impairment. *International Psychogeriatrics*, 9, 65–69. DOI 10.1017/s1041610297004717.

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401. DOI 10.1177/014662167700100306.

Roberts, R. and Knopman, D. S. (2013). Classification and epidemiology of MCI. *Clinics in Geriatric Medicine*, 29, 753–772. DOI 10.1016/j.cger.2013.07.003.

Ros, L. *et al.* (2011). Factor structure and psychometric properties of the center for epidemiologic studies depression scale (CES-D) in older populations with and without cognitive impairment. *International Journal of Aging & Human Development*, 72, 83–110. DOI 10.2190/AG.72.2.a.

Smarr, K. L. and Keefer, A. L. (2011). Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care & Research*, 63, S454–S466. DOI 10.1002/acr.20556.

The ICD-10 Classification of Mental and Behavioural Disorders (1992). World Health Organization.

Whiting, P. F. *et al.* (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155, 529–536. DOI 10.7326/0003-4819-155-8-201110180-00009.

Yesavage, J. and Sheikh, J. (1986). 9/Geriatric Depression Scale (GDS). *Clinical Gerontologist*, 5, 165–173. DOI 10.1300/J018v05n01_09.

Zung, W. W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12, 63–70. DOI 10.1001/archpsyc.1965.01720310065008.