SYSTEMATIC REVIEW

Screening instruments for cognitive impairment in older patients in the Emergency Department: a systematic review and meta-analysis

Agneta H. Calf[1], Maaike A. Pouw[1,2], Barbara C. van Munster[1,3], Johannes G. M. Burgerhof[4], Sophia E. de Rooij[1,5], Nynke Smidt[1,4]

[1]Department of Geriatrics, University Medical Center Groningen, Groningen, The Netherlands
[2]Department of Internal Medicine, Martini Hospital, Groningen, The Netherlands
[3]Department of Geriatrics, Gelre Hospitals, Apeldoorn, The Netherlands
[4]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[5]Medical Spectrum Twente, Medical School Twente, Enschede, The Netherlands

Address correspondence to: Maaike A. Pouw, University Medical Center Groningen, Department of Geriatrics, Hanzeplein 1,
PO Box 30 001, AA43, 9700 RB Groningen, The Netherlands. Tel/Fax.: +31625649286, Email: m.a.pouw@umcg.nl
†Agneta H. Calf and Maaike A. Pouw contributed equally to this manuscript.

Abstract

Background: cognitive impairment is highly prevalent among older patients attending the Emergency Department (ED) and is associated with adverse outcomes.

Methods: we conducted a systematic review and meta-analysis to evaluate the diagnostic accuracy of cognitive screening instruments to rule out cognitive impairment in older patients in the ED. A comprehensive literature search was performed in MEDLINE, EMBASE, CINAHL and CENTRAL. A risk of bias assessment using QUADAS-2 was performed.

Results: 23 articles, examining 18 different index tests were included. Only seven index tests could be included in the meta-analysis. For ruling out cognitive impairment irrespective of aetiology, Ottawa 3 Day Year (O3DY) (pooled sensitivity 0.90; (95% CI) 0.71–0.97) had the highest sensitivity. Fourteen articles focused on screening for cognitive impairment specifically caused by delirium. For ruling out delirium, the 4 A's Test (4AT) showed highest sensitivity (pooled sensitivity 0.87, 95% confidence interval (95% CI) 0.74–0.94).

Conclusions: high clinical and methodological heterogeneity was found between included studies. Therefore, it is a challenge to recommend one diagnostic test for use as a screening instrument for cognitive impairment in the ED. The 4AT and O3DY seem most promising for ruling out cognitive impairment in older patients attending the ED. The review protocol was registered in PROSPERO (CRD42018082509).

Keywords: Emergency Department, screening, cognitive impairment, dementia, delirium, older people

Key Points

- Cognitive impairment in older patients attending the Emergency Department is associated with adverse outcomes.
- Different screening tools for cognitive impairment have been developed and validated in the past decades.
- Screening for cognitive impairment enables to apply interventions and care adjustment, to prevent adverse outcomes.

Introduction

Cognitive impairment is present in 26% of the older patients attending the Emergency Department (ED) and can be caused by delirium, dementia or both [1]. Delirium is an acute and fluctuating change in mental status, characterized by impaired cognition, an altered level of consciousness and inattention and is caused by an underlying medical illness

[2]. Dementia is a chronic condition of impaired cognition, but patients with dementia can also develop delirium: delirium superimposed on dementia (DSD) [3]. Delirium and dementia are highly related syndromes: patients with dementia are at risk for delirium, but frequently dementia is not yet known, on the other hand patients with delirium can develop dementia. Recognising cognitive impairment irrespective of cause is of importance since the presence of cognitive impairment increases the risk of older patients to become hospitalised and once admitted to the hospital, they are at risk for longer hospital stay, progressive functional and cognitive decline, increased mortality and of being institutionalised, compared to patients without cognitive impairment [4–8]. Recognising delirium in the ED is of additional importance, because it is always caused by an underlying medical cause and this should be diagnosed and treated promptly. In this perspective, the altered mental status can be considered as a vital parameter [9]. Often, medical professionals are not aware of co-existing cognitive impairment and the provided (hospital) care is not suited to the demands of cognitive impaired patients [10]. Since most older patients enter the hospital through the ED, it would be favourable to recognise patients with cognitive impairment without delay in order to start interventions immediately to prevent adverse effects [7]. Unfortunately, cognitive impairment in the ED is frequently missed by health care professionals [11]. For instance, emergency physicians recognised cognitive impairment in only 38% of the older patients with either delirium or cognitive impairment without delirium [12]. An easy to apply screening tool for cognitive impairment irrespective of cause in the ED, preferably a single assessment, seems most useful [7]. Previous systematic reviews on screening instruments addressing the detection of cognitive impairment, aimed to identify delirium screening tools, focused on delirium or dementia instead of cognitive impairment in general, and others discussed a different patient population such as hospitalised patients instead of ED patients [13–15]. The aim of this systematic review and meta-analysis is to evaluate the diagnostic test accuracy of screening tools to rule out cognitive impairment in older patients in the ED. Identifying by ruling out results in a higher proportion of false positives, but simultaneously enhances the probability of patients with cognitive impairment receiving appropriate care to prevent adverse outcomes from happening.

Methods

This study was conducted according to the methods of the Cochrane Collaboration and reported according to PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses [16,17]. The review protocol was registered in PROSPERO (CRD42018082509).

Search strategy

A systematic literature review was conducted by searching the electronic databases of MEDLINE, EMBASE, CINAHL, and the Cochrane Central Register of Controlled Trials (CENTRAL), from inception to 3 March 2020. The

search string was developed in collaboration with a library information specialist (S.W., see acknowledgements) and contained search terms related to cognitive impairment, delirium, older patient population and the ED setting (See Appendix 1 for complete search strategy). The references and citations of included studies and relevant reviews on this topic were screened for potential eligible studies.

Study selection

Studies were considered eligible for review when they met the following criteria:

- Cohort study or case-control study
- Study population consisted of patients with a mean or median age 65 years or older, visiting an ED. Studies conducted in a different environment than the ED were excluded.
- The target condition was cognitive impairment irrespective of the aetiology. Ideally, the diagnosis was based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria (version III, IV, IV-R, V) made by a specialist in geriatric care [2,18–20]. The Confusion Assessment Method (CAM) and the Mini-Mental State Examination (MMSE) were accepted as a substitute gold standard because of their widely use in clinical practice [21,22].
- The index test was an instrument to assess cognition in the ED.
- The study provided sufficient data to construct a two-by-two table.

Three independent reviewers screened identified papers on title and abstract for eligibility (MP and HB screened the MEDLINE search results, AC and MP screened the results found in EMBASE, CINAHL and CENTRAL. The full text selection was performed by two reviewers independently (A.C., M.P.).

Data extraction and quality assessment

Two reviewers (A.C., M.P.) independently assessed the risk of bias and extracted data from the included studies with regard to study characteristics (design, in- and exclusion criteria); study population (age, proportion female, country); characteristics of the index test and reference standard (e.g. cut-off point) and outcome data (sensitivity, specificity, two-by-two table). In case the results of a study were reported in different articles, we treated the data as being independent when the test results of the participants, and not the participants themselves were the unit of analysis. Authors were contacted to obtain additional information in case of unreported data. The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy studies 2 (QUADAS-2) tool, which evaluates the risk of bias and applicability of diagnostic accuracy studies [23]. An additional item adapted from the QUADAS-1 criteria; "Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)?" was added to the
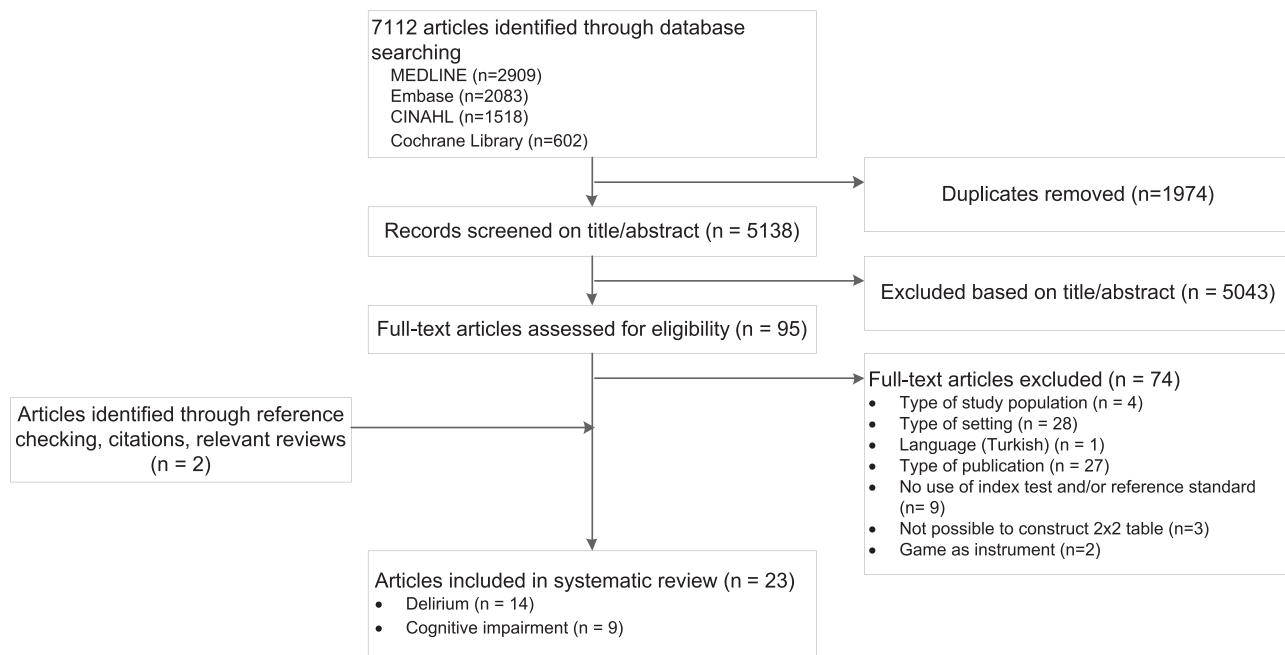
Figure 1. PRISMA flow diagram.

assessment [24]. Disagreements regarding the quality assessment were discussed in consensus meetings. The percentage of agreement and kappa statistic was calculated for the overall risk of bias.

Statistical analysis

Based on the chosen target condition of the different studies, cognitive impairment had to be categorised into caused by delirium and cognitive impairment irrespective of aetiology (i.e. delirium, dementia and DSD). The two-by-two table of each diagnostic test reported was used to calculate sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios. In case an index test was conducted twice in a study by different assessors, we selected the diagnostic test which was conducted by a health care professional (e.g. physician or nurse). The results were graphically displayed by use of coupled forest plots, representing sensitivity and specificity.

Meta-analysis was conducted if data of two or more studies reporting on the same screening instrument (index test) for the same target condition were included. Data, sensitivities and specificities were pooled if the same screening instrument was used and all inclusion criteria were met. In case an index test was performed twice or more in the same study cohort, the test results (the unit of analysis) were included in the meta-analysis only once. For this meta-analysis, a bivariate random effects model was used [25]. In order to investigate heterogeneity, we used Cochran's Q. In case the P-value of Cochran's Q was below 0.05, heterogeneity was present. Sources of heterogeneity were explored by conducting subgroup analyses including used screening instrument (index test), the used reference

standard (DSM criteria versus MMSE versus CAM), and risk of bias (per item). All statistical analyses were conducted in R, package "mada" [26].

Results

Included studies

Our literature search identified 7,112 articles. After screening the titles and abstracts of the identified articles, 95 articles were eligible for full text review (Figure 1), 74 of these articles were excluded. Main reasons for exclusion were different setting than the ED (n = 28) and type of publication (e.g. conference abstracts) (n = 27). Two studies regarding a serious game for cognitive assessment in the ED as an index test were excluded due to the chosen study methods which were not designed as diagnostic test accuracy studies. Data of these studies are provided in Appendix 2 [30–34]. Screening references of included articles and relevant reviews and citations of the included articles resulted in two additional eligible articles. In total, 23 articles were included in this systematic review. Five of these articles were part of one study design which investigated several screening instruments in the same study population [30–34]. Two other articles described two different screening instruments in the same study cohort [35,36]. One article investigated a screening instrument by use of a subanalysis on data of the same study cohort [37,38]. Nine articles evaluated >1 diagnostic test, and 37 data sets were available for the systematic review and meta-analysis.

Study characteristics

Twenty-three articles were included and described a total of 23 studies in 17 different study cohorts. The main

Table 1. Characteristics of included studies

	Country	Included in analysis (N)	Period of recruitment	Age Mean (sd)/ median (IQR)	Female (%)	Indextest, (*treshold*)	Reference standard, (*treshold*)
Cognitive impairment caused by delirium							
Baten (2018) [44]	Germany	288	2016	78 (74–82)	55	bCAM (*test+*)	DSM-V
Bédard (2019) [35]	Canada	313	2016	76.8 (7.5)	53	O3DY (*>0*)	CAM
Fabbri (2001) [39]	Brazil	100	1996–1997	73.8 (0.8)	48	CAM (*test+*)	DSM-IV
Gagné (2018) [36]	Canada	313	2016	76.8 (7.5)	52	4AT (*≥4*)	CAM
Grossmann (2017) [38]	Switzerland	286	2015	79.9 (72.4–86.7)	59	mRASS (*≠ 0*)	DSM-IV-TR
Han (2013) [30]	USA	406	2009–2012	73.5 (69–80)	50	DTS (*test+*) bCAM (*test+*)	DSM-IV-TR
Han (2014) [31]	USA	406	2009–2012	73.5 (69–80)	50	CAM-ICU (*test+*)	DSM-IV-TR
Han (2015) [32]	USA	406	2009–2012	73.5 (69–80)	50	RASS (*≠ 0*)	DSM-IV-TR
Han (2018) [33]	USA	406	2009–2012	73.5 (69–80)	50	SQ patient SQ surrogate	DSM-IV-TR
Hasemann (2018) [45]	Switzerland	286	2015	80.0 (72.2–86.8)	59	mCAM-ED (*test+*)	DSM-IV-TR
Hasemann (2019) [37]	Switzerland	286	2015	80.0 (72.2–86,8)	59	MOYB (*> 0 error*)	DSM-IV-TR
Marra (2018) [34]	USA	235	2010–2012	74 (69–79)	46	MOYB (*> 0 error*)	DSM-IV-TR
Meeberg (2016) [40]	Netherlands	53	2012	78.5 (6.9)	49	CAM-ICU (*test+*)	DSM-IV-TR
Shenkin (2019) [52]	UK	395	2015–2016	81 (77–86)	56	4AT (*≥4*) CAM (*test+*)	DSM-IV
Cognitive impairment							
Barbic (2018) [46]	Canada	117	2016	81.9 (5.7)	45	O3DY (*> 0 error*) SBT (*>4*)	MMSE (*≤ 23*)
Carpenter (2011) [47]	USA	163	2009–2010	78 (8)	61	O3DY (*> 0 error*) SBT (*≥ 4*) BAS (*<26*) cAD8 (*≥2*)	MMSE (*≤ 23*)
Carpenter (2011) [41]	USA	319	2008–2009	76 (NA)	58	SIS (*≥2*) AD8 (*≥ 2*)	MMSE (*≤ 23*)
Dyer (2016) [48]	Ireland	196	2014	78.5 (5.9)	46	AMT4 (*> 0 error*)	MMSE (*≤ 26*), CAM-ICU +
O'Sullivan (2017) [51]	Ireland	419	2015	77 (NA)	51	6-CIT (*≥10*) 4AT (*1–3*) 6-CIT (*≥10*)4AT (*≥4*)	DSM-V DSM-V
Schofield (2009) [49]	UK	520	2007	77 (NA)	NA	AMT10 AMT4	MMSE (*≤ 23*)
Wilber (2005) [42]	USA	75	2003	75.4 (6.6)	54	SIS (*≤ 4*) Mini-Cog (*test+*)	MMSE (*≤ 23*)
Wilber (2008) [43]	USA	352	2006–2007	77 (8)	63	SIS (*≤ 4*)	MMSE (*≤ 23*)
Wilding (2016) [50]	Canada	238	2010	81.9 (NA)	60	O3DY (*≤ 3*) AFT (*<15*)	MMSE (*≤ 24*)

NA: not available; bCAM: brief Confusion Assessment Method; DSM: Diagnostic and Statistical Manual of Mental Disorders; O3DY: Ottawa 3DY; CAM: Confusion Assessment Method; 4AT: 4 A's Test; mRASS: modified Richmond Sedation Scale, DTS: Delirium Triage Screen; CAM-ICU: CAM-Intensive Care Unit; RASS: Richmond Sedation Scale; SQ: single question for delirium screen; mCAM-ED: modified CAM-Emergency Department; MOYB: Months Of the Year Backwards; SBT: Short Blessed Test; MMSE: Mini-Mental State Examination; BAS: Brief Alzheimer's Screen; cAD8: caregiver Ascertain Dementia 8-item Questionnaire; SIS: Six-item Screener; AMT: Abbreviated Mental Test; TR: Text Revision;; AFT: Animal Fluency Test; 6-CIT: 6-Item Cognitive Impairment Test.

characteristics of the included studies are presented in Table 1. The index tests were conducted by physicians in 12 studies. (52%) [30–34,39–44]. In other studies, a research assistant or nurse assessed the index test [45–52]. The reference standard was assessed by a psychiatrist or geriatrician in 13 studies (57%) [30–34,37,39,40,44–46,52]. In other studies, a research assistant conducted the reference standard [35,36,41–43,47–50,53]. In one study, an ED geriatric nurse conducted the reference standard [51].

Delirium

Fourteen studies focused on cognitive impairment specifically caused by delirium, and one study investigated a screening instrument and categorised the results into chronic cognitive impairment and delirium [52]. Mean or median age of the study population included in the studies ranged from 73.5- to 81.0-years old. Forty-six to 59% of the participants were female.

Almost all studies used the DSM criteria for delirium as reference standard, two studies used the CAM. Various index tests were investigated (Table 1). Two articles reported on the brief CAM (bCAM) and a modified brief CAM (modified

bCAM), but were performed in the same cohort [30,54]. Therefore, the data were used only once in the meta-analysis.

Cognitive impairment

Nine studies reported on a screening instrument for cognitive impairment irrespective the underlying aetiology (i.e. dementia, delirium, DSD, mild cognitive impairment), one of these studies investigated a screening instrument and categorised into chronic cognitive impairment and delirium. Mean or median age of the study population included in the studies ranged from 75.4- to 81.9-years old. The percentage of females participating in the studies ranged from 45% to 63%. Eight studies used the MMSE as reference standard with cut-off values varying from ≤26 to <24 points and one study used the DSM-criteria for dementia. Ten different index tests were investigated (Table 1). The 6-CIT test and SBT appeared to be identical tests [47,48,52].

Risk of bias assessment of included studies

Results of the risk of bias assessment of the included studies according to the QUADAS-2 criteria are presented in

Cognitive impairment caused by delirium

Study	TP	FP	FN	TN	Index test	Reference standard	Sensitivity (95% CI)	Specificity (95% CI)
Grossmann 2017 [38]	14	18	6	247	(m)RASS	DSM	0.70 [0.46, 0.88]	0.93 [0.89, 0.96]
Han 2015 [32]	41	53	9	303	(m)RASS	DSM	0.82 [0.69, 0.91]	0.85 [0.81, 0.89]
Gagne 2018 [36]	44	108	5	162	4AT	CAM	0.90 [0.78, 0.97]	0.60 [0.54, 0.66]
O'Sullivan 2017 [51]	54	25	4	267	4AT	DSM	0.93 [0.83, 0.98]	0.91 [0.88, 0.94]
Shenkin 2019 [52]	37	19	12	324	4AT	DSM	0.76 [0.61, 0.87]	0.94 [0.91, 0.97]
O'Sullivan 2017 [51]	44	42	9	23	6-CIT/SBT	DSM	0.83 [0.70, 0.92]	0.35 [0.24, 0.48]
Fabbri 2001 [39]	16	3	1	80	CAM	DSM	0.94 [0.71, 1.00]	0.96 [0.90, 0.99]
Shenkin 2019 [52]	17	1	25	341	CAM	DSM	0.40 [0.26, 0.57]	1.00 [0.98, 1.00]
Han 2014 [31]	36	5	14	351	CAM-ICU	DSM	0.72 [0.58, 0.84]	0.99 [0.97, 1.00]
Meeberg 2016 [40]	12	1	0	40	CAM-ICU	DSM	1.00 [0.74, 1.00]	0.98 [0.87, 1.00]
Han 2013 [30]	49	161	1	195	DTS	DSM	0.98 [0.89, 1.00]	0.55 [0.49, 0.60]
Hasemann 2019 [37]	20	90	0	176	MOYB	DSM	1.00 [0.83, 1.00]	0.66 [0.60, 0.72]
Marra 2018 [34]	21	101	4	109	MOYB	DSM	0.84 [0.64, 0.95]	0.52 [0.45, 0.59]
Bedard 2019 [35]	16	123	3	171	O3DY	CAM	0.84 [0.60, 0.97]	0.58 [0.52, 0.64]
Han 2018 [33]	31	75	19	281	SQ patient	DSM	0.62 [0.47, 0.75]	0.79 [0.74, 0.83]
Han 2018 [33]	31	55	3	183	SQ surrogate	DSM	0.91 [0.76, 0.98]	0.77 [0.71, 0.82]
Baten 2018 [44]	30	15	16	227	bCAM	DSM	0.65 [0.50, 0.79]	0.94 [0.90, 0.96]
Han 2013 [30]	42	15	8	341	bCAM	DSM	0.84 [0.71, 0.93]	0.96 [0.93, 0.98]
Hasemann 2018 [45]	182	35	44	259	mCAM-ED	DSM	0.81 [0.75, 0.85]	0.88 [0.84, 0.92]

Cognitive impairment

Study	TP	FP	FN	TN	Index test	Reference standard	Sensitivity (95% CI)	Specificity (95% CI)
O'Sullivan 2017 [51]	69	109	13	187	4AT	DSM	0.84 [0.74, 0.91]	0.63 [0.57, 0.69]
O'Sullivan 2017 [51]	60	70	14	224	6-CIT/SBT	DSM	0.81 [0.70, 0.89]	0.76 [0.71, 0.81]
Carpenter 2011 [47]	57	36	3	67	6-CIT/SBT	MMSE	0.95 [0.86, 0.99]	0.65 [0.55, 0.74]
Barbic 2018 [46]	12	43	2	60	6-CIT/SBT	MMSE	0.86 [0.57, 0.98]	0.58 [0.48, 0.68]
Wilding 2016 [50]	29	125	3	81	AFT	MMSE	0.91 [0.75, 0.98]	0.39 [0.33, 0.46]
Schofield 2009 [49]	171	20	55	274	AMT	MMSE	0.76 [0.70, 0.81]	0.93 [0.90, 0.96]
Dyer 2016 [48]	52	3	47	94	AMT4	MMSE	0.53 [0.42, 0.63]	0.97 [0.91, 0.99]
Schofield 2009 [49]	182	35	44	259	AMT4	MMSE	0.81 [0.75, 0.85]	0.88 [0.84, 0.92]
Carpenter 2011 [47]	57	49	3	54	BAS	MMSE	0.95 [0.86, 0.99]	0.52 [0.42, 0.62]
Wilber 2005 [42]	12	9	4	50	Mini-Cog	MMSE	0.75 [0.48, 0.93]	0.85 [0.73, 0.93]
Carpenter 2011 [47]	57	50	3	53	O3DY	MMSE	0.95 [0.86, 0.99]	0.51 [0.41, 0.61]
Wilding 2016 [50]	30	56	2	150	O3DY	MMSE	0.94 [0.79, 0.99]	0.73 [0.66, 0.79]
Barbic 2018 [46]	10	45	4	58	O3DY	MMSE	0.71 [0.42, 0.92]	0.56 [0.46, 0.66]
Carpenter 2011 [41]	86	48	29	159	SIS	MMSE	0.75 [0.66, 0.82]	0.77 [0.70, 0.82]
Wilber 2005 [42]	17	8	1	48	SIS	MMSE	0.94 [0.73, 1.00]	0.86 [0.74, 0.94]
Wilber 2008 [43]	70	47	41	194	SIS	MMSE	0.63 [0.53, 0.72]	0.80 [0.75, 0.85]
Carpenter 2011 [47]	29	21	6	35	cAD8	MMSE	0.83 [0.66, 0.93]	0.63 [0.49, 0.75]
Carpenter 2011 [41]	34	17	18	63	cAD8	MMSE	0.65 [0.51, 0.78]	0.79 [0.68, 0.87]
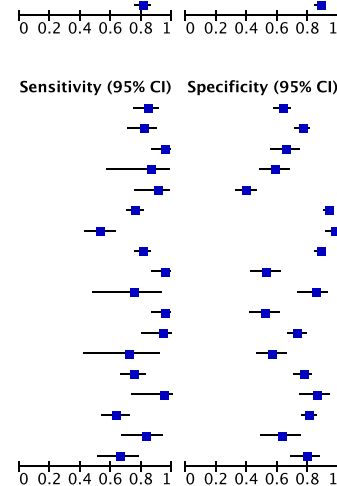
Figure 2. Coupled forest plots of diagnostic test accuracy. mRASS: modified Richmond Sedation Scale; DSM: Diagnostic and Statistical Manual of Mental Disorders; 4AT: 4 A's Test; 6-CIT: 6-Item Cognitive Impairment Test; SBT: Short Blessed Test; CAM: Confusion Assessment Method; CAM-ICU: CAM-Intensive Care Unit; DTS: Delirium Triage Screen; MOYB: Months Of the Year Backwards; SQ: single question for delirium screen to patient/surrogate; bCAM: brief CAM; mCAM-ED: modified CAM-Emergency Department; MMSE: Mini-Mental State Examination; AFT: Animal Fluency Test; AMT: Abbreviated Mental Test; BAS: Brief Alzheimer's Screen; O3DY: Ottawa 3 Day Year; SIS: Six-item Screener; cAD8: caregiver Ascertain Dementia 8-item Questionnaire.

Appendices 3 and 4. In total, 16 of the 23 articles had a high or unclear assessment of risk of bias in at least one domain. Frequent shortcomings in the studies were the absence of a consecutive or random sample of patients (20 articles, 15 studies), the person conducting the index test was not blinded for the result of the reference standard (6 articles; 5 studies) or the assessor of the reference standard was not blinded for the result of the index test (8 articles; 7 studies). The inter-rater agreement on the risk of bias assessment was excellent with an overall agreement 87% (222/253 items); kappa statistic 0.69 (95% confidence interval [CI] 0.58–0.79).

Diagnostic test accuracy

Main results

The included tests for identifying cognitive impairment caused by delirium or any type of cognitive impairment in the ED show a wide variance in sensitivity (resp. 0.62–1.00 and 0.53–0.95) and specificity (resp. 0.55–0.96 and 0.39–0.97). In Figure 2, the sensitivity and specificity per test and per study with their coupled forest plots for detection of cognitive impairment are presented.

Cognitive impairment caused by delirium

The summary estimate of the sensitivity and specificity of the 4 A's Test (4AT) was 86.9% (95% CI 73.5–94.1) and 86.9% (95% CI 59.6–96.5), respectively. For (m)RASS, the summary estimate of the sensitivity was 76.7% (95% CI 58.4–88.5) and for the specificity was 89.7% (95% CI 78.6–95.4). Pooled sensitivity and specificity per index test are shown in Appendix 5. As a result of the wide variance of index tests of the include studies screening for cognitive, conducting a meta-analysis for the other index tests was not possible. High heterogeneity was found between the results
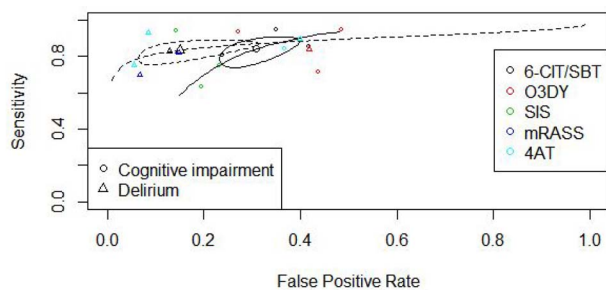
Figure 3. SROCs and estimates per index test; 6-CIT: 6-Item Cognitive Impairment Test; SBT: Short Blessed Test; O3DY: Ottawa 3 Day Year; SIS: Six-item Screener, (m)RASS: (modified) Richmond Sedation Scale, 4AT.

of the studies evaluating the accuracy of assessing delirium (Cochran's Q of 17,212; $P < 0.001$).

Cognitive impairment

The summary estimate of the sensitivity and specificity of the Ottawa 3 Day Year (O3DY) was 89.8% (95% CI 70.6–97.0) and 60.9% (95% CI 47.0–73.2), respectively. For respectively, the 6-CIT/SBT and SIS the summary estimate of the sensitivity was 89.1% (95% CI 78.2–94.9) and 71.5% (95% CI 58.5–81.8), and for the specificity was 67.2% (95% CI 55.8–76.9) and 79.2% (95% CI 75.1–82.8). Pooled sensitivity and specificity per index test are shown in Appendix 5. As a result of the wide variance of index tests of the included studies for cognitive impairment, conducting a meta-analysis for the other index tests was not possible. High heterogeneity was found between the results of the studies evaluating the accuracy of assessing cognitive impairment (Cochran's Q of 5,235, $P < 0.001$).

Exploring heterogeneity

Heterogeneity of the included studies was explored with subgroup analyses. The choice of reference standard (DSM versus MMSE versus CAM) did affect the sensitivity and specificity of the used test ($P < 0.05$). Studies using the MMSE or CAM as a reference standard showed a higher sensitivity and lower specificity compared to studies using the DSM as a reference standard. The presence of a high or unclear risk of bias, based on the QUADAS-2 criteria, did not alter the results. Furthermore, the presence of a dependency between the index test and reference standard (item adapted from the QUADAS-1 criteria) did not influence the results ($P = 0.87$). For the index tests consisting of more than two observations, we constructed a summary receiver operating characteristic (SROC; Figure 3).

Discussion

In this review, a total of 23 articles on the diagnostic test accuracy of screening instruments for cognitive impairment

in older patients in the ED have been evaluated. The ideal screening instrument has high sensitivity to rule out cognitive impairment and should be easy to integrate into daily ED practice [54]. For ruling out cognitive impairment as a global construct, the O3DY had highest sensitivity. From the studies focusing on cognitive impairment caused by delirium, the 4AT seems most promising. In a busy ED setting where time and resources are perceived as limited, preferably only one test should be assessed. Although it was not possible to include in the meta-analysis because of less than two studies reporting on the same screening instrument for the same target condition, O'Sullivan et al. described good sensitivity of the 4AT in chronic cognitive impairment (84.1%) and Bédard et al. of the use of the O3DY in delirium (84.2%) [35,51]. The 4AT and O3DY both seem to be robust enough to detect cognitive impairment, whether or not caused by delirium and since the 4AT and O3DY do not require extensive additional training of the assessor, the 4AT and O3DY seem to be the most practical screening instruments in an ED.

Some issues regarding the topic of this systematic review need to be addressed. First, most of the cognitive screening tools are designed to detect either delirium or chronic cognitive impairment such as dementia. As approximately half of the older people in the ED with dementia also have delirium, as delirium is often superimposed on dementia, perceiving this difference is not always easy, although relevant [7]. To be able to apply interventions, detection of cognitive impairment, whether it is caused by delirium, dementia or DSD, is of importance. Nonetheless, recognising delirium in the acute setting is of additional importance to detect and treat the possible causative medical disorders. As a first step, an easy to apply screening tool for cognitive impairment is useful, preferably one single assessment. An assessment such as the 4AT provides basic cognitive testing, aimed at detecting moderate–severe cognitive impairment, alongside assessment for delirium [51]. After recognising the presence of a cognitive disorder with help of a screening tool, further cognitive evaluation is needed to distinguish the aetiology at a later moment, for example within 24 h after hospitalisation. Second, in the included studies, the DSM criteria as well as the MMSE and CAM were used as reference standard. The DSM criteria are created to classify (neuro)cognitive disorders where the MMSE and CAM are created as cognitive screening instruments themselves. As shown in the subgroup analysis, the choice of reference standard did affect the results of the meta-analysis. Studies using the MMSE or CAM as a reference standard reported a higher sensitivity and lower specificity of the investigated index test compared to studies using the DSM as a reference standard. This can be explained by the fact that the index test is compared to a more extended screening instrument (i.e. MMSE or CAM), which are often developed for optimal sensitivity. We expected the more comprehensive assessment and often more trained assessors of the DSM would improve diagnostic test accuracy, but this is reflected only in improved specificity. To be able to compare and generalise results more easily, future

studies should use the same reference standard, preferably the DSM criteria.

Third, all index tests included in this review serve the same purpose of screening for cognitive impairment and consist of similar items, such as assessment of orientation, attention and memory. In addition, index tests are often based on pre-existent disease criteria, such as the DSM-criteria or are based on a shorter version of an extensive, already validated, instrument (e.g. CAM–ICU). Thus additionally to the overlap between the different index tests, there is an overlap between index tests and the reference standard. We expected overlap between index test and reference standard would result in higher diagnostic test accuracy, because of similarity in the items being tested. To overcome this incorporation bias, the item adapted from the QUADAS-1 criteria was added as one of the factors in the analysis. Although we assumed this item: "Was the reference standard independent of the index test?" would be of influence, the effect was not significant when expressed in a p-value. This can be explained by the unequal distribution of the assessment itself. Of the total of the 37 analysed datasets from the 23 different studies, this item was adjudicated negative in 35 of the cases, meaning in 35 of the cases the reference standard was not completely independent of the index test. Therefore, we cannot confirm our assumption [16]. Fourth, in the risk of bias assessment, one of the QUADAS-2 domains includes the question: "Did the study avoid inappropriate exclusions?" Most of the studies used exclusion criteria such as "severe intellectual disability" or "not able to communicate" [38,51]. Although these criteria are justly for excluding patients from participation, they could lead to exclusion of patients with severe or end-stage cognitive impairment which would lead to an underestimation of the test results.

Fifth, in a hectic and busy environment such as the ED time is of importance. In this review, we focused on diagnostic test accuracy of the different index tests, although of similar importance to clinical applicability is the administration time. For future studies, it would be of additional value to add data on administration time of the investigated index test.

Strengths and limitations

This review used a comprehensive search strategy with additional hand search of references and cross-references of included studies. Two independent reviewers screened potential studies for inclusion and extracted data, reducing the potential risk of bias. Regarding the quality assessment of the included studies, an excellent agreement was reached between the two reviewers.

One of the limitations of this review is the heterogeneity caused by the use of multiple index tests. Most of the index tests were being used in one study, resulting in a single observation per index test. Meta-analyses on all included studies were therefore not performed.

Furthermore, diagnostic tests based on a continuous variable used a cut-off value to classify the results as being either positive or negative. A cut-off value is sometimes chosen arbitrarily and does not reflect the range of cognitive impairment possible. It could occur that cognitive impairment is of clinical relevance although not meeting the pre-set cut-off value.

Conclusion

Due to clinical and methodological heterogeneity of the included studies, it is challenging to determine one diagnostic test for use as a screening instrument for cognitive impairment. In this systematic review, the 4AT or O3DY seems to be most promising for identification of cognitive impairment irrespective of aetiology in the ED. To answer the main question of this review, an accuracy study similar to the study of Carpenter et al. comparing multiple screening instruments to a single reference standard in a single population would be preferable [47]. Unfortunately, a study with the aforementioned design including all index tests, will not be feasible due to overlap of items in the tests, high burden for participating patients and too time-consuming for an ED. Heterogeneity between studies is therefore not preventable because researchers, patients and setting will differ and a review provides the highest possible grade of evidence.

References

A full list of references is available as Appendix 6.

2. American psychiatric association. Diagnostic and statistical manual of mental disorders. fifth edition. Arlington, VA: American Psychiatric Association, 2013.
18. American psychiatric association. Diagnostic and statistical manual of mental disorders. third edition. Washington, DC: American Psychiatric Association, 1980.
19. American psychiatric association. Diagnostic and statistical manual of mental disorders. fourth edition. Washington, DC: American Psychiatric Association, 1994.

20. American psychiatric association. Diagnostic and statistical manual of mental disorders, fourth edition, text revision. Washington, DC: American Psychiatric Association, 2000.

22. Folstein MF, Folstein SE, McHugh PR. Mini-mental state A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12: 189–98.

30. Han JH, Wilson A, Vasilevskis EE *et al.* Diagnosing delirium in older emergency department patients: validity and reliability of the delirium triage screen and the brief confusion assessment method. Ann Emerg Med 2013; 62: 457–65.

31. Han JH, Wilson A, Graves AJ *et al.* Validation of the confusion assessment method for the intensive care unit in older emergency department patients. Acad Emerg Med 2014; 21: 180–7.

32. Han JH, Vasilevskis EE, Schnelle JF *et al.* The diagnostic performance of the Richmond agitation sedation scale for detecting delirium in older emergency department patients. Acad Emerg Med 2015; 22: 878–82.

33. Han JH, Wilson A, Schnelle JF, Dittus RS, Ely EW. An evaluation of single question delirium screening tools in older emergency department patients. Am J Emerg Med 2018; 36: 1249–52.

34. Marra A, Jackson JC, Ely EW *et al.* Focusing on inattention: the diagnostic accuracy of brief measures of inattention for detecting delirium. J Hosp Med 2018; 13: 551.

35. Bédard C, Boucher V, Voyer P *et al.* Validation of the O3DY french version (O3DY-F) for the screening of cognitive impairment in community seniors in the emergency department. J Emerg Med 2019; 57: 59–65.

36. Gagné A, Voyer P, Boucher V *et al.* Performance of the french version of the 4AT for screening the elderly for delirium in the emergency department. Canadian Journal of Emergency Medicine 2018; 20: 903–10.

37. Hasemann W, Grossmann FF, Bingisser R *et al.* Optimizing the month of the year backwards test for delirium screening of older patients in the emergency department. Am J Emerg Med 2019; 37: 1754–7.

38. Grossmann FF, Hasemann W, Kressig RW, Bingisser R, Nickel CH. Performance of the modified Richmond agitation sedation scale in identifying delirium in older ED patients. Am J Emerg Med 2017; 35: 1324–6.

39. Fabbri RM, Moreira MA, Garrido R, Almeida OP. Validity and reliability of the portuguese version of the confusion assessment method (CAM) for the detection of delirium in the elderly. Arq Neuropsiquiatr 2001; 59: 175–9.

40. Van de Meeberg EK, Festen S, Kwant M, Georg RR, Izaks GJ, Ter Maaten JC. Improved detection of delirium, implementation and validation of the CAM-ICU in elderly emergency department patients. Eur J Emerg Med 2017; 24: 411–6.

41. Carpenter CR, DesPain B, Keeling TN, Shah M, Rothenberger M. The six-item screener and AD8 for the detection of cognitive impairment in geriatric emergency department patients. Ann Emerg Med 2011; 57: 653–61.

42. Wilber ST, Lofgren SD, Mager TG, Blanda M, Gerson LW. An evaluation of two screening tools for cognitive impairment in older emergency department patients. Acad Emerg Med 2005; 12: 612–6.

43. Wilber ST, Carpenter CR, Hustey FM. The six-item screener to detect cognitive impairment in older emergency department patients. Acad Emerg Med 2008; 15: 613–6.

44. Baten V, Busch H, Busche C *et al.* Validation of the brief confusion assessment method for screening delirium in elderly medical patients in a german emergency department. Acad Emerg Med 2018; 25: 1251–62.

45. Hasemann W, Grossmann FF, Stadler R *et al.* Screening and detection of delirium in older ED patients: performance of the modified confusion assessment method for the emergency department (mCAM-ED). A two-step tool. Intern Emerg Med 2018; 13: 915–22.

46. Barbic D, Kim B, Salehmohamed Q, Kemplin K, Carpenter CR, Barbic SP. Diagnostic accuracy of the Ottawa 3DY and short blessed test to detect cognitive dysfunction in geriatric patients presenting to the emergency department. BMJ Open 2018; 8: e019652.

47. Carpenter CR, Bassett ER, Fischer GM, Shirshekan J, Galvin JE, Morris JC. Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: brief alzheimer's screen, short blessed test, Ottawa 3DY, and the caregiver-completed AD8. Acad Emerg Med 2011; 18: 374–84.

48. Dyer AH, Briggs R, Nabeel S, O'Neill D, Kennelly SP. The abbreviated mental test 4 for cognitive screening of older adults presenting to the emergency department. Eur J Emerg Med 2017; 24: 417–22.

49. Schofield I, Stott DJ, Tolson D, McFadyen A, Monaghan J, Nelson D. Screening for cognitive impairment in older people attending accident and emergency using the 4-item abbreviated mental test. Eur J Emerg Med 2010; 17: 340–2.

50. Wilding L, Eagles D, Molnar F *et al.* Prospective validation of the Ottawa 3DY scale by geriatric emergency management nurses to identify impaired cognition in older emergency department patients. Ann Emerg Med 2016; 67: 157–63.

51. O'Sullivan D, Brady N, Manning E *et al.* Validation of the 6-item cognitive impairment test and the 4AT test for combined delirium and dementia screening in older emergency department attendees. Age Ageing 2018; 47: 61–8.

52. Shenkin SD, Fox C, Godfrey M *et al.* Delirium detection in older acute medical inpatients: a multicentre prospective comparative diagnostic test accuracy study of the 4AT and the confusion assessment method. BMC Med 2019; 17: 138.

53. Han JH, Wilson A, Graves AJ, Shintani A, Schnelle JF, Ely EW. A quick and easy delirium assessment for nonphysician research personnel. Am J Emerg Med 2016; 34: 1031–6.