Received: 6 July 2023 | Revised: 11 December 2023 | Accepted: 16 December 2023
DOI: 10.1111/jgs.18751

REVIEW ARTICLE

Journal of the American Geriatrics Society

# Delirium detection tools show varying completion rates and positive score rates when used at scale in routine practice in general hospital settings: A systematic review

Rose S. Penfold BA (Hons), BMBCh, MPH[1] | Charlotte Squires MBChB[2] | Alisa Angus MBChB[2] | Susan D. Shenkin MBChB, BSc (Hons), MSc, MD[1] | Temi Ibitoye BSc (Hons)[3] | Zoë Tieges PhD[4] | Karin J. Neufeld MD, MPH[5] | Thiago J. Avelino-Silva MD, PhD[6] | Daniel Davis BSc (Hons), MB ChB, PhD[7] | Atul Anand MB ChB, PhD[8] | Andrew D. Duckworth BSc (Hons), MBChB, MSc, FRCSEd (Tr&Orth), PhD[9] | Bruce Guthrie MB BChir, MSc, PhD[10] | Alasdair M. J. MacLullich BSc (Hons), MB ChB, PhD[3]

[1]Edinburgh Delirium Research Group, Ageing and Health and Advanced Care Research Centre, Usher Institute, University of Edinburgh, Edinburgh, UK
[2]NHS Lothian, Edinburgh, UK
[3]Edinburgh Delirium Research Group, Ageing and Health, Usher Institute, University of Edinburgh, Edinburgh, UK
[4]School of Computing, Engineering and Built Environment, Glasgow Caledonian University, Glasgow, UK
[5]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada
[6]Global Brain Health Institute, University of California, San Francisco, California, USA
[7]MRC Unit for Lifelong Health and Ageing, UCL, London, UK
[8]Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK
[9]Usher Institute, University of Edinburgh, Edinburgh, UK
[10]Advanced Care Research Centre, Usher Institute, University of Edinburgh, Edinburgh, UK

## Abstract

**Background:** Multiple short delirium detection tools have been validated in research studies and implemented in routine care, but there has been little study of these tools in real-world conditions. This systematic review synthesized literature reporting completion rates and/or delirium positive score rates of detection tools in large clinical populations in general hospital settings.

**Methods:** PROSPERO (CRD42022385166).

Medline, Embase, PsycINFO, CINAHL, and gray literature were searched from 1980 to December 31, 2022. Included studies or audit reports used a validated delirium detection tool performed directly with the patient as part of routine care in large clinical populations ($n \geq 1000$) within a general acute hospital setting. Narrative synthesis was performed.

**Results:** Twenty-two research studies and four audit reports were included. Tools used alone or in combination were the Confusion Assessment Method (CAM), 4 'A's Test (4AT), Delirium Observation Screening Scale (DOSS), Brief CAM (bCAM), Nursing Delirium Screening Scale (NuDESC), and Intensive Care Delirium Screening Checklist (ICDSC). Populations and settings varied and tools were used at different stages and frequencies in the patient journey, including on admission only; inpatient, daily or more frequently; on admission and as inpatient; inpatient post-operatively. Tool completion rates ranged from 19% to 100%. Admission positive score rates ranged from: CAM 8%–51%; 4AT 13%–20%. Inpatient positive score rates ranged from: CAM 2%–20%, DOSS

**Correspondence**
Rose S. Penfold, Edinburgh Delirium Research Group, Ageing and Health, Usher Institute, University of Edinburgh, Room S1642, Royal Infirmary of Edinburgh, 51 Little France Crescent, Edinburgh, EH16 4SA, UK
Email: rose.penfold@ed.ac.uk

**Funding information**
Wellcome Trust, Grant/Award Number: 223499/Z/21/Z

6%–42%, and NuDESC 5–13%. Postoperative positive score rates were 21% and 28% (4AT). All but two studies had moderate–high risk of bias.

**Conclusions:** This systematic review of delirium detection tool implementation in large acute patient populations found clinically important variability in tool completion rates, and in delirium positive score rates relative to expected delirium prevalence. This study highlights a need for greater reporting and analysis of relevant healthcare systems data. This is vital to advance understanding of effective delirium detection in routine care.

**KEYWORDS**
delirium, detection, geriatric assessment, hospitals, older people, routine data, systematic review

## INTRODUCTION

Delirium is an acute, severe neuropsychiatric syndrome with an overall occurrence of 23% in adult medical hospital inpatients.[1] It is associated with adverse outcomes, including higher mortality, new institutionalization, and higher risk of future dementia.[2–4]

Delirium detection is essential to deliver effective care. A delirium diagnosis prompts access to specific management pathways and facilitates accurate communication with patients and relatives. Several national guidelines and care standards recommend routine delirium detection using validated tools.[5–10] Yet delirium remains under-detected in routine care, with studies comparing clinically-documented rates of delirium with reference-standard assessments demonstrating widespread under-detection.[11–13] Hospital discharge administrative data also demonstrate under-detection and under-coding.[14]

Multiple delirium detection tools have been developed with high sensitivity and specificity in diagnostic test accuracy studies.[15,16] Yet good tool performance in validation studies, where assessments have generally been performed by research teams, does not guarantee that tools will function well in usual care. Indeed, some studies of real-world practice suggest that delirium tools are commonly not completed, and that tools may produce lower delirium positive score rates than expected.[12,17–20]

Suboptimal completion rates could indicate difficulties in performing the particular tool by clinical staff, or the tool not being attempted for various reasons including culture, staff awareness, stage of tool implementation within the given healthcare system, timepoint in the patient journey the tool is used, and so forth.[17,21–23] Considering completed tools, if the positive score rates are much lower than the expected delirium prevalence in the target population, this suggests that the sensitivity of the tool in practice might be lower than the sensitivity shown in research settings.[12,18–20]

**Key points**

- We found 22 research studies and 4 audit reports examining 6 different validated delirium detection tools in acute care
- Tool completion and positive score rates varied widely in diverse populations
- Some studies reported positive score rates far below expected for the setting, despite high tool completion

**Why does this paper matter?**

Delirium underdetection is of immense concern. This study found variable disparities between tool performance in diagnostic test accuracy studies and in performance in real-world practice. Analyzing and reporting routine healthcare systems data is vital for effective delirium detection at scale.

Lower real-world sensitivity could reflect multiple factors including staff training, clinical time constraints, timepoint or frequency of testing in the patient journey, or the operational method of tool completion. Evidence on these two metrics of *tool completion rates* and *delirium positive score rates* is crucial for clinicians and policymakers aiming to improve delirium detection because it can inform the choice of tool appropriate for the setting and clinical population.

We performed a systematic review of large-scale ($N \geq 1000$) studies and audit reports examining delirium detection tool metrics when tools are used in routine clinical practice in general hospital settings (excluding critical care). We report: (a) tool completion rate (where available), and (b) delirium positive score rate. We report contextual factors including the clinical population, stage of the patient journey, the staff discipline(s) using the tool and, where reported, information on factors that might affect tool completion.

## METHODS

The systematic review is reported according to Preferred Reporting of Items in Systematic Reviews and Meta-analysis (PRISMA) guidelines (Supplementary Table 1).[24] The protocol was prospectively registered with PROSPERO (CRD42022385166).

### Search strategy and data sources

The search strategy was developed with academic librarian support and included three main concepts: "delirium", "assess*" (and synonyms), and "clinical practice" (and synonyms). Search strategies were developed for Medline, Embase, and PsycINFO (all via Ovid), and CINAHL (via EBSCO) (Supplementary Table 2). All databases were searched from 1980 when delirium was first coded in the Diagnostic and Statistical Manual of Mental Disorders (DSM) III,[25] to December 31, 2022. Study eligibility criteria are below.

### Inclusion and exclusion criteria

Studies using any peer-reviewed methodology were included if they included adults aged ≥18 years admitted to an acute hospital setting, described the use of a specified delirium assessment tool in study participants, reported tool use in an unselected population of ≥1000 patients, and were published in English, or translatable to English using online tools. No restrictions were placed on the study type or aims, or geographical location.

Studies of patients admitted to critical care or community settings were excluded, as were studies where the assessment was not performed directly with the patient (for example, studies using Electronic Health Record (EHR) data to develop a prediction model retrospectively). Studies only reporting on tools for rating delirium severity or resolution (rather than detection) or studies where tools were not used in routine clinical practice were excluded, as were systematic reviews, meta-analyses, abstracts, letters to editors, and opinion pieces.

### Title and abstract screening

All searches were performed on December 31, 2022, with study de-duplication using Covidence software (Veritas Health Innovation, Melbourne, Australia), with manual checking by RSP of software-identified duplicates. Title and abstract screening were independently performed in Covidence by two reviewers (RSP and CS). Discrepancies between reviewers at the title/abstract screening stage were resolved through discussion, and articles with persisting discrepancies were included in a full-text review.

### Full-text review

Full-text review was performed independently in Covidence by two reviewers (RSP and AAng). Excluded full-text articles were assigned a single primary reason for exclusion (Figure 1). Disagreement between reviewers was resolved through discussion involving an additional senior reviewer (AMJM) for persisting disagreement.

For any identified abstracts (e.g., published conference abstracts), we searched the databases above for a related full-text publication. We used forward citation on all included studies to identify relevant peer-reviewed publications that may have been missed in the search algorithm.

We scoped gray literature to identify additional relevant publications, including audit reports, conference abstracts and papers not picked up in the systematic search.

### Risk of bias

Two reviewers (RSP and AAng) independently assessed studies for risk of bias (RoB) using an adapted version of the ROBINS-E quality assessment tool for observational studies.[26] Discrepancies were resolved through discussion, with an additional reviewer available if required (AMJM). Studies were assessed as having low RoB, some concerns, high RoB, or non-applicable across seven domains (confounding, measurement of exposure, participant selection, post-exposure interventions, missing data, measurement of the outcome, and selection of the reported result), with an overall judgment assigned.

### Data extraction and synthesis

Two independent reviewers (RSP and AAng) extracted data for each study on the assessment tool used, the timepoint(s) and frequency of assessment in the patient journey, and whether the study reported on the entire eligible population based on the inclusion criteria or only on patients with a completed delirium assessment. For studies reporting on the entire eligible population, we

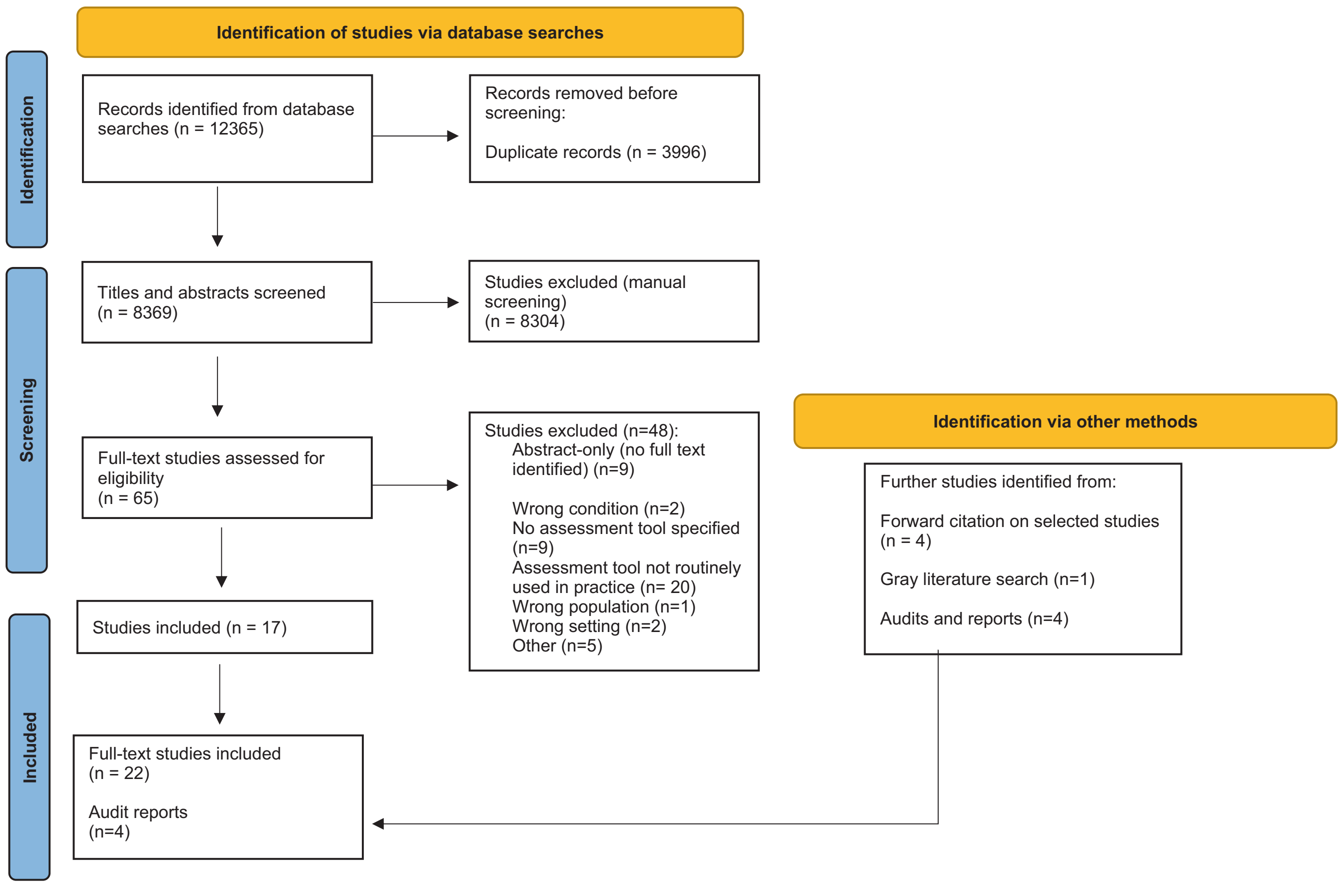


FIGURE 1 Study inclusion diagram.

extracted or calculated (if not explicitly reported) the proportion of eligible patients assessed using the tool (tool completion rate). We also extracted or calculated (if not explicitly reported) the positive score rate. If calculated, this was the number of patients with a positive score or number of positively scored observations divided by the total number of patients assessed or total number of observations.

For studies reporting on patients in both acute hospitals and excluded settings (e.g., critical care and rehabilitation facilities), we extracted data only for patients admitted to acute hospitals where possible. It is specified if the study did not report separately for the different settings.

Narrative data synthesis was performed due to anticipated heterogeneity in study populations and methodologies. Results are reported in tables and summary figures (Integrated Development for R. Rstudio, PBC, Boston, MA, USA).

# RESULTS

## Search results

After deduplication, 8369 articles were screened and 65 further evaluated. Nine articles were conference abstracts for which no related full-text study could be identified (Supplementary Table 3). Of the 56 full-text studies screened, 17 (30%) met the inclusion criteria. There was title-abstract agreement between reviewers in 8271/ 8369 (99%) of cases and 50/56 (89%) of cases at full-text review. Studies excluded after full-text review are summarized in Supplementary Table 4. No studies were excluded for language reasons. A further four papers were identified using forward citation on included full-text studies, and one paper during gray literature search.

## Study characteristics and patient populations

Twenty-two studies were included, representing over 500,000 patients (Table 1). Study denominators were either number of patients or total number of observations, precluding overall summary statistics. Studies were conducted in the United States (7), United Kingdom (5), Switzerland (5), Brazil (2), Colombia (1), New Zealand (1) and Italy (1). The average age of study participants ranged from 55–86 years. Fifteen (68%) studies only included adults at least ≥60 years; one included patients ≥50 years and six included all patients ≥18 years. Most

**TABLE 1** Summary of all included studies and patient populations.

| Author, year | Country | Study design | Type of hospital/ setting | Patient population | Total number of patients in study | Study inclusion criteria (including age and patient consent) | Study exclusion criteria | Age (mean (SD))/ median (IQR)) | Sex F (%) |
|---|---|---|---|---|---|---|---|---|---|
| Alhairadi et al., 2021 | NZ | Retrospective analysis of routine EHR data | Acute hospitals, two sites | Acute medical and geriatrics admissions | 7799 | ≥75 years | None specified | 84 (6) | 58 |
| Anand et al., 2022 | UK | Retrospective analysis of routine EHR data | Acute hospitals, two sites | Acute medical and surgical admissions (Salford only) | 82,770 admissions (31,266 patients) | ≥65 years | Not resident in Scotland (Lothian data only) | 79 (8) | 54 |
| Avelino-Silva et al., 2017 | Brazil | Prospective cohort | Acute hospital | Geriatric ward[a] | 1409 | ≥60 years, acute illness, patient consent | End of life care, incomplete data, LOS < 48 hours | 80 | 61 |
| Corradi et al., 2016 | US | Retrospective analysis of routine EHR data | Acute hospital | All inpatients | 88,206 encounters | ≥18 years with CAM scores between Sept 2012 and June 2015 | No CAM score | NA | NA |
| Di Bari et al., 2022 | Italy | Cross-sectional cohort | Academic tertiary and community hospitals | Patients admitted via the ED | 3358 | ≥75 years | Out of area, ophthalmological problems, software issue | 83 | 56 |
| Dulin et al., 2022 | US | Retrospective cohort | Academic tertiary hospital | Patients admitted from home | 93,388 admissions | ≥18 years; hospitalized through office, clinic or ED; discharged alive | Admitted not from home, admitted to ICU, died pre-discharge | 57 | 48 |
| Fuchs et al., 2020 | Switzerland | Prospective cohort | Acute university hospital | Discharged from acute medical and surgical units | 10,261 | ≥65 years | LOS < 1 day, missing data | 77 | 41 |
| Garcez et al., 2020 | Brazil | Retrospective cohort | Tertiary university hospital | Geriatric ward[a] | 1554 | ≥65 years; nonelective; has admission arousal assessment | Discharged or died <48 h | 81 | 61 |
| Han et al., 2021 | UK | Prospective cohort | Acute hospital | Patients with an acute hip fracture | 1082 | ≥60 years | None specified | 84 (9) | 73 |
| LaHue et al., 2021 | US | Retrospective cohort | Acute hospital | Admitted to specified medical and surgical wards | 22,708 | ≥50 years, hospitalized >1 day | Not admitted to and discharged from same unit | 67 (11) | 48 |

(Continues)

**TABLE 1** (Continued)

| Author, year | Country | Study design | Type of hospital/ setting | Patient population | Total number of patients in study | Study inclusion criteria (including age and patient consent) | Study exclusion criteria | Age (mean (SD))/ median (IQR)) | Sex F (%) |
|---|---|---|---|---|---|---|---|---|---|
| Lee et al., 2019 | US | Retrospective cohort | Academic medical centre | All inpatients | 12,082 episodes (9017 unique patients) | ≥65 years, admitted via ED, at least one DOSS score, no missing data | None specified | 76 | 50 |
| Lee et al., 2022 | US | Retrospective cohort | Academic medical centre | Patients admitted from ED | 7927 | ≥65 years, at least one DOSS score | None specified | 76 | 52 |
| Marquetand et al., 2021 | Switzerland | Pragmatic prospective cohort, secondary analysis | Large health care system | Acute medical and acute surgical admissions | 3076 | ≥80 years | LOS <24 h, missing data, unknown division, admission or discharge destination | 86 | 51 |
| Matharu et al., 2022 | UK | Prospective cohort | Acute hospitals, multi-site | Patients with an acute hip fracture | 135,685 | ≥60 years | Missing data for ≥1 covariate | 83 (9) | 71 |
| Peralta-Cuervo et al., 2021 | Colombia | Retrospective observational | Acute hospital | Acute medical admissions | 1599 | ≥75 years or ≥60 years with risk factor | No CAM score | 86 (9) | 57 |
| Reynish et al., 2017 | UK | Prospective cohort | Acute hospital | Acute medical admissions[a] | 10,014 | ≥65 years of age | None specified | 79 | 57 |
| Rohatgi et al., 2019 | US | Retrospective cohort study | Acute hospital | Discharged from acute medical and surgical units | 105,455 encounters | ≥18 years, admitted to any medical-surgical unit ≥24 h | >1 day in ICU, in short-stay areas, or in inpatient psychiatry unit | 60 (46–71) | 52 |
| Schubert et al., 2018 | Switzerland | Prospective cohort | Acute university hospital | Patients discharged from acute medical and surgical units | 29,278 | ≥18 years, discharged Jan –Dec 2014 | LOS <1 day | 55 (19) | 52 |
| Spiller et al., 2022 | Switzerland | Prospective cohort | Acute hospital | All hospitalized patients | 29,967 | ≥65 years or < 65 years with "delirious symptoms" or cognitive impairment | Missing screening data | 71 (12) | 41 |
| Tyas et al., 2021 | UK | Retrospective analysis of routine data | Acute hospitals, multi-site | Patients with acute hip fracture | 21,274 (2020) 22,098 (2019) | ≥60 years | None specified | NA | NA |

TABLE 1 (Continued)

| Author, year | Country | Study design | Type of hospital/ setting | Patient population | Total number of patients in study | Study inclusion criteria (including age and patient consent) | Study exclusion criteria | Age (mean (SD))/ median (IQR)) | Sex F (%) |
|---|---|---|---|---|---|---|---|---|---|
| Wong et al., 2018 | US | Retrospective cohort | Academic medical centre | Acute medical admissions | 18,223 | ≥18 years discharged Jan 1 2016–Nov 30 2017, completed assessment | Admitted with delirium, altered mental status, or requiring ICU admission | 35.9% >64 years | 52 |
| Zipser et al., 2022 | Switzerland | Pragmatic prospective cohort | Large health care system | Acute medical and surgical wards | 17,158 (surgical) 8214 (medical) | ≥18 years old, stay at least 24 h | Missing data, admission/ discharge not obtained | 56 (37–70) | 52 |

Abbreviations: CAM, Confusion Assessment Method, DOSS, Delirium Observation Screening Scale, ED, Emergency Department, ICU, Intensive Care Unit, LOS, length of stay; EHR, electronic health record.
[a]Assessments performed by specialist trained nurses.

were observational cohort studies. Four studies self-described as retrospective analyses of prospectively collected routine EHR data. Five were multi-center. All were conducted in acutely hospitalized medical and/or surgical patients, with some in specific sub-populations, such as patients admitted to a single geriatrics unit (two studies) or hospitalized with an acute hip fracture (three studies).

## Assessment tools used

Studies reported the use of six different validated assessment tools (Table 2 and Figure 2): the Confusion Assessment Method (CAM; six studies), brief-CAM (bCAM; one study), the 4 'A's Test (4AT; six studies), the Delirium Observation Screening Scale only (DOSS; four studies), DOSS or Intensive Care Delirium Screening Checklist (ICDSC; one study), the Nursing Delirium Screening Scale (NuDESC; two studies) and, in two studies, the DOSS combined with the ICDSC or electronic Patient Assessment-Acute Care (ePA-AC), a DSM-based nursing construct including functional domains shown to be associated with delirium but not validated as a detection tool.[27] Studies using tool combinations did not always report the number of observations using each tool.

## Timepoint(s) of Tool Use in Patient Journey

Delirium assessments were done at different timepoints during the patient journey. These can be classified into four categories (Table 2 and Supplementary Figure 1): admission assessment only (five studies[21,22,28–30]); on admission and as inpatient, daily or more frequently (two studies[17,31]); inpatient, daily or more frequently (11 studies[12,18–20,32–38]); inpatient post-operatively (three studies[39–41]). In one study, a single assessment was performed on ED attendance but only 31.5% of the study population were hospitalized.[42]

## Tool completion rate

Six studies reported data only on patients with a completed delirium assessment without providing a denominator (three CAM studies,[18,28,29] one DOSS,[36] and both using the NuDESC[19,33]) (Table 2). In the 16 studies reporting on the whole population eligible for assessment, the tool completion rate ranged from 19% (all inpatients ≥18 years, bCAM, although the center reported early data from a staged implementation process led by nursing champions and the completion rate was 63%

**TABLE 2** Tool name, tool characteristics, tool completion rate (proportion of patients with a completed tool assessment), delirium positive score rate (proportion of assessed patients with a positive score or positively scored assessments) and proportion of patients *unable to assess* for all included studies.

| Author, year | Tool name | Timepoint(s) of assessment(s) in patient journey | Frequency of assessment(s) | Time from admission to first assessment | Tool completion rate: No. of patients or assessments performed/no. reported as eligible for assessment (%) | Positive score rate: No. of patients or assessments with a positive score/ no. of patients or assessments (%) | Number (proportion) of patients classed as *unable to assess* (%) |
|---|---|---|---|---|---|---|---|
| Alhaidari et al., 2021 | 4AT | Admission only | One assessment only | Not specified | 6492/7799 (83) | 1154/7799 (15) | 875/7799 (11%)* *"*meaningful reason for omission*" |
| Anand et al., 2022 | 4AT | Admission only | One assessment only | Not specified | 52,965/ 82,770 (64) | 10,685/52965 (20) | 0 |
| Avelino-Silva et al., 2017 | CAM | Admission and Inpatient | On admission and Daily | <24 h of admission | 1409/1522 (93) | 379/1409 (27) (admission) 278/1409 (20) (inpatient) | Not specified |
| Corradi et al., 2016 | CAM | Inpatient | 3×/day | Not specified | Only patients assessed using tool included | 6926/88206 (8) | 15,094/88206 encounters (17) |
| Di Bari et al., 2022 | 4AT | Admission only | One assessment only | Not specified | 3188/3358 (95) | 422/3188 (13) | Not specified |
| Dulin et al., 2022 | bCAM | Admission and Inpatient | On admission and every nursing shift | Not specified | 17,769/93,388 (19)* **early in staged implementation, 63% by end of process. Range 12% (labor & delivery)-98% (medical wards)* | 2593/17,769 (15) | Not specified |
| Fuchs et al., 2020 | DOSS; ICDSC | Inpatient | 3×/day | Not specified | 10,261/10733 (96) | 3285/ 10,261 (32)* **may include patients admitted to ICU assessed using ICDSC* | 0 |
| Garcez et al., 2020 | CAM | Admission only | One assessment only | Not specified | Only patients assessed using tool included | 445/1554 (29) | 0 |
| Han et al., 2021 | 4AT | Inpatient (post-op) | One assessment only | Post-op (<24 h) | 1082/1082 (100) | 222/1082 (21) | 0 |
| LaHue et al., 2021 | NuDESC | Inpatient | 2×/day | Not specified | Not specified | 13.0% during first epoch post intervention; 12.0%, 11.7%, and 13.0% in subsequent epochs | Not specified |
| Lee et al., 2019 | DOSS | Inpatient | 2×/day | <24 h of admission | 14,527/23821 (61) | 2582/12082 (21) | Not specified |
| Lee et al., 2022 | DOSS | Inpatient | 2×/day | <24 h of admission | 7927/12082 (66) | 2008/7927 (25) | 0 |

**TABLE 2** (Continued)

| Author, year | Tool name | Timepoint(s) of assessment(s) in patient journey | Frequency of assessment(s) | Time from admission to first assessment | Tool completion rate: No. of patients or assessments performed/no. reported as eligible for assessment (%) | Positive score rate: No. of patients or assessments with a positive score/ no. of patients or assessments (%) | Number (proportion) of patients classed as *unable to assess* (%) |
|---|---|---|---|---|---|---|---|
| Marquetand et al., 2021 | DOSS; ePA-AC; ICDSC | Inpatient | 3×/day | Not specified | Only patients assessed using tool included | 1285/3076 (42)* **may include patients admitted to ICU assessed using ICDSC* | Not specified |
| Matharu et al., 2022 | 4AT | Inpatient (post-op) | One assessment only | Post-op (<1 week) | 114,886/124960 (92) | 32,244/114,886 (28) | Not specified |
| Peralta-Cuervo et al., 2021 | CAM | Admission only | One assessment only | Not specified | Only patients assessed using tool included | 816/ 1599 (51) | Not specified |
| Reynish et al., 2017 | CAM | Admission only | One assessment only | Not specified | 10,014/ 12,673 (79) | 765/ 10,014 (8) | 0 |
| Rohatgi et al., 2019 | CAM | Inpatient | Every nursing shift | Not specified | 99 | 5333/105455 (2) | Not specified |
| Schubert et al., 2018 | DOSS | Inpatient | Every 24 h | Not specified | 10,906/ 29,278 (37)* **29278 "eligible patients ≥ 18 years" included in the study but assessment only recommended for patients at "high risk" (≥65 years or < 65 years and delirious symptoms)* | 3069/ 10,906 (28) | Not specified |
| Spiller et al., 2022 | DOSS | Inpatient | 3x/day | Not specified | 29,967/ 48,840 (61) | 21,612/345,662 (6) of observations | Not specified |
| Tyas et al., 2021 | 4AT | Inpatient (post-op) | One assessment only | Post op (<72 h) | 18,511/21274 (87)<br>20,601/22098 (93) | Not specified<br>Not specified | Not specified |
| Wong et al., 2018 | NuDESC | Inpatient | 2×/day | Not specified | Only patients assessed using tool included | 878/18223 (5) of observations | Not specified |
| Zipser et al., 2022 | DOSS; ePA-AC; ICDSC | Inpatient | 3×/day | Not specified | 98 | Surgical: 2093/17,158 (12)<br>Medical: 2237/8214 (27) | Not specified |

Abbreviations: 4AT, the 4 A's Test; (b)CAM, (brief) Confusion Assessment Method; DOSS, Delirium Observation Screening Scale; ePA-AC, Electronic Patient Assessment-Acute Care; ICDSC, Intensive Care Delirium Screening Checklist; NuDESC, Nursing Delirium Screening scale.

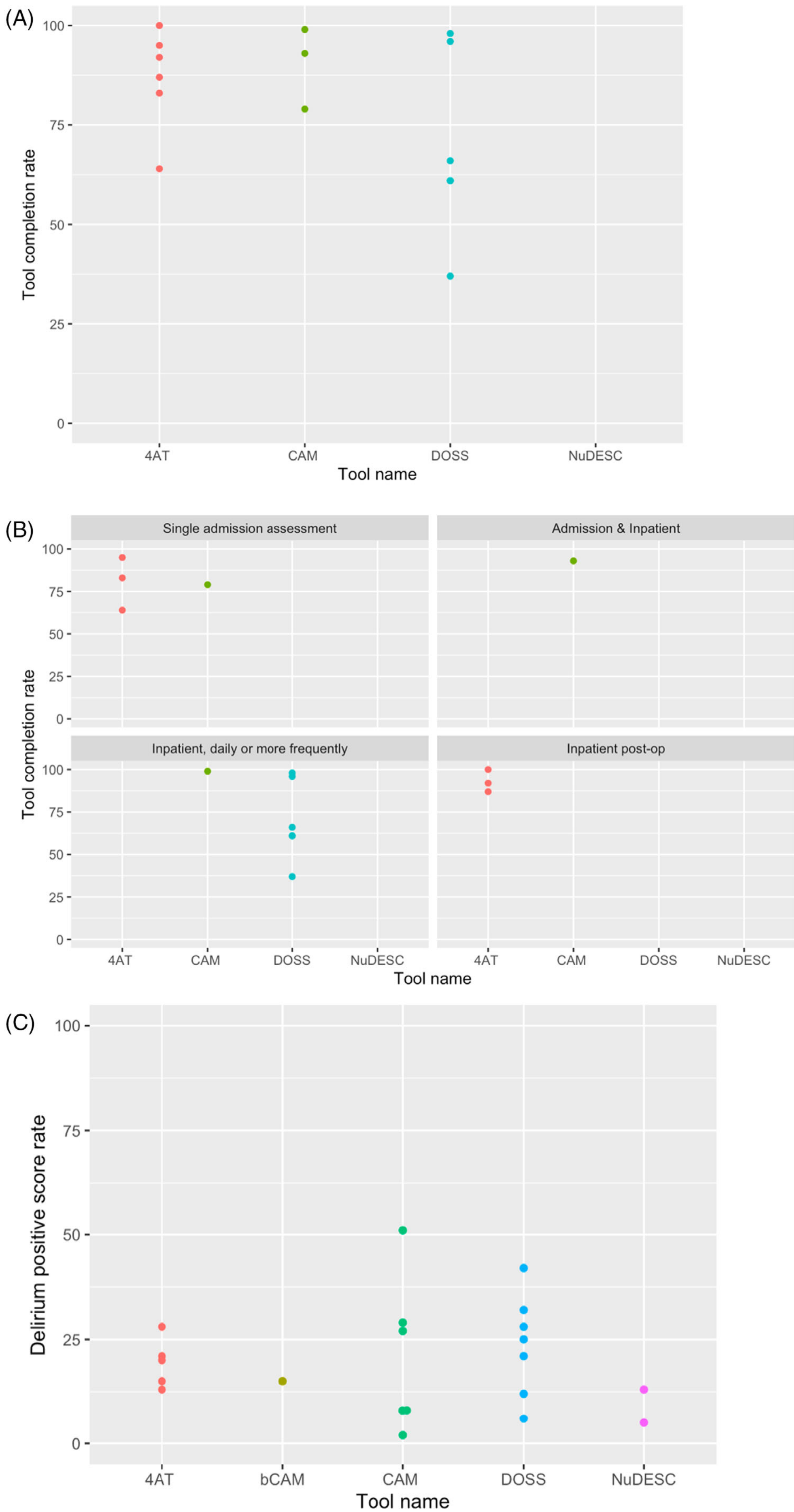


**FIGURE 2** Tool completion rates and delirium detection rates in included studies. (A) Tool completion rate by tool name. (B) Tool completion rate by tool name and timepoint(s) of assessment. (C) Delirium positive score rate (proportion of patients or assessments with a positive score) by tool name. *4AT, the 4 'A's Test; (b) CAM, (brief) Confusion Assessment Method; DOSS, Delirium Observation Screening Scale; NuDESC, Nursing Delirium Screening Scale. NB studies using DOSS in addition to the ICDSC and/or ePA-AC reported as DOSS studies.*[32,35] *Only one study reported a completion rate using the bCAM and this increased from 19% to 63% during a phased implementation process; this data is not presented in* (A) or (B).[18] *Not all included studies reported a tool completion and/or delirium detection rate and hence do not appear in both figures.*

by the end of implementation) to 100% (single orthopedic center, 4AT post-operatively) (Table 2 and Figure 2).[17,39] Two of six studies reporting use of the DOSS had completion rates above 75%[32,38] (overall range 37%[37]–98%,[38] the latter when used in combination with the ICDSC and ePA-AC); all three using the CAM which included all patients[12,30,31] (range 79%[30]–99%[12]), and five of six using the 4AT[21,39–42] (range 64%[22]–100%[39]). A 37% completion rate was calculated in one study using the DOSS, noting the study included all "eligible patients ≥18 years" but assessment was only recommended for patients deemed at high risk of delirium (≥65 years, or <65 years with delirious symptoms) as per hospital protocol.[37]

Most studies did not report the number of patients without an assessment and only three provided information on the characteristics of patients with no assessment. Corradi et al. reported 15,094 of 88,206 patient encounters (17%) as *unable to assess* using the CAM; these patients had higher mortality rates than patients with a positive delirium score.[18] Alhaidari et al. examined reasons for non-completion of the 4AT in 875 patients (11.2% of the study population), classifying these as follows: reduced patient alertness, communication barriers, pre-existing cognitive disorders, unstructured assessment completed, and prioritization of patient comfort and well-being.[21] Anand et al. reported a similar mortality rate for 31% of patients without a 4AT completed as for patients with a 4AT score of 1–3 (positive for cognitive impairment but not delirium).[22]

## Delirium positive score rates

Rates varied by tool and timepoint of assessment.

The positive score rates on a single admission assessment were reported in studies using the CAM and the 4AT (Table 2). CAM-positive rates on admission ranged from 8% in unselected older emergency admissions (specialist nurse assessment)[30] to 51% in geriatrics ward (geriatrician assessment) admissions.[29] 4AT positive rates on admission were 15% (doctor assessment)[21] and 20% (assessment by mixed disciplines).[22] Di Bari et al. reported data on a single 4AT assessment completed by mixed staff disciplines in older Emergency Department (ED) patients who were discharged (68.5%) or admitted, reporting that overall 13% of patients had a positive score.[42]

Two studies reported on admission assessment combined with inpatient repeated assessment. One study of the CAM in older geriatrics ward admissions (geriatrician assessment) found that 27% of patients had a positive score on admission and 20% on repeated inpatient assessment.[31] Another study using the bCAM in adult emergency admissions performed on admission and twice per day (nurse assessment) reported that 15% of assessments were positive.[17]

Eleven studies reported on repeated inpatient assessments with intended tool completion between 1 and 3 times per day (all nurse assessments). There were two studies using the CAM, both in unselected patients aged ≥18 years; overall the positive score rates were 8%[18] and 2%[12] (below expected rates). Three studies using the DOSS in older inpatients reported positive score rates of 21%,[34] 25%,[35] and 6%,[20] with the latter below the expected rate. A study using the DOSS in patients aged ≥18 found a positive score rate of 28%.[37] Two studies used the DOSS combined with the ICDSC and the ePA-AC, and reported an overall positive score rate of 42% in patients aged ≥80 years,[36] and 12% in surgical patients and 27% in medical patients aged ≥18 years.[38] A study using either the DOSS (non-ICU) or ICDSC (ICU) in patients aged ≥65 reported an overall positive score rate of 32%.[32] Two studies used the NuDESC. One reported a range of positive score rates in inpatients aged ≥50 across implementation phases, ranging from 5% to 13% post-implementation,[33] another in inpatients aged ≥50 years reported a positive score rate of 5%[19]; the rates of 5% are below expected rates.

Two studies reported postoperative positive score rates using the 4AT in older patients (assessment by mixed professionals) of 21%[39] and 28%.[40]

## Risk of bias

All but two studies had moderate–high risk of bias, mainly due to missing data and selection of the reported result (Supplementary Table 5). Very few studies reported on patients without a completed delirium assessment, making it difficult to determine how bias may affect results. However, given that patients without assessment may be *unable to assess*[18] or have reduced alertness, communication barriers or pre-existing cognitive disorders,[21] the observed biases may lead to delirium under-detection.

## Gray literature

A gray literature search identified 4 eligible reports (Table 3). National hip fracture audit reports from England, Wales, and Northern Ireland (National Hip Fracture Database; NHFD), Scotland, and the Republic of Ireland reported using the 4AT,[43–45] whilst a national report from Australia and New Zealand reported using the 4AT and CAM.[46] In the NHFD, over 80% of patients

TABLE 3 Published audits and reports of delirium assessment tool use at scale.

| Author, year | Country | Report type | Patient population | Total no. of patients | Tool name | Timepoint of assessment(s) in patient journey | Frequency of assessment | Tool completion rate: No. of patients or assessments performed/no. reported as eligible for assessment (%) | Positive score rate: No. of patients or assessments with a positive score/ patients or assessments (%) |
|---|---|---|---|---|---|---|---|---|---|
| Royal College of Physicians, 2018 | England, Wales & Northern Ireland | National audit (2017 data) | Admissions with acute hip fracture | 65,958 | 4AT | Inpatient post-operative | One assessment only | >80% | 25% |
| Public Health Scotland, 2022 | Scotland | National audit | Admissions with acute hip fracture | 7797 | 4AT | Admission only | Once on admission to ED, once on admission to ward within 24 hours | 72.2% | Not specified |
| National Report, 2021 | Republic of Ireland | National audit | Admissions with acute hip fracture | 3806 | 4AT | Inpatient post-operative | Day 1, day 3, on discharge | 41% assessed on day 1, 24% day 3, 35% on day of discharge | Not specified |
| ANZHFR Annual Report, 2022 | Australia & NZ | National audit | Admissions with acute hip fracture | 15,331 | Various including CAM, 4AT | Not specified | Not specified | Australia: 75% NZ: 65% | Australia: 39% NZ: 46% |

Abbreviations: 4AT, the 4 A's Test; CAM, Confusion Assessment Method; ED, Emergency Department; NZ, New Zealand.

were assessed post-operatively using the 4AT; of these, 25% had a positive score.[43]

# DISCUSSION

## Summary of main findings

This systematic review demonstrates that studies of delirium detection tools implemented in routine care show clinically important variability in the two key metrics of completion rates and delirium positive score rates. Tools were used at different stages of the patient journey, ranging from single ED or post-operative assessments to 1–3 times per day in inpatients. The CAM was successfully implemented in some contexts (e.g., as a single admission assessment in a specialist ward),[28,31] but in some studies demonstrated lower than expected positive score rates, for example when used multiple times per day on every nursing shift throughout the inpatient stay.[12,18] The 4AT was implemented as a single assessment in the ED or admission unit,[21,22,42] or postoperatively[39–41]; completion rates were high in all studies, except one substudy,[22] and positive score rates were broadly aligned with expected delirium prevalence. The DOSS demonstrated low completion rates in studies when used alone,[20,34,37] but higher overall completion rates when reported in combination with another tool (the ePA-AC or ICDSC)[32,38]; positive score rates were broadly aligned with expected delirium prevalence, except in one model development study.[20] Only two studies used the NuDESC.[19,33] Neither reported a completion rate; the positive score rate was lower than expected in both studies. Overall, these findings demonstrate that some tools perform poorly in certain routine care contexts, with low completion and/or much lower than expected positive score rates in the target clinical population.

The included studies provided limited information regarding factors that might affect the completion rate. One study reported an overall completion of 19% using the bCAM.[17] This likely reflects the fact that implementation was staged across nursing units, but completion was measured across the whole medical center early in the implementation process; notably, completion was 63% by the end of the four-year process. Furthermore, bCAM assessment was performed in all hospitalized adults, with variation in completion rates from 12% in labor and delivery wards to 98% in medical wards by the end of the staged implementation. It may be that assessment only in higher risk groups, for example, older emergency admissions, may improve completion rates. Three studies reported high 4AT completion rates (92%, 93%, and 100%) in hip fracture patients postoperatively[39–41]; all were in the UK, where delirium assessment is recommended in national guidelines and data collected as part of national audits of care (although this information was not reported in the studies).[43] Rohatgi et al., reported a CAM completion rate of 99% in the context of a dedicated, highly-resourced delirium prevention initiative.[12] Qualitative studies have suggested that multiple factors influence completion rates, including time constraints, carer availability to provide informant history, tool simplicity, user-friendliness, and staff knowledge of delirium and of the specific tool.[47–49] These factors could be targeted to improve completion rates.

Some of the variations in completion rates may be a function of the tool, as some allow patients to be classed "*unable to assess*" (e.g., the CAM and bCAM), whilst others do not (e.g., the 4AT). Inappropriate use of "*unable to assess*" may mask delirium and these patients often have adverse outcomes.[18] Unfortunately, most studies did not report the proportion of patients recorded as "*unable to assess*" or did not provide information to determine how these patients were classified in the tool completion rate. Only three studies reported specifically on patients without a completed assessment.[18,21,22] In a study using the 4AT, patients without an assessment had similar outcomes to patients with probable cognitive impairment,[22] whilst in a study using the CAM, patients "*unable to assess*" had worse outcomes than patients with delirium.[18] This concerning observation highlights the importance of elucidating barriers to delirium assessment.

Positive score rates varied widely. This variation likely reflects several factors. One is difference in true delirium occurrence rates in the populations and settings, for example, geriatrics ward versus ED attendances. Some studies reported on all inpatients aged ≥18 years,[12,17–19,37,38] whereas one study only included the oldest old.[36] Some studies reported single assessments (prevalent delirium), and others repeated assessments during the inpatient stay (a combination of prevalent and incident delirium). Estimation of real-world sensitivity of the tools for delirium detection, reflecting concordance between positive score and true delirium occurrence rates, could only be done indirectly for all but one study. This study measured delirium prevalence with assessment by trained psychiatrists in a sample of 278 patients, reporting a rate of 17%, compared to a positive score rate of 2% using the CAM in routine practice.[12] For the other studies, estimates of expected rates in the clinical populations concerned can only be made by reference to relevant epidemiologic studies. A systematic review and meta-analysis of delirium occurrence in acute admissions to secondary care aged ≥18 years demonstrated a pooled delirium admission prevalence of 15% and cumulative incidence of 9%, with an overall occurrence rate of 23% across 33 studies.[1] In the present review, studies reported positive score rates as low as 8% on admission[30] and 2% on repeated inpatient testing[12] using the CAM in acute medical and/or surgical populations. There was also variation in positive score rates between studies reporting the use of the same tool in comparable clinical populations. Evidence from national audit reports suggests that recommending a specified assessment tool may result in higher rates of delirium detection,[46,50] although this requires dedicated resources. Provision of training and/or supervision may improve detection rates[30,31]; for example, specific training in the use of the CAM is recommended for optimum performance.[51]

## Strengths and limitations

Study strengths include systematic searching and analysis, including of gray literature, by two independent reviewers following a prospectively registered protocol. Studies were screened in all languages, with no restrictions on patient demographic variables such as age, ethnicity, and sex, or study location.

Some limitations should be acknowledged. An a priori decision was made to only include studies reporting on ≥1000 patients. This was to increase the likelihood that studies reflected large-scale clinical practice, but it is possible that high-quality smaller studies reporting on tool use in unselected patient populations in routine practice were omitted. We limited the review to published full-text studies and reports only. Although we performed an additional gray literature search to identify audit reports and relevant publications, tools reported in smaller forums may have been missed. For example, nearly one-third of UK units report using the "Single Question in Delirium" (SQiD), but no eligible peer-reviewed publications, abstracts or gray literature reports using the SQiD were identified.[52] Studies reported completion rates at different stages of the patient journey, but it is possible that tool use was not fully reported in the studies. For example, some "admission only" tools may have been repeated later in the hospitalization as part of routine practice. We excluded studies in ICU settings. Delirium in the ICU is usually considered a different entity, with differences in population, tools, and expected delirium rates. This warrants a separate review. Most included studies were considered to have moderate–high RoB, highlighting the need for higher-quality implementation studies designed to assess tool performance.

## Recommendations for future clinical practice

The findings of this systematic review have important implications for clinicians and policymakers seeking to implement effective delirium assessment into practice. Tool completion rates were variable, and the findings relating to positive score rates demonstrate that certain tools substantially under-detect delirium when used routinely in some clinical contexts. When implementing a delirium assessment tool at scale, healthcare systems should plan to monitor tool completion and positive score rates, to iteratively investigate low completion or unexpectedly high or low positive score rates (choosing qualitative or quantitative methods appropriate to the context), and intervene to improve performance with repeated measurement and improvement. Implementing tools within EHRs could support routine use and measurement. Performance metrics should stimulate detailed evaluation of staff attitudes and barriers to delirium assessment through qualitative studies. Data-sharing and collaboration among healthcare providers and researchers could enable the creation of a comprehensive and up-to-date information repository on real-world assessment tool performance. This could be used to identify factors surrounding tool implementation in different contexts, to evaluate the impact of different strategies, and to inform future research and practice.

## Recommendations for future research

Though the studies included in this review provide useful information, we recommend that future studies on delirium tool implementation report a wider set of variables, including a minimum dataset (Supplementary Table 6).

# CONCLUSION

Delirium underdetection in routine care remains a major unmet clinical need. This review demonstrates that there is clinically important variability in tool completion and positive score rates relative to expected delirium rates, as well as in the proportion of patients unable to assess (where reported). Our findings are important for clinicians and policymakers seeking to improve delirium detection in acute hospital settings, and to inform future research. We encourage all healthcare systems to analyze and report their relevant data to share learning and work towards effective delirium detection at scale.

### AUTHOR CONTRIBUTIONS

All authors were involved in conception and design of the study. RSP registered the protocol to PROSPERO. SDS and ZT provided guidance on systematic review methodology. RSP, CS, and AAng performed title/

abstract screening, full-text review, quality assessment, and risk of bias analysis. AMJM was available to resolve conflicts at title/abstract screening and full-text review. RSP and AMJM drafted the initial manuscript. All authors were involved in drafting the final manuscript. All authors had access to the data, provided feedback, and were involved in the finalization of the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## CONFLICT OF INTEREST STATEMENT

AMJM is the main author of the 4AT (www.the4AT.com); the 4AT is free to download and use, and there are no current or future financial interests. The Advanced Care Research Centre is funded by Legal and General PLC as part of their corporate social responsibility (CSR) programme. The funder had no role in preparation of this editorial, and the views expressed are those of the authors.

## FUNDING INFORMATION

RSP is a fellow on the Multimorbidity Doctoral Training Programme for Health Professionals, which is supported by the Wellcome Trust [223499/Z/21/Z].

## SPONSOR'S ROLE

The financial sponsor played no role in the design, execution, analysis, and interpretation of data, or writing of the study.

## ORCID

*Rose S. Penfold* https://orcid.org/0000-0001-7023-7108
*Charlotte Squires* https://orcid.org/0000-0003-0527-4302
*Susan D. Shenkin* https://orcid.org/0000-0001-7375-4776
*Temi Ibitoye* https://orcid.org/0000-0001-8571-988X
*Zoë Tieges* https://orcid.org/0000-0002-3820-3917
*Karin J. Neufeld* https://orcid.org/0000-0002-2395-1883
*Thiago J. Avelino-Silva* https://orcid.org/0000-0001-9347-0519
*Daniel Davis* https://orcid.org/0000-0002-1560-1955
*Atul Anand* https://orcid.org/0000-0002-6428-4554
*Bruce Guthrie* https://orcid.org/0000-0003-4191-4880
*Alasdair M. J. MacLullich* https://orcid.org/0000-0003-3159-9370

## TWITTER

*Rose S. Penfold* *rosespenfold*

## REFERENCES

1. Gibb K, Seeley A, Quinn T, 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;49:352-360.
2. Wilson JE, Mart MF, Cunningham C, et al. Delirium. *Nat Rev Dis Primers*. 2020;6:90.
3. Burton JK, Guthrie B, Hapca SM, Cvoro V, Donnan PT, Reynish EL. Living at home after emergency hospital admission: prospective cohort study in older adults with and without cognitive spectrum disorder. *BMC Med*. 2018;16:231.
4. Davis DH, Muniz-Terrera G, Keage HA, et al. Association of Delirium with Cognitive Decline in late life: a neuropathologic study of 3 population-based cohort studies. *JAMA Psychiatry*. 2017;74:244-251.
5. American Geriatrics Society expert panel on postoperative delirium in older a. American Geriatrics Society abstracted clinical practice guideline for postoperative delirium in older adults. *J Am Geriatr Soc*. 2015;63(1):142-150.
6. National Institute for Health and Clinical Excellence. *Delirium: Prevention, Diagnosis and Management*. **Volume 2017**. NICE; 2010.
7. Australian Commission on Safety and Quality in Health Care. *Delirium Clinical Care Standard. **Volume** 2021*. Australian Commission on Safety and Quality in Health Care; 2021. https://www.safetyandquality.gov.au/sites/default/files/2021-11/delirium_clinical_care_standard_2021.pdf
8. Khachaturian AS, Hayden KM, Devlin JW, et al. International drive to illuminate delirium: a developing public health blueprint for action. *Alzheimers Dement*. 2020;16:711-725.
9. Han JH, Schnelle JF, Ely EW. The relationship between a chief complaint of "altered mental status" and delirium in older emergency department patients. *Acad Emerg Med*. 2014;21:937-940.
10. Peden CJ, Miller TR, Deiner SG, Eckenhoff RG, Fleisher LA. Improving perioperative brain health: an expert consensus review of key actions for the perioperative care team. *Br J Anaesth*. 2021;126:423-432.
11. Bellelli G, Nobili A, Annoni G, et al. Under-detection of delirium and impact of neurocognitive deficits on in-hospital mortality among acute geriatric and medical wards. *Eur J Intern Med*. 2015;26:696-704.
12. Rohatgi N, Weng Y, Bentley J, 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, 1421-1430.e8.
13. Lange PW, Lamanna M, Watson R, Maier AB. Undiagnosed delirium is frequent and difficult to predict: results from a prevalence survey of a tertiary hospital. *J Clin Nurs*. 2019;28:2537-2542.
14. Ibitoye T, So S, Shenkin SD, et al. Delirium is under-reported in discharge summaries and in hospital administrative systems: a systematic review. *Delirium*. 2023. doi:10.56392/001c.74541
15. Perez-Ros P, Martinez-Arnau FM. Delirium assessment in older people in emergency departments. A literature review. *Diseases*. 2019;7:7.
16. Keenan CR, Jain S. Delirium. *Med Clin North Am*. 2022;106:459-469.

17. Dulin JD, Zhang J, Marsden J, Mauldin PD, Moran WP, Kalivas BC. Association of delirium screening on hospitalized adults and postacute care utilization: a retrospective cohort study. *Am J Med Sci*. 2022;364:554-564.
18. Corradi JP, Chhabra J, Mather JF, Waszynski CM, Dicks RS. Analysis of multi-dimensional contemporaneous EHR data to refine delirium assessments. *Comput Biol Med*. 2016;75:267-274.
19. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *JAMA Netw Open*. 2018;1:e181018.
20. Spiller TR, Tufan E, Petry H, et al. Delirium screening in an acute care setting with a machine learning classifier based on routinely collected nursing data: a model development study. *J Psychiatr Res*. 2022;156:194-199.
21. Alhaidari AAO, Matsis KP. Barriers to completing the 4AT for delirium and its clinical implementation in two hospitals: a mixed-methods study. *Eur Geriatr Med*. 2022;13:163-172.
22. Anand A, Cheng M, Ibitoye T, Maclullich AMJ, Vardy E. Positive scores on the 4AT delirium assessment tool at hospital admission are linked to mortality, length of stay and home time: two-Centre study of 82,770 emergency admissions. *Age Ageing*. 2022;51(3). doi:10.1094/ageing/afac051
23. Friedman JI, Li L, Kirpalani S, et al. A multi-phase quality improvement initiative for the treatment of active delirium in older persons. *J Am Geriatr Soc*. 2021;69(1):216-224.
24. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (3rd ed.)*. APA Press; 1980.
26. ROBINS-E Development Group (Higgins JMR, Rooney A, Taylor K, Thayer K, Silva R, Lemeris C, Akl A, Arroyave W, Bateson T, Berkman N, Demers P, Forastiere F, Glenn B, Hróbjartsson A, Kirrane E, LaKind J, Luben T, Lunn R, McAleenan A, McGuinness L, Meerpohl J, Mehta S, Nachman R, Obbagy J, O'Connor A, Radke E, Savović J, Schubauer-Berigan M, Schwingl P, Schunemann H, Shea B, Steenland K, Stewart T, Straif K, Tilling K, Verbeek V, Vermeulen R, Viswanathan M, Zahm S, Sterne J). Risk of bias in non-randomized studies—of exposure (ROBINS-E). Launch Version, 1 June 2022, 2022.
27. Bode L, Isler F, Fuchs S, et al. The utility of nursing instruments for daily screening for delirium: delirium causes substantial functional impairment. *Palliat Support Care*. 2020;18:293-300.
28. Garcez FB, Jacob-Filho W, Avelino-Silva TJ. Association between level of arousal and 30-day survival in acutely ill older adults. *J Am Med Dir Assoc*. 2020;21:493-499.
29. Peralta-Cuervo AF, Garcia-Cifuentes E, Castellanos-Perilla N, Chavarro-Carvajal DA, Venegas-Sanabria LC, Cano-Gutierrez CA. Delirium prevalence in a Colombian hospital, association with geriatric syndromes and complications during hospitalization. *Rev Esp Geriatr Gerontol*. 2021;56:69-74.
30. Reynish EL, Hapca SM, De Souza N, Cvoro V, Donnan PT, Guthrie B. Epidemiology and outcomes of people with dementia, delirium, and unspecified cognitive impairment in the general hospital: prospective cohort study of 10,014 admissions. *BMC Med*. 2017;15:140.
31. Avelino-Silva TJ, Campora F, Curiati JA, Jacob-Filho W. Association between delirium superimposed on dementia and mortality in hospitalized older adults: a prospective cohort study. *PLoS Med*. 2017;14:e1002264.
32. Fuchs S, Bode L, Ernst J, Marquetand J, von Kanel R, Bottger S. Delirium in elderly patients: prospective prevalence across hospital services. *Gen Hosp Psychiatry*. 2020;67:19-25.
33. LaHue SC, Maselli J, Rogers S, et al. Outcomes following implementation of a hospital-wide, multicomponent delirium care pathway. *J Hosp Med*. 2021;16:397-403.
34. Lee S, Harland K, Mohr NM, et al. Evaluation of emergency department derived delirium prediction models using a hospital-wide cohort. *J Psychosom Res*. 2019;127:109850.
35. Lee S, Okoro UE, Swanson MB, Mohr N, Faine B, Carnahan R. Opioid and benzodiazepine use in the emergency department and the recognition of delirium within the first 24 hours of hospitalization. *J Psychosom Res*. 2022;153:110704.
36. Marquetand J, Bode L, Fuchs S, et al. Risk factors for delirium are different in the very old: a comparative one-year prospective cohort study of 5,831 patients. *Front Psych*. 2021;12:655087.
37. Schubert M, Schürch R, Boettger S, et al. A hospital-wide evaluation of delirium prevalence and outcomes in acute care patients - a cohort study. *BMC Health Serv Res*. 2018;18:550.
38. Zipser CM, Spiller TR, Hildenbrand FF, et al. Discharge destinations of delirious patients: findings from a prospective cohort study of 27,026 patients from a large health care system. *J Am Med Dir Assoc*. 2022;23:1322-1327.e2.
39. Han TS, Lisk R, Osmani A, et al. Increased association with malnutrition and malnourishment in older adults admitted with hip fractures who have cognitive impairment and delirium, as assessed by 4AT. *Nutr Clin Pract*. 2021;36:1053-1058.
40. Matharu GS, Shah A, Hawley S, et al. The influence of mode of anaesthesia on perioperative outcomes in people with hip fracture: a prospective cohort study from the National hip Fracture Database for England, Wales and Northern Ireland. *BMC Med*. 2022;20:319.
41. Tyas B, Wilkinson M, Singisetti K. Effect of Covid-19 on best practice care of hip fracture patients: an analysis from the National Hip Fracture Database (NHFD). *Surgeon*. 2021;19:e298-e303.
42. Di Bari M, Giordano A, Tonarelli F, et al. Estimating prognosis and frailty in persons aged >/=75 years in the emergency department: further validation of dynamic silver code. *J Am Med Dir Assoc*. 2022;23:87-91.
43. Royal College of Physicians. *Improving Understanding: the National hip Fracture Database Report on 2021*. RCP; 2022.
44. Public Health Scotland Scottish Hip Fracture Audit. 2022.
45. National Office of Clinical Audit. *Irish Hip Fracture Database*. 2020. National Office of Clinical Audit; 2020. https://www.noca.ie/audits/irish-hip-fracture-database.
46. Australian and New Zealand Hip Fracture Registry. Annual report of hip fracture care 2022. 2022 http://anzhfr.org/registry-reports/
47. Chary AN, Castilla-Ojo N, Joshi C, et al. Evaluating older adults with cognitive dysfunction: a qualitative study with emergency clinicians. *J Am Geriatr Soc*. 2022;70(2):341-351.

48. Johansson YA, Tsevis T, Nasic S, et al. Diagnostic accuracy and clinical applicability of the Swedish version of the 4AT assessment test for delirium detection, in a mixed patient population and setting. *BMC Geriatr*. 2021;21:568.
49. Dos Santos FCM, Rêgo AS, Montenegro WS, et al. Delirium in the intensive care unit: identifying difficulties in applying the confusion assessment method for the intensive care unit (CAM-ICU). *BMC Nurs*. 2022;21:323.
50. Farrow L, Hall A, Wood AD, et al. Quality of Care in hip Fracture Patients: the relationship between adherence to National Standards and improved outcomes. *J Bone Joint Surg Am*. 2018;100:751-757.
51. Inouye SK. *The Confusion Assessment Method (CAM): Training Manual and Coding Guide*. Hospital Elder Life Program; 2003.
52. Tieges Z, Lowrey J, MacLullich AMJ. What delirium detection tools are used in routine clinical practice in the United Kingdom? Survey results from 91% of acute healthcare organisations. *Eur Geriatr Med*. 2021;12:1293-1298.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Supplementary Table 1.** PRISMA checklist.
**Supplementary Table 2.** Full search strategies for each database.
**Supplementary Table 3**. Gray literature: abstract-only and conference presentations.
**Supplementary Table 4.** List of excluded studies after full-text review.
**Supplementary Table 5.** Risk of bias in included studies.
**Supplementary Table 6.** Reporting standards for delirium assessment tool implementation studies.
**Supplementary Figure 1**. Timepoint(s) of delirium detection tool used in the patient journey.

**How to cite this article:** Penfold RS, Squires C, Angus A, et al. Delirium detection tools show varying completion rates and positive score rates when used at scale in routine practice in general hospital settings: A systematic review. *J Am Geriatr Soc*. 2024;1-17. doi:10.1111/jgs.18751.