Chapter 2

Attention, arousal and other rapid bedside screening instruments for delirium in older patients: a systematic review of test accuracy studies

D.W.P. Quispel-Aggenbach, RN,
G.A. Holtman, PhD,
H.A.H.T. Zwartjes, MSc,
S.U. Zuidema, MD, PhD,
H.J. Luijendijk, MD, PhD.

Age and Ageing. 2018;47(5): 644-653. DOI: https://doi.org/10.1093/ageing/afy058

Abstract

Objective
Delirium occurs frequently in frail patients but is easily missed. Screening with a rapid, easy-to-use, and highly sensitive instrument might help improve recognition. The aim of this study was to review attention, arousal and other rapid bedside screening instruments for delirium in older patients.

Methods
A literature search was performed in PubMed, PsycINFO, and Embase. We scrutinized forward citations in Google Scholar, and references of included articles and prior reviews. We included studies among older patients that investigated the sensitivity and specificity of delirium screening instruments that could be administered in 3 minutes or less, and did not require surrogate information. We extracted study characteristics, risk of bias, sensitivity, and specificity.

Results
We identified 27 studies among 4766 patients in hospitals and nursing homes. They tested many different single and several combined screening instruments. Prevalence of delirium varied between 4% and 57%. Only one study scored a low risk of bias on all domains. Sensitivity varied between 17% and 100%, and specificity between 38% and 99%. Of the 22 tests with sensitivity ≥90%, seven also had specificity ≥80% in older patients in general. These results were approximately reproduced for the Observational Scale of Level of Arousal (OSLA) and Richmond Agitation and Sedation Scale (RASS): sensitivity and specificity were >80%.

Conclusion
Two arousal tests -OSLA and RASS- had reproduced high sensitivity and specificity in older patients. Nurses can administer these tests during daily interaction with patients. Test accuracy studies about rapid screening tools for delirium superimposed on dementia were scarce.

1. Introduction

Delirium is a serious neuropsychiatric disorder with potentially severe consequences such as longer hospital stay, poor cognitive and functional recovery, increased risk of nursing home placement and death (1). It occurs in 10-40% of older patients in hospitals and nursing homes (2,3). Frailty, age above 80 years and the presence of dementia increase the risk of delirium (4).

Around one third of delirium cases go undetected (5). The overlap with dementia and depression might hinder recognition, as might a history of psychiatric disease (6). Lack of trained health care professionals can also contribute to failure in identifying delirium (7). Screening frail older persons regularly may help detect delirium more quickly and has been advised in many guidelines (8). The goal of screening for a disease is to identify persons that are at increased risk of having that disease in a large population (triage). If screening tests are applied to detect diseases that are easily missed, as is the case with delirium, they need to be very sensitive (9–12). Usually, a screening test cannot be used to make a definitive diagnosis, because vital diagnostic information has not been collected (13). Subsequently, screen-positive patients need to receive a diagnostic work-up to confirm the diagnosis (9,12,14). Diagnostic tests need to be very specific (9–12). Ideally, relatively untrained personnel can perform a screening test quickly, easily, and as part of routine of their clinical practice.

A number of instruments have been developed to screen for delirium such as the DOSS, the CRS and the DSI (see list of abbreviations below). In addition, diagnostic tools for delirium such as the CAM and the DRS-R98 have been used to screen for delirium (15,16). These tests cover all diagnostic and many supporting criteria for delirium, including (surrogate) information about acute onset and fluctuation that patients with cognitive disorders cannot provide reliably. All of the above instruments require a lot of time to administer regularly. In addition, some screening tools such as the DOSS have not been validated in patients with dementia (17). The CAM and DRS-R98 require training, expertise and experience to be administered correctly. It is likely that the lack of an easy-to-use and rapid screening tool for delirium has hampered the implementation of regular screening (18).

In recent years, several screening tools with a test time of 3 minutes or less have been developed and validated. Such instruments may allow screening of many patients in relatively little time. The aim of this study was to review the sensitivity and specificity of rapid screening instruments for delirium in older patients.

Abbreviations	
4AT	4 Attention Tests
ALOC	Altered level of consciousness
AMT	Abbreviated Mental Test
AMT-4	4-point Abbreviated Mental Test
AMT-10	10-point Abbreviated Mental Test
CAM	Confusion Assessment Method
CDT	Clock Drawing Test
Cog-4	Cognitive examination derived from NIH Stroke Scale (NIHSS)
CRS	Confusion Rating Scale
DCT	Digit Cancellation Test
DCT-1	Digit Cancellation Test with a 1-digit matrix
DCT-2	Digit Cancellation Test with a 2-digit matrix

DOSS	Delirium Observation Screening Scale
DRS-R98	Delirium Rating Scale Revised 1998
DSB	Digit Span Backwards
DSI	Delirium Symptom Interview
DSF	Digit Span Forward
DST	Digit Span Test (DSF+DSB)
DTS	Delirium Triage Screen
DW	Day of the week
GAR	Global Attentiveness Rating
GCS	Glasgow Coma Scale
HDS	Hierarchic Dementia Scale
IPT	Intersecting Pentagons Test
Lunch BW	Lunch spelled backwards
MDAS	Memorial Delirium Assessment Scale
MOTYB	Months of the year recited backwards
MOTYF	Month of the year recited forwards
mRASS	Modified Richmond Agitation and Sedation Scale
Nu-DESC	Nursing Delirium Screening Scale
OSLA	Observational Scale of Level of Arousal
O3DY	Ottowa Day, Date, WORLD BW and Year
RADAR	Recognizing Acute Delirium As part of your Routine
RASS	Richmond Agitation and Sedation Scale
SAVEAHAART	S-A-V-E-A-H-A-A-R-T
Serial 3s	Serial threes subtraction test
Serial 7s	Serial sevens subtraction test
SSB	Spatial span backwards test
SS	Spatial span forward test
SQeeC	The Simple Query for Easy Evaluation of Consciousness
World BW	World spelled backwards

2. Methods

2.1 Search strategy and selection criteria

Two authors performed an independent literature search (DWPQ, HJL). First, they searched PubMed, Embase and PsycINFO with the search terms "delirium, acute confusion, encephalopathy, clouding of consciousness, toxic psychosis", " tool, test, instrument, assessment, questionnaire, interview, diagnostic, screening", and "sensitivity, specificity, accuracy, validity, reliability, predictive value, likelihood-ratio" (see online appendix). Secondly, they scrutinized references of the selected articles and four prior reviews (15,16,19,20). Thirdly, they performed a forward citation search in Google Scholar for each included article. Finally, they asked the authors of the included studies per email whether they knew unpublished studies. If title or abstract suggested that the study investigated the test accuracy of a rapid screening instrument for delirium, the full (un)published paper – if available – was obtained. Two authors assessed the papers independently (DWPQ, HJL) for eligibility. The search was finalized in December 12th 2017.

Studies were selected if they met the following inclusion criteria: a bedside screening instrument for delirium was tested; administration time was < 3 minutes as reported in the included or another article; the study reported sensitivity and specificity of a screening tool; and the study was

performed in patients aged 60 years or older. Exclusion criteria were: (index) tests to diagnose delirium (CAM, DRS-R98) or delirium tremens, or to rate the severity of delirium (MDAS) or the accompanying cognitive impairment (CTD); tests based on surrogate information because it generally takes more than 3 minutes to reach a caregiver and administer the test, and retrieving surrogate information is often unsuccessful (21); tests based on symptoms elicited during history taking; tests part of establishing the reference standard diagnosis; and studies performed in patients on mechanical ventilation. No restriction was made with respect to year of publication or language.

2.2 Data-extraction

Two authors (DWPQ, HJL or GAH, HJL) independently extracted the following study characteristics: setting, number of participants, prevalence of delirium and of dementia, the index test (screening instrument), the administrator and test-time of the index test, and reference standard (criteria used to diagnose delirium).

They also assessed risk of bias with the QUADAS-2 tool (22). This tool consists of four domains: patient selection, index test, reference standard, and flow of patients and timing of the index test and reference standard. In addition, the tool requires the assessment of the applicability of the patient population, index test and target condition. Risk of bias and applicability concerns were scored as low or high, or unclear if information was missing. We modified the assessment to fit the specifics of our review (see online appendix).

Finally, the test accuracy of the screening instruments in terms of sensitivity, specificity were extracted for all patients and patients diagnosed with dementia, as well as inter-rater reliability. Sensitivity and specificity concerned patient level data (not per assessment) and current delirium (if measured during a period, we used the day with highest delirium prevalence) for the tester with the lowest level of training in psychiatric assessment (in case of multiple testers) and the cut-off with highest sensitivity (in case of multiple cut-offs). When information about study characteristics or results was missing in the publication, we requested the author to provide it. Differences in data-extraction and risk of bias assessments were resolved in consensus meetings.

2.3 Statistical analysis

We presented the reported sensitivity, specificity, and inter-rater reliability of the tests in all patients, and patients with dementia. We found that confidence intervals around sensitivity and specificity were missing for a number of studies. Therefore, we extracted the raw data of these studies (number of true positives, false positives, true negatives, false negatives) and calculated the 95% confidence intervals with STATA 14.0. Results were not pooled across studies.

2.4 Declaration of sources of funding

The Dutch Ministry of Health supported this work (grant number 325414). The sponsor had no role in its design or conduct, interpretation of results, and reporting.

3. Results

The literature search yielded 6077 hits. The search in the online bibliographies yielded 84 potentially eligible articles, the forward citation search 67, references of reviews and articles 101, and responses of 18 authors 13. After exclusion of duplicates we assessed 68 full-texts for eligibility. Finally, 27 studies were included that were reported in 31 publications (figure 1) (9–

11,14,23–49). Most excluded studies did not report test accuracy of a rapid test (see online appendix for references).

3.1 Study characteristics

The studies investigated 1 to 20 different single or combined tests. MOTYB was studied most often (seven studies). Table 1 presents the key characteristics of the study designs. The setting was mostly a geriatric, surgical or acute care ward, or emergency department of a hospital. One study was performed in a consultation-liaison psychiatry service, one in a hospice and four studies in a nursing home. The number of participants varied between 14 and 500. The prevalence of delirium varied between 4% and 57%.

Table 2 shows the results of the risk of bias assessment. One of the 27 included studies had a low risk of bias on all items (30). Twelve studies scored reasonably well with only one or two domains with a high or unclear risk of bias. Fourteen studies scored a high risk of bias for selection of patients due to exclusion criteria that we deemed inappropriate such as previous diagnosis of dementia (25,34,44) or psychiatric illness (37,39,44,48), expected hospital stay of ≤2 days (11,24), patients in rehabilitation, respite care (23), ophthalmological, or gynaecological wards (46), and being too unwell or cognitively incapable to consent to participation (10,26,35,45). One study enrolled patients in office hours only (24) and another excluded patients older than 80 years (25). In addition, almost all studies requested patients to provide informed consent before inclusion, which might have led to exclusion of relatively severe cases of delirium. Significant heterogeneity existed in the professional background of the individuals performing the index-tests. Applicability concerns were low for most populations, screening tests and target conditions.

3.2 Test accuracy

Table 3 presents the sensitivity and specificity of the rapid screening instruments for delirium. Most were attention or level of arousal assessment tests. The test-time varied from 7 seconds for RADAR to 3 minutes for combinations of tests per assessment. All tests were described as easy and requiring minimal training (up to 45 minutes) and minimal clinical experience. The articles described how the tests needed to be rated and which cut-offs to use (see online appendix for content of tests).

Twenty-six studies reported results for mixed groups of patients with and without dementia. The sensitivity of single and combinations of tests varied between 17% and 100%. Twenty-two instruments had a sensitivity of 90% or higher. Of these tests, only the RASS + Lunch BW (DTS) had a lower confidence interval limit above 90%. The specificity of the tests varied between 14% and 100%. Of the tests with sensitivity of 90% or more only the AMT-4, DCT-2, GAR, MOTYB, OSLA, RASS and ‘writing name and address’ had specificity of 80% or more. Sensitivity results were reproduced for AMT-4, OSLA and RASS, but sensitivity and specificity results (approximately) only for OSLA and RASS (see online appendix).

Nine studies reported test accuracy of screening tools in patients with dementia. Sensitivity varied from 21% to 100%, and specificity from 15% to 96%. Eight tests had a sensitivity of 90% or higher, but only the OSLA + SAVEAHAART showed specificity of 80% or higher. None of the findings in patients with dementia have been reproduced consistently. In both groups ‘older patients in general’ and ‘patients with dementia’, six tests had high sensitivity of 90% or higher, but none had specificity of 80% or higher.

In general, confidence intervals around sensitivity and specificity were wide in most studies, indicating insufficiently large study populations. Most studies did not report inter-rater reliability, but if reported, it was generally high.

4. Discussion

We performed a systematic review of rapid and easy-to-administer screening instruments for delirium in older patients. The tools took 3 minutes or less to administer. The AMT-4, DCT-2, GAR, OSLA, RASS and 'writing name and address' had sensitivity above 90% and specificity above 80% in older patients in general. The OSLA + SAVEAHAART performed well in those with dementia.

4.1 Promising tests

Successful implementation of a screening delirium tool is affected by the administration time, the training required, the burden posed to the patient, and its appropriateness in the clinical setting it is used (12,16,21). To minimize the burden of screening on professionals, patients and resources, and maximise the number of cases found, we and other authors propose a two-step approach (12,30,35,47,50). A highly sensitive tool is needed in the first step to detect as many possible cases of delirium as possible (few false-negative cases), and a highly specific tool in the second step to make definitive diagnoses (few false-positive cases).

Most tests with sensitivity of 90% or more and specificity of 80% or more either require observation of level of arousal (GAR, RASS, OSLA), a combination of such a test with an attention test (OSLA + SAVEAHAART), or multiple cognitive tests (AMT-4, DCT-2). Sensitivity results were reproduced for AMT-4, OSLA and RASS, but specificity results only for OSLA and RASS in general older populations. Remarkably, both latter tests are level of arousal tests. The OSLA + SAVEHAART might perform well in terms of sensitivity and specificity in patients with dementia. Hence, level of arousal also seems to distinguish delirium from dementia. However, these study results have not always been reproduced and study populations were sometimes small.

There was no apparent relationship between the reported test accuracies and risk of bias, delirium criteria used, and prevalence of delirium. Naturally, reported test accuracies need to be interpreted with caution because most studies reported those that would correctly classify most patients, delirious or not. In other words, high sensitivity was not always the aim, and would have been achieved if lower specificity had been accepted. With high applicability of patient populations, index tests and target condition, indirectness is not a serious concern.

Our review complements the findings of a prior study about single-item screening questions for delirium (12). Such questions are short too but probe (subjective) symptoms of delirium such as confusion and hallucinations with the patient, surrogate or a health professional (11,33,51). Sensitivity was mostly poor, but specificity sometimes very high. Other reviews about delirium instruments did not focus on short tests and did not capture the recently published tests (7,15,16,19).

4.2 Methodological challenges

Performing a diagnostic test accuracy study in patients with delirium might be challenging. All studies required patients or their legal representatives to provide informed consent. It is likely that patients with delirium and their families will not give permission due to lack of cognitive and decisional capacity as easily as patients without delirium and their families (52). As a result,

patients with (severe) delirium may not have been represented sufficiently in the study populations. The use of exclusion criteria such as “included only in office hours” (9), “an expected hospital stay of ≤2 days” (11), “dementia with MMSE 10 or less” (26) and “not able to speak English” (9,35) might have negatively influenced the number and diversity of included delirium cases too (20). In addition, exclusion of patients with dementia might have led to overestimation of specificity, because symptoms of severe dementia overlap considerably with symptoms of delirium (26,33,35,37,43).

All studies were performed in hospitalized patients, except four studies that tested RADAR in nursing home patients and one study in a hospice (25,37,39,41,48). In one hospital and the nursing home studies, delirium prevalence was 4 to 7%, lower than estimates from prior observational studies (2). Cases may have been missed (53). Additional studies are needed in nursing homes and hospices. Due to the overlap between delirium and (neuropsychiatric symptoms of) dementia, diagnostic expertise is needed to ensure valid reference diagnoses.

When performing a diagnostic test study in delirium, researchers need to consider how to score untestable patients. In some studies, patients were excluded if they were considered too ill or too drowsy. Many of these patients might have had a delirium. Twelve studies reported that untestable patients were considered screen-positive (9-11,14,25,27,29,36,39–41,46). We agree with this approach. Delirium will probably be missed less often if untestable patients are scored as screen-positive.

4.3 Strengths and limitations

Strength of our study was that we performed a broad search with no restriction related to publication year or language. We used the internationally accepted QUADAS-2 tool to assess risk of bias in diagnostic test accuracy studies. There were two independent data extractors and they used a consensus procedure for disagreements. Our review meets the PRISMA criteria for reporting a review.

As we chose to exclude tests based on surrogate information, some relatively quick (diagnostic) tools were excluded, such as the bCAM (50), 3D-CAM (54), Nu-DESC (55) and 4AT (56). They seemed to perform (very) well in general older patient populations and patients with dementia. Serious games and mobile computerized tests present interesting options too (57,58). Another limitation of our study is that our results are not generalizable to non-elderly or ICU patients, because we did not include studies in such patients.

Finally, test accuracy studies do not measure outcomes of implementing screening tools. Professionals have reported that they do not always believe that screening will lead to better treatment (59). This is conceivable in younger patients with a clear underlying disease. Delirium in frail older patients, who often have multiple modifiable predisposing and precipitating conditions, would probably remain undetected and inadequately treated if it is not diagnosed (60,61). A diagnosis is also important for adequate psycho-education of patients, relatives and caregivers.

4.4 Conclusion

We identified 27 studies that investigated test accuracy of rapid and easy-to-administer bedside delirium screening instruments in older patients. All except one study had at least one source of potential bias. Two tests had high sensitivity and high specificity in more than one study among older hospitalized patients: the OSLA and RASS. Tests of arousal seemed to perform well in patients with dementia too, but results need to be reproduced in larger populations and long-term

care settings. The advantage of rapid and frequent screening by non-specialized personnel will be that only screen-positive patients need an extensive diagnostic work-up by a medical specialist.

5. Statements

5.1 Acknowledgments
None.

5.2 Statement of Ethics
The medical ethics committee of Erasmus Medical Centre Rotterdam, an academic research institute that we consulted, has approved the study protocol.

5.3 Disclosure Statement
The authors have no conflicts of interest to declare.

5.4 Author Contributions
Daisy Quispel wrote the manuscript. Sytse Zuidema provided feedback on the manuscript. Hendrika Luijendijk designed the study, supervised the study and the statistical analysis, and helped write the manuscript.

5.5 Data sharing
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References chapter 2

1. Witlox J, Eurelings LSM, Jonghe JFM, Kalisvaart KJ, Eikelenboom P & Van Gool WA. Delirium in Elderly Patients and the Risk of Postdischarge Mortality. JAMA. 2010;304(4):443–51.
2. Siddiqi N, House A & Holmes J. Occurrence and outcome of delirium in medical in-patients: A systematic literature review. Age Ageing. 2006;35(4):350–64.
3. De Lange E, Verhaak PFM & Van Der Meer K. Prevalence, presentation and prognosis of delirium in older people in the population, at home and in long term care: A review. Int J Geriatr Psychiatry. 2013;28(2):127–34.
4. Fong T, Tulebaev S & Inouye SK. Delirium in elderly adults: diagnosis, prevention and treatment. Nat Rev Neurol. 2009;5(4):210–20.
5. Wahid ZU. Delirium in Hospitalized Patients : Underdiagnosed and Undertreated Missed Opportunities in Preventive Care. J Natl Med Assoc. 2004;96(4):414.
6. Kishi Y, Kato M, Okuyama T, Hosaka T, Mikami K, Meller W & Kathol R. Delirium: patient characteristics that predict a missed diagnosis at psychiatric consultation. Gen Hosp Psychiatry. 2007;29(5):442–5.
7. Wong CL, Holroyd-Leduc J, Simel D & Straus Se. et al. Does This Patient Have Delirium? Value of Bedside Instruments. JAMA 2005;293(7):987–97.
8. National Clinical Guideline Centre for Acute and Chronic Conditions. Delirium: diagnosis, prevention and management. NICE Clin Guidel. 2010;7:10–4.
9. Han JH, Wilson A, Vasilevskis EE, Shintani A,Schnelle JF, Dittus RS & Ely E. 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(5):457–65.
10. O'Regan NA, Ryan DJ, Boland E, Connoly W, Mc Glade C, Leonard M & Timmons S. et al. Attention! A good bedside test for delirium? J Neurol Neurosurg Psychiatry. 2014;85(10):1122–31.
11. Fick DM, Inouye SK, Guess J, Ngo LH, Jones RN, Saczynski JS & Marcantonio ER. Preliminary Development of an Ultrabrief Two-Item Bedside Test for Delirium. J Hosp Med. 2015;10(10):645–50.
12. Hendry K, Hill E, Quinn TJ, Evans J & Stott DJ. Single screening questions for cognitive impairment in older people: a systematic review. Age Ageing. 2015;44(2):322-6.
13. Streiner DL. Diagnosing Tests: Using and Misusing Diagnostic and Screening Tests. J Pers Assess 2003;81(3):209–19.
14. Adamis D, Meagher D, Murray O, O'Neill D, O'Mahony E, Mulligan O & Mc GArthy G. Evaluating attention in delirium: A comparison of bedside tests of attention. Geriatr Gerontol Int 2016;16(9):1028–35.
15. Adamis D, Sharma N, Whelan PJP & Macdonald AJ. Delirium scales: A review of current evidence. Aging Ment Health 2010;14(5):543–55.
16. Grover S, Kate N. Assessment scales for delirium: A review. World J Psychiatry . 2012;2(4):58.
17. Schuurmans MJ, Deschamps PI, Markham SW, Shortridge-Bagget LM & Duursma SA. The Measurement of Delirium : Review of Scales. Res theory Nurs Pract An Int J. 2003;17(3):207–24.
18. Hasemann W, Tolson D, Godwin J, Spirig R, Frei IA & Kressig RW. A before and after study of a nurse led comprehensive delirium management programme (DemDel) for older acute care in-patients with cognitive impairment. International journal of nursing studies. 2016;53: 27-38.
19. Tieges Z, Brown LJE & MacLullich AMJ. Objective assessment of attention in delirium: A narrative review. Int J Geriatr Psychiatry. 2014;29(12):1185–97.

20. De J, Wand APF. Delirium Screening: A Systematic Review of Delirium Screening Tools in Hospitalized Patients. Gerontologist. 2015;55(6):1079–99.
21. O'Sullivan D, Brady N, Manning E, O'Shea E, O'Grady S, O'Regan N & Timmons S. 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. 2017;47(1):61-68.
22. Whiting PF, Rutjes AWS, Westwood ME, Mallet S, Deeks JJ, Reitsma JB & Bossuyt PMM. Research and Reporting Methods Accuracy Studies. Ann Intern Med. 2011;155(4):529–36.
23. Jitapunkel S, Pillay I & Ebrahim S. Delirium in New Admitted Elderly Patients: A Prospective study. QJM: An international Journal of Medicine. 1992;83(1):307–14.
24. Pompei P, Foreman M, Cassel CK, Alessi C & Cox D. Detecting Delirium Among Hospitalized Older Patients. Arch Intern Med. 1995;155(3):301–7.
25. Macleod A, Whitehead L. Dysgraphia and terminal delirium. Palliat Med. 1997;11(2):127–32.
26. O'Keeffe ST & Gosney MA. Assessing Attentiveness in Older Hospital Patients: Global Assessment Versus Tests of Attention. JAGS. 1997;45(4):470–3.
27. Adamis D, Reich S, Treloar A, Macdonald AJ & Martin FC. Dysgraphia in elderly delirious medical inpatients. Aging Clin Exp Res. 2006;18(4):334–9.
28. Bryson GL, Wyand A, Wozny D, Rees L, Taljaard M & Nathan H. The clock drawing test is a poor screening tool for postoperative delirium and cognitive dysfunction after aortic repair. Can J Anesth/J Can Anesth. 2011;58(3):267–74.
29. Leung JLM, Lee GTH, Lam YH, Chan RC & Whu JY. The use of the Digit Span Test in screening for cognitive impairment in acute medical inpatients. Int Psychogeriatr. 2011;23(10):1569–74.
30. Chester JG, Harrington MB, Rudolph JL & VA Delirium Working Group. Serial administration of a modified richmond agitation and sedation scale for delirium screening. J Hosp Med. 2012;7(5):450–3.
31. Emerson G, Carlson R, Nicolson S & Han J. The clinical utility of the clock drawing test in detecting delirium in older emergency department patients. Ann Emerg Med. 2014;64(4):S5.
32. Han JH, Vasilevskis EE, Schnelle JF, Shintani A, Dittus RS, Wilson A & Ely EW. The Diagnostic Performance of the Richmond. Acad Emerg Med. 2015;22(7):878–82.
33. Lees R, Corbet S, Johnston C, Moffitt E, Shaw G & Quin TJ. Test accuracy of short screening tests for diagnosis of delirium or cognitive impairment in an acute stroke unit setting. Stroke. 2013;44(11):3078–83.
34. Tieges Z, McGrath A, Hall RJ & MacLullich AM. Abnormal level of arousal as a predictor of delirium and inattention: An exploratory study. Am J Geriatr Psychiatry. 2013;21(12):1244–53.
35. Lin HS, Eeles E, Pandy S, Pinsker D, Brasch C, & Yerkovich S. Screening in delirium: A pilot study of two screening tools, the Simple Query for Easy Evaluation of Consciousness and Simple Question in Delirium. Australas J Ageing. 2015;34(4):259–64.
36. Shoaib A, Kaehr E & Malmstrom T. Mental status (MS) Vital sign: A Pictorial Facial Scale as a Screening Tool for delirium. In: Mental status (MS) Vital sign: A Pictorial Facial Scale as a Screening Tool for delirium. JAGS; 2015. s116.
37. Voyer P, Champoux N, Desrosiers J, Landreville P, Cusker J, Monette J & Carmichael PH. Recognizing acute delirium as part of your routine [RADAR]: a validation study. BMC Nurs. 2015;14(1):19.
38. Voyer P, Champoux N, Desrosiers J , LAnddreville P, Monette J, Savoie M & Bédard A. Assessment of inattention in the context of delirium screening: one size does not fit all! Int Psychogeriatrics. 2016;28(8):1293–301.
39. Bilodeau C, Voyer P. Radar : un outil valide pour le repérage du syndrome confusionnel aigu

(delirium) en résidences intermédiaires. NPG Neurol - Psychiatr - Geriatr. 2016;17(98):144–51.
40. Hendry K, Quinn TJ, Evans J, Scortichini V, Miller H,Burns j & Stott D. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. Age Ageing. 2016;45(6):832–7.
41. Koop R, Notter J. Delirium in nursing homes. Res Diss Master Sci Nurs / Adv Heal Care. 2016.
42. Leonard M, O'Connel H, Williams O, Awan F, Exton C, O'Conner M & Meagher DJ. Attention , vigilance and visuospatial function in hospitalized elderly medical patients : Relationship to neurocognitive diagnosis. J Psychosom Res. 2016;90:84–90.
43. O'Regan NA, Maughan K, Liddy N, Fitzgerald J, Adamis D, Molloy DW & Timmons S. Five short screening tests in the detection of prevalent delirium: diagnostic accuracy and performance in different neurocognitive subgroups. Int J Geriatr Psychiatry. 2017;32(12):1440–9.
44. Bédard C, Voyer P, Eagles D, Boucher V, Pelletier M, Gouin E & Emond M. LO57: Validation of the Ottawa 3DY in community seniors in the ED. CJEM. 2017;19(S1):S47.
45. Dyer AH, Briggs R, Nabeel S, O'Neill D & Kennely SP. The Abbreviated Mental Test 4 for cognitive screening of older adults presenting to the Emergency Department. Eur J Emerg Med. 2017;24(6):417–22.
46. 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(9):1324–6.
47. Hasemann W, Grossmann FF, Stadler R, Bingisser R, Breill D, Hafner M & Nickel CH. 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. 2017;13(6):915-922.
48. Pelletier I. L'utilisation du RADAR en centre d'hébergement et de soins de longue durée (CHSLD) pendant sept journées consécutives. 2017.
49. Richardson SJ, Davis DHJ, Bellelli G, Hasemann W, Meagher D, Kreisel SK & Morandi A. Detecting delirium superimposed on dementia: diagnostic accuracy of a simple combined arousal and attention testing procedure. Int Psychogeriatrics. 2017;29(10):1585–93.
50. Han JH, Wilson A, Graves AJ, Shintani A, Schenlle JF & Ely E. A quick and easy delirium assessment for nonphysician research personnel. Am J Emerg Med. 2016;34(6):1031–6.
51. Sands MB, Dantoc BP, Hartshorn A, Ryan CJ & Lujic S. Single Question in Delirium (SQiD): testing its efficacy against psychiatrist interview, the Confusion Assessment Method and the Memorial Delirium Assessment Scale. Palliat Med. 2010;24(6):561–5.
52. Downing L, Caprio T & Lyness JM. Geriatric psychiatry review: differential diagnosis and treatment of the 3 D's - delirium, dementia, and depression. J Fam Pract. 1993;37(6):325–8.
53. Laurila JV, Pitkala KH & Strandberg TE & Tilvis RS. Confusion assessment method in the diagnostics of delirium among aged hospital patients: would it serve better in screening than as a diagnostic instrument? Int J Geriatr Psychiatry. 2002;17(12):1112-9.
54. Marcantonio ER, Ngo LH, O'Connor M, Jones RN, Crane PK, Metzger ED & Inouye SK. 3D-CAM: Derivation and validation of a 3-minute diagnostic interview for CAM-defined delirium: A cross-sectional diagnostic test study. Ann Intern Med. 2014;161(8):554–61.
55. Gaudreau JD, Gagnon P, Harel F & Roy M. Impact on delirium detection of using a sensitive instrument integrated into clinical practice. Gen Hosp Psychiatry. 2005;27(3):194–9.
56. Bellelli G, Morandi A, Davis DHJ, Mazzola P, Tuco R, Gentile S & MacLullich AM. Validation of the 4AT, a new instrument for rapid delirium screening: A study in 234 hospitalised older people. Age Ageing. 2014;43(4):496–502.
57. Tieges Z, Stíobhairt A, Scott K, Suchorab K, Weir A, Parks S & MacLullich A. Development

of a smartphone application for the objective detection of attentional deficits in delirium. Int Psychogeriatrics. 2015;27(8):1251–62.
58. Tong T, Chignell M, Tierney MC & Lee JS. Test-Retest Reliability of a Serious Game for Delirium Screening in the Emergency Department. Front Aging Neurosci. 2016;8:1251–62.
59. Gesin G, Russell BB, Lin AP, Northon HJ, Evans Sl & Devlin JW. Impact of a delirium screening tool and multifaceted education on nurses' knowledge of delirium and ability to evaluate it correctly. Am J Crit Care. 2012;21(1):e1-e11.
60. Inouye SK, Cahrpentier PA. Precipitating Factors for Delirium in Hospitalized Elderly Persons. JAMA. 1996 Mar 20;275:852-857.
61. Stroomer-van Wijk AJM, Jonker BW, Kok RM, van der Mast RC & Luijendijk HJ. Detecting delirium in elderly outpatients with cognitive impairment. Int Psychogeriatrics. 2016;28(8):1303–11.

Figure 1 Flow diagram of literature search and selection

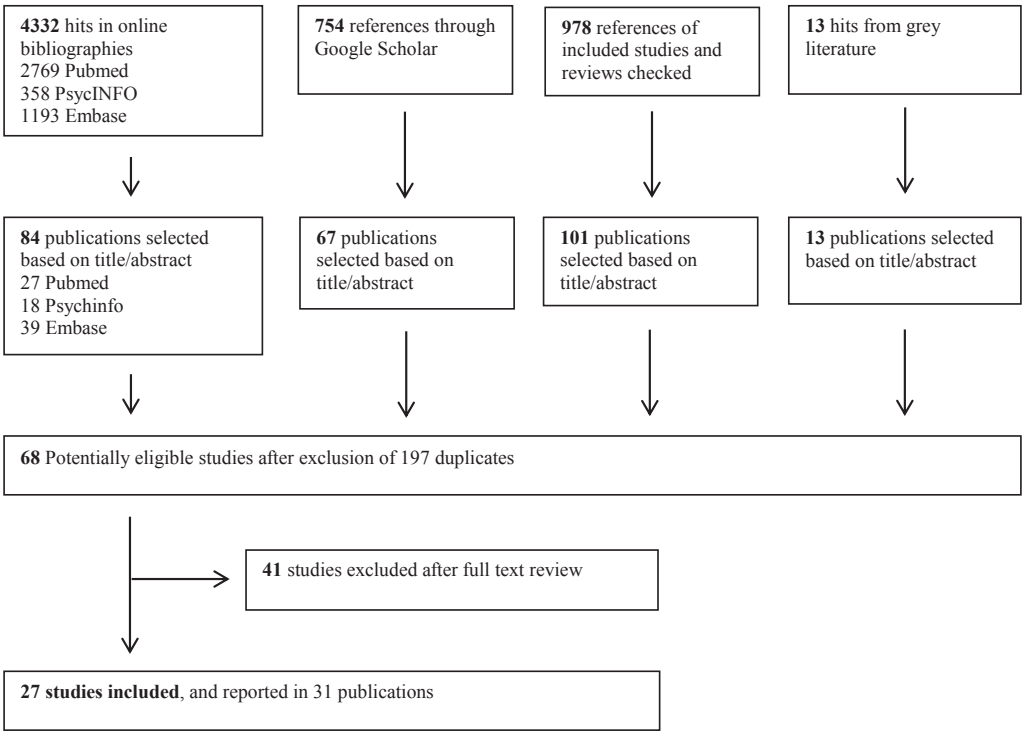

Table 1 Characteristics of included studies

Study	Index test	Tester	Reference standard	Study population			
				N	Setting (ward, type patients)	Delirium prevalence, %	Dementia prevalence, %
Jitapunkul 1992(23)	AMT-10	Researcher	DMS-III	184	Acute geriatric	22	18
Pompei 1995(24)	DSF, Vigilance A, DSF+Vigilance A	Research assistants	DSM-III-R	432	Medical and surgical	15	NT
Macleod 1997(25)	Writing name & address	Speech-language therapist	DSM-IV & DRS	20	Terminal cancer, hospice	NA (case-control)	0
O'Keeffe 1997(26)	GAR, DSF, DSB, Vigilance A, DCT-1, DCT-2	Geriatrician	DSM-III	90	Acute geriatric	21	24
Adamis 2006 (27)	Signature MMSE sentence	Neuropsychologist	CAM with DRS	94	Elderly Care Unit, hospital	32	NR
Bryson 2011(28)	CDT	Nurses trained in psychometric testing	CAM	88	Abdominal aorta surgery	26	NA
Leung 2011(29)	DSF, DSB, DSF+DSB	Nurses	DSM-IV	144	Acute medical and geriatric	18	22
Chester 2012(30)	mRASS	Nurse	DSM-IV	95	Tertiary VA hospital	11	NR
Han 2013(9) Emerson 2014(31) Han 2015(32)	RASS + Lunch BW (DTS) CDT RASS	Research assistant Emergency physician Research assistant	DSM-IV-TR	406	Emergency department	12	6
Lees 2013(33)	AMT-4, AMT-10, CDT, Cog-4, GCS	Medical student	CAM	111	Acute stroke	11	41
Tieges 2013(34)	OSLA, RASS	Graduate psychologist	CAM	30	Hip fracture	33	NA

O'Regan 2014(10)	MOTYB, SSF5, SSF5 then MOTYB	Junior medical staff	DSM IV	133*	One hospital$	NR*	NR*
Fick 2015(11)	20 single items of 3D-CAM and pairs of items, ALOC	Research assistants	DSM IV	201	General and geriatric medicine	21	28
Lin 2015(35)	SQeeC	Geriatrician in training	DSM IV	100	General medicine	12	30
Shoaib 2015(36)	Pictorial Facial Scale	Nurses/ nurse aids	DSM-IV	55	Acute geriatric	26	NT
Voyer 2015(37) Voyer 2016(38)	RADAR 10 items from HDS	Nurse or research assistant Research assistent	DSM-IV-TR (with CAM) DSM-V with CAM	142 51	Acute care hospital and nursing home	15 4	4 71
Adamis, 2016(14)	DST (DSF+DSB), Vigilance A, Serial 7s, MOTYB	Medical students (5th year)	CAM	200	Geriatric unit	17	63
Bilodeau 2016(39)	RADAR	Nurse-assistants	DSM-V	31	Nursing home	3	100
Hendry 2016 (40)	AMT-10, AMT-4, MOTYB	Nurse	DSM-V	500	Geriatric unit	19	32
Koop 2016(41)	RADAR	Nurse-assistants	CAM	14	Rehabilitation ward of nursing home	7	7
Leonard 2016(42)	World BW, MOTYB, SSF, SSB, Vigilance A, Vigilance B, CDT, IPT, and combinations	Trained raters/ psychiatrists	DRS-R98-severity >=15 or DSM-IV	193	Consultation-liaison psychiatry service	57	51
O'Regan 2016(43)	CDT, SSF, MOTYB, IPT	Medical expert	DRS-R98	470	Emergency department	39	25

Bedard 2017(44)	O3DY	Research assistant	CAM	305	Emergency department of 4 hospitals	NR	NR
Dyer 2017(45)	AMT-4	Research assistant	CAM-ICU	220	Emergency department	13	24
Grossmann 2017(46,47)	mRASS, MOTYB	Nurses	DSM-IV-TR	298	Emergency department	7	14
Pelletier 2017 (48)	RADAR	Nurse-assistants	DSM-V (with CAM)	45	Nursing home	4	93
Richardson 2017 (49)	OSLA, SAVEAHAART, OSLA+SAVEAHA ART	Delirium experts	DSM-V	114	Acute and rehabilitation hospitals	46	52

NR stands for not reported; NT for not tested; * subgroup of patients aged 69 or older; ** determined in patients with delirium and a random sample of patients without delirium (n=40); $ all wards except ED, ICU, and isolation rooms; NA not applicable (all or most patients with dementia excluded at entry)

Table 2 Risk of bias and applicability of included study

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient population	Index test	Target condition
Jitapunkul 1992	H	L	L	U	L	L	L
Pompei 1995	H	L	H	H	L	L	L
Macleod 1997	H	H	U	L	H	L	L
O'Keeffe 1997	L	H	L	H	L	L	L
Adamis 2006	L	L*	L	H	L	L	L
Bryson 2011	L	H	H	L	H	L	L
Leung 2011	H	H	L	L	L	L	L
Chester 2012	L	L	L	L	L	L	L
Han 2013	H	L	L	H	L	L	L
Lees 2013	L	L	U	L	H	L	L
Tieges 2013	H	H	H	U	H	L	L
O'Regan 2014	L	U	H	H	L	L	L
Fick 2015	H	L	L	L	L	L	H
Lin 2015	L	H	L	L	L	L	L
Shoaib 2015	U	L	L	U	L	L	L
Voyer 2015	H	L*	U*	L	L	H&^	L
Adamis 2016	H	H	H	U	L	L	L
Bilodeau 2016	H	U	U	U	L	H&	L
Hendry 2016	L	H	L	H	H	L	L
Koop 2016	L	L	U	U	L	H&	L
Leonard 2016	L	H	U	H	L	L	L
O'Regan 2016	L	H	H	H	L	L	L
Bedard 2017	H	H	H	U	L	L	L
Dyer 2017	U	U	H	L	L	L	L
Grossmann 2017	H	L	L	H	L	L	L
Pelletier 2017	H	U	U	U	L	H&	L
Richardson 2017	U	H	H	U	L	L	L

H stands for high, L for low and U for unclear; * high for tests other than GAR (Adamis 2006) or RADAR (Voyer 2016); & RADAR required 2 or more medication administrations per day; ^low for HDS

Table 3 Test characteristics of rapid screening instruments for delirium

Study	Test (cut-off)	Test time, minutes	Sensitivity, % (95% CI)		Specificity, % (95% CI)		Interrater Reliability, % agreement (kappa)
			All patients	In dementia	All patients	In dementia	
Jitapunkul 1992	AMT-10 (<8)	< 2	92 [78-98]	NR	65 [56-73]	NR	NT
Pompei 1995	DSF (<5)	< 2	34 [22-48]$	NT	90 [87-93]	NT	NT
	Vigilance A (> 2 errors)		61 [47-74]	NT	77 [73-81]	NT	NT
	DSF + Vigilance A (both failed)		26 [15-40]	NT	97 [95-99]	NT	NT
Macleod 1997	Writing name and address	< 1	100 [69-100]	NA	100 [69-100]	NA	NT
O'Keeffe 1997	GAR (<7)	Each <2	94 [73-100]	NR	99 [92-100]	NR	ICC = 0.83
	DSF (NR)		NR	NR	NR	NR	NR
	DSB (< 4)		83 [59-96]	NR	96 [88-99]	NR	NR
	Vigilance A (> 2 errors)		83 [59-96]	NR	83 [72-91]	NR	NR
	DCT-1 (NR)		NR	NR	NR	NR	NR
	DCT-2 (< 9 in 2 trials)		94 [73-100]	NR	87 [77-94]	NR	NR
Adamis 2006	Signature (abnormal)	< 0.5	54 [32-76]	NR	88 [76-96]	NR	NT
	MMSE sentence (abnormal)	< 0.5	NR	NR	NR	NR	
Bryson 2011	CDT (≤18)	2	66 (38-85)	NA	64 (52-76)	NA	NT
Leung 2011	DSF (<8)	< 2	58 [37-77]	NR	72 [63-80]	NR	NR
	DSB (<3)	< 2	81 [61-93]	NR	63 [53-71]	NR	NR
	DSF + DSB (NR)	3	NR	NR	NR	NR	NR
Chester 2012	mRASS (≠ 0)	< 0.5	64 (52-76)	NR	93 (90-96)	NR	98 (0.48)
Han 2013	RASS (≠ 0) + Lunch BW (>1 error)	< 1	98 (90-100)	NR	56 (51-61)	NR	89 (0.79)
Emerson 2014	CDT (abnormal)	< 2	94 [84-99]	NR	44 [39-49]	NR	NR (0.84)
Han 2015	RASS (≠ 0)	< 0.3	84 (74-94)	NR	88 (84-91)	NR	NR (0.63)
Lees 2013	AMT-4 (<4)	Each	83 (52-98)	NR	61 (51–71)	NR	NT
	AMT-10 (<8)	< 2	75 (43-95)	NR	61 (51–71)	NR	NT
	CDT (<3 on 0-3 scale)		67 (22-96)	NR	38 (28–49)	NR	NT

	Cog-4 (>0)		70 (35–93)	NR	44 (35–55)	NR	NT
	GCS (eye open<4 & verbal response<5)		17 (02–48)	NR	81 (71–88)	NR	NT
Tieges 2013	OSLA (>3)	< 1	90 [56-100]	NA	90 [68-99]	NA	NT
	RASS (≠ 0)	< 1	90 [56-100]	NA	85 [62-97]	NA	NT
O'Regan 2014	MOTYB (1 error up to July)	< 2	84 (68-94)	88 (68-97)	90 (82-95)	71 (29-96)	NR
	SSF (<5)	1-1.5	95 (82-99)	NR	58 (48-68)	NR	NR
	SSF (<5), then MOTYB (1 error up to July)	< 3	81 (65-92)	NR	91 (83-96)	NR	NR
Fick 2015	MOTYB (1 error) #	< 2	83 (69-93)	89 (72-98)	69 (61-76)	61 (41-78)	NR
	DSB (<4)	< 2	83 (69-93)	86 (67-96)	52 (44-60)	54 (34-72)	
	Day of week (an inadequate answer)	< 1	71 (55-84)	75 (55-89)	92 (87-96)	75 (55-89)	
	Day of week + MOTYB (1 fail) &	< 3	93 (81- 99)	96 (82-100)	48 (40-56)	43 (24-63)	
	Day of week + DSB (1 fail)	< 3	93 (81-99)	93 (76-99)	48 (40-56)	39 (22-59)	
	DSB + MOTYB (1 fail)	< 3	93 (81-99)	93 (76-99)	42 (34-50)	39 (22-59)	
	ALOC	< 2	19 (9-34)	21 (8-41)	99 (96-100)	NR	
Lin 2015	SQeeC (an inadequate answer)	1	83 (52-98)	83 (36-99)	81 (72-89)	59 (36-79)	NT
Shoaib 2015	Pictorial Facial Scale (≠ 0)	< 1	86 [57-98]	NT	67 [49-80]	NT	76 (.41)
Voyer 2015	RADAR 1-4x daily (>0 item present)	0.5	65 (43-84)	NR	71 (64-78)	NR	82-98 (.34-.79)
	RADAR 3-4x daily (>0 item present)	0.5	73 (39-94)	71 (29-96)	67 (57-76)	43 (26-61)	
Voyer 2016	Serial 7s (failure)*	< 1	96 (78-100)	NR	14 (9-20)	NR	72-100 (.30-1.00)
	Serial 3s (failure)	< 1	87 (66-97)	NR	47 (39-55)	NR	
	MOTYB (failure)	< 1	83 (61-95)	63 (24-91)	63 (55-70)	79 (71-86)	
	Days of week B (failure)	< 1	48 (27–69)	NR	85 (79–90)	NR	
	Counting 93>85 (failure)	< 1	48 (27–69)	NR	85 (79–90)	NR	
	Counting 10>1 (failure)	< 1	30 (13–53)	NR	87 (81–92)	NR	
	MOTYF (failure)	< 1	26 (10–48)	NR	89 (84–94)	NR	
	Days of week F (failure)	< 1	26 (10–48)	NR	90 (84–94)	NR	
	Counting 1>10 (failure)	< 1	17 (5–39)	NR	91 (86–95)	NR	
	Counting 10 objects (failure)	< 1	17 (5–39)	NR	94 (89–97)	NR	
Adamis 2016	DST (>1)	< 2	74 (55-87)	NT	62 (54-69)	NT	90% during pre-study training
	Vigilance A (>2 errors)	< 2	82 (65-93)	NT	60 (52-68)	NT	
	Serial 7s (>2 errors)	< 2	91 (75-98)	NT	46 (38-54)	NT	
	MOTYB (up to July >0 error)	< 2	82 (65-93)	NT	66 (58-73)	NT	

Bilodeau 2016	RADAR (>0 item present)	7 sec	NA	100 (3-100)	NA	77 (58-90)	94-99 (.76-1.00)
Hendry 2016	AMT-10 (<5)	< 2	87 (77–93)	NR	64 (58–69)	NR	NT
	AMT-4 (<4)	< 2	93 (85–97)	NR	54 (48–59)	NR	
	MOTYB (<6)	2	91 (83–96)	NR	50 (44–55)	NR	
Koop 2016	RADAR (>0 item present)	7 sec	100 [3-100]$	NR	69 [39-91]	NR	90 (.08)
Leonard 2016	World BW (>0 error)	< 1	90 (83–95)	NR	41 (30–52)	NR	90% during pre-study training
	MOTYB (>1 error up to July)	< 2	75 (67–83)	NR	70 (59–80)	NR	
	SSF (<5)	1-1.5	75 (66–83)	NR	56 (45–68)	NR	
	SSB (<3)	1-1.5	77 (68–84)	NR	58 (47–69)	NR	
	Vigilance A (<27)	1-1.5	75 (66–83)	NR	73 (72–90)	NR	
	Vigilance B (<18)	1.5	94 (88–98)	NR	56 (45–67)	NR	
	CDT (<6 on Sunderland rating)	2 min	72 (63–81)	NR	64 (52–74)	NR	
	IPT (<4)	2	71 (61–79)	NR	73 (61–82)	NR	
	MOTYB + Vigilance A (1 fail)	< 3	91 (84–96)	NR	59 (48–70)	NR	
O'Regan 2016	CDT (< 10 on 15-point scale)	Each	81 (72–88)	NR	63 (57–69)	NR	NR
	SSF (< 5)	< 2	90 (84–94)	90 (75–97)	41 (35–47)	25 (3–43)	
	MOTYB (> 0 error up to July)		85 (78–90)	83 (68–93)	58 (52–64)	33 (19–51)	
	IPT (> 0 error)		93 (86–96)	87 (69–96)	40 (34–46)	15 (6–32)	
	20>1 (NR)		70 (62–77)	66 (49–79)	69 (63–74)	44 (27–62)	
Bedard 2017	O3DY(<4)	< 1	85 (62-97)	NR	58 (52-64)	NR	NR
Dyer 2017	AMT-4 (<4)	<= 1	92 [75-99]	100 [82-100]	82 [75-87]	52 [37-67]	NT
Grossmann 2017	mRASS (≠0)	0.5	70 (48-85)	55 (28-79)	93 (90-96)	83 (66-93)	NT
	MOTYB in 30s (> 2 errors or >1 error & > 30s)	0.5	95 (76-99)	100 (74-100)	86 (81-90)	63 (46-78)	
Pelletier 2017	RADAR (>0 item present)	< 1	100 (16-100)	NR	72 (59-86)	NR	94-97 (.44-.70)
Richardson 2017	OSLA (>3)	<1	85 [72-93]	74 [55-88]	82 [71-91]	96 [82-100]	NT
	SAVEAHAART (>3 errors)	< 1	90 [79-97]	84 [66-95]	64 [51-76]	73 [51-87]	
	OSLA + SAVEAHAART (> 9)	< 2	84 [72-93]	94 [79-99]	97 [89-100]	92 [77-99]	

[ ] CI in squared brackets were calculated with data in article; ICC for intraclass correlation coefficient, NA for not applicable, NR for not reported.
^ sensitivity and specificity of DTS and RASS administered by research assistant (almost similar when administered by physician); $ score closest to day of delirium diagnosis; # top three single items; & top three pairs of items; * Because the items are arranged in descending order of difficulty, researchers had to assume that participants had succeeded in the easier items and failed the more difficult ones, even if the research assistant had not necessarily administered them.

Appendix 1. Search string

EMBASE

#1 delirium:ab,ti OR 'acute confusion*':ab,ti OR encephalopathy:ab,ti OR 'clouding of consciousness':ab,ti OR 'toxic psychosis':ab,ti

#2 tool* OR test* OR instrument OR assessment OR questionnaire OR interview OR diagnostic OR screening

#3 sensitivity OR specificity OR accuracy OR validity OR reliability OR 'predictive value*' OR 'likelihood-ratio*'

#1 AND #2 AND #3

#1 AND #2 AND #3 AND [embase]/lim AND [humans]/lim AND ([adult]/lim OR [middle aged]/lim OR [aged]/lim OR [very elderly]/lim)

PsychINFO

#1 delirium OR 'acute confusion OR encephalopathy OR ‘clouding of consciousness' OR 'toxic psychosis' in ABSTRACT

#2 tool OR test OR instrument OR assessment OR questionnaire OR interview OR diagnostic OR screening in ABSTRACT

#3 sensitivity OR specificity OR accuracy OR validity OR reliability OR 'predictive value' OR 'likelihood-ratio' in ABSTRACT

Pubmed (incl Medline)

(((delirium[Title/Abstract] OR acute confusion$[Title/Abstract] OR encephalopathy[Title/Abstract] OR clouding-of-consciousness[Title/Abstract] OR toxic-psychosis[Title/Abstract])) AND

(tool$ OR test$ OR instrument OR assessment OR questionnaire OR interview OR diagnostic OR screening)) AND

(sensitivity OR specificity OR accuracy OR validity OR reliability OR predictive-value$ OR likelihood-ratio$)

Appendix 2. Modifications to QUADAS-2 tool

We made a number modifications to the assessments with the QUADAS-2 tool in line with the purpose of our review.

The modifications for risk of bias were:

Patient selection was scored as high risk of bias if patients had been excluded that were testable and could have had delirium, e.g. patients with dementia; otherwise no modifications.

Index test was as high risk of bias if the testers had a background in psychiatric assessment (psychiatrist, psychologist, geriatrician); otherwise no modifications.

Classification of delirium (reference standard) was scored as high risk of bias if a person without a background in psychiatric assessments had made the diagnoses; otherwise no modifications.

Timing was scored as high risk of bias if index and reference test were not assessed within 24 hours of each other; otherwise no modifications.

The modifications for applicability concerns were:

Patient population was scored as high applicability concern if the study was performed in a subset of patients, i.e. older patients with stroke, older patients without dementia

Index test was scored as high applicability concern if the test was applicable to a subset of patients only, i.e. patients that received medication at least twice a day

Target condition was scored as high applicability concern if the study concerned a subset of delirium types, i.e. hyperactive delirium only.

Appendix 3. References of excluded studies

1. Adamis D, Morrison C, Treloar A, Mac Donald AJ & Martin FC. The performance of the Clock Drawing Test in elderly medical inpatients: does it have utility in the identification of delirium? J Geriatr Psychiatry Neurol. 2005;18(3):129–33.
2. Anthony JC, LeResche LA, Von Korff MR, Niaz U & Folstien MF. Screening for delirium on a general medical ward: The tachistoscope and a global accessibility rating. Gen Hosp Psychiatry. 1985;7(1):36–42.
3. Arsène O & Lassaunière L. Evaluation des troubles cognitifs et epistage du delirium chez les patients cancéreux recevant de la morphine. Press Med 2000;29(40):2207–12.
4. Baranowski SL & Patten SB. The Predictive Value of Dysgraphia and Constructional Apraxia for Delirium in Psychiatric Inpatients. Can J Psychiatry . 2000;45(1):75–8.
5. Bellelli G, Morandi A, Davis DHJ, Mazolla P, Turco R, Gentile S & Maclullich AM. Validation of the 4AT, a new instrument for rapid delirium screening: A study in 234 hospitalised older people. Age Ageing. 2014;43(4):496–502.
6. De J, Wand APF, Smerdely PI & Hunt GE. Validating the 4A's test in screening for delirium in a culturally diverse geriatric inpatient population. Int J Geriatr Psychiatry. 2017;32(12): 1322-1329.
7. Erkinjuntti T, Sulkava R, Wikström J & Autio L. Short Portable Mental Status Questionnaire as a Screening Test for Dementia and Delirium Among the Elderly. J Am Geriatr Soc. 1987; 35(5):412–6.
8. Gagnon P, Allard P, Mâsse B & DeSerres M. Delirium in terminal cancer: a prospective study using daily screening, early diagnosis, and continuous monitoring. J Pain Symptom Manage. 2000;19(6):412–26.
9. Gaudreau JD, Gagnon P, Harel F, Tremblay A & Roy MA. Fast, systematic, and continuous delirium assessment in hospitalized patients: The nursing delirium screening scale. J Pain Symptom Manag. 2005;299(4):368–75.
10. Gaudreau JD, Gagnon P, Harel F & Roy MA. Impact on delirium detection of using a sensitive instrument integrated into clinical practice. Gen Hosp Psychiatry. 2005;27(3):194–9.
11. Hart RP, Best AM, Sessler CN & Levenson JL. Abbreviated cognitive test for delirium. J Psychosom Res. 1997;43(4):417–23.
12. Han J, Wilson A, D J, Schnelle J, Dittus RS & Ely E. An Evaluation of Single-Item Delirium Screening Questions in Older Emergency Department Patients. Acad Emerg Med. 2018;36(7):1249-1252.
13. Hare M, Wynaden D, McGowan S & Speed G. Assessing cognition in elderly patients presenting to the emergency department. Int Emerg Nurs. 2008;16(2):73–9.
14. Hendry K, Quinn TJ, Evans JJ & Stott DJ. Informant single screening questions for delirium and dementia in acute care – a cross-sectional test accuracy pilot study. BMC Geriatr. 2015;17(1):1-6.
15. Infante MT, Pardini M, Balestrino M, Finocchi C, Malfatto L, Bellelli G, & Serrati C. et al. Delirium in the acute phase after stroke: comparison between methods of detection. Neurol Sci. 2017;38(6):1101–4.
16. Jiang J, Shulman RW. Evaluation of a New Screening Tool for Delirium in Hospitalized Seniors: The Sour Seven. Am J Geriatr Psychiatry. 2013;21(3):S114.
17. Jitapunkel S, Pillay I &, Ebrahim, S. The Abbreviated Mental Test: Its Use and Validity. Age Ageing. 1991;20(5):332–6.
18. Kahn RL, Goldfarb AI, Pollack M & Peck A. A. Brief objective measures for the determination of metal status in the aged. Am J Psychiatry. 1960;117(4):326–8.

19. Lewallen K. Improving Nurses Knowledge of Delirium and Implementation of the R.A.D.A.R. Screening Tool at two Post Acute Care Facilities in the United States. Belmont University, 2015.
20. Lowery DP, Wesnes K, Brewster N & Ballard C. Quantifying the association between computerised measures of attention and confusion assessment method defined delirium: a prospective study of older orthopaedic surgical patients, free of dementia. Int J Geriatr Psychiatry. 2008;23(12):1253–60.
21. Marcantonio ER, Ngo LH, O'Connor M, Jones RN, Crane PK, Metzger ED, & Inouye SL. 3D-CAM: Derivation and validation of a 3-minute diagnostic interview for CAM-defined delirium: A cross-sectional diagnostic test study. Ann Intern Med. 2014;161(8):554–61.
22. Martins S, Simões MR & Fernandes L. Delirium in older adults: assessment scales review. Eur Psychiatry. 2011;26(S2):486.
23. Morandi A, Han JH, Meagher D, Vasilevskis E, Cerejeira J, Hasemann W & Bellelli G. Detecting Delirium Superimposed on Dementia : Evaluation of the Diagnostic Performance of the Richmond Agitation and Sedation Scale. J Am Med Dir Assoc. 2016;17(9):828-33.
24. Nagly SJ. Predicting and preventing confusion in your patients. J Gerontol Nurs 1986;12(3):27–31.
25. Neufeld KJ, Leoutsakos JS, Sieber FE, Joshi D, Wanamaker BL, Rios-Robles J & Needham DM. Evaluation of two delirium screening tools for detecting post-operative delirium in the elderly. Br J Anaesth 2013;111(4):612–8.
26. Ni Chonchubhair Á & O'Keeffe S. Postoperative cognitive dysfunction in the elderly. In: Postoperative cognitive dysfunction in the elderly. The Lancet 1998;351(9119):1888–9.
27. Alcoba Pérez T, Ciria Poza S, Carracedo Catalán C, Fernández, AG & Vidal JM. Valoración de la concordancia entre la escala CAM-ICU y la nursing delirium screening scale en el postoperatorio de cirugía cardiaca en una unidad de críticos. Enferm Intensiva SEEIUC 2014;25(3):100–6.
28. Pfeiffer E. A Short Portable Mental Status Questionnaire for the Assessment of Organic Brain Deficit in Elderly Patients. JAGS 1975;23(10): 433–41.
29. Pfeifer GM. A Promising Brief Assessment for Delirium in Older Adults. Am J Nurs 2015;115(12):12.
30. Pisani MA, Araujo KLB, Van Ness PH, Zhang Y, Ely E & Inouye SK. A research algorithm to improve detection of delirium in the intensive care unit. Crit Care 2006;10(4):R121.
31. O'Sullivan D, Brady N, Manning E, O'Shea E, O"Grady S, O'Reagan N & Timmons S. 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 2017;47(1);11–7.
32. Radtke FM, Franck M, Schneider M, Seeling M, Heinz A & Spies CD. Comparison of three scores to screen for delirium in the recovery room. Br J Anaesth 2008;101(3):338–43.
33. Sands MB, Dantoc BP, Hartshorn A, Ryan CJ & Lujic S. Single Question in Delirium (SQiD): testing its efficacy against psychiatrist interview, the Confusion Assessment Method and the Memorial Delirium Assessment Scale. Palliat Med.2010;24(6):561–5.
34. Salih SA, Paul S, Klein K, Lakhan P & Gray L. Screening for delirium within the interrai acute care assessment system. J Nutr Heal Aging 2012;16(8):695–700.
35. Sarhill N, Walsh D, Nelson KA, LeGrand & Davis MP. Assessment of delirium in advanced cancer: the use of the bedside confusion scale. Am J Hosp Palliat Care 2001;18(5): 335–41.

36. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal PV, Keane KA & Elswijk RK. The Richmond Agitation-Sedation Scale: Validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med 2002;166(10):1338–44.
37. Simon SE, Bergmann MA, Jones RN, Murphy KM, Orav EJ & Marcantonia ER. Reliability of a Structured Assessment for Nonclinicians to Detect Delirium Among New Admissions to Postacute Care. J Am Med Dir Assoc 2006;7(7):412–5.
38. Slor CJ, Adamis D, Jansen RW, Meagher DJ, Witlox J, Houdijk AP & de Jonghe JF. Validation and psychometric properties of the Delirium Motor Subtype Scale in elderly hip fracture patients. Arch Gerontol Geriatr 2014;58(1):140–4.
39. Voyer P, Émond M, Boucher V, Carmichae, PH, Juneau L, Vu TM & Bouchard G. RADAR: a rapid detection tool for signs of delirium (6th vital sign) in emergency departments. Canadian Journal of Emergency Nursing. 2017;40(2):37-43
40. Williams MA. Delirium/acute confusional states: evaluation devices in nursing. Int Psychogeriatr 1991;3(2):301–8.
41. Yeo C, Lim WS, Chan M, Ho X, Han HC & Chong MS. Severe Impairment Rating Scale: A Useful and Brief Cognitive Assessment Tool for Advanced Dementia for Nursing Home Residents. Am J Alzheimer's Dis Other Dementiasr 2016;31(1):87–96.

Appendix 4. Content of rapid screening tests in this review

AMT: A Mental Test-score is used to assess mental impairment. The patient is asked to answer questions or perform a task. Each correct answer is awarded with one point. The original test has 26 questions testing memory and orientation developed by Hodkinson in 1972. The test is later shortened to 10 questions and to 4 questions. The AMT-4 covers: Age, Date of birth, Place, and Year. (1)

Clock-Drawing Test (CDT): The patient is asked to draw a circle, in some tests de circle is given. Than the patient is asked to position the numbers of the clock correctly (1 to 12). The scoring system assesses three domains: correctly drawn clock shape, all numbers in the correct position, and ‘‘hands of the clock set to the correct time’’ – each domain scoring one point if completed properly. There are three different scoring systems: Scoring according to clinical gestalt (normal or abnormal), CAMDEX scoring method ranging from 0 (no clock face drawn) to 3 (perfect clock; and the Schulman scoring method ranging from 0 (no clock face drawn) to 5 (perfect clock). Patients who did not attempt to perform the Clock Draw Test were rated as abnormal in the clinical gestalt rating method and assigned a “0” for the CAMDEX and Schulman scoring methods regardless of the reason (2,3).

Cog-4: The COG-4 is a cognitive examination suitable to be used in strokes. Four items corresponding to the NIHSS items on orientation, executive function, language and inattention are been used. These components individually score 0–2 or 0–3 points giving the Cog-4 scale, a score that ranges from 0 to 9 points. The Cog-4 can be used to detect severe cognitive impairment (4).

Confusion Rating Scale: The Confusion Rating Scale is developed to assess confusional behavior that nurses could easily identify. The Confusion Rating Scale rates four dimensions of confusional behavior: 1) disorientation to place, time, or recognition of persons assessed by spontaneous remarks of the patient or by informal questioning by the caregiver; 2) communication that is unrelated or inappropriate to the situation or unusual for the person or lack of communication when that had not been the persons usual pattern, 3) behaviors inappropriate to the situation such as pulling at tubes or attempting to get out of bed when that was contraindicated; and 4) the presence of illusions or hallucinations. The intensity of each type of behavior on an eight-hour shift is rated on a scale ranging from 0 which means the behavior is never present during the shift to 2 which means the behavior is present during the shift(5).

Counting 1 to 10: The patient is asked to count from 1 to 10. This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale (HDS)(6)

Counting 10 to 1: The patient is asked to count back from 10 to 1. This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale (HDS)(6).

Counting 93 to 85: The patient is asked to count back from 93 to 85. This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale (HDS)(6)

Counting objects: The patient is asked to count objects. This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale (HDS)(6).

Day of the week: The patient is asked which day of the week it is(6).

Days of the week backward: The patient is asked to recite the days of the week backwards. This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale (HDS)(6).

Days of the week forward: The patient is asked to recite the days of the week forward. This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale (HDS)(6).

Digit Cancellation Test: This test uses two different matrixes. Each matrix consists of 13 lines with 10 digits from 0 to 9. The subject is asked to cross out specified target digits on each matrix within 45 seconds. In the DCT10 the person has to cross one target digit, in the DCT20 two target digits. One point is awarded for a correct response, of which there are 10 in DCT10 and 20 in DCT20. A negative score is subtracted for each incorrect response such that a subject crossing out every digit will receive a total score of zero on a matrix(7)

Digit Span Test: The Digit Span Test refers to the combination of two tests: the digit span forward and the digit span backwards as described in the Montreal Cognitive Assessment (MoCA) scale (7). Patients were asked to repeat a sequence of five numbers (forwards) followed by a sequence of three numbers to repeat backwards. Total score ranged from 0 to 2 according to how many elements of the test were correctly completed; if they performed correctly in both tasks, a maximum score of 2 was given; if they completed only one test, a score of 1 was given and if they were unable to complete either task, a score of 0 was assigned(8).

DTS: The DTS Delirium Triage consists consisted of two components the level of consciousness - measuring the Richmond Agitation Sedation Scale (RASS) and Attention by spelling the word “LUNCH” backwards. The RASS is an arousal scale commonly used in the intensive care unit to assess level of sedation. It ranges from −5 (unarousable) to +4 (combative), and 0 indicates a normal level of alertness. If a patient had a RASS other than 0, or made >1 error on the “LUNCH” backwards test, then the DTS was considered positive(9).

GAR: After a conversation between patient and physician (minimum of 2 minutes) the global attentiveness of the patient is rated on an uninterrupted 10-cm visual analog scale. A high rating was determined if the observer could easily keep the patient engaged throughout the interview; a low rating was recorded if the patient could not be aroused from stupor or was so agitated that no conversation was possible(10).

GCS: The Glasgow Coma Scale was developed for assessing the depth and duration of impaired consciousness and coma. Three aspects of behaviour are independently measured—motor responsiveness, verbal performance, and eye opening. A patient is assessed against the criteria of the scale, and the resulting points give a patient score between 3 (indicating deep unconsciousness) to 15, indicating fully alert(4).

Global clinical subjective rating of attention: Patients were rated according to a brief (30–60 s) conversation about their hospitalization and treatment. Afterwards the clinican rated the conversation ranging 0 to 3((0) for those who were fully alert and attentive to (3) for those with severe difficulty focusing and/or sustaining attention such that an interview was severely affected(10).

Intersecting Pentagons test: This is a geometric coping test. The patient is asked to copy a picture of two intersecting pentagrams on a blank page. The patients is rewarded 1 to 6 points depending on how correctly he copies the picture(11).

LUNCH backwards test: the patient is asked to spell the word lunch backwards.

Memorial Delirium Assessment Scale (MDAS): is an instrument designed to measure the severity of delirium. It captures behavioral manifestations and cognitive deficits. Impairment can be measured in 10 domains: awareness, orientation, short term memory, digit span, attention capacity, organizational thinking, perceptual disturbance, delusions, psychomotor activity and sleep-wake cycle. Items are rated on a 4-point scale from 0 (none) to 3 (severe) depending on the level of impairment(12).

Month of the year backwards (MOTYB): The patient is asked to repeat the months of the year backwards. He starts with December. If he can produce 6 corrects months (till July) the test is scored as passed(13).

Months of the year forward: The patient is asked to repeat the moth of the year forwards. This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale(6).

mRASS: The RASS is an observational instrument that has been validated in the intensive care unit setting for objectively determining level of sedation. A brief open-ended question was added by Chester to improve assessment of attention. This question is scored as passed or failed(14).

Pictorial Facial Scale: The Pictorial Facial Scale (the “Mental Status vital sign”)is based on pictures of patient faces of varying mental states connected to a numerical scale.. It can be scored -4 to + 5(15).

OSLA: The Observational Scale of Level of Arousal (OSLA) is used to assess level of arousal in delirium. The OSLA is scored after a brief interaction with the patient, and it does not necessarily require a verbal response. Higher scores indicate an abnormal level of arousal. It consists of four domains: Eye opening, Eye contact, Posture, and Movement(16).

The Ottowa 3DY Scale: The patient is asked the date, the day, world spelled backwards and year. He is also asked to spell the word WORLD backwards(17).

RADAR: The RADAR is a tool based on observation by nurses while administering medication. Afterwards a three question scoring-list is scored with yes or no(18).

RASS: The RASS is an arousal scale commonly used in intensive care units to assess for depth of sedation but has been incorporated into several delirium assessments to assess for level of consciousness. The scoring varies from +4 "combative" to -5 "unarousable"(19).

S-A-V-E-A-H-A-A-R-T: Attention can be measured using a vigilance task, with participants signalling each time an “A” is heard when the sequence of ten consecutive letters “SA-V-E-A-H-A-A-R-T” was read out, each letter 3 seconds apart. Errors were counted when a patient failed to signal on the letter “A” or when a patient signalled on any letter other than “A”(20).

Serial sevens subtraction test: This task was scored according to the system outlined in the MoCA(7) and MMS(13). Participants were asked to sequentially subtract 7 starting at 100,and were scored 0–3 according to the number of successful trials: no correct subtractions (0); one correct subtraction (1); two or three correct subtractions (2); and four or five correct subtractions(11).

SSB: The Serial Span Backwards is an attention-test. The person has to copy the squares the investigator touches on a white card with blue squares. He has to touch them in a reversed order. Two trails are conducted and the best trail is used. If the patient is unable to complete a sequence of three or more numbers he failed the test(21).

SSF: The Serial Span Forward test was performed using an A5-sized piece of white card with eight red squares. The investigator taps out a predetermined sequence for the patient to replicate. The test begins with a sequence of two squares and is increased in number with each correct iteration, up to a maximum sequence of seven. Two attempts were allowed at each level using different predetermined sequences. Patients who are unable to correctly repeat a sequence of five are considered to have failed the test(21,22)).

Serial 3 subtraction test: Serial threes subtraction test: Participants are asked to subtract three This attention test is drawn from the Concentration subscale from the Hierarchic Dementia Scale(6).

SQeeC: The SQeeC consists of two simple questions: ‘Name a place you would like to visit that you have never been before?’ and ‘How would you make the journey?’ These two questions examine the contemplative state or one’s ability to reflect, imagine and appreciate and the ability to problem solving. This test gives a demonstration of consciousness. If this demonstration is successful the patient does not have a delirium(23).

Vigilance “A” test: is one of the tests of MoCA scale(7). A list of 29 letters with the letter “A” included on 11 occasions was presented to the patient and they were asked to indicate each time the letter “A” was mentioned. Scores ranged from 0 to 1 depending on whether patients could complete the test without making more than two errors(24).

Vigilance “B” test: Is similar to the Vigilance “A” test except there are two letters included (the C and the E). The patient has to indicate when one of the letters is mentioned. Scores are calculated by subtracting commissions from correct responses (scored double) and rated as unable to engage with the test(0), score 1-9(1), score 10-18(2), score 19-26(3), score > 27(22).

World Backwards-test: This test is derived from the MMSE(11). The patients is asked to spell the word WORLD backwards. Each letter spelled correctly or each correction made without help is rewarded by one point. If a patient is unable to recite a letter this letter is given to him so the patient can make an accurate attempt on the next letter.

Writing test: In this test the patient is asked to write his name and address. There are no universally acceptable and accurate methods of quantifying dysgraphic errors; impairments of writing were assessed using the Delirium Writing Test in which writing ability is assessed on items such as reluctance to write and motor, spatial, syntactical, and spelling disorders(25).

References appendix 4

1. Hodkinson HM. Evaluation of a mental test score for assessment of mental impairment in the elderly. Age Ageing. 1972;1(4):233–8.
2. Roth M, Tym E, Mountjoy CQ, Huppert FA, Hendrie H, Verma S & Goddard R. CAMDEX: a standardized instrument for the diagnosis of mental disorders in the elderly with special reference to the early detection of dementia. Br J Psychiatry. 1986;149(6):698–709.
3. Shulman KI. Clock-drawing: is it the ideal cognitive screening test? Int J Geriatr Psychiatry. 2000;15(6):548–61.
4. Lees R, Corbet S, Johnston C, Moffitt E, Shaw G & Quinn TJ.Test accuracy of short screening tests for diagnosis of delirium or cognitive impairment in an acute stroke unit setting. Stroke. 2013;44(11):3078–83.
5. Inouye S, Van Dyck C, Alessi C, Balkin S, Segal A. & Horwitz RI. Clarifying confusion: The confusion assessment method: A new method for detection of delirium. Annals of Internal Medicine. 1990;113(12):941–8.
6. Voyer P, Champoux N, Desrosiers J , LAnddreville P, Monette J, Savoie M & Bédard A. Assessment of inattention in the context of delirium screening: one size does not fit all! Int Psychogeriatrics. 2016;28(8):1293–301.
7. Sala S Della, Laiacona M, Spinnler H & Ubezio C. A cancellation test: its reliability in assessing attentional deficits in Alzheimer's disease. Psychol Med. 1992;22(4):885-901.
8. Nasreddine Z, Phillips N, Bédirian V , Charbonneau S, Whitehead V, Colin I & Chertkov H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–9.
9. Han JH, Wilson A, Vasilevskis EE, Shintani A,Schnelle JF, Dittus RS & Ely E. 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(5):457–65.
10. O'Keefe ST & Gosney MA. Assessing Attentiveness in Older Hospital Patients: Global Assessment Versus Tests of Attention. JAGS 1997;45(4):470–3.
11. 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(3):189–98.
12. Fadul N, Kaur G, Zhang T, Palmer JL & Bruera E. Evaluation of the memorial delirium assessment scale (MDAS) for the screening of delirium by means of simulated cases by palliative care health professionals. Support Care Cancer 2007 ;12(11);15:1271–6.
13. Fick DM, Inouye SK, Guess J, Ngo L, Jones RN, Saczynski JS & Marcantonio ER. Preliminary Development of an Ultrabrief Two-Item Bedside Test for Delirium. J Hosp Med 2015;10(10):645–50.
14. Chester JG, Harrington MB, Rudolph JL & VA Delirium working group. Serial administration of a modified richmond agitation and sedation scale for delirium screening. J Hosp Med 2012;7(5):450–3.
15. Shoaib A, Kaehr E & Malmstrom T. Mental status (MS) Vital sign: A Pictorial Facial Scale as a Screening Tool for delirium. In Mental status (MS) Vital sign: A Pictorial Facial Scale as a Screening Tool for delirium. JAGS 2015. S116.
16. Tieges Z, McGrath A, Hall RJ & Maclullich AMJ. Abnormal level of arousal as a predictor of delirium and inattention: An exploratory study. Am J Geriatr Psychiatry 2013;21(12):1244–53.
17. Molnar FJ, Wells GA & Mcdowell I. The Derivation and Validation of the Ottawa 3D and Ottawa 3DY Three- and Four-Question Screens for Cognitive Impairment. Clinical

Medical Insights. Geriatrics 2008;2:1–11.
18. Voyer P, Champoux N, Desrosiers J, Landreville P, Cusker J, Monette J & Carmichael PH. Recognizing acute delirium as part of your routine [RADAR]: a validation study. BMC Nurs. 2015;14(1):19.
19. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal PV, Keane KA & Elswijk RK. The Richmond Agitation-Sedation Scale: Validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med. 2002;166(10):1338–44.
20. Richardson SJ, Davis DHJ, Bellelli G, Hasemann W, Meagher D Kreisel DH & Morandi A. Detecting delirium superimposed on dementia: diagnostic accuracy of a simple combined arousal and attention testing procedure. Int Psychogeriatrics 2017;29(10):1585–93.
21. Hart RP, Best AM, Sessler CN & Levenson JL. Abbreviated cognitive test for delirium. J Psychosom Res 1997;43(4):417–23.
22. Leonard M,)Çonnel H, Williams O, Awan F, Exton C, O'Çonnor M & Meagher DJ. Attention , vigilance and visuospatial function in hospitalized elderly medical patients: Relationship to neurocognitive diagnosis. J Psychosom Res 2016;90:84–90.
23. Lin HS, Eeles E, Pandy S, Pinks er D, Brasch C & Yerkovich S. Screening in delirium: A pilot study of two screening tools, the Simple Query for Easy Evaluation of Consciousness and Simple Question in Delirium. Australas J Ageing 2015;34(4):259–64.
24. Lim J & Dinges D. Sleep Deprivation and Vigilant Attention. Ann N Y Acad Sci 2008;1129(1):305–22.
25. Macleod A & Whitehead L. Dysgraphia and terminal delirium. Palliat Med. 1997;11(2):127–32.

Appendix 5. Rapid delirium screening test characteristics and risk of bias

Study	Test (cut-off)	Sensitivity, % (95% CI)		Specificity, % (95% CI)		Risk of bias			
		All patients	In dementia	All patients	In dementia	Patient selection	Index test	Reference standard	Flow & timing
Jitapunkul 1992	AMT-10 (<8)	92 [78-98]	NR	65 [56-73]	NR	H	L	L	U
Pompei 1995	DSF (<5) Vigilance A (> 2 errors) DSF + Vigilance A (both failed)	34 [22-48]$ 61 [47-74] 26 [15-40]	NT NT NT	90 [87-93] 77 [73-81] 97 [95-99]	NT NT NT	H	L	H	H
Macleod 1997	Writing name and address	100 [69-100]	NA	100 [69-100]	NA	H	H	U	L
O'Keeffe 1997	GAR (<7) DSF (NR) DSB (< 4) Vigilance A (> 2 errors) DCT-1 (NR) DCT-2 (< 9 in 2 trials)	94 [73-100] NR 83 [59-96] 83 [59-96] NR 94 [73-100]	NR NR NR NR NR NR	99 [92-100] NR 96 [88-99] 83 [72-91] NR 87 [77-94]	NR NR NR NR NR NR	L	H	L	H
Adamis 2006	Signature (abnormal) MMSE sentence (abnormal)	54 [32-76] NR	NR NR	88 [76-96] NR	NR NR	L	L**	L	H
Bryson 2011	CDT (≤18)	66 (38-85)	NA	64 (52-76)	NA	L	H	H	L
Leung 2011	DSF (<8) DSB (<3) DSF + DSB (NR)	58 [37-77] 81 [61-93] NR	NR NR NR	72 [63-80] 63 [53-71] NR	NR NR NR	H	H	L	L
Chester 2012	mRASS (≠ 0)	64 (52-76)	NR	93 (90-96)	NR	L	L	L	L
Han 2013 Emerson 2014 Han 2015	RASS (≠ 0) + Lunch BW (>1 error) CDT (abnormal) RASS (≠ 0)	98 (90-100) 94 [84-99] 84 (74 -94)	NR NR NR	56 (51-61) 44 [39-49] 88 (84-91)	NR NR NR	H	L	L	H
Lees 2013	AMT-4 (<4) AMT-10 (<8) CDT (<3 on 0-3 scale)	83 (52-98) 75 (43-95) 67 (22-96)	NR NR NR	61 (51–71) 61 (51–71) 38 (28–49)	NR NR NR	L	L	U	L

	Cog-4 (>0)	70 (35–93)	NR	44 (35–55)	NR				
	GCS (eye open<4 & verbal response<5)	17 (02–48)	NR	81 (71–88)	NR				
Tieges 2013	OSLA (>3)	90 [56-100]	NA	90 [68-99]	NA	H	H	H	U
	RASS (≠ 0)	90 [56-100]	NA	85 [62-97]	NA				
O'Regan 2014	MOTYB (1 error up to July)	84 (68-94)	88 (68-97)	90 (82-95)	71 (29-96)	L	U	H	H
	SSF (<5)	95 (82-99)	NR	58 (48-68)	NR				
	SSF (<5), then MOTYB (1 error up to July)	81 (65-92)	NR	91 (83-96)	NR				
Fick 2015	MOTYB (1 error) #	83 (69-93)	89 (72-98)	69 (61-76)	61 (41-78)	H	L	L	L
	DSB (<4)	83 (69-93)	86 (67-96)	52 (44-60)	54 (34-72)				
	Day of week (an inadequate answer)	71 (55-84)	75 (55-89)	92 (87-96)	75 (55-89)				
	Day of week + MOTYB (1 fail) &	93 (81- 99)	96 (82-100)	48 (40-56)	43 (24-63)				
	Day of week + DSB (1 fail)	93 (81-99)	93 (76-99)	48 (40-56)	39 (22-59)				
	DSB + MOTYB (1 fail)	93 (81-99)	93 (76-99)	42 (34-50)	39 (22-59)				
	ALOC	19 (9-34)	21 (8-41)	99 (96-100)	NR				
Lin 2015	SQeeC (an inadequate answer)	83 (52-98)	83 (36-99)	81 (72-89)	59 (36-79)	L	H	L	L
Shoaib 2015	Pictorial Facial Scale (≠ 0)	86 [57-98]	NT	67 [49-80]	NT	U	L	L	U
Voyer 2015	RADAR 1-4x daily (>0 item present)	65 (43-84)	NR	71 (64-78)	NR	H	L**	U**	L
	RADAR 3-4x daily (>0 item present)	73 (39-94)	71 (29-96)	67 (57-76)	43 (26-61)				
Voyer 2016	Serial 7s (failure)*	96 (78-100)	NR	14 (9-20)	NR				
	Serial 3s (failure)	87 (66-97)	NR	47 (39-55)	NR				
	MOTYB (failure	83 (61-95)	63 (24-91)	63 (55-70)	79 (71-86)				
	Days of week B (failure)	48 (27–69)	NR	85 (79–90)	NR				
	Counting 93>85 (failure)	48 (27–69)	NR	85 (79–90)	NR				
	Counting 10>1 (failure)	30 (13–53)	NR	87 (81–92)	NR				
	MOTYF (failure)	26 (10–48)	NR	89 (84–94)	NR				
	Days of week F (failure)	26 (10–48)	NR	90 (84–94)	NR				
	Counting 1>10 (failure)	17 (5–39)	NR	91 (86–95)	NR				
	Counting 10 objects (failure)	17 (5–39)	NR	94 (89–97)	NR				
Adamis 2016	DST (>1)	74 (55-87)	NT	62 (54-69)	NT	H	H	H	U
	Vigilance A (>2 errors)	82 (65-93)	NT	60 (52-68)	NT				
	Serial 7s (>2 errors)	91 (75-98)	NT	46 (38-54)	NT				

	MOTYB (up to July >0 error)	82 (65-93)	NT	66 (58-73)	NT				
Bilodeau 2016	RADAR (>0 item present)	NA	100 (3-100)	NA	77 (58-90)	H	U	U	U
Hendry 2016	AMT-10 (<5) AMT-4 (<4) MOTYB (<6)	87 (77–93) 93 (85–97) 91 (83–96)	NR NR NR	64 (58–69) 54 (48–59) 50 (44–55)	NR NR NR	L	H	L	H
Koop 2016	RADAR (>0 item present)	100 [3-100]$	NR	69 [39-91]	NR	L	L	U	U
Leonard 2016	World BW (>0 error) MOTYB (>1 error up to July) SSF (<5) SSB (<3) Vigilance A (<27) Vigilance B (<18) CDT (<6 on Sunderland rating) IPT (<4) MOTYB + Vigilance A (1 fail)	90 (83–95) 75 (67–83) 75 (66–83) 77 (68–84) 75 (66–83) 94 (88–98) 72 (63–81) 71 (61–79) 91 (84–96)	NR NR NR NR NR NR NR NR NR	41 (30–52) 70 (59–80) 56 (45–68) 58 (47–69) 73 (72–90) 56 (45–67) 64 (52–74) 73 (61–82) 59 (48–70)	NR NR NR NR NR NR NR NR NR	L	H	U	H
O'Regan 2016	CDT (< 10 on 15-point scale) SSF (< 5) MOTYB (> 0 error up to July) IPT (> 0 error) 20>1 (NR)	81 (72–88) 90 (84–94) 85 (78–90) 93 (86–96) 70 (62–77)	NR 90 (75–97) 83 (68–93) 87 (69–96) 66 (49–79)	63 (57–69) 41 (35–47) 58 (52–64) 40 (34–46) 69 (63–74)	NR 25 (3–43) 33 (19–51) 15 (6–32) 44 (27–62)	L	H	H	H
Bedard 2017	O3DY (<4)	85 (62-97)	NR	58 (52-64)	NR	H	H	H	U
Dyer 2017	AMT-4 (<4)	92 [75-99]	100 [82-100]	82 [75-87]	52 [37-67]	U	U	H	L
Grossmann 2017	mRASS (≠0) MOTYB (> 2 errors & 30s or >1 error & > 30 s)	70 (48-85) 95 (76-99)	55 (28-79) 100 (74-100)	93 (90-96) 86 (81-90)	83 (66-93) 63 (46-78)	H	L	L	H
Pelletier 2017	RADAR (>0 item present)	100 (16-100)	NR	72 (59-86)	NR	H	U	U	U
Richardson 2017	OSLA (>3) SAVEAHAART (>3 errors) OSLA + SAVEAHAART (> 9)	85 [72-93] 90 [79-97] 84 [72-93]	74 [55-88] 84 [66-95] 94 [79-99]	82 [71-91] 64 [51-76] 97 [89-100]	96 [82-100] 73 [51-87] 92 [77-99]	U	H	H	U

[ ] CI in squared brackets were calculated with data in article; NA for not applicable, NR for not reported; H stands for high, L for low and U for unclear;
^ sensitivity and specificity of DTS and RASS administered by research assistant (almost similar when administered by physician); $ score closest to day of delirium diagnosis; # top three single items; & top three pairs of items; * Because the items are arranged in descending order of difficulty, researchers had to assume that participants had succeeded in the easier items and failed the more difficult ones, even if the research assistant had not necessarily administered them; ** high for tests other than GAR (Adamis 2006) or RADAR (Voyer 2016)

Appendix 6. Point estimates of sensitivity and specificity of reviewed tests across studies

Rapid screening test	Older patients in general			In patients with dementia		
	Studies, n *	Sensitivity, range	Specificity, range	Studies, n*	Sensitivity, range	Specificity, range
Tests reported in more studies						
AMT-4	3	83 – 93	54 – 82	1	100	52
AMT-10	3	75 – 92	61 – 65			
CDT	5	66 – 94	38 – 64			
DSB	3	81 – 83	52 – 96	1	86	54
DSF	2	34 – 58	72 – 90			
IPT	2	71 – 93	40 – 73	1	87	15
MOTYB	8	75 – 95	58 – 90	5	63 – 100	33 – 79
mRASS	2	64 – 70	93 – 93	1	55	83
OSLA	2	85 – 90	82 – 90	1	74	96
RADAR	2	100 – 100	69 – 72	1	100	77
RASS	2	84 – 90	85 – 88			
Serial 7s	2	91 – 96	14 – 46			
SSF	3	75 – 95	41 – 58	1	90	25
Vigilance A	4	61 – 83	60 – 83			
Tests reported in one study						
ALOC	1	19	99	1	21	
Cog-4	1	70	44			
Counting 10>1	1	30	87			
Counting 93>85	1	48	85			
Counting 1>10	1	19	91			
Counting 10 objects	1	17	94			
Day of week	1	71	92	1	75	75
Days of week B	1	48	85			
Days of week F	1	26	90			
DCT-2	1	94	87			
DST	1	74	62			
GAR	1	94	99			
GCS	1	17	81			
MOTYB in 30s	1	95	86	1	100	63
MOTYF	1	26	89			
O3DY	1	85	58			
Pictorial Facial Scale	1	86	67			
RADAR 1-4x daily	1	65	71			
RADAR 3-4x daily	1	73	67	1	71	43
SAVEAHAART	1	90	64	1	84	73
Serial 3s	1	87	47			
Signature	1	54	88			
SSB	1	77	58			
SqeeC	1	83	81	1	83	59
Vigilance B	1	94	56			
World BW	1	90	41			
Writing name and address	1	100	100			

Counting 20>1	1	70	69	1	66	44
Combination of tests reported in one study						
Day of week + DSB	1	93	48	1	93	39
Day of week + MOTYB	1	93	48	1	96	43
DSB + MOTYB	1	93	42	1	93	39
DSF + Vigilance A	1	26	97			
MOTYB + Vigilance A	1	91	59			
OSLA + SAVEAHAART	1	84	97	1	94	92
RASS + Lunch BW	1	98	56			
SSF, then MOTYB	1	81	91			

* The number of studies that reported test accuracy of a test was not always the number of studies that investigated it; DCT1, MMSE and DSF + DSB were investigated once, but test accuracy was not reported.