medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Diagnostic Accuracy of the 4AT for delirium detection: systematic review and meta-analysis.

Authors

Zoë Tieges[1,2], Alasdair M. J. MacLullich[1], Atul Anand[1,3], Claire Brookes[4], Marica Cassarino[4], Margaret O'Connor[5], Damien Ryan[6], Thomas Saller[7], Rakesh C. Arora[8,9], Yue Chang[8], Kathryn Agarwal[10,11], George Taffet[10,11], Terence Quinn[12], Susan. D. Shenkin[1], and Rose Galvin[4]

Affiliations

[1]Geriatric Medicine, Edinburgh Delirium Research Group, Usher Institute, University of Edinburgh, Scotland, UK

[2]School of Health and Life Sciences, Glasgow Caledonian University, Scotland, UK

[3]Centre for Cardiovascular Science, University of Edinburgh, Scotland, UK

[4]School of Allied Health, Faculty of Education and Health Sciences, Ageing Research Centre, Health Research Institute, University of Limerick

[5]Department of Ageing and Therapeutics, University Hospital Limerick, Dooradoyle, Limerick, Ireland

[6]Retrieval, Emergency and Disaster Medicine Research and Development Unit (REDSPoT), Emergency Department, University Hospital Limerick, Dooradoyle, Limerick, Ireland, and Graduate Entry Medical School, University of Limerick, Ireland.

[7]Department of Anaesthesiology, University Hospital, LMU Munich, Munich, Germany

[8]Department of Surgery, Section of Cardiac Surgery, Max Rady College of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada

[9]Cardiac Sciences Program, St. Boniface Hospital, Winnipeg, Manitoba, Canada

[10]Section of Geriatrics, Baylor College of Medicine, Houston, Texas, USA

[11]Houston Methodist Hospital, Houston, Texas, USA

[12]Institute of Cardiovascular Medicine, University of Glasgow, Glasgow, Scotland, UK

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Corresponding author:

Dr. Zoë Tieges

Edinburgh Delirium Research Group

Geriatric Medicine

Usher Institute

University of Edinburgh

Room S1642, Royal Infirmary of Edinburgh, 51 Little France Crescent

Edinburgh, EH16 4SA, UK

Tel. +44 (0) 131 242 6481

Email: zoe.tieges@ed.ac.uk

Key points:

- The 4AT is a short delirium assessment tool that is widely used internationally in clinical practice.
- This systematic review and meta-analysis of diagnostic accuracy studies of the 4AT included 3721 observations in 17 studies from nine countries.
- Studies recruited from a range of settings including the Emergency Department, medical, stroke, and surgical wards.
- The 4AT had a pooled sensitivity 0.88 and pooled specificity of 0.88.
- The methodological quality of studies varied but was moderate to good overall.

Declaration of sources of funding

This work was supported by the Wellcome Trust-University of Edinburgh Institutional Strategic Support Fund. Grant no. IS3-T06/03

Conflict of Interest

AMJM led the design of the 4AT in 2011 (with others, see www.the4AT.com); note that 4AT is free to download and use. SDS and AA provided comments on its development.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ABSTRACT

Objective: Detection of delirium in hospitalised older adults is recommended in national and international guidelines. The 4 'A's Test (4AT) is a short (<2 min) instrument for delirium detection that is used internationally as a standard tool in clinical practice. We performed a systematic review and meta-analysis of diagnostic test accuracy of the 4AT for delirium detection.

Methods: We searched MEDLINE, EMBASE, PsycINFO, CINAHL, clinicaltrials.gov and the Cochrane Central Register of Controlled Trials, from 2011 (year of 4AT release on the website www.the4AT.com) until 21 December 2019. Inclusion criteria were: older adults (≥65y); diagnostic accuracy study of the 4AT index test when compared to delirium reference standard (standard diagnostic criteria or validated tool). Methodological quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. Pooled estimates of sensitivity and specificity were generated from a bivariate random effects model.

Results: 17 studies (3721 observations) were included. Settings were acute medicine, surgery, a care home, and the emergency department. Three studies assessed performance of the 4AT in stroke. The overall prevalence of delirium was 24.2% (95% CI 17.8-32%; range 10.5-61.9%). The pooled sensitivity was 0.88 (95% CI 0.80-0.93) and the pooled specificity was 0.88 (95% CI 0.82-0.92). Excluding the stroke studies, the pooled sensitivity was 0.86 (95% CI 0.77-0.92) and the pooled specificity was 0.89 (95% CI 0.83-0.93). The methodological quality of studies varied but was moderate to good.

Conclusions: The 4AT shows good diagnostic test accuracy for delirium in the 17 available studies. These findings support its use in routine clinical practice in delirium detection.

PROSPERO Registration number CRD42019133702.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Key words: delirium, 4AT, systematic review, screening tool, older patients, dementia

List of abbreviations

3D-CAM: 3-Minute Diagnostic Assessment for Delirium using the Confusion Assessment Method algorithm

4AT: 4 'A's Test

bCAM: brief Confusion Assessment Method

CAM-ICU: Confusion Assessment Method -Intensive Care Unit

CAM: Confusion Assessment Method

DRS-R98: Delirium Rating Scale-Revised 98

DSM: Diagnostic and Statistical Manual of Mental Disorders

IQCODE: Informant Questionnaire on Cognitive Decline in the Elderly

MMSE: Mini-Mental Status Examination

NICE: National Institute for Health and Clinical Excellence

PRISMA-DTA: Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies - Version 2

SIGN: Scottish Intercollegiate Guidelines Network

STARD: Standards for Reporting Diagnostic Accuracy Studies

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Delirium is a serious acute neuropsychiatric disorder of consciousness, attention and cognition triggered by general medical conditions, drugs, surgery, or a combination of causes. It manifests through acute and fluctuating cognitive, psychomotor and perceptual disturbances which develop over hours to days. Delirium is common in hospitalised older adults, with a recent meta-analysis of 33 studies of medical inpatients finding an overall delirium occurrence of 23% (95% CI 19-26%) [1]. It is also common in surgical patients, in care homes and palliative care settings [2]. Delirium is associated with significant adverse outcomes including functional decline and mortality, and patient and carer distress [3, 4].

Detection of delirium at the earliest possible time point is important for several reasons, including prompting the search for acute triggers, gaining access to recommended treatment pathways, in managing delirium-associated risks such as falls, in identifying and treating distress, in providing prognostic information, and in communicating the diagnosis to patients and carers. Detection has been recommended in multiple guidelines including the Scottish Intercollegiate Guidelines Network (SIGN) guidelines on delirium [5]. More than 30 delirium assessment tools exist, though these vary considerably in purpose and clinical applicability [6, 7]. Categories of tools include: those intended for use at first presentation or at other points when delirium is suspected; regular use (that is, daily or more frequently) in monitoring for new onset delirium in inpatients; ‘ultra-brief’ screening tools; intensive care unit tools; measurement of delirium severity; informant-based; detailed phenomenological assessment.

The 4 ‘A’s Test or 4AT was developed as a short delirium assessment tool intended for clinical use in general settings at first presentation and when delirium is suspected. It was initially published on a dedicated website in 2011 [8]. It consists of four items: an item assessing level of alertness, a test of orientation (the Abbreviated Mental Test–4, comprising 4 orientation

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

questions), a test of attention (Months Backward test); and an item ascertaining acute change or fluctuating course [Appendix 1]. The first diagnostic test accuracy study in general settings was published in 2014 [9]. Since publication 4AT performance has been evaluated in multiple studies [10]. The 4AT has become a standard tool in clinical practice [11, 12] and it is recommended in guidelines and pathways [5, 13]. Two prior systematic reviews of general delirium assessment tools included the 4AT, but could only cite the original general validation study because of the date of the reviews [6, 7]. A systematic review of delirium detection in stroke patients published in 2019 included three studies that had evaluated the 4AT post-stroke reporting sensitivities from 0.90 to 1.00 and specificities from 0.65 to 0.86 [14].

Here we report a systematic review and meta-analysis of studies that have evaluated the diagnostic test accuracy of the 4AT in older adults in all available care settings.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

METHODS

The methods and search strategy were documented in advance and published in the PROSPERO database (available at http://www.crd.york.ac.uk/PROSPERO/ with registration number CRD42019133702). The review and meta-analysis was conducted in compliance with the principles in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [15], and reported using the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines [16].

Search strategy and selection criteria

An inclusive search strategy was developed with a medical librarian. The validated delirium search syntax produced by the National Institute for Health and Clinical Excellence (NICE) clinical guidance for delirium was used to identify delirium [Appendix 2: search strategy]. The following databases were searched: MEDLINE® (OVID), EMBASE (OVID), PsycINFO (EBSCO), CINAHL (EBSCO), clinicaltrials.gov and the Cochrane Central Register of Controlled trials from 2011 (the year the 4AT was published online) to 21 December 2019. We conducted forward citation searches of included articles and checked reference lists of included articles for further articles of potential relevance. We contacted delirium experts from international delirium-focused organisations to identify relevant published or unpublished data and searched relevant conference proceedings.

Studies were included if they met the following criteria: (1) age ≥65; (2) attempted to examine the diagnostic accuracy of the 4AT for detection of delirium; (3) reference standard assessment of delirium made using standardised diagnostic criteria or a validated tool; and (4) cross-sectional, retrospective or prospective cohort design. If identified studies included adults both younger and older than the threshold age, the study authors were contacted to enquire about the possibility to access data on the older adults only. Studies in patients with delirium tremens were excluded.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Data extraction

Titles and abstracts were independently screened for inclusion by individuals in pairs of review authors (C.B. and R.G., and Z.T. and A.A.). Full-text screens were carried out independently by two review authors (Z.T. and A.M.). The reviewer pairs performed data extraction independently, resolving disagreement by discussion, or by involving a fifth review author (S.S.) where necessary.

Data were extracted on: type of study; setting; study population; patient demographics; prevalence of delirium; co-morbid illness or illness severity if reported, details of 4AT administration (timing, assessors etc.) and the reference standard; statistics used including adjustments made, and study conclusions. Test accuracy data were extracted to a two-by-two table (number of true positives, false positives, true negatives and false negatives for the 4AT). Study authors were contacted for further information on index and reference test results if insufficient data were provided to perform statistical analyses.

Risk of bias assessment

Studies were assessed for methodological quality by two independent review authors (R.G. and Z.T.) using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. Narrative summaries were generated describing risk of bias (high, low, or unclear) and concerns regarding applicability. As part of a tailoring phase of the QUADAS-2 tool, the item on the threshold used was omitted because the design of the 4AT pre-specifies the threshold to be used for delirium detection (cut-off ≥ 4). For the item on the appropriate interval between index test and reference standard, the interval was set to a maximum of three hours [Appendix 3: Assessment of methodological quality with the QUADAS-2 tool].

Statistical analysis

Meta-analyses were completed using Diagnostic Test Accuracy Meta-Analysis software, version 1.21 (https://crsu.shinyapps.io/dta_ma/) [17]. Pooled estimates of delirium prevalence were calculated using random effects models (*meta* package in R [18]). The primary outcome of

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

interest was the identification of delirium (presented as a dichotomous yes/no variable) by a reference standard (i.e. Diagnostic and Statistical Manual of Mental Disorders (DSM)) or a validated diagnostic tool such as the Confusion Assessment Method (CAM) [19]. Summary estimates of sensitivity and specificity with 95% confidence intervals (CI) were calculated using a bivariate random effects model. Receiver operating characteristic (ROC) curves were used to plot summary estimates of sensitivity and specificity.

A sensitivity analysis was performed including only those studies which were deemed to have an overall low risk of bias (that is, high study quality). Pre-planned subgroup analyses were also conducted to investigate clinical heterogeneity across studies: (i) excluding studies in patients with stroke, because of the potential influence of aphasia on the test, [20, 21], to assess test accuracy of the 4AT in non-stroke populations, (ii) analysing separately for studies using (a) a clinical reference standard (e.g. DSM) or (b) a validated assessment tool (e.g. the Confusion Assessment Method (CAM)). A post-hoc subgroup analysis was conducted to compare diagnostic accuracy of the English 4AT versus the translated versions.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

RESULTS

Study identification

We identified 853 records from our initial search and 3 records from conference abstracts (Figure 1). A total of 780 records remained after initial deduplication. Following title and abstract screening, 21 records had full-text review and 16 articles were included reporting 17 different studies [9, 10, 22-35]. The main reason for exclusion of articles was that studies were not designed as a diagnostic accuracy study of the 4AT and/or did not include data that allowed derivation of diagnostic test accuracy data. One conference proceeding reported two separate studies [25]. Two authors provided study data on subgroups of older patients [23, 27, 36].

Study characteristics

A summary of the characteristics of the included studies is provided in Table 1. The number of study participants ranged between 49 [26] and 785 [10]. The prevalence of delirium across the studies was 24.2% (95% CI 17.8-32%), varying between 10.5% [10] to 61.9% [28]. Eight studies validated a translated version of the 4AT in Italian, Persian, Thai, Russian, French, Norwegian and German. Two studies used a modified 4AT where the months of the year backwards test was replaced by the days of the week backwards to assess inattention [25, 32]; this modification does not affect the threshold scoring for delirium versus no delirium in the tool. Studies were conducted in inpatient general medical or geriatric medical wards, acute stroke units, emergency departments and post-operative care units, and nursing homes, in eleven countries. In one study in Australia 39% of participants were non-English speakers and required an interpreter during the assessment [28].

Study quality

The methodological quality of studies varied but was moderate to good overall. Potential for bias in studies was generally low, but where present was due to the selection of participants (excluding patients unable to give consent or those with dementia, n=2), the timing between the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

reference standard assessment and the 4AT (not reported (n=6) or exceeding the maximum interval of 3 hours (n=2)), and the blinding of assessments (unblinded raters (n=2) or blinding status unclear (n=3)) (Table 1 and Figure 2). Seven papers were of higher concern (rated high or unclear risk of bias across three areas). Nine studies were considered low risk overall.

Myrstad et al. [26] used a reference standard based on the whole length of stay (median 5 days) whereas the 4AT was performed in the first 24 hours of the admission; this could have led to a reduced sensitivity of the 4AT as some delirium arises after the first 24 hours. Hendry et al. [35] administered the 4AT as part of a larger cognitive test battery, therefore the index rater had knowledge of the participant's mental status beyond that elicited by the 4AT assessment that could have affected 4AT scoring. Gagné et al. [24] repeatedly administered the 4AT and the combined results were incorporated in the reported sensitivity and specificity. Asadollahi et al. [22] administered the 4AT only to those patients who had delirium according to DSM-5 criteria.

Diagnostic test accuracy

All 17 studies were included in the meta-analysis. The 4AT had a pooled sensitivity for detecting delirium of 0.88 (95% CI 0.80-0.93) and a pooled specificity of 0.88 (95% CI: 0.82 - 0.92), indicating good diagnostic accuracy of the 4AT as a tool to identify individuals at high risk of delirium (Table 2 and Figure 3). These estimates were broadly consistent across studies with the exception of two studies reporting lower sensitivities (both studies had a high or unclear risk of bias) [22, 26]. A sensitivity analysis including 9 studies with overall low risk of bias resulted in comparable summary estimates of sensitivity (0.87, 95% CI 0.84, 0.90) and specificity (0.88, 95% CI 0.81, 0.93) [Appendix 5].

The planned subgroup analysis excluding three studies in stroke patients resulted in similar summary estimates of sensitivity (0.86; 95% CI 0.77-0.92) and specificity (0.89; 95% CI 0.83, 0.93), suggesting robust results across populations (results of the other subgroup analyses are presented in (Table 2; results of the other subgroup analyses are presented in Appendix 4]. Three studies reported findings in subset of patients with known dementia, with

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

sensitivities of 0.94, 0.86, and 0.92 and specificities of 0.65, 0.71, and 0.79, in the Bellelli et al. [9], De et al. [28], and O'Sullivan et al. [34], studies, respectively.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

DISCUSSION

Statement of principal findings

This systematic review identified 17 studies involving 3721 patients evaluating the diagnostic test accuracy of the 4AT for detection of delirium in older patients (≥65y) across eleven countries, a variety of care settings and in multiple languages. The prevalence of delirium was 24.2% (N=945), ranging from 10.5%-61.9%. Pooled sensitivity and specificity were 0.88 and 0.88, respectively, indicating good accuracy. Notably, the sensitivity and specificity were balanced. Similar estimates were demonstrated when subgroup analyses were performed based on study quality and population type.

Results in the context of the current literature

Delirium detection remains a major challenge, with recent studies continuing to show underdetection [37]. An important factor in improving detection is the availability of validated assessment tools usable in clinical practice. The 4AT now has a substantial evidence base which is supportive of its use for this purpose. Coupled with this is also emerging evidence of implementation of the 4AT scale in routine clinical practice, for example in data from the National Hip Fracture Database which assesses the clinical care of >95% hip fracture patients in England, Wales and Northern Ireland. In 2018, 25% of approximately 60,000 4AT assessments (92% of all patients) performed in the 7 postoperative days (the audit period) were positive [11]. Though diagnostic accuracy data were not collected, these data suggest that the 4AT may be detecting the expected level of delirium.

There are many other tools in the literature; however those with profiles of intended use similar to the 4AT with more than two published diagnostic accuracy studies are the CAM, the 3D-CAM, and the brief CAM (bCAM) [6, 35, 38-44]. The CAM was first published in 1990 and is a widely used tool in research and clinical practice. There are 23 published CAM diagnostic test accuracy studies involving a total of 2629 patients [10, 38], with sensitivities of 0.09-1.0 and

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

specificities of 0.84-1.0 reported. There is limited published information on its performance in routine clinical care. One recent large study found a sensitivity of 0.27 [39] though the CAM was scored without the recommended preceding interview and cognitive testing. Alternative tools include the 3D-CAM, a 20-item variant of the CAM that takes 2-5 minutes to complete (median 3 minutes) [40], and the bCAM, a 2 minute, 4-item variant of the CAM designed and validated for use in the emergency department [41]. Both of these tools show generally good performance in published diagnostic test accuracy studies, with reported 3D-CAM sensitivities of 0.85-1.0 and specificities of 0.88-0.97 [40, 45-47], and reported bCAM sensitivities of 0.65-0.84 and specificities of 0.87-0.97 [35, 41-44]. To our knowledge there are currently no published clinical implementation data for these tools.

Our review provides evidence that the 4AT has good diagnostic test accuracy for identification of delirium, with a body of validation data comparable to the CAM. The 4AT has some advantages over the CAM and 3D-CAM, being shorter and simpler, and not requiring special training. Notably, the 4AT had a higher sensitivity than and similar specificity to the CAM in a recent STARD-compliant randomised controlled trial [10]. However, currently the 4AT lacks diagnostic accuracy data in palliative care settings and the community. The number of studies examining its performance in patients with known dementia is relatively small; the three studies presented in this review found lower specificity in delirium superimposed on dementia [9, 28, 34]. As with other delirium tools, studies on clinical implementation of the 4AT are relatively lacking. These kinds of studies might expose training needs or other challenges in implementation such as lower sensitivity when used in routine practice.

Strengths and weaknesses of the study

This is the first meta-analysis of 4AT diagnostic test accuracy studies. Our findings were broadly consistent across different care settings and languages. We published the protocol in advance, and we used systematic and robust methods including using a comprehensive search strategy, and independent reviewers to identify, select, appraise and synthesise relevant

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

studies. The selected studies originated from nine countries, and eight were conducted with a translated version of the tool. Thus, the findings of the review demonstrate good generalisability. The methodological quality of the studies was moderate to good, despite some uncertainty in relation to the conduct of the 4AT in four studies. The two studies showing low sensitivities both had high risk of bias. Due to the data of the studies included in this review, it was not possible to perform sensitivity analyses to determine the impact of time interval between tests and this should be the subject of further studies. In one study [30] we were unable to analyse data from those aged 65 or older separately; however the number of observations in patients <65 was likely low and would have a small potential impact on the overall results. Also, the Cochrane guidelines recommend the use of a single reference standard in order to prevent bias or ambiguity, but we included studies using either DSM-IV, DSM-5 or CAM as reference standard to maximise comprehensiveness.

Areas for further research

Methodological deficiencies related to the timings of the reference standard and 4AT identified in this review, as well as lack of adherence to the STARD guidelines, should be better addressed in future validation studies. Studies evaluating the 4AT in other settings and in patients with dementia, preferably taking into account the severity of dementia, are required. Clinical implementation studies evaluating 4AT performance including completion rates as well as diagnostic accuracy in routine clinical practice are also needed.

Conclusion

This meta-analysis quantifies the diagnostic accuracy of the 4AT. The psychometric performance is good and coupled with its simplicity and brevity, the present findings support ongoing adoption and evaluation of the 4AT in routine clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Gibb K, Seeley A, Quinn T, Siddiqi N, Shenkin S, Rockwood K, et al. The consistent burden in published estimates of delirium occurrence in medical inpatients over four decades: a systematic review and meta-analysis study. Age Ageing 2020; 49: 352-60.
2. Koirala B, Hansen BR, Hosie A, Budhathoki C, Seal S, Beaman A, et al. Delirium point prevalence studies in inpatient settings: A systematic review and meta-analysis. J Clin Nurs 2020.
3. Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in Older Persons: Advances in Diagnosis and Treatment. JAMA 2017; 318: 1161-74.
4. Williams ST, Dhesi JK, Partridge JSL. Distress in delirium: causes, assessment and management. Eur Geriatr Med 2020; 11: 63-70.
5. Scottish Intercollegiate Guidelins Network (SIGN). Risk reduction and management of delirium. Edinburgh 2019. https://www.sign.ac.uk/sign-157-delirium. Accessed 10 June 2020.
6. De J, Wand AP. Delirium Screening: A Systematic Review of Delirium Screening Tools in Hospitalized Patients. The Gerontologist 2015; 55: 1079-99.
7. van Velthuijsen EL, Zwakhalen SM, Warnier RM, Mulder WJ, Verhey FR, Kempen GI. Psychometric properties and feasibility of instruments for the detection of delirium in older hospitalized patients: a systematic review. Int J Geriatr Psychiatry 2016; 31: 974-989.
8. The 4 “A”s Test . www.the4AT.com. Accessed 10 June 2020.
9. Bellelli G, Morandi A, Davis DH, Mazzola P, Turco R, Gentile S, et al. Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. Age Ageing 2014; 43: 496-502.
10. Shenkin SD, Fox C, Godfrey M, Siddiqi N, Goodacre S, Young J, et al. Delirium detection in older acute medical inpatients: a multicentre prospective comparative

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

diagnostic test accuracy study of the 4AT and the confusion assessment method. BMC Med 2019; 17: 138.

11. National Hip Fracture Database (NHFD). Annual Report 2019. https://www.nhfd.co.uk/20/hipfractureR.nsf/docs/2019Report. Accessed 10 June 2020.
12. Vardy E, Collins N, Grover U, Thompson R, Bagnall A, Clarke G, et al. Use of a digital delirium pathway and quality improvement to improve delirium detection in the emergency department and outcomes in an acute hospital. Age Ageing May 2020. Online ahead of print.
13. Royal College of Physicians. NEWS2: Additional implementation guidance. 2020. https://www.rcplondon.ac.uk/projects/outputs/news2-additional-implementation-guidance. Accessed 10 June 2020.
14. Mansutti I, Saiani L, Palese A. Detecting delirium in patients with acute stroke: a systematic review of test accuracy. BMC Neurol. 2019; 19: 310.
15. Deeks JJBP, Gatsonis C. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0.0 London, UK: The Cochrane Collaboration; 2013.
16. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the P-DTAG, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA 2018; 319: 388-396.
17. Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. BMC Med Res Methodol 2019; 19: 81.
18. Balduzzi S, Rucker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid-Based Ment Health 2019; 22: 153-160.
19. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann Int Med 1990; 113: 941-948.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

20. Shaw R, Drozdowska B, Taylor-Rowan M, Elliott E, Cuthbertson G, Stott DJ, et al. Delirium in an Acute Stroke Setting, Occurrence, and Risk Factors. Stroke 2019; 50: 3265-3268.

21. Shaw RC, Walker G, Elliott E, Quinn TJ. Occurrence Rate of Delirium in Acute Stroke Settings: Systematic Review and Meta-Analysis. Stroke 2019; 50: 3028-3036.

22. Asadollahi A, Saberi M, Entezari M, Hoseini Z, Hasani SA, Saberi LF. Iranian version of 4AT, an Instrument for rapid delirium screening for later life. Int J Adv Appl Sci 2016; 3: 33-38.

23. Chang Y, Oravec N, Kent D, Nugent K, Cornick A, Hiebert B, et al., editors. Validation of the “4AT” delirium screening tool in a cardiac surgery patient population. 9th Annual Meeting of the American Delirium Society; 2019; Boston, USA.

24. Gagné A-J, Voyer P, Boucher V, Nadeau A, Carmichael P-H, Pelletier M, et al. Performance of the French version of the 4AT for screening the elderly for delirium in the emergency department. CJEM 2018; 20: 903-910.

25. Kazim R, MacLullich AMJ, Taffet GE, Agarwal KS, editors. 4AT Screening Tool used by Nurses is Both Sensitive and Specific for Detection of Delirium in Hospitalized Elders. 7th Annual Meeting of the American Delirium Society; 2016; Nashville, Tennessee, USA.

26. Myrstad M, Watne LO, Johnsen NT, Bors-Lind E, Neerland BE. Delirium screening in an acute geriatric ward by nurses using 4AT: results from a quality improvement project. Eur Geriatr Med 2019; 10: 667-671.

27. Saller T, MacLullich AMJ, Schäfer ST, Crispin A, Neitzert R, Schüle C, et al. Screening for delirium after surgery: validation of the 4 A's test (4AT) in the post-anaesthesia care unit. Anaesthesia 2019; 74: 1260-1266.

28. 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: 1322-1329.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

29. Kuladee S, Prachason T. Development and validation of the Thai version of the 4 'A's Test for delirium screening in hospitalized elderly patients with acute medical illnesses. Neuropsychiatr Dis Treat 2016; 12: 437-443.

30. Infante M, Pardini M, Balestrino M, Finocchi C, Malfatto L, Bellelli G, et al. Delirium in the acute phase after stroke: comparison between methods of detection. Neurol Sci 2017; 38: 1101-1104.

31. Kutlubaev MA, Bikbulatova LF, Akhmadeeva LR. Early diagnosing of delirium in the elderly with acute stroke. Adv Gerontol 2015; 28: 493-499.

32. Al-Jumayli M, Hiner JA, Taffet GE, Agarwal KS. Use of the 4AT to detect delirium in older hospitalized cancer patients. J Am Geriatr Soc 2018; 66: S33.

33. 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. 2013; 44: 3078-3083.

34. O'Sullivan D, Brady N, Manning E, O'Shea E, O'Grady S, N OR, 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: 61-68.

35. Hendry K, Quinn TJ, Evans J, Scortichini V, Miller H, Burns J, et al. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. Age Ageing 2016; 45: 832-837.

36. Saller T, MacLullich AMJ, Perneczky R. The 4AT - an instrument for delirium detection for older patients in the post-anaesthesia care unit. Anaesthesia 2020; 75: 410.

37. Geriatric Medicine Research Collaborative. Delirium is prevalent in older hospital inpatients and associated with adverse outcomes: results of a prospective multi-centre study on World Delirium Awareness Day. BMC Med 2019; 17: 229.

38. Heinrich TW, Kato H, Emanuel C, Denson S. Improving the Validity of Nurse-Based Delirium Screening: A Head-to-Head Comparison of Nursing Delirium-Screening Scale and Short Confusion Assessment Method. Psychosomatics 2019; 60: 172-178.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

39. Rohatgi N, Weng Y, Bentley J, Lansberg MG, Shepard J, Mazur D, et al. Initiative for Prevention and Early Identification of Delirium in Medical-Surgical Units: Lessons Learned in the Past Five Years. Am J Med 2019; 132: 1421-1430.

40. Marcantonio ER, Ngo LH, O'Connor M, Jones RN, Crane PK, Metzger ED, et al. 3D-CAM: derivation and validation of a 3-minute diagnostic interview for CAM-defined delirium: a cross-sectional diagnostic test study. Ann Int Med 2014; 161: 554-561.

41. 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: 1031-1036.

42. Han JH, Wilson A, Vasilevskis EE, Shintani A, Schnelle JF, Dittus RS, 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: 457-465.

43. Baten V, Busch HJ, Busche C, Schmid B, Heupel-Reuter M, Perlov E, 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: 1251-1262.

44. Wilson JE, Boehm L, Samuels LR, Unger D, Leonard M, Roumie C, et al. Use of the brief Confusion Assessment Method in a veteran palliative care population: A pilot validation study. Palliat Support Care 2019; 17: 569-573.

45. Mu DL, Ding PP, Zhou SZ, Liu MJ, Sun XY, Li XY, et al. Cross-cultural adaptation and validation of the 3D-CAM Chinese version in surgical ICU patients. BMC Psychiatry 2020; 20: 133.

46. Olbert M, Eckert S, Morgeli R, Kruppa J, Spies CD. Validation of 3-minute diagnostic interview for CAM-defined Delirium to detect postoperative delirium in the recovery room: A prospective diagnostic study. Eur J Anaesthesiol 2019; 36: 683-687.

47. Kuczmarska A, Ngo LH, Guess J, O'Connor MA, Branford-White L, Palihnich K, et al. Detection of Delirium in Hospitalized Older General Medicine Patients: A Comparison of the 3D-CAM and CAM-ICU. J Gen Intern Med 2016; 31: 297-303.

Study ID	Country	Language 4AT	Study Design	Patient population	Setting	Total N (sample size)	N (%) delirium	Sex	Mean age (years)	Reference Standard	Details of 4AT administration (timing, assessor etc.)
Al-Jumayli et al., 2018 [32]	USA	English	Retrospective, quality improvement study	Haematology, oncology patients age 70 or greater	Hematology-Oncology unit in Hospital (non-ICU)	95 patients; 160 hospital admissions over 1 year	20 (12.5%)	Not reported directly. For delirium: 53.7%	No delirium 76.2; Delirium 77.1 (SD not reported	Chart review by 2-3 physicians	Administered by nurses every shift
Asadollahi et al. 2016 [22]	Iran	Persian			Nursing homes and daily care centres	293	164 (56%)	0% female	69.3 ± 1.47	DSM-5 by psychiatrist	24h after reference standard assessment by psychiatrist (blinded)
Bellelli et al. 2014 [9]	Italy	Italian	Prospective consecutive patient study	Older hospital inpatients	Acute geriatric and rehabilitation wards	234	29 (12.4%)	64.1% female	83.9 ± 6.1	Structured reference standard assessment, based on DSM-IV-TR by geriatrician	15-30 minutes before reference standard by geriatrician (blinded)
Chang et al., 2019 [23]	Canada	English	Prospective study	Post-operative cardiac surgery patients	Postoperative cardiac surgery ward	91 out of 137 patients (237 observations)	34 (24.8%) (whole sample)	35.7% female (whole sample)	67.5 (9.0) (whole sample)	DSM-5 by trained reference rater	Within 2h by researchers (blinded)
De et al. 2016 [28]	Australia	English	Prospective study	Older hospital patients	Geriatric and orthogeriatric hospital wards	257	159 (61.9%)	56.8% female	85 ± 7.3	DSM-5 by geriatrician	Within 30 min by nurses (blinded)
Gagné et al., 2018 [24]	Canada	French		Older Emergency Department patients	Emergency Departments	319	49 (15.4%)	53.3% female	76.8 (7.4)	CAM by researchers	Same time by researchers (not blinded)
Hendry et al., 2016 [35]	UK	English	Prospective consecutive patient study	Older hospital patients	Geriatric hospital wards	500	93 (18.6%) definite delirium; 104 (20.8%) possible delirium	87% female	83.1 ± 6.7	DSM-5 by geriatrician (using checklist)	Within 2h by researcher (blinded)
Infante et al. 2017 [30]	Italy	Italian	Prospective study	Stroke and TIA patients	Acute stroke unit	100	32 (32%)	Not reported	79 (median), range 19-93, IQR not	DSM-5 by neurologist; diagnoses afterwards reviewed independently by two neurologists	Same day by neurologist (not blinded)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

									reported		
Kazim et al., 2016 [25]	USA	English	Retrospective quality improvement study	General Medical and surgical acute care unit patients	Acute care units (not ICU or psychiatry). Study 1: Academic Center; Study 2: Community hospitals	Study 1: 310; Study 2: 188	Study 1: 53 (17%) Study 2: 17(9%)	Not reported	Study 1: age 78; Study 2: 80.8 (SD not reported)	Chart review tool	Administered by nurses every shift
Kuladee et al., 2016 [29]	Thailand	Thai	Cross-sectional study	Older hospital patients	General Medical Ward	97	24 (24.7%)	49.5% female	73.6 ± 8.17	DSM-4-TR (and using Thai Delirium Rating Scale) by psychiatrist	Within 30 min by nurses (blinded)
Kutlubaev et al., 2016 [31]	Russia	Russian	Prospective study	Hyperacute older stroke patients	Neurovascular department	73	33 (45.2%)	71.2% female	74 (SD not reported)	DSM-4 by neurologist	Patient examined twice at interval of 12-24h by neurologist
Lees et al. 2013 [33]	UK	English	Prospective consecutive study	Acute stroke inpatients	Stroke unit	108	58 (16.6%)	51.8% female	74 (median) IQR 64-85	CAM and case notes extraction by medical student	Same day by medical student (blinded)
Myrstad et al., 2019 [26]	Norway	Norwegian	Retrospective, quality improvement study	Older hospital patients	Acute geriatric ward	49	21 (42.8%)	54.2% female	87 (SD not reported), range 68-99	Diagnosed retrospectively according to DSM-5 using chart-based method (over whole admission, mean/median approximately 5 days, up to 15 days)	Within 24 hours of admission by nurses (blinded)
O'Sullivan et al., 2018 [34]	Ireland	English	Prospective non-consecutive study	Older Emergency Department patients	Emergency Department	350	58 (16.6%)	Not reported	77 (median, IQR not reported)	DSM-5 by geriatrician; Use of DRS-R98, MMSE, IQCODE	Within 3h by researcher (blinded)
Saller et al., 2019 [36]	Germany	German	Prospective consecutive study	Elective surgical patients	Recovery room	143 out of 543 patients	15 (10.5%)	54.6% female	73 (SD not reported) IQR 68–76, range 65-96	DSM-5 by medical doctors; Use of CAM-ICU	A few minutes before reference standard assessment by researchers
Shenkin et al., 2019 [10]	UK	English	Prospective study	Acute medical patients	Emergency room and acute geriatric wards	395 out of 785 patients	49 (12.4%)	54.2.% female	81.4 (6.4)	DSM-5 by researchers. Use of structured reference standard assessment using	Within 2h by researchers (blinded). Order of 4AT and reference

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

										DRS-R98, attention and memory tests, and arousal scales	standard randomised

Table 1: Characteristics of included studies

Notes. CAM = Confusion Assessment Method; DSM-4 = Diagnostic and Statistical Manual of Mental Disorders, 4th edition; DSM-4-TR = Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text revision; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, 5th edition; IQR = interquartile range; SD = standard deviation; CAM= Confusion Assessment Method; CAM-ICU = Confusion Assessment Method-Intensive Care Unit; DRS-R98 = Delirium Rating Scale-Revised 98; MMSE = Mini-Mental Status Examination; IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly.

The Infante et al. [30] study included patients aged under 65.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Application of 4AT	No. of studies (observations)	Sensitivity (95% CI)	Specificity (95% CI)
All studies	17 (3721)	0.88 (0.80-0.93)	0.88 (0.82-0.92)
Sensitivity analysis (low risk of bias)	9 (2252)	0.87 (0.84-0.90)	0.88 (0.81-0.93)
Subgroup analysis (excluding stroke)	14 (3440)	0.86 (0.77-0.92)	0.89 (0.83-0.93)

Table 2. Summary estimates of sensitivity and specificity.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

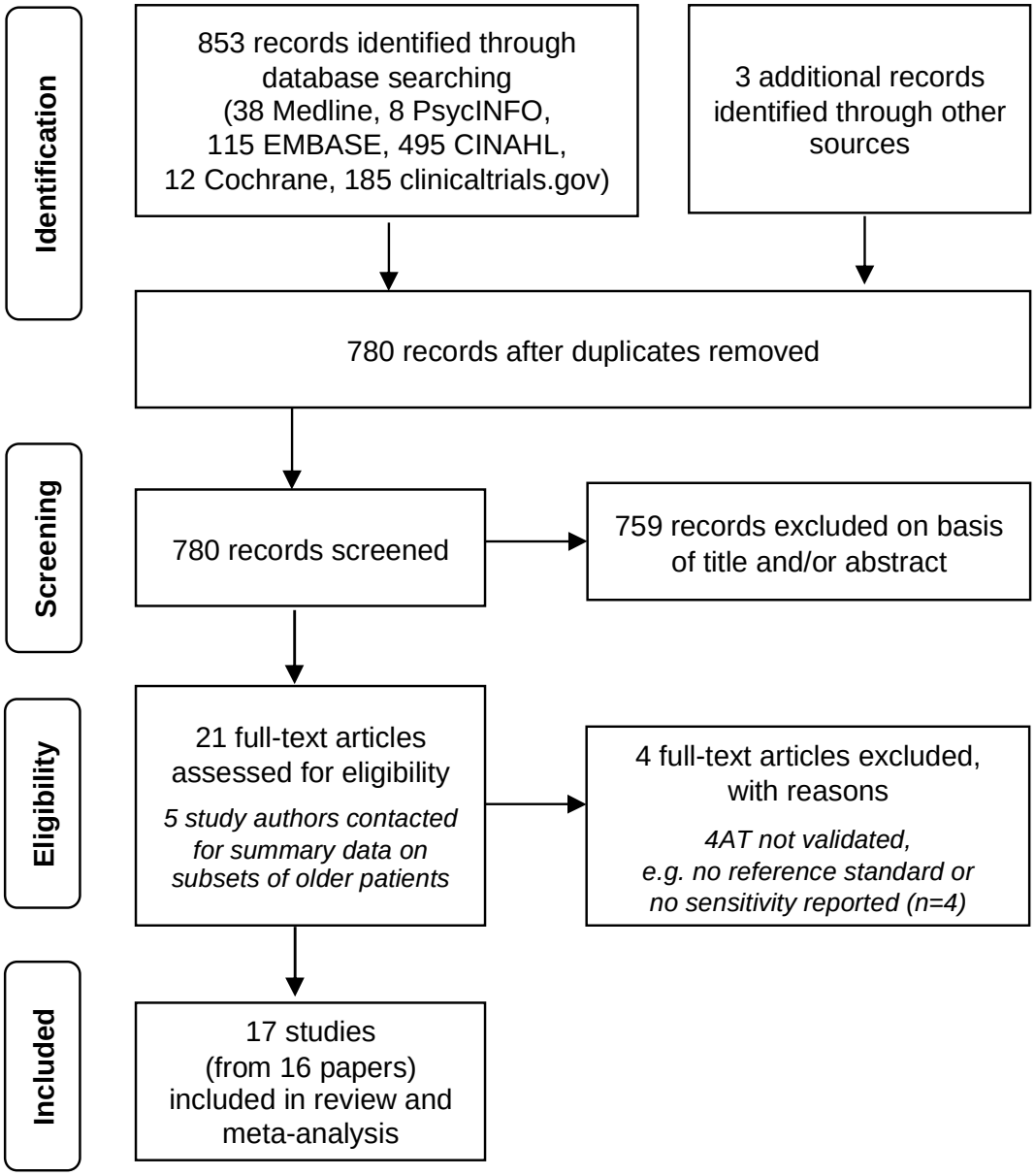


Figure 1. PRISMA flow chart diagram for the search and study selection process. PRISMA: *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

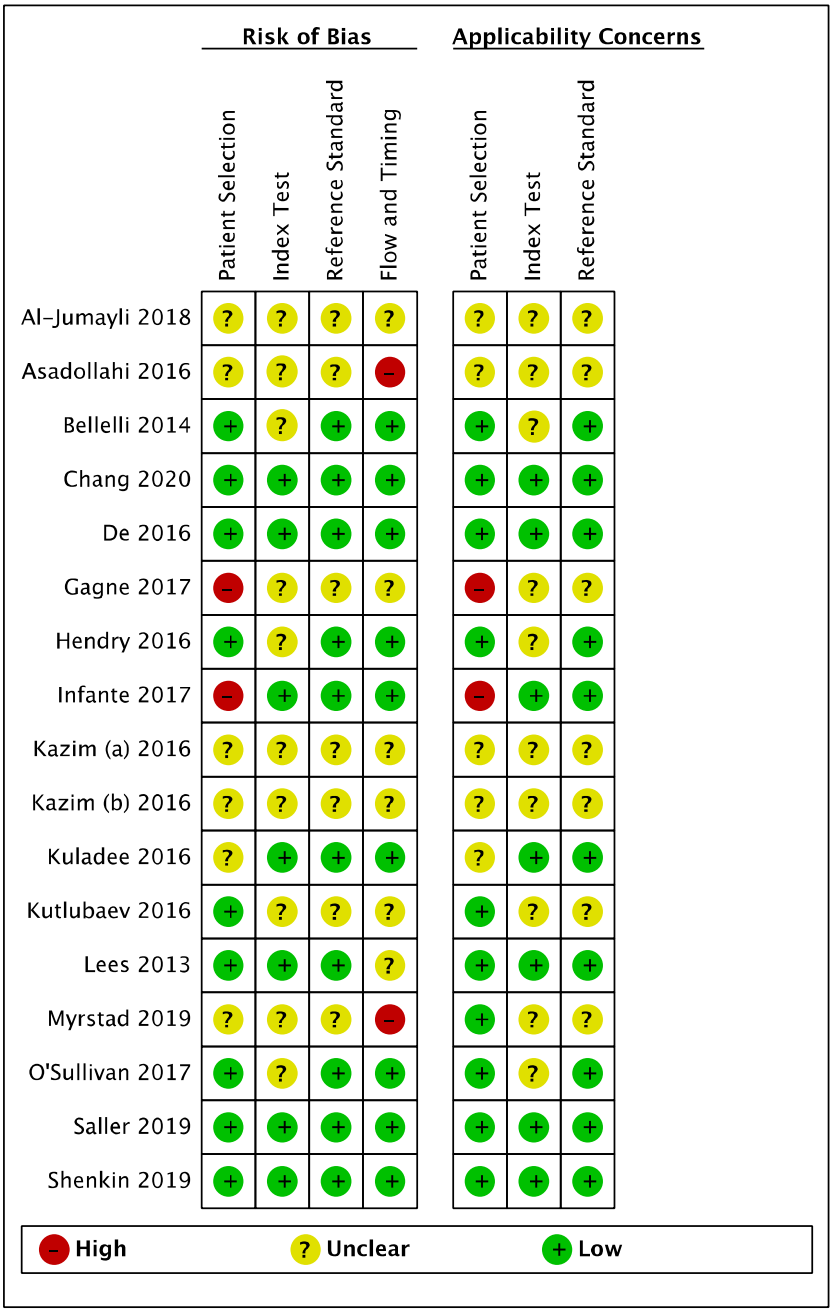


Figure 2. Risk of bias and applicability concerns summary (see [Supplementary File X] for QUADAS-2 assessment criteria).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

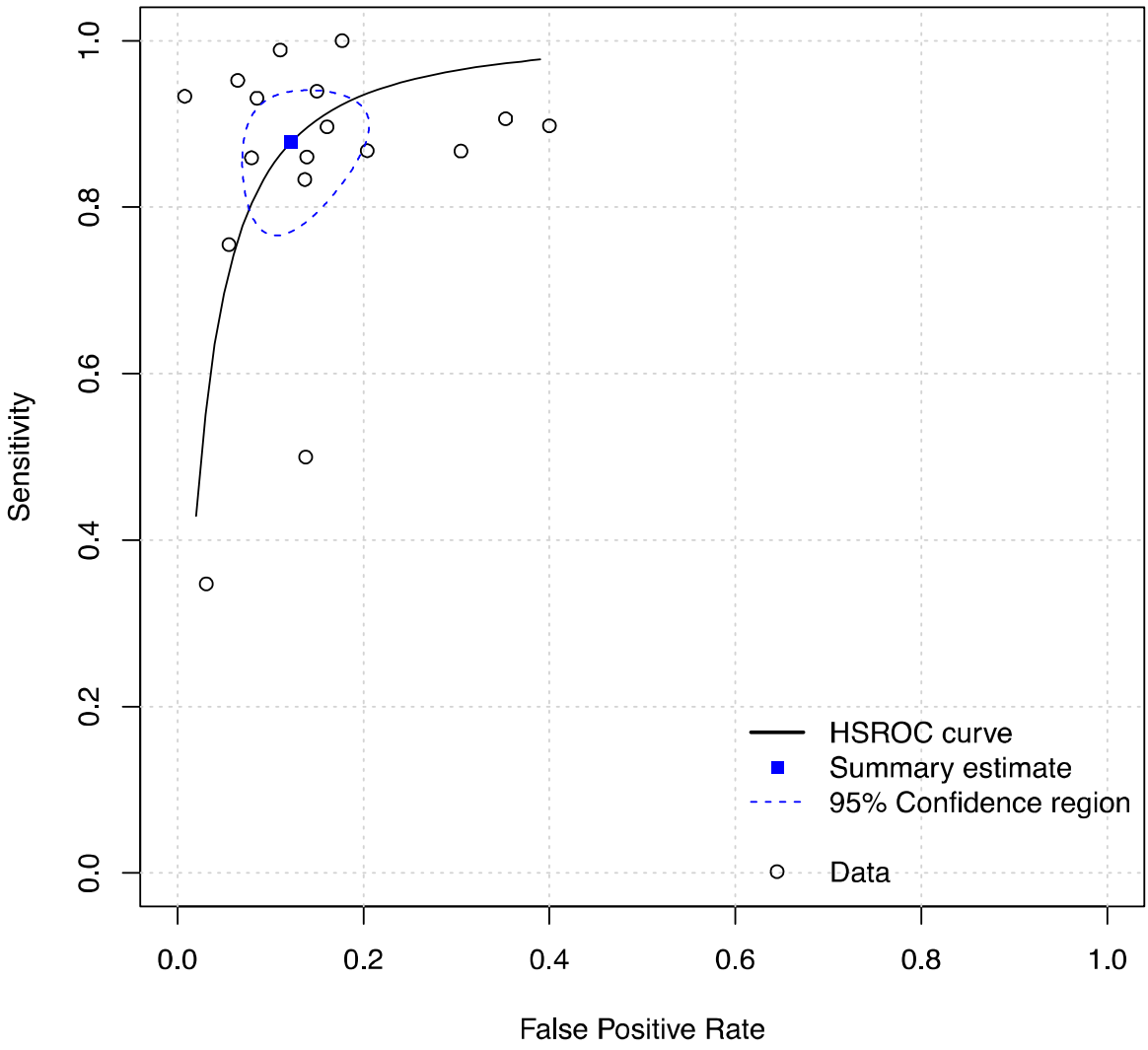


Figure 3. Hierarchical Summary Receiver Operating Characteristic (HSROC) curve of the 4AT for identifying individuals with delirium: the bivariate summary estimates (solid ellipses), with the corresponding 95% confidence ellipses (dashed lines).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Appendix 1. The 4AT

4AT

Assessment test for delirium & cognitive impairment

(label)

Patient name:

Date of birth:

Patient number:

Date: **Time:**

Tester:

CIRCLE

[1] ALERTNESS

This includes patients who may be markedly drowsy (eg. difficult to rouse and/or obviously sleepy during assessment) or agitated/hyperactive. Observe the patient. If asleep, attempt to wake with speech or gentle touch on shoulder. Ask the patient to state their name and address to assist rating.

Normal (fully alert, but not agitated, throughout assessment)	**0**
Mild sleepiness for <10 seconds after waking, then normal	**0**
Clearly abnormal	**4**

[2] AMT4

Age, date of birth, place (name of the hospital or building), current year.

No mistakes	**0**
1 mistake	**1**
2 or more mistakes/untestable	**2**

[3] ATTENTION

Ask the patient: "Please tell me the months of the year in backwards order, starting at December." To assist initial understanding one prompt of "what is the month before December?" is permitted.

Months of the year backwards	Achieves 7 months or more correctly	**0**
	Starts but scores <7 months / refuses to start	**1**
	Untestable (cannot start because unwell, drowsy, inattentive)	**2**

[4] ACUTE CHANGE OR FLUCTUATING COURSE

Evidence of significant change or fluctuation in: alertness, cognition, other mental function (eg. paranoia, hallucinations) arising over the last 2 weeks and still evident in last 24hrs

No	**0**
Yes	**4**

4 or above: possible delirium +/- cognitive impairment
1-3: possible cognitive impairment
0: delirium or severe cognitive impairment unlikely (but delirium still possible if [4] information incomplete)

4AT SCORE ☐

GUIDANCE NOTES Version 1.2. Information and download: **www.the4AT.com**

The 4AT is a screening instrument designed for rapid initial assessment of delirium and cognitive impairment. A score of 4 or more *suggests* delirium but is not diagnostic: more detailed assessment of mental status may be required to reach a diagnosis. A score of 1-3 suggests cognitive impairment and more detailed cognitive testing and informant history-taking are required. A score of 0 does not definitively exclude delirium or cognitive impairment: more detailed testing may be required depending on the clinical context. Items 1-3 are rated *solely on observation of the patient at the time of assessment.* Item 4 requires information from one or more source(s), eg. your own knowledge of the patient, other staff who know the patient (eg. ward nurses), GP letter, case notes, carers. The tester should take account of communication difficulties (hearing impairment, dysphasia, lack of common language) when carrying out the test and interpreting the score.

Alertness: Altered level of alertness is very likely to be delirium in general hospital settings. If the patient shows significant altered alertness during the bedside assessment, score 4 for this item. **AMT4 (Abbreviated Mental Test - 4):** This score can be extracted from items in the AMT10 if the latter is done immediately before. **Acute Change or Fluctuating Course:** Fluctuation can occur without delirium in some cases of dementia, but marked fluctuation usually indicates delirium. To help elicit any hallucinations and/or paranoid thoughts ask the patient questions such as, "Are you concerned about anything going on here?"; "Do you feel frightened by anything or anyone?"; "Have you been seeing or hearing anything unusual?"

© 2011-2014 MacLullich, Ryan, Cash

Note: The 4AT can be downloaded from www.the4at.com in different languages.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Appendix 2: Search strategy

MEDLINE Search terms

"4 A's test".mp.

2. "four A's Test".mp.

3. "4 'A's Test".mp.

4. "4 A-T".mp.

5. "four A-T".mp.

6. "4-A Test".mp.

7. "4 A's scale".mp.

8. "four 'A's test".mp.

9. '4AT'.mp.

10. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9

11. limit 10 to yr="2011 -Current"

12. deliri$.ti,ab.

13. (acute adj2 (confusion$ or "brain syndrome" or "brain failure" or "psycho-organic syndrome" or "organic psychosyndrome")).mp.

14. (terminal$ adj restless$).mp.

15. toxic confus$.mp.

16. delirium/

17. confusion/

18. 12 or 13 or 14 or 15 or 16 or 17

19. 11 and 18

Note: The delirium search strategy was taken from:

https://www.nice.org.uk/guidance/cg103/documents/delirium-appendix-c-search-strategies2.

The strategies for other databases are available on request.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Conference proceedings from the following professional societies were searched: Scottish Delirium Association; European Delirium Association (EDA); American Delirium Society (ADS); and Australasian Delirium Association (ADA). Members of the EDA, ADS and ADA were also contacted via email and twitter to identify relevant published or unpublished data.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Appendix 3.

Supplementary Table 3. Assessment of methodological quality with the QUADAS-2 tool.

Domain	**Patient selection**	**Index test**	**Reference standard**	**Flow and timing**
Description	Describe methods of patient selection: describe included patients (prior testing, presentation, intended use of index test and setting).	Describe the index test and how it was conducted and interpreted.	Describe the reference standard and how it was conducted and interpreted.	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2 x 2 table (refer to flow diagram): describe the time interval and any interventions between index test(s) and reference standard.
Signalling questions (yes, no, unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without	Is the reference standard likely to correctly classify the target	Was there an appropriate interval between index test(s) and

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	Was a case-control design avoided? Did the study avoid inappropriate exclusions?	knowledge of the results of the reference standard? Was the person who administered 4-AT not trained/not expert in delirium?	condition? Were the reference standard results interpreted without knowledge of the results of the index test?	reference standard (3h maximum)? Did all patients receive the same reference standard? Were all patients included in the analysis?
Risk of bias	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or its interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	—

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

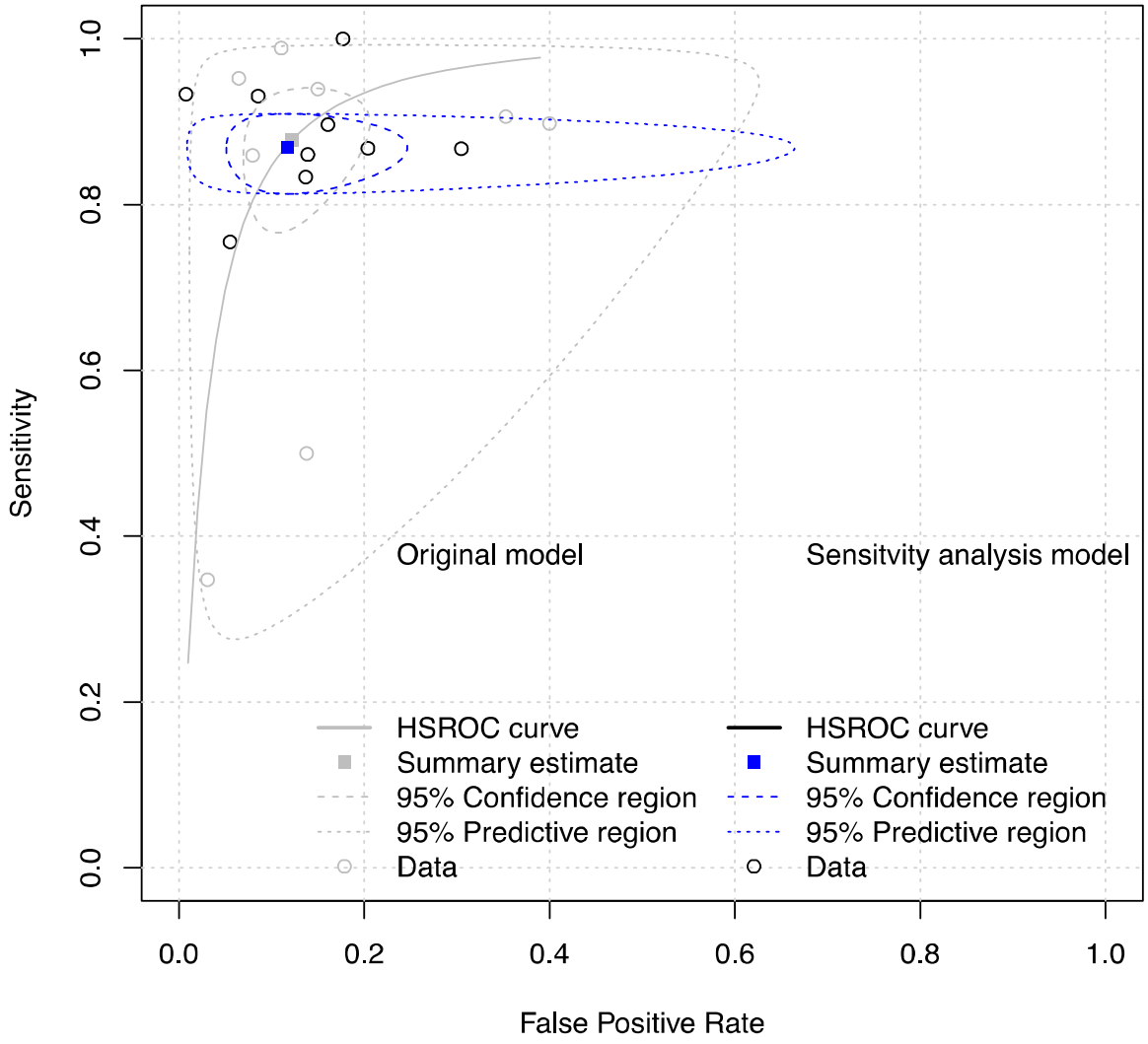
Appendix 4.

Supplementary Table 4. Summary estimates of sensitivity and specificity per subgroup.

Subgroup	No. of studies (observations)	Sensitivity (95% CI)	Specificity (95% CI)
Studies using a clinical reference standard (e.g. DSM)	12 (2659)	0.83 (0.73-0.90)	0.89 (0.82-0.93)
Studies using a validated assessment tool	4 (1062)	0.95 (0.84-0.94)	0.86 (0.83-0.92)
Studies using the English 4AT	9 (2413)	0.91 (0.80-0.94)	0.88 (0.83-0.92)
Studies using a translated version of the 4AT	8 (1308)	0.83 (0.66-0.92)	0.88 (0.75-0.95)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128280; this version posted June 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Appendix 5.

Supplementary Figure 4. Hierarchical Summary Receiver Operating Characteristic (HSROC) curve of the 4AT for identifying individuals with delirium: results from sensitivity analysis.

The sensitivity analysis included 9 studies rated as low risk of bias. Hierarchical Summary Receiver Operating Characteristic (HSROC) curve of the 4AT for identifying individuals with delirium: the bivariate summary estimates (solid ellipses), with the corresponding 95% prediction ellipses (outer dotted lines) and 95% confidence ellipses (inner dashed lines).