

# Delirium Screening in the Emergency Department: A Systematic Review and Meta-analysis

**Qian Zhang**

Lanzhou University

**Meixi Chen**

Lanzhou University

**Liangying Hou**

Lanzhou University

**Ziqi Guo**

Lanzhou University

**Qing Zhang**

Lanzhou University

**Baoshan Di**

Gansu Provincial Hospital

**Long Ge** (✉ [gelong2009@163.com](mailto:gelong2009@163.com))

Lanzhou University

---

## Research article

**Keywords:** Delirium, Diagnosis, Emergency department, Systematic review

**Posted Date:** October 30th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-97690/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Delirium is a complex syndrome characterized by a disturbance in attention and awareness, with a prevalence of 10-20% in patients admitted to the Emergency Department (ED). Screening tools have been developed to identify delirium in the ED, but their accuracy of screening remains unclear. To address this challenge, we conducted a comprehensive meta-analysis to systematically review the accuracy of delirium screening tools currently being used to assess ED patients.

**Methods:** PubMed, PsycINFO, EMBASE, and the Cochrane Library were searched. Studies involving ED inpatients which compared diagnostic tools with the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria as a reference standard were included. Two reviewers independently screened the studies, extracted data, and assessed the quality of studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 scale. We conducted a conventional meta-analysis for each screening tool. Then we used network meta-analysis method to calculate the relative sensitivity and specificity among the diagnostic tests. The diagnostic accuracies were then ranked through the superiority index.

**Results:** Thirteen studies included six screening tools. The pooled sensitivity and specificity for the Confusion Assessment Method (CAM) were 0.71 and 0.98, and for 4AT (Arousal, Attention, Abbreviated Mental Test 4, Acute change) were 0.83 and 0.93, respectively. The other four tools used were only reported in one or two studies. Their sensitivity ranged from 0.70 to 1.00, and their specificity ranged from 0.64 to 0.99. Moreover, network meta-analysis indicated that the CAM and 4AT had a greater superiority index and a higher diagnostic accuracy.

**Conclusions:** The available data suggested that both the CAM and 4AT can be used as efficient screening tools for the ED patients.

## 1. Introduction

Delirium is a neurology conative disorder characterized by acute onset, disturbed consciousness and fluctuated course. In the emergency department (ED)<sup>[1]</sup> the prevalence reported as high as 10–20%, and 8–25% of old patients in ED present with delirium<sup>[2]</sup>. Studies have confirmed that patients with delirium tend to have poor outcomes including increased length of hospital stay, medical complications, increased risk of falls and higher mortality<sup>[3][4]</sup>.

Although delirium is prevalent and associated with adverse outcomes, there is still three out of four patients missed delirium detection by bedside nurses and medical staff<sup>[5][6]</sup>. Especially in the ED, the assessment of delirium is rarely done due to the high volume of patients and tense time demands on providers<sup>[7]</sup>. In the United States, emergency physicians miss about 75% cases of delirium each year<sup>[8]</sup>. Delirium screening is still a challenge for the ED staff. As the center of modern healthcare, ED should provide appropriate and rapid treatment in the first time<sup>[9]</sup>. Therefore, it suggests a need for screening tools. Clinical practice guidelines recommend that a valid tool for delirium assessment is a crucial component in the detection of delirium<sup>[10]</sup>. An accurate screening tool could identify high-risk patients to reduce or prevent delirium occurrence and reduce the burden of delirium<sup>[11]</sup>.

Currently, several screening tools have been designed to support the assessment of delirium, but the screening in the ED has not been uniformly recognized<sup>[12]</sup>. Different screening tools have a variety of sensitivities and specificities<sup>[13]</sup>. The time needed to complete the assessments also adds to the complexity of delirium detection<sup>[14]</sup>. Different guidelines provide different recommendations. The Scottish Intercollegiate Guidelines Network (SIGN)<sup>[15]</sup> recommends that in the ED, the 4AT (Arousal, Attention, Abbreviated Mental Test 4, Acute change) tool should be used for identifying delirium. The National Institute for Health and Care Excellence (NICE)<sup>[16]</sup> suggests that short Confusion Assessment Method (short CAM) should be routinely used to diagnosis delirium. Consequently, it is still not certain which screening tools to use in the ED.

Several systematic reviews have been conducted to summarize the finding of delirium screening tools in the ED, but they did not suggest which screening tool is better. Ewan<sup>[17]</sup> et al. summarized the results for delirium assessment and concluded that there is variability in screening methodology, the procedures to obtain consent and the methodological quality. A validated screening method is urgently needed to identify delirium in the early time. Lamantia<sup>[12]</sup> et al. concluded that there is still a lack of validated delirium screening tools in the ED. José<sup>[18]</sup> et al. conducted a systematic search and found that the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) is the most widely used instrument, but not the most suitable for the ED. Although they all gave a comprehensive description of the screening tools for delirium in the ED, some studies focused on the effectiveness of screening tools rather than accuracy, and they did not conduct a quantitative meta-analysis to compare the different screening tools. With the emergence of new evidence, a systematic evaluation and meta-analysis is necessary. In addition, a pairwise meta-analysis could not provide a whole picture about the screening tools and assess the relevant diagnostic accuracy of different tools. Network meta-analysis has been used widely in interventional studies, that could

compare the relevant effect of three or more interventions at the same time and effectively ranked the interventions to select the optimal treatment plan even in the absence of direct comparisons<sup>[19]</sup>.

This study aimed to evaluate the accuracy of different screening tools for ED patients by using a network meta-analysis method, and to rank different methods of assessment using the superiority index (SI).

## 2. Methods

We conducted a systematic review and network meta-analysis according to PRISMA (preferred reporting items for systematic reviews and meta-analysis) statements for diagnostic test accuracy<sup>[20]</sup>. The study protocol was registered on the PROSPERO registry (CRD42020153618).

### 2.1 Search strategy

PubMed, PsycINFO, EMBASE, and the Cochrane Library were searched from inception to December 2019. The search strategies were developed by QZ and guided by LG, who is an experienced evidence-based medicine researcher. The search terms were “delirium”, “acute confusion”, “diagnosis”, “sensitivity” and “specificity”. Completed details of the search strategies can be found in the Supplement Table 1. The references of relevant systematic reviews and meta-analyses were also searched to identify potential studies.

Table 1  
Study characteristic

Study Year	Country	Sample size	Gender (M/F)	Age	Index test	Reference standard	Sensitivity	Specificity	Prevalence
Verena 2018	Germany	288	129/159	78.00	CAM	DSM-V	0.65	0.94	0.16
Jayita 2016	Australia	257	111/146	86.00	4AT	DSM-V	0.87	0.80	0.62
Florian 2017	Swiss	285	117/168	79.90	RASS	DSM IV	0.71	0.93	0.07
Jin 2015	America	406	204/202	73.00	RASS	DSM IV	0.84	0.88	0.12
Jin 2014	America	406	204/202	73.50	CAM-ICU	DSM IV	0.72	0.99	0.12
Jin 2013	America	406	204/202	73.50	DTS	DSM-IV	0.98	0.55	0.12
Jin* 2013	America	406	204/202	73.50	CAM	DSM-IV	0.84	0.96	0.12
Wolfgang 2017	Switzerland	286	118/169	80.02	CAM	DSM-IV	0.90	0.98	0.07
Sanchai 2016	Thailand	97	49/48	73.60	4AT	DSM-IV	0.83	0.86	0.25
MacLulich 2019	Scotland,EngLand	392	349/436	81.40	4AT	DSM-IV	0.76	0.94	0.13
MacLulich* 2019	Scotland,EngLand	384	349/436	81.40	CAM	DSM-IV	0.4	1.00	0.11
O'Sullivan 2017	Ireland	378	206/213	77.00	6-CIT	DSM-V	0.83	0.87	0.14
O'Sullivan 2017	Ireland	360	206/213	77.00	4AT	DSM-V	0.93	0.91	0.17
Shenkin 2019	UK	421	349/436	81.40	4AT	DSM-IV	0.76	0.94	0.13
Shenkin 2019	UK	420	349/436	81.40	CAM	DSM-IV	0.40	1.00	0.13
Van 2016	Ireland	478	248/230	78.50	CAM-ICU	DSM-IV	1.00	0.98	0.23
Renato 2001	Brasil	100	52/48	73.80	CAM	DSM-IV	0.94	0.96	0.17

## 2.2 Study selection

We included studies that met the following criteria: (1) population limited to ED patients; (2) index tests that included at least one delirium assessment tool for diagnosed patients (e.g., 4AT, CAM), which was compared with the reference standards (Diagnostic and Statistical Manual of Mental Disorders IV or V (DSM-IV or DSM-V). (3) sufficient information to calculate the crucial values to perform diagnostic analysis<sup>[22]</sup> including true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values; and (4) cohort or cross-sectional designs. Only studies published in English were included. We excluded editorials, commentaries, as well as pilot, case report, and duplicated studies.

EndNote X9 was used to manage the initial search records; after removing duplicate records, the remaining records were imported to Rayyan<sup>[23]</sup>, a free mobile app and web for systematic reviews. Two reviewers (ZQG and QZ) independently screened the titles and abstracts of all identified records. We then downloaded the full texts of the potential records to further review them for inclusion. Disagreements were resolved by discussion or through consultation with a third reviewer (LG).

## 2.3 Data extraction

Two reviewers (QZ and MXC) extracted data independently from a pre-designed data extraction form using Microsoft Excel 2019 (Microsoft, Redmond, WA, USA, www. Microsoft. Com). We collected data including their study characteristics (e.g. year of publication, surname of the first author, country where the research was conducted, reference standard, index tests used), patient characteristics (sample size, male/female, mean age, diagnostic method used, duration of the interventions) and outcomes (TP, FP, FN, TN).

If the studies involved more than one assessor (either physicians or nurses) who used the screening tools, we included the one with the highest sensitivity in our meta-analysis. Data extraction was performed independently by two reviewers (QZ and MXC). Conflicts were resolved by consensus or consultation with a third reviewer (LG).

## 2.4 Quality evaluation

Two reviewers (ZQG and QZ) independently assessed the risk of bias for each study as low, moderate, or high using criteria adapted from Quality Assessment of Diagnostic Accuracy Studies 2<sup>[24]</sup> (QUADAS-2). This method comprises four domains: patient selection, index test, reference standard, and flow and timing. The answer risk for bias and applicability was rated as “no”, “yes”, or “unclear”. We conducted the quality evaluation for each test method. For example, one study both assessed the CAM and 4AT on the same patients would have two QUADAS-2 assessments. Conflicts were resolved by discussion. Unified results were solved by consulting a third reviewer (LG).

## 2.5 Statistical analysis

### 2.5.1 Pairwise meta-analysis

All statistical analyses were performed with STATA version 15.1 (Stata Corporation, College Station, TX, USA) with the programs “*midas*”. We calculated the pooled accuracy estimates (sensitivity, specificity, positive and negative likelihood ratios (LRs), diagnostic odds ratio (DOR) across studies with 95% confidence intervals (CIs) using the bivariate mixed-effects regression model. The summary receiver operator characteristic (SROC) curve was plotted and the pooled area under the curve (AUC) value was calculated<sup>[25]</sup>.

Univariable meta-regression and subgroup analysis were planned to further explore the potential sources of heterogeneity<sup>[26]</sup>. A  $P < 0.05$  for sensitivity or specificity was used to determine whether there was a statistically significant difference in sensitivity, specificity, or both among the levels of a particular covariate.

Deeks' funnel plot asymmetry test was used to assess publication bias. An unequal distribution in the visual funnel plot or a  $P$ -value of  $< 0.05$  was considered to indicate statistically significant bias.

### 2.5.2 Geometry of the network

We drew a network plot show the geometry of the evidence. In the network plot, the research and evaluation size are proportional to the number of tested nodes, and the number of direct comparisons between tests is proportional to the thickness of the lines between the nodes. Evaluation of the presence of at least one test and at least one other from the remaining tests was done to assess network connectivity<sup>[27]</sup>.

### 2.5.3 Indirect comparison between competing diagnostic tests

We used the analysis of variance model in the R software V.3.4.1 (R Core Team, Vienna, Austria) to calculate the relative diagnostic outcomes between index tests, such as relative sensitivity, relative specificity, relative diagnostic odds ratio and superiority index.

### 3. Results

#### 3.1 Study selection

The search yielded 4904 records, after removing duplication, of which 4765 were excluded after title and abstract screening. Out of the 139 articles restricted the medical setting to the ED assessed, we excluded 100 records. Finally, 13 studies with 3023 participants were included in the analysis (Fig. 1) and the reasons for exclusion are presented in the Supplement Table 2.

Table 2  
Meta-analysis results of CAM and 4AT

Tests	Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood Ratio (95% CI)	Negative Likelihood Ratio (95% CI)	Diagnostic OR (95% CI)
CAM	0.71 (0.50–0.86)	0.98 (0.95–0.99)	39.70 (16.40–96.10)	0.30 (0.16–0.55)	135.00 (58.00–314.00)
4AT	0.83 (0.77–0.88)	0.93 (0.86–0.94)	9.20 (6.30–13.50)	0.18 (0.13–0.25)	51.00 (34.00–76.00)

#### 3.2 Study characteristics and quality

We identified six different tools to screen for delirium. Figure 2 presents the network plot. Six studies (2650 patients) evaluated the CAM, five studies (2343 patients) evaluated the 4AT, two studies (884 patients) evaluated the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU), two studies (691 patients) evaluated the Richmond Agitation Sedation Scale (RASS), one study (406 patients) evaluated the Delirium Triage Screen (DTS) and one study (419 patients) evaluated the 6-Item Cognitive Impairment Test (6-CIT). The sensitivity ranged from 0.40 to 1.00 and the specificity ranged from 0.55 to 1.00. The prevalence of delirium reported ranged from 7–23%.

The characteristics of the different studies are shown in the Table 1. Thirteen included studies were conducted in 9 different countries, 11 studies (84.6%, n = 2148) were published after 2015. 3 studies (23.0%) were multicenter studies. The sample size varied from 97 to 785 and the mean age of participants ranged from 73.00 to 86.00. The duration of the studies varied from 11 days to 32 months. Four studies stated their sources of funding support. 3 studies (23.0%) used the DAM-V as the reference standard conducted by emergency physician.

#### 3.3 Quality of the studies

Figure 3 summarizes the results of the QUADAS-2 assessment. The Supplement Table 4 displays each quality assessment results of the individual studies. Four studies (30.7%) had a low risk of bias. Four studies (23.5%) stated clear patient selection criteria. Most unclear risk of bias related to not using consecutive or random sampling. Eleven studies (64.7%) clearly defined the threshold for a pre-specified index test and did not know the outcomes of the reference standard. Twelve studies (70.6%) used a standard reference to properly classify the target condition without the knowledge of the index test. Nine studies (52.9%) followed the correct flow and timing.

#### 3.4 Results of meta-analysis

Some assessment tools only include one or two studies, so just studies that evaluated CAM and 4AT were included in the conventional meta-analysis. Table 2 presents the results of the meta-analysis.

##### 3.4.1 CAM

The pooled sensitivity and specificity of the CAM were 0.71 (95% CI: 0.50–0.86) and 0.98 (95% CI: 0.95–0.99), respectively (Supplement Fig. 1). The pooled DOR was 134.53 (95% CI: 57.66–313.90) (Supplement Fig. 2). The SROC plot showed an area under the curve (AUC) of 0.96 (95% CI: 0.94 – 0.98) (Fig. 4). The representation of accuracy estimates from each study in the SROC space was not a typical "shoulder arm" pattern, which indicated that no threshold effects were found. Supplement Fig. 3 shows the pre-test and post-test probability. The pooled positive likelihood ratio (+ LR) was 39.71 (95% CI: 16.41– 96.12) and the negative likelihood ratio (-LR) was 0.30

(95% CI: 0.16–0.55) (Supplement Fig. 4). Supplement Fig. 5 shows the Deeks' funnel plot, indicating that no significant publication bias was found among six studies ( $t = 1.27$ ,  $P = 0.27$ ).

### 3.4.2 4AT

The pooled sensitivity and specificity of the 4AT were 0.83 (95% CI: 0.77–0.88) and 0.93 (95% CI: 0.86–0.94), respectively (Supplement Fig. 6). The pooled DOR was 50.64 (95% CI: 38.33–75.70) (Supplement Fig. 7). The SROC was 0.93 (95% CI: 0.90–0.95). (Fig. 5). The representation of accuracy estimates from each study in a SROC space was not a typical "shoulder arm" pattern, which indicated that no threshold effects were found. Supplement Fig. 8 shows the pre-test and post-test probability. The pooled + LR was 9.22 (95% CI: 6.28–13.53) and the -LR was 0.18 (95% CI: 0.13– 0.25) (supplement Fig. 9).

### 3.5 Results of the network meta-analysis

The results of the network meta-analysis are presented in Table 3 and supplement Fig. 10. Ranking probability showed that the CAM and 4AT had the same highest sensitivity followed by RASS and the CAM had the highest specificity followed by 4AT. Also, the CAM had the highest superiority index, suggesting higher diagnostic accuracy, followed by the 4 AT and RASS. The order of superiority index is the same as the first three assessment tools of sensitivity and specificity.

Table 3  
Results of network meta-analysis

Tests	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)	Superiority index (95% CI)	Rank
CAM	1.00 (1.00–1.00)	1.00 (1.00–1.00)	4.98 (0.20–11.00)	1
4AT	1.00 (0.91–1.10)	1.00 (1.00–1.00)	4.91 (0.14-11.00)	2
CAM-ICU	0.96 (0.73-1.00)	0.96 (0.69-1.00)	1.84 (0.11-9.00)	4
RASS	0.97 (0.73-1.00)	0.98 (0.82-1.00)	2.18 (0.11-11)	3
DTS	0.89 (0.48-1.00)	0.91 (0.48-1.00)	0.89 (0.09-7.00)	6
6-CIT	0.91 (0.50-1.00)	0.91 (0.46-1.00)	0.98 (0.09-7.00)	5

## 4. Discussion

This systematic review and network meta-analysis on the diagnostic accuracy of delirium screening tools is the first attempt to summarize the evidence of recent studies. Our main finding is that in the ED, the 4AT showed higher sensitivity (0.83), and the CAM has higher specificity (0.98) and higher diagnostic accuracy indicating a good overall diagnostic performance.

It is clear from this study that delirium is common among the ED patients with the prevalence of 7–23%. All the included participants were over 65 years old. Age as one of the risk factors for delirium makes it more common among the elderly. Due to the special nature of the ED, most studies do not limit the types of diseases. Different assessors could lead to different sensitivity and specificity. When choosing an instrument, training should also be considered before use. Most of the studies were conducted by trained nurses or geriatricians. Evelien<sup>[29]</sup> et al. compared the 4AT with the CAM, special training on the CAM were conducted while the 4AT did not because the specific training for the 4AT is not required. These factors may affect the sensitivity and specificity of the screening tools.

Although the CAM has good accuracy as the routine use for delirium screening, it may not be ideally suited for the ED due to the time taken. The assessment of the CAM is based on only four cardinal elements: 1) an acute onset of mental status changes of fluctuating course; 2) inattention; 3) disorganized thinking; 4) an altered level of consciousness. The patient is diagnosed as delirious if he has both

features 1 and 2, and either feature 3 or 4<sup>[30]</sup>. Before assessing delirium with the CAM, it usually takes 5–10 minutes to complete bedside interviews and short cognitive tests before the screening. Wolfgang<sup>[31]</sup> et al. developed modified Confusion Assessment Method for the Emergency Department (mCAM-ED), which takes about 3.2 min. Differently, the 4AT is a newly developed screening tool for rapid initial assessment of delirium ([www.the4AT.com](http://www.the4AT.com)). It is brief (generally < 2 min) makes it more suitable for delirium screening particularly in the ED which there is a limited time to perform diagnosis.

Through the network meta-analysis, the CAM has higher specificity than 4AT, which is consistent with the results of our meta-analysis. It is worth noting that all existing screening tools were included to validate our findings.

## 4.1 Comparison with other studies

Michael<sup>[12]</sup> and Pilar<sup>[32]</sup> conducted systematic reviews to describe the delirium screening tools, but they provide limited evidence to compare the accuracy of each screening tools. Unlike these previous researches, we included more recent published studies to perform a pairwise meta-analysis and network meta-analysis to draw our conclusion. We did a detailed analysis of the included studies and provide evidence for the selection of screening tools for the ED staff.

## 4.2 Strengths and Limitations

This review provides a comprehensive outline of the use of delirium screening instruments in the ED. Our research has the following advantages: (1) in the review, we used a comprehensive search with explicit inclusion and exclusion criteria; (2) study selection and data extraction were conducted independently. (3) we assessed the risk of bias of the individual studies, which increased the validity of our conclusions; (4) we used the advanced meta-analysis methods - indirect comparison - to compare different screening tools and rank their relevant superiority. However, there are some limitations in our research that only studies published in English were included. It is worth noting that four screening tools in our network meta-analysis only included one or two studies, this may lead to unstable outcomes.

## Conclusion

The present meta-analysis suggests that both CAM and 4AT show better diagnostic accuracy in detecting delirium for the ED patients, however, 4AT shows less screening time when compared with the CAM. More high-quality studies are needed to assess the accuracy of other delirium screening tools.

## Abbreviations

ED: Emergency Department; DSM: Diagnostic and Statistical Manual of Mental Disorders; QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies; CAM: Confusion Assessment Method; 4AT: Arousal, Attention, Abbreviated Mental Test 4, Acute change; SIGN: Scottish Intercollegiate Guidelines Network; NICE: National Institute for Health and Care Excellence; CAM-ICU: Confusion Assessment Method for the Intensive Care Unit; SI: superiority index; PRISMA: preferred reporting items for systematic reviews and meta-analysis; TP: true positive; FP: false positive; TN: true negative; FN: false negative; AUC: area under the curve; SROC: summary receiver operator characteristic; RASS: Richmond Agitation Sedation Scale; DTS: Delirium Triage Screen; 6-CIT: 6-Item Cognitive Impairment Test; mCAM-ED: modified Confusion Assessment Method for the Emergency Department.

## Declarations

### 5.1 Ethical approval

For this type of study formal consent is not required.

### 5.2 Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### 5.3 Competing interests

The authors declare that they have no competing interests.

## 5.4 Funding

None

## 5.5 Authors' contributions

Long Ge, and Baoshan Di planned and designed the current study. Qian Zhang and Meixi Chen extracted data. Ziqi Guo and Qian Zhang assessed the risk of bias of included studies. Qian Zhang, Ziqi Guo and Baoshan Di performed the data analysis and initial interpretation. Qian Zhang, Liangying Hou, Qing Zhang and Long Ge wrote the article and revised it. All authors revised critically for important intellectual content and approved the final version to be submitted.

## 5.6 Acknowledgements

Not applicable

## Consent to participate

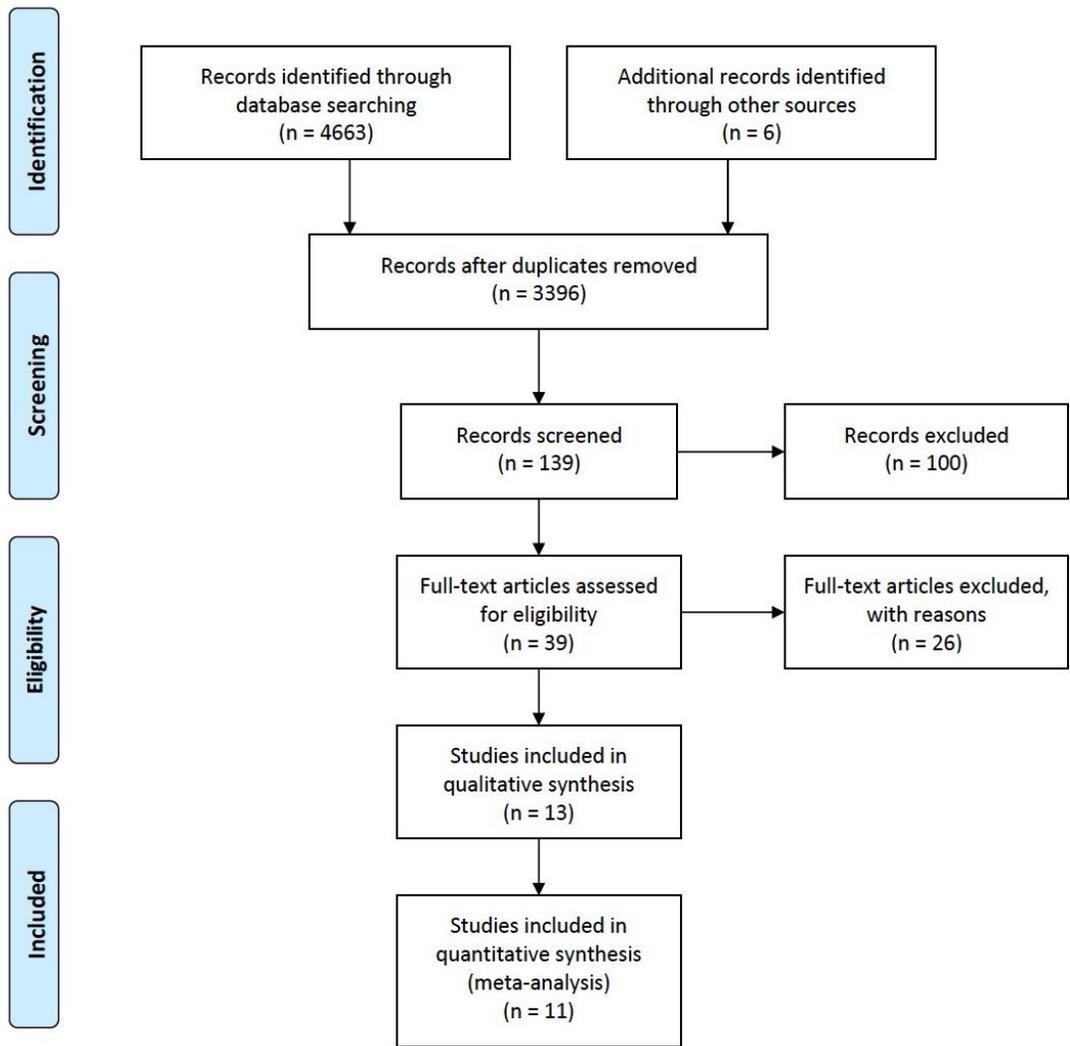
Written informed consent for publication was obtained from all participants.

## References

1. Meagher D, Trzepacz PT. Phenomenological distinctions needed in DSM-V: delirium, subsyndromal delirium, and dementias. *J Neuropsychiatry Clin Neurosci.* 2007;19(4):468-470.
2. Nguyen PV, Pelletier L, Payot I, Latour J. The Delirium Drug Scale is associated to delirium incidence in the emergency department. *Int Psychogeriatr.* 2018;30(4):503-510.
3. Woodhouse R, Burton JK, Rana N, Pang YL, Lister JE, Siddiqi N. Interventions for preventing delirium in older people in institutional long-term care. *Cochrane Database Syst Rev.* 2019;4(4).
4. Neto AS, Nassar AP Jr, Cardoso SO, et al. Delirium screening in critically ill patients: a systematic review and meta-analysis. *Crit Care Med.* 2012;40(6):1946-1951.
5. Han JH, Zimmerman EE, Cutler N, et al. Delirium in older emergency department patients: recognition, risk factors, and psychomotor subtypes. *Acad Emerg Med.* 2009;16(3):193-200.
6. Hustey FM, Meldon SW, Smith MD, Lex CK. The effect of mental status screening on the care of elderly emergency department patients. *Ann Emerg Med.* 2003;41(5):678-684.
7. Wahid ZU. Delirium in Hospitalized Patients: Underdiagnosed and Undertreated Missed Opportunities in Preventive Care. *J Natl Med Assoc.* 2004;96(4, 2010):414.
8. Han JH, Shintani A, Eden S, et al. Delirium in the emergency department: an independent predictor of death within 6 months. *Ann Emerg Med.* 2010;56(3):244-252.
9. Pérez-Ros P, Martínez-Arnau FM. Delirium Assessment in Older People in Emergency Departments. A Literature Review. *Diseases.* 2019;7(1):14.
10. Delaney M, Pepin J, Somes J. Emergency Department Delirium Screening Improves Care and Reduces Revisits for the Older Adult Patient. *J Emerg Nurs.* 2015;41(6):521-524.
11. Martins S, Fernandes L. Delirium in elderly people: a review. *Front Neurol.* 2012;3:101.
12. LaMantia MA, Messina FC, Hobgood CD, et al. Screening for delirium in the emergency department: a systematic review. *Ann Emerg Med.* 2014;63(5):551-560.e2.
13. Delaney M, Pepin J, Somes J. Emergency Department Delirium Screening Improves Care and Reduces Revisits for the Older Adult Patient. *J Emerg Nurs.* 2015;41(6):521-524.
14. Carpenter CR, Shelton E, Fowler S, et al. Risk factors and screening instruments to predict adverse outcomes for undifferentiated older emergency department patients: a systematic review and meta-analysis. *Acad Emerg Med.* 2015;22(1):1-21.
15. Davis D, Searle SD, Tsui A. The Scottish Intercollegiate Guidelines Network: risk reduction and management of delirium. *Age Ageing.* 2019 Jul 1;48(4):485-488.

16. Delirium: prevention, diagnosis and management. London: National Institute for Health and Care Excellence (UK); 2019.
17. Barron EA, Holmes J. Delirium within the emergency care setting, occurrence and detection: a systematic review. *Emerg Med J*. 2013;30(4):263-268.
18. Mariz J, Costa Castanho T, Teixeira J, et al. Delirium Diagnostic and Screening Instruments in the Emergency Department: An Up-to-Date Systematic Review. *Geriatrics (Basel)*. 2016;1(3):22.
19. Pan B, Ge L, Xun YQ, Ya JC, Cai YG, Xue, H. Exercise training modalities in patients with type 2 diabetes mellitus: a systematic review and network meta-analysis. *Int J Behav Nutr Phys Act*. 2018;15(1):72.
20. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA Statement. *JAMA*, 2018, 319(4): 388-396.
21. Hutton B, Salanti G, Caldwell DM. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. 2015;162(11):777–784.
22. Ma X, Zeng H, Zhang J, et al. New practical methods to obtain crucial data in performing diagnostic meta-analysis of the published literature. *J Evid Based Med*. 2018 Feb;11(1):56-63.
23. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
24. Whiting PF, Rutjes AW, Westwood ME. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.
25. Hamza TH, Arends LR, van Houwelingen HC, et al: Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Med Res Methodol* 2009, 9:73-73.
26. Lijmer JG, Bossuyt PMM, Heisterkamp SH: Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002, 21(11):1525-1537.
27. Ge L, Pan B, Song F, Ma J, Zeraatkar D, Zhou, J. Comparing the diagnostic accuracy of five common tumour biomarkers and CA19-9 for pancreatic cancer: a protocol for a network meta-analysis of diagnostic test accuracy. *BMJ Open*. 2017;7(12).
28. Nyaga VN, Aerts M, Arbyn M. ANOVA model for network meta-analysis of diagnostic test accuracy data. *Stat Methods Med Res*. 2018;27(6):1766–1784.
29. Shenkin SD, Fox C, Godfrey M, et al. Delirium detection in older acute medical inpatients: a multicentre prospective comparative diagnostic test accuracy study of the 4AT and the confusion assessment method. *BMC Med*. 2019;17(1):138.
30. MacLulich AM, Shenkin SD, Goodacre S, et al. The 4 'A's test for detecting delirium in acute medical patients: a diagnostic accuracy study. *Health Technol Assess*. 2019;23(40):1-194. doi:10.3310/hta23400
31. Hasemann W, Grossmann FF, Stadler R, et al. Screening and detection of delirium in older ED patients: performance of the modified Confusion Assessment Method for the Emergency Department (mCAM-ED). A two-step tool. *Intern Emerg Med*. 2018;13(6):915-922.
32. Pérez-Ros P, Martínez-Arnau FM. Delirium Assessment in Older People in Emergency Departments. A Literature Review. *Diseases*. 2019;7(1):14.

## Figures



**Figure 1**

PRISMA summary of evidence search and selection

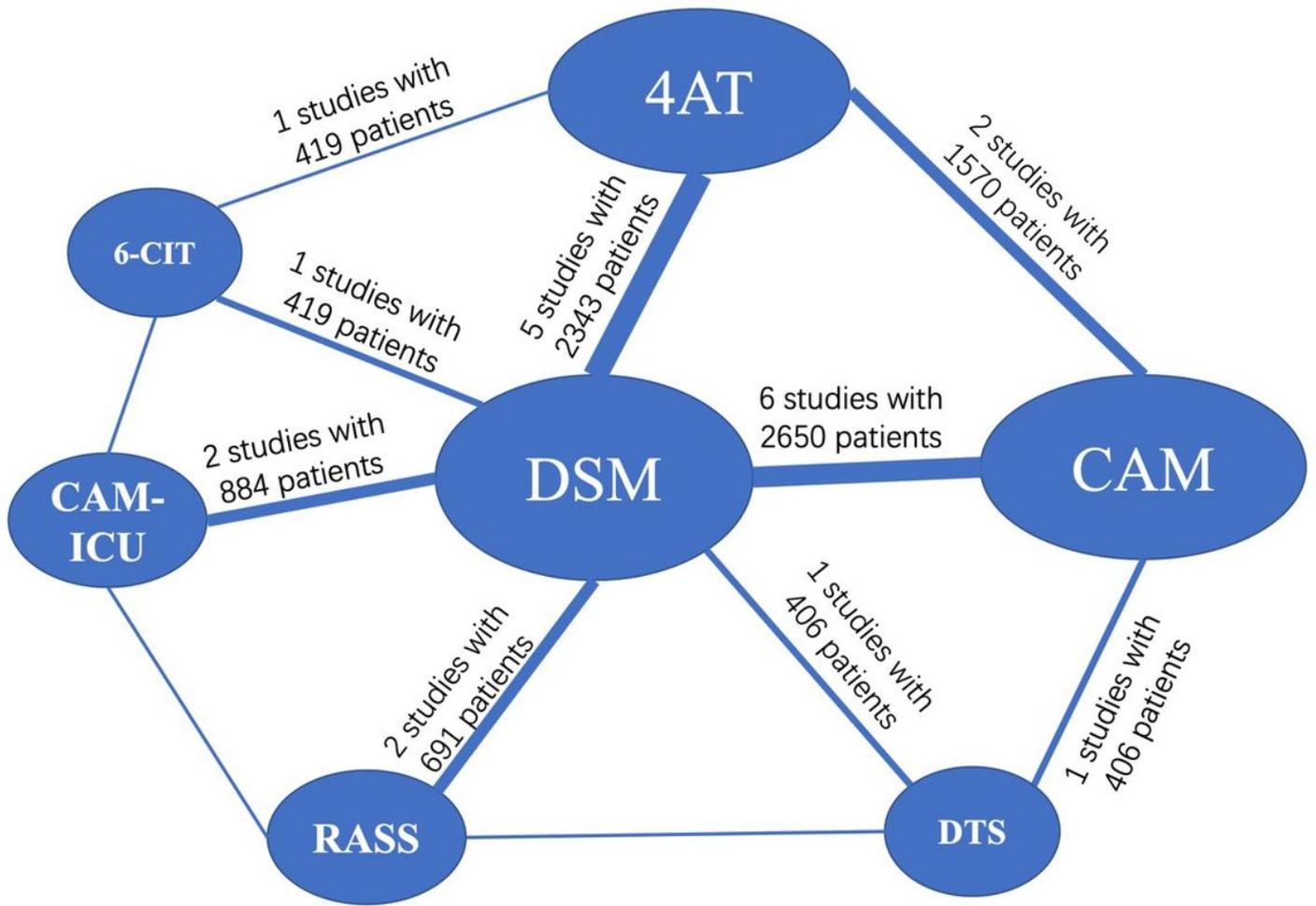


Figure 2

Network plot

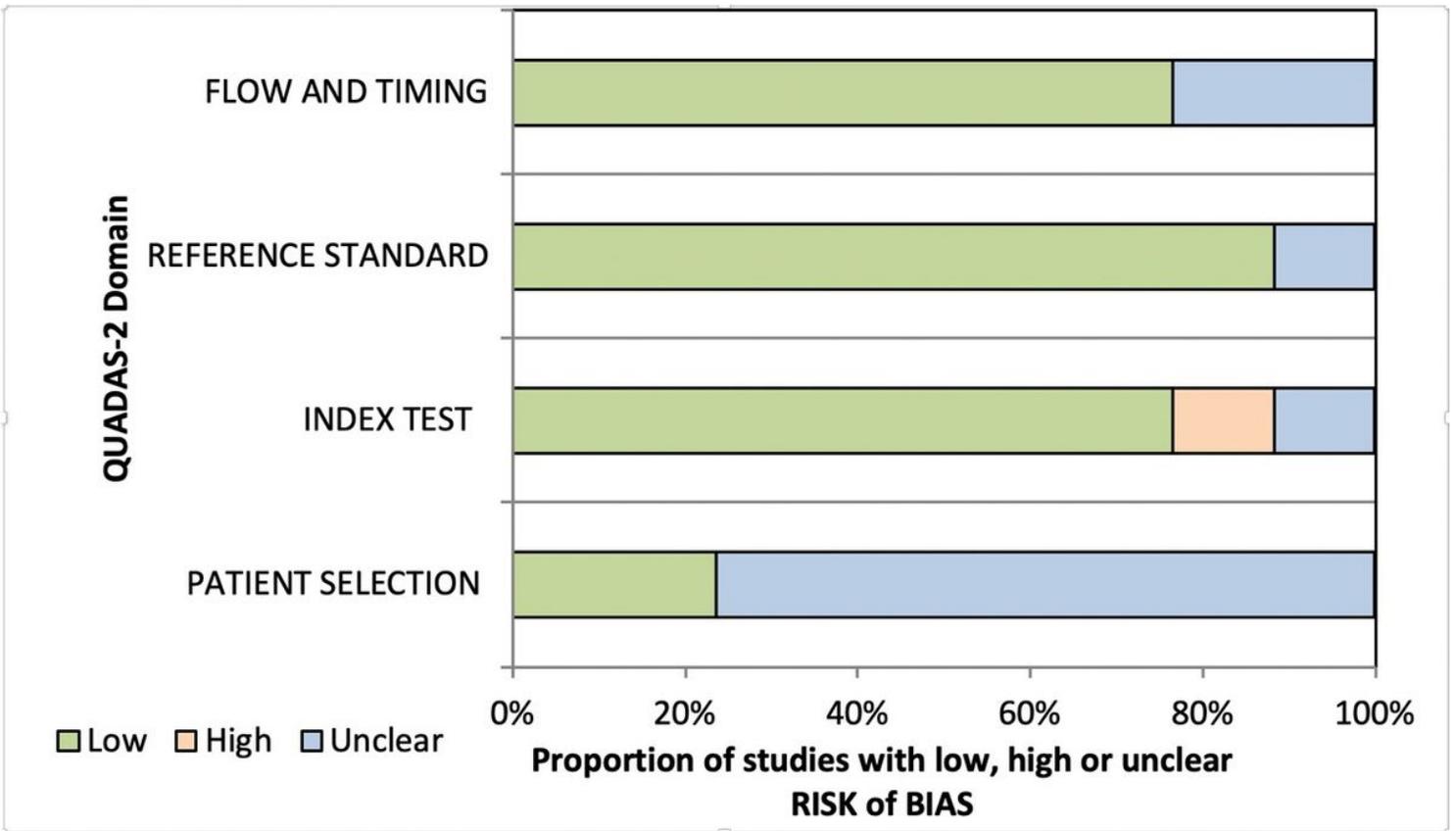


Figure 3

risk of bias

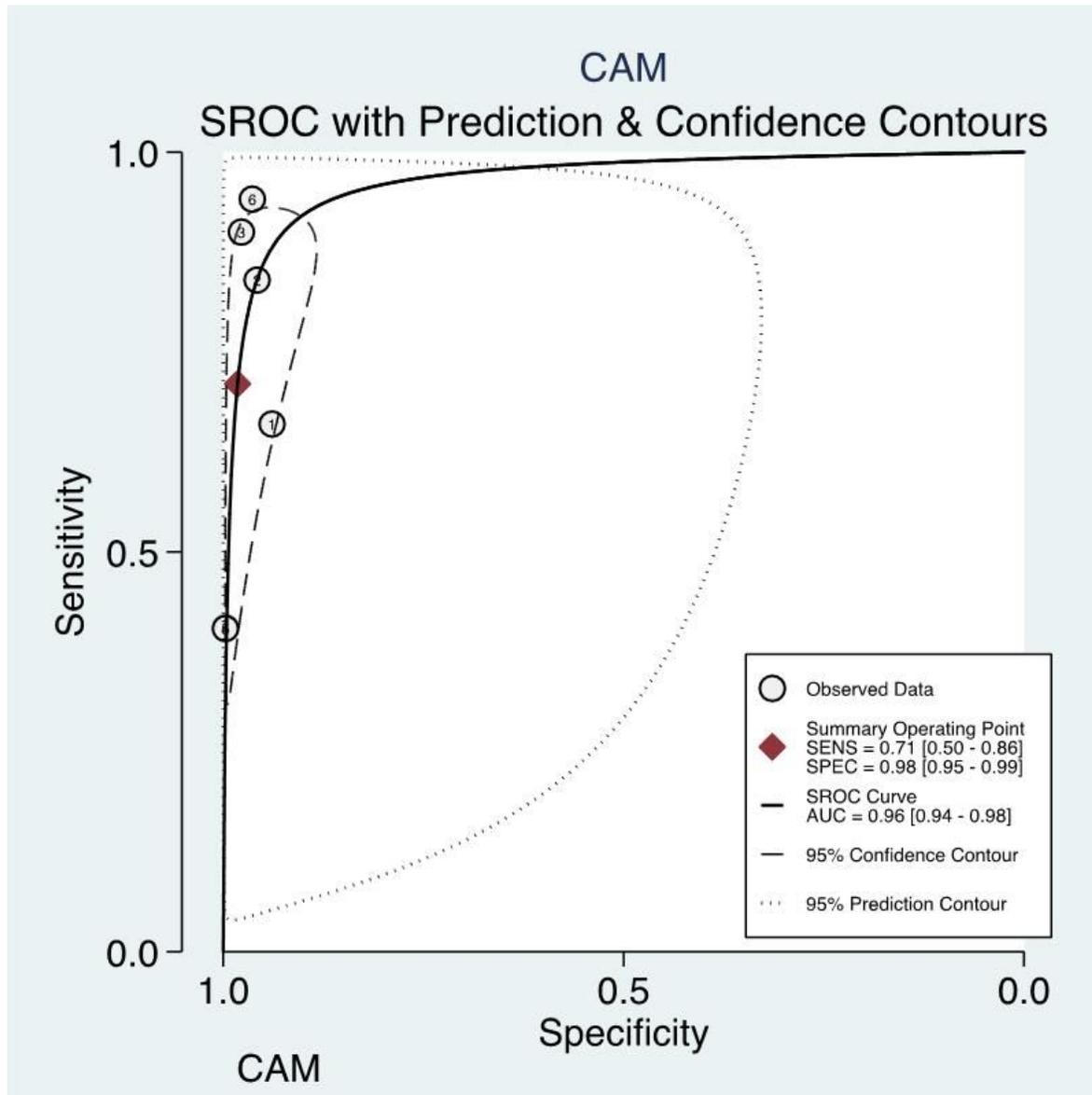
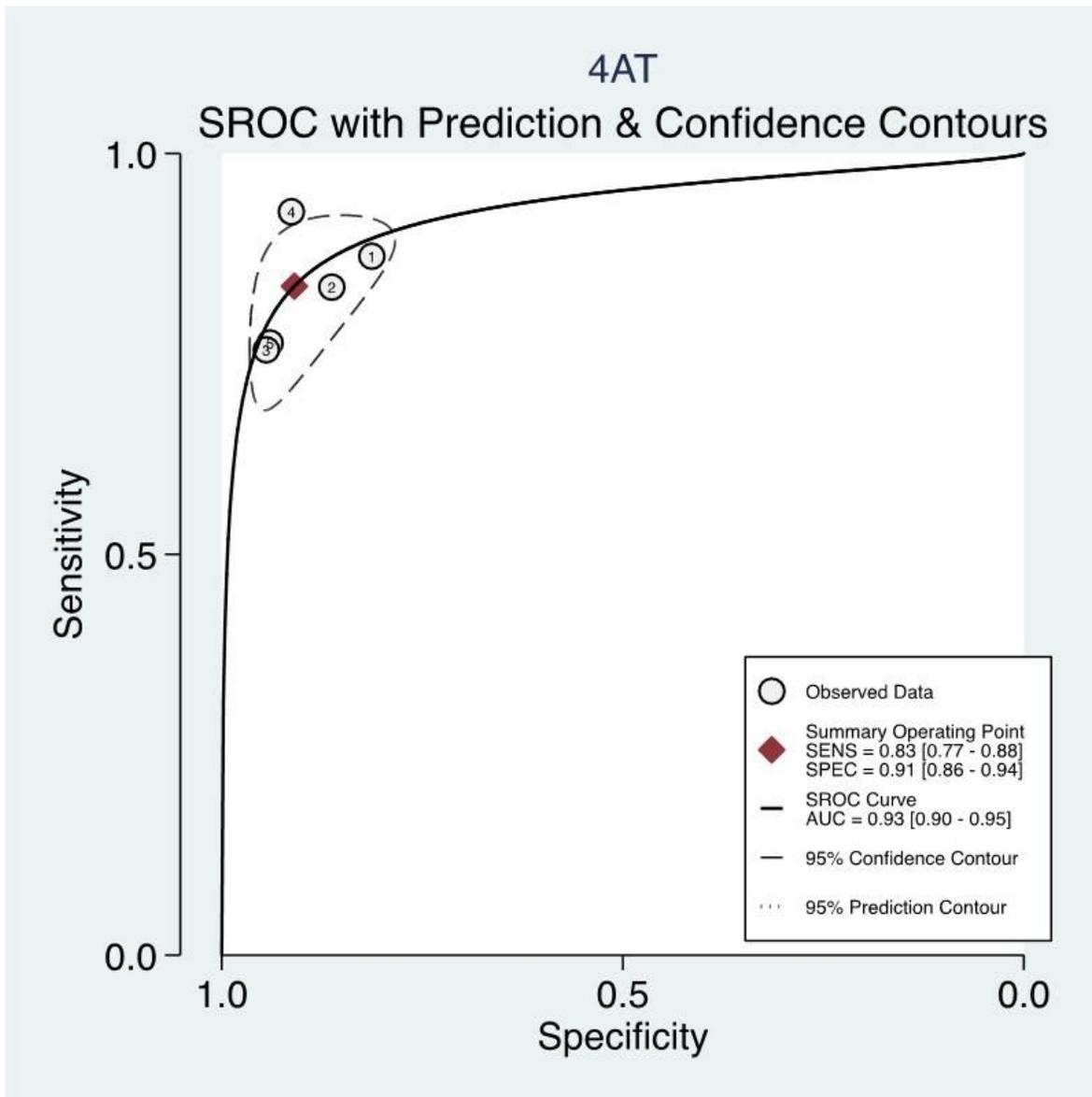


Figure 4

ROC Curve of the CAM



**Figure 5**

ROC Curve of the 4AT

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryAA201368.docx](#)