The 4 'A's test for detecting delirium in acute medical patients: a diagnostic accuracy study

Alasdair MJ MacLullich,[1]* Susan D Shenkin,[1] Steve Goodacre,[2] Mary Godfrey,[3] Janet Hanley,[4] Antaine Stíobhairt,[1] Elizabeth Lavender,[3] Julia Boyd,[5] Jacqueline Stephen,[5,6] Christopher Weir,[5,6] Allan MacRaild,[7] Jill Steven,[7] Polly Black,[7] Katharina Diernberger,[6,7] Peter Hall,[5] Zoë Tieges,[1] Christopher Fox,[8] Atul Anand,[1] John Young,[9] Najma Siddiqi[10,11,12] and Alasdair Gray[6,7]

[1]Geriatric Medicine, Division of Health Sciences, University of Edinburgh, Edinburgh, UK
[2]Emergency Medicine, University of Sheffield, Sheffield, UK
[3]Health and Social Care, Leeds Institute of Health Sciences, School of Medicine, Faculty of Medicine and Health, University of Leeds, Leeds, UK
[4]School of Health and Social Care, Edinburgh Napier University, Edinburgh, UK
[5]Edinburgh Clinical Trials Unit, University of Edinburgh, Edinburgh, UK
[6]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
[7]Emergency Medicine Research Group (EMERGE), NHS Lothian, Edinburgh, UK
[8]Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK
[9]Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK
[10]Psychiatry, University of York, York
[11]Hull York Medical School, York, UK
[12]Bradford District Care NHS Foundation Trust, Bradford, UK

*Corresponding author a.maclullich@ed.ac.uk

Declared competing interests of authors: none

Published August 2019
DOI: 10.3310/hta23400

Scientific summary

The 4 'A's test: a diagnostic accuracy study

Health Technology Assessment 2019; Vol. 23: No. 40
DOI: 10.3310/hta23400

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Delirium is significantly underdetected in the acute hospital, and there is a need for short, easy-to-administer screening tools that can be used by non-experts. The 4 'A's Test (4AT), a four-item bedside assessment, was developed for this purpose. This project assessed whether or not further development of the 4AT was needed, and then tested the 4AT's diagnostic accuracy compared with that of the Confusion Assessment Method (CAM), an alternative, older test with longer administration time. A health economics analysis was also undertaken.

Objectives

Phase 1

- What is the usability of the 4AT in clinical practice?
- Are any changes needed in the current structure and scoring of the 4AT?

Phase 2

The primary objective was:

- What is the diagnostic accuracy of the 4AT for delirium detection?

The secondary objectives were to:

- compare the performance of the 4AT with that of the CAM
- examine the performance of the cognitive test items in the 4AT in detecting general cognitive impairment
- determine if the 4AT predicts length of stay, new institutionalisation and mortality at 12 weeks
- examine the contribution of individual items of the 4AT to overall delirium diagnosis
- estimate the costs of delirium in the study population.

Methods

Phase 1

Surveys

We created two online surveys (A and B). We invited health-care practitioners across multiple health-care settings via social media and by e-mail. Items were presented in a fixed order and included multiple choice, ranking, five-point Likert scale and open comment response formats. Survey A considered general aspects of the knowledge and the attitudes to delirium assessment in medical practitioners, nurses, occupational therapists and physiotherapists in the UK. Survey B was tailored to those health-care practitioners who had previously used the 4AT to identify any potential changes to instrument.

Qualitative study

We performed a qualitative, mixed-methods study comprising interviews with health professionals and observations of practice in different locations along the patient journey; and in selected acute wards in

three sites: two English sites where the 4AT was not in use, and one Scottish site where the 4AT was widely used. Interviews and ethnographic observations were analysed using grounded theory.

Current use in clinical practice

We explored current clinical use of the 4AT by searching for published studies and online reports, looking for the presence of the 4AT in clinical guidelines, pathways and websites, and contacting networks of clinicians active in delirium, and making requests via social media.

Phase 2

Diagnostic test accuracy study

We performed a multicentre prospective diagnostic test accuracy study. We recruited older patients in emergency departments (EDs) or in acute general medical wards in three UK sites from October 2015 to December 2016. Inclusion criteria were being aged ≥ 70 years, and having been admitted to the ED within the past 12 hours or to the medical unit within the past 96 hours. Exclusion criteria were being in a coma, having an acute life-threatening illness and having severe communication impairment. Consent was sought from patients, or from legal proxies for patients who lacked the capacity to consent. Each patient underwent a reference standard delirium diagnostic assessment, and was then randomised to assessment (by a different researcher) with either (1) the 4AT (range 0–12; score of ≥ 4 indicates 'possible delirium') or (2) the CAM (scores are either negative or positive). The order of reference standard and 4AT/CAM was computer-randomised. We assessed previous cognitive impairment using the Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE). Length of stay, institutionalisation, and mortality at 12 weeks were measured.

A statistical analysis plan was agreed prior to database lock, blinded to randomised allocations. We calculated positive and negative predictive values, sensitivity and specificity [with exact binomial 95% confidence intervals (CIs)] for the 4AT versus the delirium reference standard. We constructed a receiver operating characteristic (ROC) curve. We tested differences in proportions and 95% CIs for each of sensitivity, specificity, positive and negative predictive values using Fisher's exact test. The overall performance of 4AT and CAM was summarised using Youden's Index (sensitivity minus false positive rate) and the odds ratio of sensitivity to specificity.
We used logistic regression modelling to predict mortality, and Kaplan–Meier curves and Cox proportional hazards models (adjusted for age, gender and presence of dementia) to predict length of stay in hospital.

Health economic analysis

Cost analysis using data collected within the study

Research nurses collected information about patients' use of hospital services from the Health Economics Questionnaire at the 12-week outcome assessment and about length of stay on different hospital wards from the clinical notes. We calculated the observed mean (95% CIs) of costs occurring within the 12 weeks. Unit costs were assigned to inpatient stays specific to the admitting specialty. Inpatient, day-case, outpatient or rehabilitation costs were summed, and then costs from community services were calculated and added to generate the total cost per patient. Alternative costing perspectives were taken for Scotland and England.

Cost-effectiveness analysis generated through a health economic model

We developed a health economic model to compare the cost-effectiveness of 4AT and CAM: a decision tree with pathways representing reference standard diagnosis of delirium and non-delirium, divided into branches dictated by sensitivity and specificity. The model was parameterised with costs and quality-adjusted life-years (QALYs) to calculate the incremental cost-effectiveness ratio (ICER).

Results

Phase 1

Surveys

In survey A, 83% of respondents ($n = 2306$: medical, 47.7%; nursing, 29.0%; occupational therapy, 11.7%; physiotherapy, 11.5%) reported that awareness of delirium was increasing among hospital staff. Only 20% of respondents reported delirium detection rates of > 80% in their unit. Respondents stated that 64% of units had guidelines for delirium detection, although only 20% reported that these were 'almost always/always followed'. The vast majority (91%) of respondents agreed or strongly agreed that formal diagnosis of delirium is important, 93% agreed or strongly agreed that distinguishing between delirium and dementia is important, and 89% agreed or strongly agreed that delirium treatment improves patient care.

In survey B ($n = 100$: medical, 64%; nursing, 32%; occupational therapy, 3%; physiotherapy, 1%), 69% of respondents reported that the 4AT was routinely used in their clinical area, and 52% used the 4AT frequently. Respondents generally viewed the 4AT as a useful, rapid and practical tool. No consistent problems with the content or administration of the 4AT emerged.

Qualitative study

We undertook 19 interviews and 10 informant discussions in the 4AT 'virgin' sites, and 23 interviews in the experienced site, in EDs, medical assessment units (MAUs), elderly assessment units (EAUs)/frailty assessment units and surgical assessment units (SAUs).

Surgical assessment unit staff generally had a low level of understanding and awareness of delirium. Knowledge and understanding of the significance of delirium, including its impact on patient distress and outcomes, were critical to engaging staff in delirium detection. A simple, easy-to-use standardised tool, such as the 4AT, was felt to be useful. Staff needed to understand what action would result from completing a delirium detection tool.

In the ED, the place of work and the impact of the admission event and process on patients made a determination of 'acute confusion' difficult and uncertain. In the MAU, the detection of delirium was considered important, but the busy environment was felt to be a barrier to screening for and managing delirium. In the experienced site, nursing staff completed the 4AT, and medical staff conducted further formal cognitive testing. The lack of clarity about how the tool was used to inform action planning was a factor in the priority given to completing it. In the EAU/frailty assessment unit, staff appeared more knowledgeable about delirium. Collaboration between medical and nursing staff led to a culture in which delirium identification and action was owned by staff across disciplines as part of routine practice.

Current use in clinical practice

By December 2017, the 4AT was present in multiple local, national and international policy, guidelines, advisory documents and websites, and it was used in clinical practice in a range of medical and surgical units in the UK and other countries. The URL www.the4AT.com was accessed 61,269 times in the year beginning June 2016.

Phase 2

Diagnostic test accuracy study

A total of 843 patients were recruited from 5696 screened, and two withdrew; therefore, 841 had data for analysis, of whom 785 had a positive or negative reference standard assessment [delirium present, $n = 95$ (12.1%); delirium absent, $n = 690$ (87.9%)]. Delirium prevalence was 12.1% as assessed by the reference standard. Those with delirium were older (mean age 83.5 years vs. 81.1 years), more likely to be from a minority ethnic group (4.2% vs. 1.2%) and more likely to have a diagnosis of dementia (26.3% vs. 6.7%).

Primary objective

What is the diagnostic accuracy of the 4AT for delirium detection?

The 4AT had an area under the ROC curve of 0.90. A positive 4AT score (> 3) had specificity of 95% (95% CI 92% to 97%) and sensitivity of 76% (95% CI 62% to 87%), with a positive predictive value (PPV) of 66% (95% CI 52% to 78%) and a negative predictive value (NPV) of 96% (95% CI 94% to 98%) for detecting delirium.

Secondary objectives

To compare the performance of the 4AT with that of the CAM

Among the 384 patients who had a valid CAM, the prevalence of delirium was 11% ($n = 42$) as assessed by the reference standard, and 5% ($n = 18$) as assessed using the CAM. Among the 392 patients who had a valid 4AT score, the prevalence of delirium was 12.5% ($n = 49$) as assessed by the reference standard, and 14% ($n = 56$) as assessed using the 4AT. The CAM had a specificity of 100% (95% CI 98% to 100%) and a sensitivity of 40% (95% CI 26% to 57%) in the subset of participants whom it was possible to assess using this. Youden's Index (sensitivity minus specificity, plus one) was 0.40 for CAM and 0.70 for 4AT. The odds ratio of sensitivity to specificity was 232 for the CAM (95% CI 30 to 1812) and 53 for the 4AT (95% CI 24 to 117).

To examine the performance of the cognitive test items in the 4AT in detecting general cognitive impairment

The Abbreviated Mental Test – 4 (AMT4) and attention (Months Backwards test) cognitive test items of the 4AT did show some relationships – high specificity but low to moderate sensitivity – with general cognitive impairment as ascertained in the reference standard assessment by a documented prior diagnosis of dementia and/or a positive IQCODE score [specificity AMT4 score of 2: 97% (95% CI 94% to 98%); attention score of 2: 98% (95% CI 96% to 99%); sensitivity AMT4 score of 2: 47% (95% CI 32% to 62%; attention score of 2: 23% (95% CI 12% to 38%)], and a high NPV but low PPV, for a diagnosis of dementia (a documented history, or an IQCODE score of ≥ 3.44).

To determine if the 4AT predicts length of stay, new institutionalisation, and mortality at 12 weeks

Patients with a 4AT score of > 3 had longer lengths of stay (median 5 days, interquartile range 2.0–14.0 days) than those with a score of ≤ 3 (median 2 days, interquartile range 1.0–6.0 days). The hazard ratio of time to discharge was 0.64 (95% CI 0.46 to 0.88). Positive 4AT scores were also significantly associated with mortality: 16.1% of those with a 4AT score of > 3 had died by 12-week follow-up, compared with 9.2% of those with a 4AT score of ≤ 3. The odds ratio of mortality for those with a 4AT score of > 3 was higher, at 2.00 (95% CI 0.85 to 4.70). New institutionalisation within 12 weeks was too infrequent to be analysed.

To examine the contribution of individual items of the 4AT to overall delirium diagnosis

The alertness test item of the 4AT showed high specificity (99%, 95% CI 98% to 100%) and low sensitivity (31%, 95% CI 18% to 45%) for delirium. Two or more patient errors (including untestable) on the AMT4 had high specificity (96%, 95% CI 94% to 98%) and low sensitivity (41%, 95% CI 27% to 56%) for delirium. An attention (Months Backwards test) score of 1 or 2 (< 7 months correct, refuses to start or untestable) versus 0 (≥ 7 months correct) had specificity of 79% (95% CI 74% to 83%) and sensitivity of 71% (95% CI 57% to 83%). A score of 2 (untestable) versus 1 or 0 had identical performance to a positive score (4) on the alertness item. The acute change/fluctuating course item had high specificity (96%, 95% CI 93% to 98%) and low sensitivity (76%, 95% CI 61% to 87%) for delirium. Almost all of those who scored positively on the 4AT scored positively on 'acute change or fluctuating course'.

Health economic analysis

To model the costs of delirium in the study population

Using data collected during the study, a diagnosis of delirium was associated with higher costs regardless of the costing perspective or diagnosis method. Using original researcher delirium ascertainment, costs in Scotland for the initial stay were £2810 (95% CI £2734 to £2886) with delirium and £1277 (95% CI £1267 to £1287) without, and costs at 12 weeks were £7559 (95% CI £7362 to £7755) with delirium and £4215 (95% CI £4175 to £4254) without. Costs in England for the initial stay were £2810 (95% CI £2734 to £2886) with delirium and £1277 (95% CI £1267 to £1287) without, and costs at 12 weeks were £5216 (95% CI £5107 to £5326) with delirium and £4320 (95% CI £4295 to £4346) without. A similar pattern emerged for patients with delirium as assessed by centrally ascertained reference standard diagnosis: costs for patients with delirium at 12 weeks (Scotland £2934, 95% CI £2862 to £3006; England £2934, 95% CI £2862 to £3006) were higher than for those without delirium at 12 weeks (Scotland £1239, 95% CI £1229 to £1249; England £4264, 95% CI £4239 to 4290).

Cost-effectiveness analysis generated through a health economic model

The results of the economic model were found to be highly dependent on underlying assumptions that were informed by expert opinion. In the base-case analysis, the difference in costs between the 4AT and the CAM was –£90.35, representing lower 12-week health-care costs for the 4AT. The difference in QALYs was –0.00053, representing very similar health outcomes using the 4AT to those using the CAM over 12 weeks. The base-case ICER was £170,533 (meaning that for each QALY lost, £170,533 would be saved), but the ICER varied considerably depending on model assumptions.

In the scenario analysis with English costs at the best estimate of parameters from expert elicitation, the difference in costs was –£61.52. The difference in QALYs remained the same as in the base case, yielding an ICER of £116,133. As in the base case, this would be considered cost-effective at conventional thresholds.

A scenario analysis was undertaken using the lowest and highest values of parameters estimated from expert opinion elicitation. At the lowest estimates, the ICER was £24,289 per QALY gained, which is within a range in which the 4AT may be considered cost-effective. At the highest estimates, the 4AT dominates the CAM, being more effective and less expensive.

The model estimated the 12-week cost incurred by patients with a false-positive diagnosis of delirium to be £4653. For patients with a false-negative diagnosis, the estimated cost was £8955.70.

Conclusions

The 4AT is already widespread, and growing, in international clinical use, and it is cited in many guidelines, pathways and websites. This study supports its ongoing clinical use as a rapid delirium assessment instrument for older acute hospital patients. It detects delirium with high specificity and reasonable sensitivity. Subscores of the 4AT (an AMT4 score of 2 or an attention score of 2) have high specificity but low to moderate sensitivity for general cognitive impairment. However, the qualitative studies indicate that improving delirium detection relies not only on an effective assessment tool but also on adequate training in delirium and its assessment, and an understanding of the action plan resulting from a positive delirium diagnosis. Thus, the 4AT has to be embedded into wider education about delirium and dementia, with the test result directly affecting a patient's care pathway in a way that is visible to staff involved. Further research should address the real-world implementation of delirium assessment. Delirium is associated with higher costs, with undetected delirium possibly costing more, although this needs to be studied further. The 4AT shows clear potential to be cost-effective, but further research is needed to prove this definitively. Improving the detection of delirium has the potential to result in large cost savings within the health-care system.

Trial registration

This trial is registered as ISRCTN53388093.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.819

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 11/143/01. The contractual start date was in May 2017. The draft report began editorial review in December 2017 and was accepted for publication in September 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).