BMJ Open

Protocol for validation of the 4AT, a rapid screening tool for delirium: a multicentre prospective diagnostic test accuracy study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015572
Article Type:	Protocol
Date Submitted by the Author:	21-Dec-2016
Complete List of Authors:	Shenkin, Susan; University of Edinburgh, Geriatric Medicine Fox, Chris; Norwich Medical School, Department of Psychological Sciences Godfrey, Mary; University of Leeds, Academic Unit of Elderly Care and Rehabilitation; Leeds Institute of Health Sciences Siddiqi, Najma; University of York, Psychiatry, Hull York Medical School, York and Bradford District Care NHS Foundation Trust, Bradford, UK Goodacre, Steve; University of Sheffield, Medical Care Research Unit Young, John; Leeds University Hospital, Elderly Care and Rehabilitation Anand, Atul; University of Edinburgh, Geriatric Medicine Gray, Alasdair; NHS Lothian, Emergency Medicine Research Group (EMERGE) Smith, Joel; University of Oxford, Health Economics Research Centre, Nuffield Department of Population Health Ryan, Tracy; NHS Lothian, Old Age Psychiatry Hanley, Janet; Edinburgh Napier University, Health and Social Care MacRaild, Allan; NHS Lothian, Emergency Medicine Research Group (EMERGE) Steven, Jill; NHS Lothian, Emergency Medicine Research Group (EMERGE) Black, Polly; NHS Lothian, Emergency Medicine Research Group (EMERGE) Boyd, Julia; Edinburgh Clinical Trials Unit Weir, Christopher; University of Edinburgh, MRC Hub for Trials Methodology Research MacLullich, Alasdair; University of Edinburgh, Geriatric Medicine Unit
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Geriatric medicine, Neurology, Mental health
Keywords:	Delirium & cognitive disorders < PSYCHIATRY, diagnosis, screening

Protocol for validation of the 4AT, a rapid screening tool for delirium: a multicentre prospective diagnostic test accuracy study

Corresponding author:

Prof Alasdair MJ MacLullich

Edinburgh Delirium Research Group, Geriatric Medicine

Room S1642

Department of Clinical and Surgical Sciences

The University of Edinburgh, Royal Infirmary of Edinburgh

51 Little France Crescent, Edinburgh, EH16 4SB

Tel: +44(0)131 242 6481

Email: a.maclullich@ed.ac.uk

Keywords: delirium, confusion, hospital, diagnostic test accuracy

Word count (excl title page, refs, tables, figure): 5,464

Authors:

Susan D Shenkin, Geriatric Medicine, University of Edinburgh, Edinburgh, UK

Christopher Fox, Old Age Psychiatry, University of East Anglia, Norfolk UK

Mary Godfrey, Health and Social Care, Institute of Health Sciences, University of Leeds, Leeds, UK

Najma Siddiqi, Psychiatry, University of York, York, Hull York Medical School, York and Bradford District Care NHS Foundation Trust, Bradford, UK

Steve Goodacre, Emergency Medicine, University of Sheffield, Sheffield, UK

John Young, Elderly Care and Rehabilitation, University of Leeds, Leeds, UK

Atul Anand, Cardiovascular Sciences and Geriatric Medicine, University of Edinburgh, Edinburgh, UK

Alasdair Gray, Professor of Emergency Medicine, Emergency Medicine Research Group (EMERGE), Department of Emergency Medicine, NHS Lothian, Edinburgh, UK

Joel Smith, Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Tracy Ryan, Old Age Liaison Psychiatry, NHS Lothian, Edinburgh, UK

Janet Hanley, Health and Social Care, Edinburgh Napier University, Edinburgh, UK

Allan MacRaild, Emergency Medicine Research Group Edinburgh (EMERGE), NHS Lothian, Edinburgh, UK

Jill Steven, Emergency Medicine Research Group Edinburgh (EMERGE), NHS Lothian, Edinburgh, UK

Polly L Black, Emergency Medicine Research Group Edinburgh (EMERGE), NHS Lothian, Edinburgh, UK

Julia Boyd, Edinburgh Clinical Trials Unit, University of Edinburgh, Edinburgh, UK

Christopher J Weir, Edinburgh Clinical Trials Unit, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

Alasdair MJ MacLullich, Geriatric Medicine, University of Edinburgh, Edinburgh UK

Competing interests:

This work was funded by National Institute of Health Research Technology Assessment Programme (NIHR DTA) grant number 11/143/01 (PI: AMJM; CO-I: SG, JSmith, AA, MG, SDS, TR, NS, CF, JH, CJW, AG). JB and CW were supporrted in this work by NHS Lothian via the Edinburgh Clinical Trials Unit. AA has received

personal fees from Abbott Diagnostics, outside the submitted work. SG is funded by NIHR to undertake his role as NIHR HTA Clinical Evaluation and Trials Board Chair. SDS is funded by NRS (NHS Research Scotland) to undertake her role as NRS Ageing Specialty Group Lead (Scotland).

Data sharing:

Not applicable: protocol paper

ABSTRACT

Introduction: Delirium is a severe neuropsychiatric syndrome of rapid onset, commonly precipitated by acute illness. It is common in older people in the emergency department and acute hospital, but greatly under-recognised in these and other settings. Delirium and other forms of cognitive impairment, particularly dementia, commonly co-exist. There is a need for a rapid delirium screening tool that can be administered by a range of professional level healthcare staff to patients with sensory or functional impairments in a busy clinical environment, which also incorporates general cognitive assessment. We developed the '4As Test' (4AT) for this purpose.

This study's primary objective is to validate the 4AT against a reference standard. Secondary objectives include (a) comparing the 4AT with another widely used test (the Confusion Assessment Method (CAM)); b) determining if the 4AT is sensitive for general cognitive impairment; c) assessing if 4AT scores predict outcomes; including d) a health economic analysis.

Methods and analysis: 900 patients aged 70 or over in Emergency Departments or acute general medical wards will be recruited in three sites (Edinburgh, Bradford, Sheffield) over 18 months. Each patient will undergo a reference standard delirium assessment and will be randomised to assessment with either the 4AT or the CAM. At 12 weeks outcomes (length of stay, institutionalisation, and mortality) and resource utilisation will be collected by a questionnaire, and via the electronic patient record.

Ethics and dissemination: Ethical approval was granted in Scotland and England. The study involves administering tests commonly used in clinical practice. The main ethical issues are the essential recruitment of people without capacity. Dissemination is planned via publication in high impact journals, presentation at conferences, social media and the website www.the4AT.com.

Registration details: International standard randomised controlled trial number (ISRCTN) 53388093. UK Clinical Research Network ID: 19502

Strengths and limitations of this study:

Strengths:

- Pre-registered protocol of multicentre study to validate brief screening tool (4AT) for delirium
- Validation against a reference standard
- Comparison with another widely used test (Confusion Assessment Method)
- Includes health economic analysis

Limitations:

- Protocol for ongoing study
- Use of 4AT in clinical practice has increased world-wide while study open for recruitment
- Different ethical and legal frameworks in Scotland and England

INTRODUCTION

Background

Delirium is a severe and distressing neuropsychiatric syndrome which is characterised by acute deterioration in attention and other mental functions. The diagnostic criteria are, in summary: a disturbance of consciousness (that is, reduced ability to focus, sustain or shift attention), and a change in cognition. The mental status deterioration develops over short periods of time (usually hours to days) and it tends to fluctuate[1,2]. Delirium is commonly precipitated by acute illness, trauma, or the side-effects of medications. The presence of a 'medical condition' is part of the Diagnostic and Statistical Manual for Mental Disorders, 4th and 5th Edition (DSM-IV, DSM-V) criteria. Delirium is extremely common: it affects at least 15% of patients in acute hospitals, and is more common in older people[3-5]. It is independently associated with many poor outcomes[6-10]. Delirium is also a marker of current dementia[6,11] and is associated with acceleration of existing dementia[12]. In older patients without dementia, an episode of delirium strongly predicts future dementia risk[7,13]. The economic burden of delirium derived from 2008 US data estimates the one-year health care costs to be $38-$152 billion[13] but to date there are no exact investigations of the costs associated with delirium.

Detection of delirium is essential because it indicates acute systemic or central nervous system illness, physiological disturbance and drug intoxication or withdrawal. Failure to detect delirium in the acute setting is associated with worse outcomes[14]. Specific management of delirium is of obvious and immediate benefit to patients in many clinical situations, e.g. in reversing opioid toxicity, treatment of peripheral infections which have

presented with delirium, alleviating distress caused by delusions and hallucinations[15], and in prompting more thorough assessment of symptoms[16].

More broadly, detecting cognitive impairment in general (delirium, dementia, depression, learning disability, etc.) is a prerequisite for high quality care because of the multiple immediate implications of cognitive impairment for patients and staff, including: ensuring adequate communication with patients and their families, doing careful assessment of capacity to provide consent for clinical procedures, avoiding giving treatments contrary to the law because of lack of consent, alleviating distress more readily, avoiding unnecessary bed transfers, and prompting delirium prevention including a detailed drugs review. Detection of dementia has recently been highlighted in the Dementia Commissioning for Quality and Innovation framework in operation in NHS England[17].

In general medical and Emergency Department settings delirium is grossly under-detected: at least two-thirds of cases are missed[5,18,19]. It is unclear why detection rates are so low. Evidence from surveys and workshops have raised several possibilities, including general ignorance about delirium, lack of awareness of its importance, uncertainty about discriminating delirium from dementia, and lack of time for assessment in the acute setting[20-24]. The lack of a very rapid, simple, and validated screening tool is a major factor in the under-detection of delirium.

Many delirium assessment instruments have been developed that operationalise the standard diagnostic criteria for delirium, but these have largely remained research tools. The most commonly advocated screening tool for use in routine clinical care, the short Confusion Assessment Method (CAM)[25], has satisfactory sensitivity and specificity in trained hands

but takes around 10 minutes to complete because it requires a cognitive assessment like the Modified Mini-Cog[26,27] to be done first. The CAM also requires the rater to make subjective judgement of mental status. Subjective judgements are less reliable, often more time-consuming, and more difficult for staff (particularly non-specialists) than simple objective measures with clearly-defined cut-points[27].

The problem of some patients being 'untestable' is likely to be another important factor in delirium under-detection: many patients in acute settings are too unwell, sleepy, or agitated to undergo cognitive testing or even interview[28-31]. Most screening tools do not make explicit how these patients should be classified. The result is that mental status assessments are often simply left uncompleted in most 'untestable' patients, and no diagnosis, and often no specific treatment, is applied. This lack of a diagnosis can be harmful[14].

Finally, given the time pressures in acute settings, it is challenging to implement a separate delirium screening instrument in addition to any existing general cognitive screening instruments. The lack of a combined instrument allowing screening for both general cognitive impairment and delirium may therefore contribute to the lack of specific delirium detection. Early diagnosis of delirium using evidence-based diagnostic tools offers a means for improved outcomes and more efficient resource allocation decisions.

Rationale for the Study

Given the multiple constraints of the acute environment, the range of staff that might be expected to screen for delirium, the common co-existence of delirium and dementia, and the heterogeneity of patients, we determined the requirements for a screening tool (Table 1).

Table 1: Requirements for a screening tool for delirium for use in the acute hospital environment

Short (less than 2 minutes)
Easy to learn
Easy to administer and score
Can be used by professional-level healthcare staff from a variety of disciplines
Allows scoring of patients who are too drowsy or agitated to undergo cognitive testing or clinical interview
Takes account of informant history
Can be administered through written questions to people with severe hearing impairment
Can be administered to patients with visual impairments
Does not require subjective judgements based on interview
Combines delirium screening with general cognitive screening
Does not need a quiet environment for administration
Does not require physical responses such as drawing figures or clocks

There are multiple instruments for delirium screening, diagnosis, severity assessment, and monitoring[32-35]. Before deciding to design a new screening tool, we therefore examined each of the available tools against the above criteria, focusing on screening tools such as the CAM. We also searched the literature systematically, including conference proceedings, books, and book chapters, for any newly-published tools as well as to examine the study data for each tool. Most scales were excluded on grounds of duration alone. The remaining scales lacked features such as general cognitive screening, and other important features. We thus found that, in late 2010, no existing tool fulfilled the above requirements, and because of this

we decided to design a new test. This conclusion was supported by the NICE Guidelines on Delirium[6] which emphasised the need for research on a screening tool for delirium suitable for routine use.

The subsequent design process involved scrutiny of each of the nearly 30 published delirium assessment tools, evaluating the performance of each, including subtests, in published studies and, in most cases, through direct clinical or research experience of their use. Because we had decided to incorporate general cognitive screening into the new instrument, to avoid the need to have separate instruments for cognitive screening and delirium screening, we also reviewed the broader literature on brief tests for general cognitive impairment (including dementia). In the context of designing a screening tool for the acute hospital, it is important to note that delirium generally causes cognitive impairment detectable on the kinds of tests used for dementia screening[36,37]. Therefore, abnormal test results may indicate delirium and/or dementia (as well as other causes of cognitive impairment, such as learning disability).
It is clinically essential to know if any such impairment is acute, that is, delirium, but also important to identify underlying general (acute or chronic) cognitive impairment. A tool designed exclusively to detect cognitive impairment will not lead to delirium detection without another step, and a tool designed only to detect delirium may miss general cognitive impairment. In this light, we decided that the 4AT should include cognitive screening sensitive to general cognitive impairment, but also including items on altered level of alertness and change in mental status, both of which are strong indicators of delirium.
The first version of the 4AT was drafted and tested informally by colleagues, changes were made based on feedback, and updated versions tested again. After several iterations involving 20 doctors and nurses of varying levels of experience, the final version was produced. An initial audit in 30 inpatients comparing clinical use of 4AT with independent reference

standard DSM-IV assessment found 100% sensitivity (CI 69-100%) and 90% specificity (CI 68-99%). A subsequent validation study in Italy involving 234 consecutively recruited older hospitalised patients found that the 4AT had a sensitivity of 89.7% and specificity 84.1% for delirium[38]. The area under the receiver operating characteristic curves for delirium diagnosis was 0.93. Since the 4AT was launched, locally and through the www.the4AT.com website, it has been adopted in clinical units in several centres in the UK and internationally.

Thus, in 2014 there was encouraging evidence that the 4AT has value as a tool for delirium detection in routine practice. This evidence came from several sources: one published study, audits in several sites, informal feedback, adoption in clinical practice by several clinical units globally, and a recent web based survey focused specifically on 4AT provided evidence supporting its use. Since this study was designed, other validation studies have been published, with favourable results, however these included specific clinical populations (e.g. stroke[39]), languages (Thai[40],) had relatively small numbers[41] or validated assessments against clinical assessment rather than research reference standard assessment[42]. Therefore, a formal, large validation study is necessary to provide definitive evidence of the diagnostic accuracy of the 4AT.

Comparison with the CAM is also of value, because the CAM is in use in some clinical units and thus information on how the 4AT performs in relation to the CAM will help clinicians decide which tool is suitable for their particular context. Further information on how the 4AT performs as a cognitive screening tool, its ability to predict outcomes, and how each item of the 4AT contributes to its diagnostic accuracy will also provide important guidance to clinicians. Finally, understanding the economic costs and benefits of using the 4AT and the CAM will help providers in service pathway decisions.

Study objectives

The **primary objective** of the study is to determine the diagnostic accuracy of the 4AT for delirium detection versus the reference standard of a DSM-IV diagnosis.

The **secondary objectives** are:

(a) to compare performance of the 4AT and the Confusion Assessment Method (CAM);

(b) to determine if the 4AT is an adequately sensitive tool for detecting general cognitive impairment as judged against a documented history of dementia and/or the Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE);

(c) to determine if 4AT scores predict important outcomes such as length of stay, institutionalisation, and mortality, up to 12 weeks;

(d) to determine the performance of individual items of the 4AT, e.g. how accurate is altered level of alertness alone as a predictor of delirium diagnosis?;

(e) to assess the 4AT total score as a measure of delirium severity;

(f) to estimate the delivery costs of the 4AT and CAM as a function of their diagnostic performance up to 12 weeks as well as modelling longer term resource consequences.

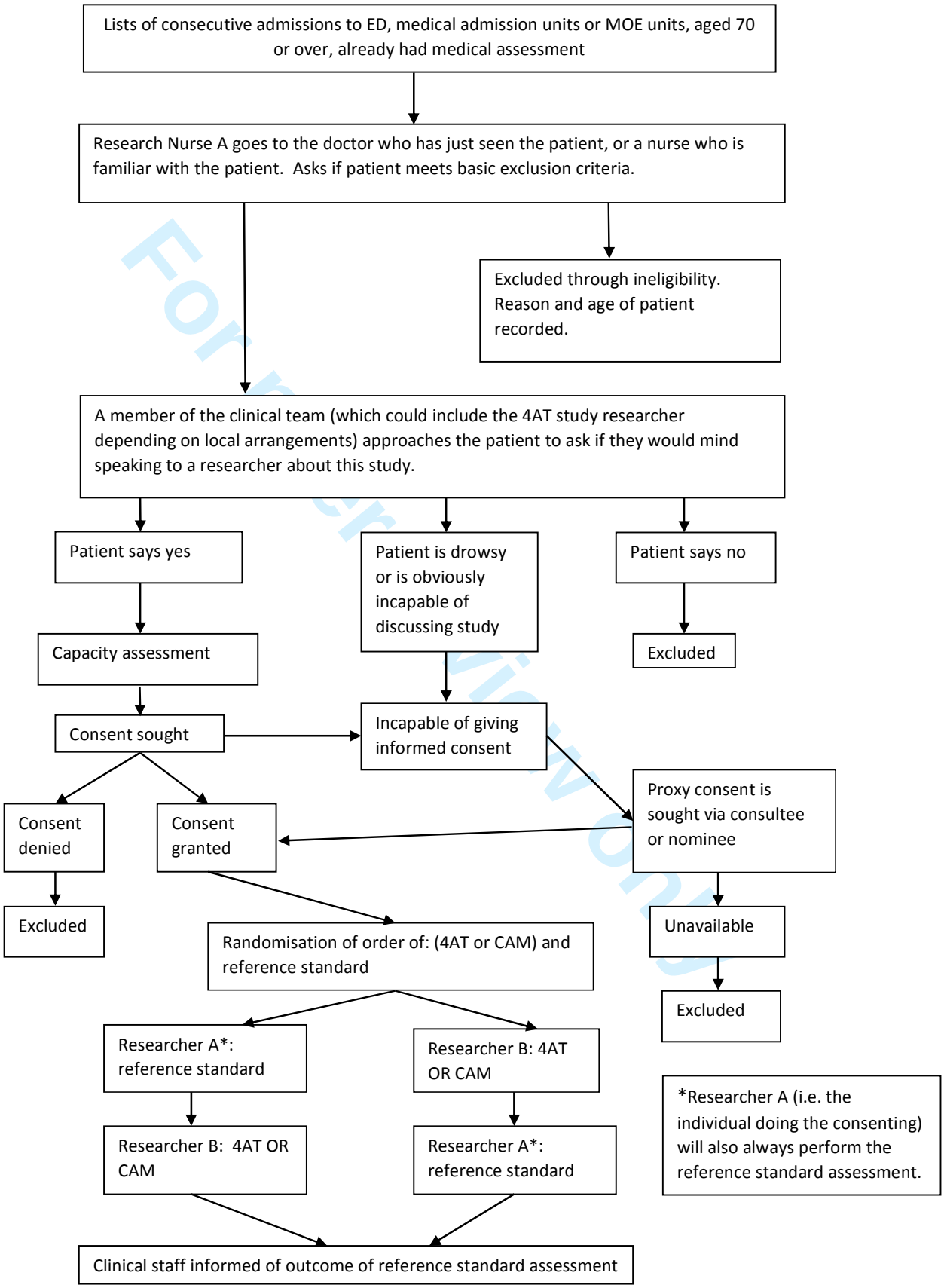
METHODS AND ANALYSIS

Study overview

900 patients aged 70 or over in Emergency Departments or acute general medical wards will be recruited in three sites (Edinburgh, Bradford, Sheffield). Study recruitment commenced on 19th October, 2015. Recruitment will complete in December 2016, with final follow-up data collection and locking of the database in March 2017. The assessments are: (a) a reference standard delirium assessment lasting up to 20 minutes, and (b) either the 4AT or the CAM (lasting up to 10 minutes). The reference standard and 4AT or CAM assessments will take place within a maximum of two hours of each other, with a target interval of 15 minutes. The team will invite an appropriate informant to complete a questionnaire on participant's pre-admission cognitive function. This will be completed within 4 weeks of the patient being recruited to the study assuming an appropriate individual is available.

At 12 weeks the team will also administer a 10 minute resource use questionnaire (face to face in hospitalised patients, or by telephone when possible), and will access each recruited patient's medical records at 12 weeks to ascertain a set of key clinical outcomes including length of stay, institutionalisation, and mortality, as well as to derive further information on resource utilisation. The study flowchart is shown in Figure 1.The study has been registered: International standard randomised controlled trial number (ISRCTN) 53388093. UK Clinical Research Network ID: 19502

Figure 1: study overview flowchart



Inclusion criteria

- Aged 70 or over
- Acutely admitted to the Emergency Department (ED) (within 12 hours of attending) or acute general medical and geriatrics units (within 96 hours of admission to the ward). For ED patients, we will only recruit from those patients who were brought in by ambulance as an emergency or through their general practitioner.

Exclusion criteria

- Acute life-threatening illness requiring time-critical intervention e.g. ST-elevation myocardial infarction; septic shock; severe pulmonary oedema.
- Coma ('Unresponsive' on the AVPU scale[43])
- Unable to communicate in English or severe dysphasia.

Identification of participants

The participant screening strategy in the initial protocol stated that patients will be recruited between 08.00 and 22.00. A list of potentially eligible patients will be generated in batches at the start of each recruitment period and initial eligibility screening will be carried out by clinical staff (including clinical research nurses embedded in the clinical team). Then, in alphabetical order, in each batch consent/agreement from patient (or proxy/consultee) will be sought by a study researcher. Numbers of those (a) initially potentially eligible (b) screened as non-eligible by clinical staff and (c) declining to take part will be recorded.

This recruitment strategy was modified in the last 5 months of recruitment because preliminary analyses suggested that patients at lower risk of delirium (that is, those not requiring capacity assessment) were more likely to be recruited than those with impaired capacity. Thus, to allow for some oversampling of patients at higher risk of delirium a

pragmatic approach was adopted. From the batches of patients identified as in the original strategy, patients considered at higher risk of delirium on clinical grounds (for example, older age, likely to be admitted, higher degree of ongoing acute and chronic illnesses) were approached first, rather than by alphabetical order.

Assessing capacity and obtaining informed consent

Informed consent will be sought by a trained researcher using a combined informal capacity assessment/consent process[44]. Both verbal and written information will be provided about the study, using a style and format suitable for the participant group (i.e. for varying levels of capacity). The researcher will ask the potential participant to recount the study which will be used, with the treating team views, to assess capacity to consent. For participants judged to have capacity, consent will be sought for:

(a) Conducting assessments as specified in the study information sheets

(b) Accessing health records for information relevant to outcomes and health service use

(c) Recording these data in secure study databases

It will be made clear to participants that they are under no obligation to take part, their usual care will not be affected by their decision, and they can withdraw consent without giving a reason. Once participants are enrolled in the study they will be given a sheet with contact details for the research team and instructions on what to do if they wish to withdraw consent or require further information.

Lack of capacity to consent

It is essential that this study recruits patients which reflect the target clinical population. This means that we must recruit patients with delirium in the same proportion as in the clinical population. We will seek consent/agreement from legal proxies, consultees or other legal

representatives. Where the potential participant is deemed to lack capacity to consent, recruitment will proceed under the provisions of the Mental Capacity Act, 2005 in England or Adults with Incapacity (Scotland) Act, 2000. The clinical team will be asked to identify an appropriate personal or nominated consultee, guardian, welfare attorney or nearest relative. Because of differing legal requirements in Scotland and England, the details of the processes in each nation are given in Appendix 1.

Interventions to be measured

Assessments will be carried out by researchers fully trained in background information on delirium, the features of delirium, and each rating scale. Training is carried out using written, video and bedside training until competence in all aspects of the assessments is achieved.

The 4 "A"s Test (4AT)

The 4AT (see www.the4AT.com) comprises 4 items. Item 1 concerns an observational assessment of level of alertness. The next 2 items are brief cognitive tests: the Abbreviated Mental Test – 4 (AMT4) which asks the patient to state their age, their date of birth, the current year, and the place they are in; and attention testing with Months Backwards, in which the patient is asked to state the months of year in reverse order, starting with December. Only items 1-3 are done at the bedside, and the typical duration is under 2 minutes. Item 4 concerns acute change in mental status, a core diagnostic feature of delirium; this information is obtained from the casenotes or the GP letter or from an informant.

Short Confusion Assessment Method (CAM)

The CAM is a diagnostic algorithm in which the tester rates the following four features as positive or negative: 1. Acute Change and Fluctuating Course; 2. Inattention; 3. Disorganised

Thinking; and 4. Altered Level of Consciousness. The CAM scoring process requires that Features 1 and 2 are both positive; if they are positive then Features 3 and 4 are assessed and if one of Features 3 or 4 is positive, then the whole CAM is positive. The tester scores the features by a combination of interview with the patient, cognitive testing (the CAM requires that a cognitive test is performed before the features are scored), examining the casenotes, and seeking informant history if required. Note that the questionnaires used to assess cognition are not specified by the CAM manual, though some suggested tests are provided. Feature 1 is assessed by the same process as Item 4 in the 4AT. Feature 2 is assessed by the tester giving a positive or negative rating to the question, "Did the patient have difficulty focusing attention, for example, being easily distractible, or having difficulty keeping track of what was being said?" Feature 3 is assessed by the tester giving a positive or negative rating to the question, "Was the patient's thinking disorganized or incoherent, such as rambling or irrelevant conversation, unclear or illogical flow of ideas, or unpredictable switching from subject to subject?" Feature 4 is similar to item 1 in the 4AT. In this study, for the pre-CAM cognitive assessment we will use a set of questions covering the cognitive domains represented in the suggested tests in the CAM manual, including Days of the Week Backwards, counting from 20 down to 1, orientation questions, three-word recall, and clockdrawing. All of these questions are used in routine clinical practice at the bedside.

Reference Standard Assessment

Reference standard assessment: This will be centred on the Delirium Rating Scale-Revised-98 (DRSR98)[45], requiring inspection of case notes, speaking to staff who know the patient, or speaking to the patient's relatives or others who know them (with patient consent). As per the instruction manual, the DRS-R98 will be supplemented by short neuropsychological tests of attention and other domains, including Digit Span[27], the Observational Scale for Level

of Arousal[46], the Richmond Agitation Sedation Scale[47] and the DelApp objective attentional assessment[48]. We will also record any formal prior diagnosis of dementia and Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE)[49] scores. The DRS-R98 and supporting tests will be used to inform a binary ascertainment of delirium based on DSM-IV criteria. The final DSM-IV ascertainment of delirium will be based on a standardised process with final verification by the Chief Investigator, blind to the 4AT or CAM results. The panel of supporting tests, and the way the data are coded will be designed such that the performance of the 4AT can also be evaluated against the DSM-5 criteria[2]. The reference standard assessment will take approximately 15-20 minutes.

Ordering of assessments

All patients will undergo a reference standard assessment for delirium by the researcher who conducted the capacity assessment and consenting process. A different researcher will also ask each patient to undergo either the 4AT or the CAM. The reason that the researcher doing the capacity assessment and consenting process must also do the reference standard assessment is that the capacity and consenting process provides information to the tester over and above the normal 4AT or CAM testing process. This is not a concern for the reference standard assessment, which is aimed at providing a thorough assessment so as to optimise diagnostic accuracy. The order of these two assessments ([4AT or CAM assessment] and reference standard assessment) will be randomly allocated immediately after consenting, as will the assignment to either the 4AT or CAM. Each patient will receive the reference standard assessment by the same researcher who did the capacity and consenting process. The 4AT or CAM will be performed by a different researcher. When possible the IQCODE will then be administered to a person who knows the patient well (within 4 weeks of the patient joining the study).

Randomisation procedure

The allocation sequence will be created using computer-generated random numbers. Participants will be randomised in a 1:1 ratio to be assessed using the 4AT or CAM experimental assessment. The order in which they receive the reference standard and experimental assessment will also be randomised in a 1:1 ratio. Randomisation will be stratified by study site with block allocation. The randomised allocations will be concealed until they are assigned as the randomisation system will be web-based and require a personal log-in and password. Once randomisation has been performed neither the researchers nor the participant will be blinded to the allocation as both will be aware of the assessments conducted and the order in which they are performed.

Outcome measurements (what, when, how)

Note that the outcome measurements for the primary study (the reference standard) are performed at almost the same time as the 4AT and CAM. The only subsequent data collection is capturing clinical outcomes at 12 weeks. This will be achieved through searching electronic patient records, telephone calls with participants, or face-to-face-interviews if still in hospital. The information gleaned at the 12 week point will at times be generalised due to participant recall or availability of full electronic records.

1) Primary outcome measure:

Diagnostic accuracy of the 4AT versus the reference standard delirium diagnosis

2) Secondary outcome measures:

(a) 4AT versus CAM in relation to reference standard delirium diagnosis

(b) Performance of 4AT cognitive test items (AMT4 and Months Backwards) in detecting longer-term cognitive impairment as detected by the IQCODE

(c) 4AT total scores as a predictor of the following clinical outcomes as determined at 12 weeks post-test: length of stay, falls, institutionalisation (as assessed by proportion of patients newly admitted to care homes or awaiting care homes at that time) and mortality

(d) Performance of individual items of the 4AT in relation to reference standard delirium diagnosis

(e) We will assess the 4AT total score as a measure of delirium severity.

(f) The primary output from the health economic analysis will be a comparison of the service delivery costs associated with the diagnostic accuracy of alternative (4AT vs. CAM vs. reference standard) triage tools for delirium.

Coding and recording assessments

The experimental assessments of delirium will be the 4AT and the CAM. The 4AT has a total possible score of 12: items (1) and (4) can score 0 or 4; items (2) and (3) can score 0, 1, or 2[12]. 4AT data will be used for the primary objective as a binary outcomes, with 0-3 scores giving a 'no delirium' classification, and 4-12 scores giving a 'delirium' classification; for the secondary objectives, continuous scoring, from 0-12, will be studied as a possible severity indicator, and scores of 1-3 (indicating cognitive impairment but not delirium) can be studied against other assessments of chronic cognitive impairment). The CAM will scored as delirium present or absent according to the algorithm. The 4AT, CAM scoring and reference standard scoring will be recorded on a paper Case Report Form.

Patient resource-use will be derived from medical records, including the 'TrakCare' (InterSystems Corporation, Cambridge, MA, USA) electronic patient record system, where available, as well as via patient or carer self-report. The self-report resource-use questionnaire will include questions regarding inpatient health and social care utilisation with a maximum recall period of 16 weeks. The self-report resource use questionnaire will be developed

specifically for the study for use by patient or proxy respondent using guidance from the Database of Instruments for Resource Use Measurement[51]. Administration of the questionnaire will be conducted at 12 weeks by one of the researchers in the study team, face-to-face where patients are still hospitalised, or via telephone. Data from the questionnaire will be recorded on a paper Case Report Form.

The data on all the Case Report Forms will be transcribed into a secure database by the researchers or a suitably qualified member of the research team. This will be conducted using Edinburgh Clinical Trials Unit Standard Operating Procedures. Quality checking will be performed in 10% of Case Report Forms.

Sample size calculation

450 patients will be randomised to assessment by 4AT and 450 to CAM. We will recruit sufficient patients to account for attrition; though we do not expect significant attrition because the recruitment, consenting and assessment process takes place over a small number of hours, in a single episode. Of the 450 patients within each assessment arm, 15% (67) would be expected to have delirium. The specificity of the triage tool would be estimated based on the 85% (383) without delirium, while the sensitivity would be estimated from the 67 with delirium. Based on analysis using the normal approximation to the binomial distribution, the two-sided 95% confidence interval widths for the specificity and sensitivity would be as shown in the table for a range of levels of diagnostic test performance.

Table 2 Precision of specificity, sensitivity estimation

Parameter	True level of parameter	95% confidence interval width
Specificity	0.5	±0.050
Specificity	0.7	±0.046
Specificity	0.9	±0.030
Sensitivity	0.5	±0.120
Sensitivity	0.7	±0.110

Sensitivity	0.9	±0.072

It will therefore be possible to estimate the specificity precisely and the sensitivity with moderate precision. The precision in estimating negative predictive value would be expected to be similar to that for specificity; for positive predictive value it would be expected to be similar to that for sensitivity. For the secondary objective of comparing 4AT and CAM, based on analysis by continuity corrected chi-squared test, we have 83% power to detect a difference in specificity of 0.1, assuming a null hypothesis of specificity=0.70 for both tests and a two-sided 5% significance level. The corresponding difference detectable for sensitivity (null hypothesis sensitivity=0.7) would be 0.224 with 80% power.

Data analysis plan

The detailed statistical analysis plan will be agreed prior to database lock and will be prepared by individuals blinded to the randomised allocations.

Primary objective:

(a) 4AT vs reference standard: the diagnostic accuracy of 4AT versus the reference standard will be assessed using positive and negative predictive values, sensitivity and specificity. The exact binomial 95% confidence interval will be reported for each measure. A receiver operating characteristic (ROC) curve will be constructed to verify that the proposed cut point of greater than 3 on the 4AT score is appropriate. The area under the ROC curve and its 95% confidence interval will be reported.

Secondary objectives:

(a) 4AT vs CAM: differences in each of sensitivity, specificity, positive and negative predictive values between 4AT and CAM will be tested by Fisher's exact test and quantified by the difference in the two proportions (4AT-CAM) and its 95% confidence interval. To aid comparison of 4AT and CAM, the overall performance of each will also be summarised

using Youden's Index (sensitivity minus false positive rate) and the odds ratio of sensitivity to specificity.

(b) Performance of the 4AT cognitive screening items: is the 4AT an adequately sensitive tool for detecting general cognitive impairment as judged against a documented history of dementia and/or the Informant Questionnaire for Cognitive Decline in the Elderly? This objective will be addressed using the same methods as for the primary objective.

(c) 4AT vs clinical outcomes: as assessment of criterion validity, we will assess the performance of the 4AT in predicting length of stay, institutionalisation, and mortality, up to 12 weeks. Descriptive statistics of clinical outcomes will be presented for the groups with and without 4AT scores above the cut point of 3. The relationship between 4AT and each of mortality and institutionalisation will be analysed via logistic regression modelling; Kaplan-Meier curves and the Cox proportional hazards model will be used to assess 4AT as a predictor of hospital length of stay. The logistic regression and Cox models will adjust for age, gender and presence of dementia.

(d) Individual items: we will conduct analyses examining performance of individual items of the 4AT, e.g. is altered level of alertness alone a good predictor of delirium diagnosis? (Methods as per primary objective);

(e) Delirium severity: we will assess the 4AT total score as a measure of delirium severity by calculating the Spearman correlation between 4AT and DRS-R98 scores and its 95% confidence interval.

Full details of the proposed statistical analyses for the primary objective and secondary objectives (a) to (e) will be documented in a statistical analysis plan (SAP) which will include details of methods for calculating derived variables, any sensitivity and subgroup analyses, and approaches to testing the assumptions in the statistical analyses. The SAP will outline the plan for validation of the statistical analysis.

Individuals with missing data for the reference diagnostic test will be removed from formal statistical analysis. Where any items of the CAM or 4AT were not able to be assessed, an overall delirium diagnosis will still be derived where possible based on the items which have been recorded. There will be no other imputation of missing delirium diagnoses.

(f) Delivery costs of the triage tools: We will estimate the delivery costs and subsequent resource consequences associated with the triage tools as a function of sensitivity and specificity from the perspective of the UK National Health Service. Potential resource consequences may include additional diagnostic procedures (e.g. more detailed cognitive screening and brain imaging), altered management as well as re-admissions. Healthcare resource use will be derived from medical records, including the 'TrakCare' (InterSystems Corporation, Cambridge, MA, USA) electronic patient record system, where available, as well as via patient or carer self-report. Monetary values will be attached to resource use, training and labour costs as well as the indirect costs of delivering each diagnostic tool using standard NHS pay and price estimates. Generalised linear models will be used to analyse 12 week cumulative costs which will inform longer-term resource consequences within a decision analytic model.

ETHICS

This study was granted ethical approval in Scotland REC 15/SS/0071 and England REC 15/YH/0317.

Study oversight

Study oversight is through the Trial Steering Committee, which will meet every four months during the study. The Trial Steering Committee comprises two independent lay representatives, three independent experts (one of whom is the Chair of the Committee), the PI, the study statistician, and representatives from the Edinburgh Clinical Trials Unit.

Data Protection

Data will be collected and handled in line with sponsor and Edinburgh Clinical Trials Unit Standard Operating Procedures and NHS Trust policies. All electronic data will be link-anonymised.

DISCUSSION

This study was designed to validate the 4AT against a reference standard assessment, as well as compare it to another widely used test for assessment of delirium. Since the initial study design, the 4AT has been widely adopted nationally and internationally. The 4AT has been incorporated into routine practice in multiple international centres, both in paper and electronic format, many centres reporting 10,000 uses of the tool. The website www.the4AT.com has had an increasing degree of traffic, and the 4AT has been translated into multiple languages. The 4AT is also included in several national guidelines and position statements internationally as a recommended tool and it has been validated in other studies[38-42] but it is still essential that it is further tested in a large study. It is also essential to consider how well it identifies other types of cognitive impairment, relates to future outcomes, and its health economic impact.

In the initial design and implementation of the study, the main challenging aspects have been: 1) considering both Scottish and English legal and ethical framework to ensure that patients without capacity are included. Ethical approval for inclusion of these patients was granted though recruitment of this patient group proved difficult from the outset for several reasons. Firstly, the narrow boundaries for the screening and identification strategy. This was addressed in a subsequent protocol amendment to aim for oversampling of patients at risk of developing delirium. Secondly, the availability of an appropriate individual to provide consent on behalf of the participant (i.e. a personal or nominated consultee, guardian, welfare

attorney or nearest relative). Thirdly, a reluctance to consent due to perceived burden on participant. Persuading relatives of the value, importance and necessity of research even in clinically unwell patients demands a particular skillset from researchers and involved perseverance and excellent communication in order to achieve recruitment targets.

2) Recruitment and training of staff, with some staff moving to different posts, and new staff being recruited and requiring training; in each case detailed training supported by reading materials and practice sessions was provided.

CONCLUSION

The 4AT study aims to assess the validity of this rapid delirium screening tool that can be administered by a range of professional level healthcare staff to patients with sensory or functional impairments in a busy clinical environment, which also incorporates general cognitive assessment. We will also assess the later functional outcomes of people with and without delirium, and the health economic implications. The overall aim is to improve detection, and therefore management and outcomes, of this important and devastating condition.

REFERENCE LIST

1. American Psychiatric Association. Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* 4th ed. Washington, DC: American Psychiatric Association; 2000.
2. American Psychiatric Association, American Psychiatric Association. *Diagnostic and statistical manual of mental disorders : DSM-V.* 5th ed. Washington, DC: American Psychiatric Association; 2013

3. Siddiqi N, House AO, Holmes JD. Occurrence and outcome of delirium in medical in-patients: a systematic literature review. *Age and Ageing.* 2006;35(4):350-364.

4. Goldberg SE, Whittamore KH, Harwood RH, et al. The prevalence of mental health problems among older adults admitted as an emergency to a general hospital. *Age Ageing.* 2012;41(1):80-86.

5. Collins N, Blanchard MR, Tookman A, et al.. Detection of delirium in the acute hospital. *Age and Ageing.* 2010;39(1):131-135.

6. Delirium: diagnosis, prevention and management. 2010; www.nice.org.uk/CG103.

7. Davis DH, Muniz Terrera G, Keage H, et al. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. *Brain.* 2012;135(Pt 9):2809-2816.

8. MacLullich AM, Beaglehole A, Hall RJ, et al. Delirium and long-term cognitive impairment. *Int Rev Psychiatry.* 2009;21(1):30-42.

9. Witlox J, Eurelings LSM, de Jonghe JFM, et al. Delirium in Elderly Patients and the Risk of Postdischarge Mortality, Institutionalization, and Dementia: A Meta-analysis. *JAMA.* 2010;304(4):443-451.

10. 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 e241.

11. Anderson D. *Who Cares Wins.* London: Royal College of Psychiatrists; 2005.

12. Fong TG, Jones RN, Shi P, et al. Delirium accelerates cognitive decline in Alzheimer disease. *Neurology.* 2009;72(18):1570-1575.

13. Leslie DL, Marcantonio ER, Zhang Y, et al.. One-year health care costs associated with delirium in the elderly population. *Archives of Internal Medicine.* 2008;168(1):27-32.

14. Kakuma R, du Fort GG, Arsenault L, et al. Delirium in older emergency department patients discharged home: Effect on survival. *Journal of the American Geriatrics Society.* 2003;51:443-450.

15. Partridge JS, Martin FC, Harari D, et al. The delirium experience: what is the effect on patients, relatives and staff and what can be done to modify this? *Int J Geriatr Psychiatry.* 2012. 2013 28(8):804-12. doi: 10.1002/gps.3900 [Epub 2012 Oct 30]

16. Adunsky A, Levy R, Mizrahi E, et al. Exposure to opioid analgesia in cognitively impaired and delirious elderly hip fracture patients. *Arch Gerontol Geriatr.* 2002;35(3):245-251.

17. Department of Health Using the Commissioning for Quality and Innovation (CQUIN) payment framework: Guidance on new national goals for 2012-13. Department of Health, 2012.

18. Elie M, Rousseau F, Cole M, et al. Prevalence and detection of delirium in elderly emergency department patients. *Canadian Medical Association Journal.* 2000;163:977-981.

19. 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.

20. Davis D, MacLullich A. Understanding barriers to delirium care: a multicentre survey of knowledge and attitudes amongst UK junior doctors. *Age Ageing.* 2009;38(5):559-563.

21. Teodorczuk A, Reynish E, Milisen K. Improving recognition of delirium in clinical practice: a call for action. *BMC Geriatr.* 2012;12:55.

22. Kennelly SP, Morley D, Coughlan T, et al. Knowledge, skills and attitudes of doctors towards assessing cognition in older patients in the emergency department. *Postgrad Med J.* 2012.

23. Steis MR, Fick DM. Delirium superimposed on dementia: accuracy of nurse documentation. *J Gerontol Nurs.* 2012;38(1):32-42.

24. Steis MR, Fick DM. Are nurses recognizing delirium? A systematic review. *J Gerontol Nurs.*2008;34(9):40-48.

25. Inouye SK, van Dyck CH, Alessi CA, et al. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med.* 1990;113(12):941-8.

26. Borson S, Scanlan J, Brush M, et al. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry.* 2000;15(11):1021-1027.

27. Simon SE, Bergmann MA, Jones RN, et al. Reliability of a structured assessment for nonclinicians to detect delirium among new admissions to postacute care. *Journal of the American Medical Directors Association.* 2006;7(7):412-415.

28. Chester JG, Beth Harrington M, Rudolph JL. Serial administration of a modified Richmond Agitation and Sedation Scale for delirium screening. *J Hosp Med.* 2012;7(5):450-453.

29. Yates C, Stanley N, Cerejeira JM, et al. Screening instruments for delirium in older people with an acute medical illness. *Age Ageing.* 2009;38(2):235-237.

30. Harwood DM, Hope T, Jacoby R. Cognitive impairment in medical inpatients. I: Screening for dementia--is history better than mental state? *Age Ageing.* 1997;26(1):31-35.

31. Eeles EM, White S, Bayer T. Inability to complete cognitive assessment is a marker of adverse risk. *Age Ageing.* 2009;38(5):633-634; author reply 634-635.

32. Hall RJ, Meagher DJ, MacLullich AMJ. Delirium detection and monitoring outside the ICU. *Best Pract Res Clin Anaesthesiol.* 2012;26(3):367-383.

33. Kean J, Ryan K. Delirium detection in clinical practice and research: critique of current tools and suggestions for future development. *J Psychosom Res.* 2008;65(3):255-259.

34. Adamis D, Sharma N, Whelan PJ, et al. Delirium scales: A review of current evidence. *Aging Ment Health.* 2010;14(5):543-555.

35. Wong CL, Holroyd-Leduc J, Simel DL, et al. Does this patient have delirium?: value of bedside instruments. *JAMA.* 2010;304(7):779-786.

36. Ringdal GI, Ringdal K, Juliebo V, et al. Using the Mini-Mental State Examination to screen for delirium in elderly patients with hip fracture. *Dement Geriatr Cogn Disord.* 2011;32(6):394-400.

37. O'Keeffe ST, Mulkerrin EC, et al. Use of serial Mini-Mental State Examinations to diagnose and monitor delirium in elderly hospital patients. *J Am Geriatr Soc.*2005;53(5):867-870.

38. Bellelli G, Morandi A, Davis DH, et al. Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age and ageing.* 2014;43(4):496-502.

39. Lees R, Corbet S, Johnston C, et al. Test accuracy of short screening tests for diagnosis of delirium or cognitive impairment in an acute stroke unit setting. *Stroke*. 2013;44:3078-83

40. Kuladee S, Prachason T. Development and validation of the Thai version of the 4 'A's Test for delirium screening in hospitalized elderly patients with acute medical illnesses. *Neuropsychiatr Dis Treat.* 2016; 12:437-43

41. De J, Wand AP, Smerdely PI, et al. Validating the 4A's test in screening for delirium in a culturally diverse geriatric inpatient population. *Int J Geriatr Psychiatry*. 2016 Oct 20. [Epub ahead of print]

42. Hendry K, Quinn TJ, Evans J, et al. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. *Age and Ageing.* 2016 Aug 8. [Epub ahead of print]

43. McNarry AF, Goldhill DR. Simple bedside assessment of level of consciousness: comparison of two simple assessment scales with the Glasgow Coma scale. *Anaesthesia.* 2004;59(1):34-37.

44. Adamis D, Martin FC, Treloar A, Macdonald AJD. Capacity, consent, and selection bias in a study of delirium. *Journal of Medical Ethics.* 2005;31(3):137-143.

45. Trzepacz PT, Mittal D, Torres R, et al. Validation of the Delirium Rating Scale-revised-98: Comparison with the Delirium Rating Scale and the Cognitive Test for Delirium. *Journal of Neuropsychiatry and Clinical Neurosciences.* 2001;13(2):229-242.

46. Tieges Z, McGrath A, Hall RJ, et al. Abnormal level of arousal as a predictor of delirium and inattention: an exploratory study. *The American journal of geriatric psychiatry:official journal of the American Association for Geriatric Psychiatry.* 2013;21(12):1244-1253.

47. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med.* 2002;166(10):1338-1344.

48. Tieges ZS, Scott, K, Suchorab K et al. Development of a smartphone application for the objective detection of attentional deficits in delirium *International Psychogeriatrics.* 2015 Aug;27(8):1251-62. doi: 10.1017/S1041610215000186. [Epub 2015 Mar 6]

49. Jorm AF. A Short-Form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) - Development and Cross-Validation. *Psychol. Med.* 1994;24(1):145-153. 43.

50. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc.* 1975;23(10):433-441.

51. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials:a systematic review of studies funded by the UK Health Technology Assessment program.*Value Health.* 2010;13(8):867-872.

AUTHORS' CONTRIBUTIONS: All authors were involved in the design of the protocol for the study. AMJM initiated the conception of the study. AM, JSteven, PLB were involved in protocol design and acquisition of data; CJW initiated the design of the statistical analysis plan. JSmith initiated the design of the economic analysis. SDS and AMJM drafted the paper and all authors revised it critically for important intellectual content. All authors have approved the final approval of the version to be published, and agree to be accountable for all aspects of the work. AMJM is the guarantor.

ACKNOWLEDGMENTS

Thanks to the Trial Steering Committee.

FUNDING

This work was funded by National Institute of Health Research Technology Assessment Programme (NIHR DTA) grant number 11/143/01 (PI: AMJM; CO-I: SG, JSmith, AA, MG, SDS, TR, NS, CF, JH, CJW, AG). JB and CW were supporrted in this work by NHS Lothian via the Edinburgh Clinical Trials Unit. AA has received personal fees from Abbott Diagnostics, outside the submitted work. SG is funded by NIHR to undertake his role as NIHR HTA Clinical Evaluation and Trials Board Chair. SDS is funded by NRS (NHS Research Scotland) to undertake her role as NRS Ageing Specialty Group Lead (Scotland).

Appendix 1: Detailed consent procedures for incapacity in Scotland and England

Scotland

An appropriate legal proxy (that is, a guardian, welfare attorney, nearest relative, but not a member of the clinical team) will be approached by a member of the clinical team (potentially including researchers who are part of the clinical team) to be asked if they would be willing to consider hearing about a study involving the patient, and to potentially give consent on their behalf. If the proxy assents to hearing more about the study, the study team member responsible for consent will provide the proxy with information about: why they are being approached; the role of a proxy, explanation that acting as a proxy is voluntary; details of the study (as would be given to a participant with capacity). The proxy will be asked for advice on whether the participant should take part in the study and what, in their opinion, the participant's views and feelings would have been on taking part in the project had they retained capacity. Consent forms will be signed when the proxy is physically present. If no appropriate legal proxy can be identified within 96 hours, the patient will not be recruited to the study. This is because in Scotland patients with incapacity cannot be included in studies of nonmedicinal treatments unless there is a guardian, welfare attorney or nearest relative available to give consent.

England

If the patient is incapacitated at study entry then a personal consultee (usually a friend or relative) will be consulted and their opinion sought. The approach used will be similar to that detailed in the previous section when consulting legal proxies in Scotland. If the personal consultee agrees that their friend/relative can enter the study then we would ask them to sign a declaration form. If a personal consultee is not available for consultation then the treating doctor (who will be independent of the research team and of appropriate seniority),

will be asked to act as the nominated consultee and advise on inclusion in the study. If agreement is given it will be recorded on the declaration form.

All Trial Participants (England and Scotland)

All patients who lack mental capacity at the time of enrolment will be approached for consent to remain in the trial at the earliest opportunity once they regain capacity. Research staff have planned contact with study patients on the day of enrolment and only on one further occasion at 12 weeks when they will collect questionnaire data from the patient. If research staff become aware the patient has regained capacity while in hospital then written consent from the patient will be sought at this time. It is likely that in many cases the first contact by the research team will be at 12 weeks either in person or by phone.

The patient will be given the opportunity to either withdraw or remain in the study at this time. If the patient chooses to withdraw from the study they will be given the option of allowing/not allowing the use of data already collected. A patient information sheet will be posted to participants who wish to remain on the study and other patients on request.

If patients have not regained capacity at 12 weeks they will remain in the study based on the advice of the consultee or legal proxy.

BMJ Open

Protocol for validation of the 4AT, a rapid screening tool for delirium: a multicentre prospective diagnostic test accuracy study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-015572.R1
Article Type:	Protocol
Date Submitted by the Author:	16-Mar-2017
Complete List of Authors:	Shenkin, Susan; University of Edinburgh, Geriatric Medicine Fox, Chris; University of East Anglia Faculty of Medicine and Health Sciences, Old Age Psychiatry Godfrey, Mary; University of Leeds, Health and Social Care, Institute of Health Sciences Siddiqi, Najma; University of York, Psychiatry, Hull York Medical School, York and Bradford District Care NHS Foundation Trust, Bradford, UK Goodacre, Steve; University of Sheffield, Emergency Medicine Young, John; University of Leeds Anand, Atul; University of Edinburgh, Geriatric Medicine and Cardiovascular Sciences Gray, Alasdair; NHS Lothian, Emergency Medicine Research Group (EMERGE) Smith, Joel; University of Oxford, Health Economics Research Centre, Nuffield Department of Population Health Ryan, Tracy; NHS Lothian, Old Age Psychiatry Hanley, Janet; Edinburgh Napier University, Health and Social Care MacRaild, Allan; NHS Lothian, Emergency Medicine Research Group (EMERGE) Steven, Jill; NHS Lothian, Emergency Medicine Research Group (EMERGE) Black, Polly; NHS Lothian, Emergency Medicine Research Group (EMERGE) Boyd, Julia; University of Edinburgh, Edinburgh Clinical Trials Unit Weir, Christopher; University of Edinburgh, Edinburgh Clinical Trials Unit; University of Edinburgh MacLullich, Alasdair; University of Edinburgh, Geriatric Medicine Unit
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Geriatric medicine, Neurology, Mental health
Keywords:	Delirium & cognitive disorders < PSYCHIATRY, diagnosis, screening

Protocol for validation of the 4AT, a rapid screening tool for delirium: a multicentre prospective diagnostic test accuracy study

Corresponding author:

Prof Alasdair MJ MacLullich

Edinburgh Delirium Research Group, Geriatric Medicine

Room S1642

Department of Clinical and Surgical Sciences

The University of Edinburgh, Royal Infirmary of Edinburgh

51 Little France Crescent, Edinburgh, EH16 4SB

Tel: +44(0)131 242 6481

Email: a.maclullich@ed.ac.uk

Keywords: delirium, confusion, hospital, diagnostic test accuracy

Word count (excl title page, refs, tables, figure): 5,464

Authors:

Susan D Shenkin, Geriatric Medicine, University of Edinburgh, Edinburgh, UK

Christopher Fox, Old Age Psychiatry, University of East Anglia, Norfolk UK

Mary Godfrey, Health and Social Care, Institute of Health Sciences, University of Leeds, Leeds, UK

Najma Siddiqi, Psychiatry, University of York, York, Hull York Medical School, York and Bradford District Care NHS Foundation Trust, Bradford, UK

Steve Goodacre, Emergency Medicine, University of Sheffield, Sheffield, UK

John Young, Elderly Care and Rehabilitation, University of Leeds, Leeds, UK

Atul Anand, Cardiovascular Sciences and Geriatric Medicine, University of Edinburgh, Edinburgh, UK

Alasdair Gray, Professor of Emergency Medicine, Emergency Medicine Research Group (EMERGE), Department of Emergency Medicine, NHS Lothian, Edinburgh, UK

Joel Smith, Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Tracy Ryan, Old Age Liaison Psychiatry, NHS Lothian, Edinburgh, UK

Janet Hanley, Health and Social Care, Edinburgh Napier University, Edinburgh, UK

Allan MacRaild, Emergency Medicine Research Group Edinburgh (EMERGE), NHS Lothian, Edinburgh, UK

Jill Steven, Emergency Medicine Research Group Edinburgh (EMERGE), NHS Lothian, Edinburgh, UK

Polly L Black, Emergency Medicine Research Group Edinburgh (EMERGE), NHS Lothian, Edinburgh, UK

Julia Boyd, Edinburgh Clinical Trials Unit, University of Edinburgh, Edinburgh, UK

Christopher J Weir, Edinburgh Clinical Trials Unit, University of Edinburgh, Edinburgh, UK; Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

Alasdair MJ MacLullich, Geriatric Medicine, University of Edinburgh, Edinburgh UK

ABSTRACT

Introduction: Delirium is a severe neuropsychiatric syndrome of rapid onset, commonly precipitated by acute illness. It is common in older people in the emergency department and acute hospital, but greatly under-recognised in these and other settings. Delirium and other forms of cognitive impairment, particularly dementia, commonly co-exist. There is a need for a rapid delirium screening tool that can be administered by a range of professional level healthcare staff to patients with sensory or functional impairments in a busy clinical environment, which also incorporates general cognitive assessment. We developed the '4As Test' (4AT) for this purpose.

This study's primary objective is to validate the 4AT against a reference standard. Secondary objectives include (a) comparing the 4AT with another widely used test (the Confusion Assessment Method (CAM)); b) determining if the 4AT is sensitive for general cognitive impairment; c) assessing if 4AT scores predict outcomes; including d) a health economic analysis.

Methods and analysis: 900 patients aged 70 or over in Emergency Departments or acute general medical wards will be recruited in three sites (Edinburgh, Bradford, Sheffield) over 18 months. Each patient will undergo a reference standard delirium assessment and will be randomised to assessment with either the 4AT or the CAM. At 12 weeks outcomes (length of stay, institutionalisation, and mortality) and resource utilisation will be collected by a questionnaire, and via the electronic patient record.

Ethics and dissemination: Ethical approval was granted in Scotland and England. The study involves administering tests commonly used in clinical practice. The main ethical issues are the essential recruitment of people without capacity. Dissemination is planned via publication in high impact journals, presentation at conferences, social media and the website www.the4AT.com.

Registration details: International standard randomised controlled trial number (ISRCTN) 53388093. UK Clinical Research Network ID: 19502

Strengths and limitations of this study:

- The study protocol involved seeking a representative sample of older acute medical patients in the Emergency Department and acute medical wards. A detailed, structured reference standard with explicit and reproducible methods is used to assess the features of delirium and reach a diagnosis.
- Two different rating scales, the 4 "A"s Test (4AT) and the Confusion Assessment Method are being evaluated in similar groups of patients.
- Reference standard and index assessments were performed blinded to each other.
- A limitation of the study is that participants or legal proxies were required to give consent and thus the sample was selected.

INTRODUCTION

Background

Delirium is a severe and distressing neuropsychiatric syndrome which is characterised by acute deterioration in attention and other mental functions. The diagnostic criteria are, in summary: a disturbance of consciousness (that is, reduced ability to focus, sustain or shift attention), and a change in cognition. The mental status deterioration develops over short periods of time (usually hours to days) and it tends to fluctuate[1,2]. Delirium is commonly precipitated by acute illness, trauma, or the side-effects of medications. The presence of a 'medical condition' is part of the Diagnostic and Statistical Manual for Mental Disorders, 4th and 5th Edition (DSM-IV, DSM-5) criteria. Delirium is extremely common: it affects at least 15% of patients in acute hospitals, and is more common in older people[3-5]. It is independently associated with many poor outcomes[6-10]. Delirium is also a marker of current dementia[6,11] and is associated with acceleration of existing dementia[12]. In older patients without dementia, an episode of delirium strongly predicts future dementia risk[7,13]. The economic burden of delirium derived from 2008 US data estimates the one-year health care costs to be $38-$152 billion[13] but to date there are no exact investigations of the costs associated with delirium.

Detection of delirium is essential because it indicates acute systemic or central nervous system illness, physiological disturbance and drug intoxication or withdrawal. Failure to detect delirium in the acute setting is associated with worse outcomes[14]. Specific management of delirium is of obvious and immediate benefit to patients in many clinical situations, e.g. in reversing opioid toxicity, treatment of peripheral infections which have

presented with delirium, alleviating distress caused by delusions and hallucinations[15], and in prompting more thorough assessment of symptoms[16].

More broadly, detecting cognitive impairment in general (delirium, dementia, depression, learning disability, etc.) is a prerequisite for high quality care because of the multiple immediate implications of cognitive impairment for patients and staff, including: ensuring adequate communication with patients and their families, doing careful assessment of capacity to provide consent for clinical procedures, avoiding giving treatments contrary to the law because of lack of consent, alleviating distress more readily, avoiding unnecessary bed transfers, and prompting delirium prevention including a detailed drugs review. Detection of dementia has recently been highlighted in the Dementia Commissioning for Quality and Innovation framework in operation in NHS England[17].

In general medical and Emergency Department settings delirium is grossly under-detected: at least two-thirds of cases are missed[5,18,19]. It is unclear why detection rates are so low. Evidence from surveys and workshops have raised several possibilities, including general ignorance about delirium, lack of awareness of its importance, uncertainty about discriminating delirium from dementia, and lack of time for assessment in the acute setting[20-24]. The lack of a very rapid, simple, and validated screening tool is a major factor in the under-detection of delirium.

Many delirium assessment instruments have been developed that operationalise the standard diagnostic criteria for delirium, but these have largely remained research tools. The most commonly advocated screening tool for use in routine clinical care, the short Confusion Assessment Method (CAM)[25], has satisfactory sensitivity and specificity in trained hands

but takes around 10 minutes to complete because it requires a cognitive assessment like the Modified Mini-Cog[26,27] to be done first. The CAM also requires the rater to make subjective judgement of mental status. Subjective judgements are less reliable, often more time-consuming, and more difficult for staff (particularly non-specialists) than simple objective measures with clearly-defined cut-points[27].

The problem of some patients being 'untestable' is likely to be another important factor in delirium under-detection: many patients in acute settings are too unwell, sleepy, or agitated to undergo cognitive testing or even interview[28-31]. Most screening tools do not make explicit how these patients should be classified. The result is that mental status assessments are often simply left uncompleted in most 'untestable' patients, and no diagnosis, and often no specific treatment, is applied. This lack of a diagnosis can be harmful[14].

Finally, given the time pressures in acute settings, it is challenging to implement a separate delirium screening instrument in addition to any existing general cognitive screening instruments. The lack of a combined instrument allowing screening for both general cognitive impairment and delirium may therefore contribute to the lack of specific delirium detection. Early diagnosis of delirium using evidence-based diagnostic tools offers a means for improved outcomes and more efficient resource allocation decisions.

Rationale for the Study

Given the multiple constraints of the acute environment, the range of staff that might be expected to screen for delirium, the common co-existence of delirium and dementia, and the heterogeneity of patients, we determined the requirements for a screening tool (Table 1).

Table 1: Requirements for a screening tool for delirium for use in the acute hospital environment

Short (less than 2 minutes)
Easy to learn
Easy to administer and score
Can be used by professional-level healthcare staff from a variety of disciplines
Allows scoring of patients who are too drowsy or agitated to undergo cognitive testing or clinical interview
Takes account of informant history
Can be administered through written questions to people with severe hearing impairment
Can be administered to patients with visual impairments
Does not require subjective judgements based on interview
Combines delirium screening with general cognitive screening
Does not need a quiet environment for administration
Does not require physical responses such as drawing figures or clocks

There are multiple instruments for delirium screening, diagnosis, severity assessment, and monitoring[32-35]. Before deciding to design a new screening tool, we therefore examined each of the available tools against the above criteria, focusing on screening tools such as the CAM. We also searched the literature systematically, including conference proceedings, books, and book chapters, for any newly-published tools as well as to examine the study data for each tool. Most scales were excluded on grounds of duration alone. The remaining scales lacked features such as general cognitive screening, and other important features. We thus found that, in late 2010, no existing tool fulfilled the above requirements, and because of this

we decided to design a new test. This conclusion was supported by the NICE Guidelines on Delirium[6] which emphasised the need for research on a screening tool for delirium suitable for routine use.

The subsequent design process involved scrutiny of each of the nearly 30 published delirium assessment tools, evaluating the performance of each, including subtests, in published studies and, in most cases, through direct clinical or research experience of their use. Because we had decided to incorporate general cognitive screening into the new instrument, to avoid the need to have separate instruments for cognitive screening and delirium screening, we also reviewed the broader literature on brief tests for general cognitive impairment (including dementia). In the context of designing a screening tool for the acute hospital, it is important to note that delirium generally causes cognitive impairment detectable on the kinds of tests used for dementia screening[36,37]. Therefore, abnormal test results may indicate delirium and/or dementia (as well as other causes of cognitive impairment, such as learning disability).
It is clinically essential to know if any such impairment is acute, that is, delirium, but also important to identify underlying general (acute or chronic) cognitive impairment. A tool designed exclusively to detect cognitive impairment will not lead to delirium detection without another step, and a tool designed only to detect delirium may miss general cognitive impairment. In this light, we decided that the 4AT should include cognitive screening sensitive to general cognitive impairment, but also including items on altered level of alertness and change in mental status, both of which are strong indicators of delirium.
The first version of the 4AT was drafted and tested informally by colleagues, changes were made based on feedback, and updated versions tested again. After several iterations involving 20 doctors and nurses of varying levels of experience, the final version was produced. An initial audit in 30 inpatients comparing clinical use of 4AT with independent reference

standard DSM-IV assessment found 100% sensitivity (CI 69-100%) and 90% specificity (CI 68-99%). A subsequent validation study in Italy involving 234 consecutively recruited older hospitalised patients found that the 4AT had a sensitivity of 89.7% and specificity 84.1% for delirium[38]. The area under the receiver operating characteristic curves for delirium diagnosis was 0.93. Since the 4AT was launched, locally and through the www.the4AT.com website, it has been adopted in clinical units in several centres in the UK and internationally.

Thus, in 2014 there was encouraging evidence that the 4AT has value as a tool for delirium detection in routine practice. This evidence came from several sources: one published study, audits in several sites, informal feedback, adoption in clinical practice by several clinical units globally, and a recent web based survey focused specifically on 4AT provided evidence supporting its use. Since this study was designed, other validation studies have been published, with favourable results, however these included specific clinical populations (e.g. stroke[39]), languages (Thai[40],) had relatively small numbers[41] or validated assessments against clinical assessment rather than research reference standard assessment[42]. Therefore, a formal, large validation study is necessary to provide definitive evidence of the diagnostic accuracy of the 4AT.

Comparison with the CAM is also of value, because the CAM is in use in some clinical units and thus information on how the 4AT performs in relation to the CAM will help clinicians decide which tool is suitable for their particular context. Further information on how the 4AT performs as a cognitive screening tool, its ability to predict outcomes, and how each item of the 4AT contributes to its diagnostic accuracy will also provide important guidance to clinicians. Finally, understanding the economic costs and benefits of using the 4AT and the CAM will help providers in service pathway decisions.

Study objectives

The **primary objective** of the study is to determine the diagnostic accuracy of the 4AT for delirium detection versus the reference standard of a DSM-IV diagnosis.

The **secondary objectives** are:

(a) to compare performance of the 4AT and the Confusion Assessment Method (CAM);

(b) to determine if the 4AT is an adequately sensitive tool for detecting general cognitive impairment as judged against a documented history of dementia and/or the Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE);

(c) to determine if 4AT scores predict important outcomes such as length of stay, institutionalisation, and mortality, up to 12 weeks;

(d) to determine the performance of individual items of the 4AT, e.g. how accurate is altered level of alertness alone as a predictor of delirium diagnosis?;

(e) to assess the 4AT total score as a measure of delirium severity;

(f) to estimate the delivery costs of the 4AT and CAM as a function of their diagnostic performance up to 12 weeks as well as modelling longer term resource consequences.

METHODS AND ANALYSIS

Study overview

900 patients aged 70 or over in Emergency Departments or acute general medical wards will be recruited in three sites (Edinburgh, Bradford, Sheffield). Study recruitment commenced on 19th October, 2015. Recruitment will complete in December 2016, with final follow-up data collection and locking of the database in March 2017. The assessments are: (a) a reference standard delirium assessment lasting up to 20 minutes, and (b) either the 4AT or the CAM (lasting up to 10 minutes). The reference standard and 4AT or CAM assessments will take place within a maximum of two hours of each other, with a target interval of 15 minutes. The results of the reference standard assessment were recorded in the casenotes and communicated to the clinical team after the index assessments had been completed and recorded. The team will invite an appropriate informant to complete a questionnaire on participant's pre-admission cognitive function. This will be completed within 4 weeks of the patient being recruited to the study assuming an appropriate individual is available.

At 12 weeks the team will also administer a 10 minute resource use questionnaire (face to face in hospitalised patients, or by telephone when possible), and will access each recruited patient's medical records at 12 weeks to ascertain a set of key clinical outcomes including length of stay, institutionalisation, and mortality, as well as to derive further information on resource utilisation. The study flowchart is shown in Figure 1.The study has been registered: International standard randomised controlled trial number (ISRCTN) 53388093. UK Clinical Research Network ID: 19502

Inclusion criteria

- Aged 70 or over
- Acutely admitted to the Emergency Department (ED) (within 12 hours of attending) or acute general medical and geriatrics units (within 96 hours of admission to the ward). For ED patients, we will only recruit from those patients who were brought in by ambulance as an emergency or through their general practitioner.

Exclusion criteria

- Acute life-threatening illness requiring time-critical intervention e.g. ST-elevation myocardial infarction; septic shock; severe pulmonary oedema.
- Coma ('Unresponsive' on the AVPU scale[43])
- Unable to communicate in English or severe dysphasia.

Identification of participants

The participant screening strategy in the initial protocol stated that patients will be recruited between 08.00 and 22.00. A list of potentially eligible patients will be generated in batches at the start of each recruitment period and initial eligibility screening will be carried out by clinical staff (including clinical research nurses embedded in the clinical team). Then, in alphabetical order, in each batch consent/agreement from patient (or proxy/consultee) will be sought by a study researcher. Numbers of those (a) initially potentially eligible (b) screened as non-eligible by clinical staff and (c) declining to take part will be recorded.

This recruitment strategy was modified in the last 5 months of recruitment because preliminary analyses suggested that patients at lower risk of delirium (that is, those not requiring capacity assessment) were more likely to be recruited than those with impaired capacity. Thus, to allow for some oversampling of patients at higher risk of delirium a

pragmatic approach was adopted. From the batches of patients identified as in the original strategy, patients considered at higher risk of delirium on clinical grounds (for example, older age, likely to be admitted, higher degree of ongoing acute and chronic illnesses) were approached first, rather than by alphabetical order.

Assessing capacity and obtaining informed consent

Informed consent will be sought by a trained researcher using a combined informal capacity assessment/consent process[44]. Both verbal and written information will be provided about the study, using a style and format suitable for the participant group (i.e. for varying levels of capacity). The researcher will ask the potential participant to recount the study which will be used, with the treating team views, to assess capacity to consent. For participants judged to have capacity, consent will be sought for:

(a) Conducting assessments as specified in the study information sheets

(b) Accessing health records for information relevant to outcomes and health service use

(c) Recording these data in secure study databases

It will be made clear to participants that they are under no obligation to take part, their usual care will not be affected by their decision, and they can withdraw consent without giving a reason. Once participants are enrolled in the study they will be given a sheet with contact details for the research team and instructions on what to do if they wish to withdraw consent or require further information.

Lack of capacity to consent

It is essential that this study recruits patients which reflect the target clinical population. This means that we must recruit patients with delirium in the same proportion as in the clinical population. We will seek consent/agreement from legal proxies, consultees or other legal

representatives. Where the potential participant is deemed to lack capacity to consent, recruitment will proceed under the provisions of the Mental Capacity Act, 2005 in England or Adults with Incapacity (Scotland) Act, 2000. The clinical team will be asked to identify an appropriate personal or nominated consultee, guardian, welfare attorney or nearest relative. Because of differing legal requirements in Scotland and England, the details of the processes in each nation are given in Appendix 1.

Interventions to be measured

Assessments will be carried out by researchers fully trained in background information on delirium, the features of delirium, and each rating scale. Training is carried out using written, video and bedside training until competence in all aspects of the assessments is achieved.

The 4 "A"s Test (4AT)

The 4AT (see www.the4AT.com) comprises 4 items. Item 1 concerns an observational assessment of level of alertness. The next 2 items are brief cognitive tests: the Abbreviated Mental Test – 4 (AMT4) which asks the patient to state their age, their date of birth, the current year, and the place they are in; and attention testing with Months Backwards, in which the patient is asked to state the months of year in reverse order, starting with December. Only items 1-3 are done at the bedside, and the typical duration is under 2 minutes. Item 4 concerns acute change in mental status, a core diagnostic feature of delirium; this information is obtained from the casenotes or the GP letter or from an informant.

Short Confusion Assessment Method (CAM)

The CAM is a diagnostic algorithm in which the tester rates the following four features as positive or negative: 1. Acute Change and Fluctuating Course; 2. Inattention; 3. Disorganised

Thinking; and 4. Altered Level of Consciousness. The CAM scoring process requires that Features 1 and 2 are both positive; if they are positive then Features 3 and 4 are assessed and if one of Features 3 or 4 is positive, then the whole CAM is positive. The tester scores the features by a combination of interview with the patient, cognitive testing (the CAM requires that a cognitive test is performed before the features are scored), examining the casenotes, and seeking informant history if required. Note that the questionnaires used to assess cognition are not specified by the CAM manual, though some suggested tests are provided. Feature 1 is assessed by the same process as Item 4 in the 4AT. Feature 2 is assessed by the tester giving a positive or negative rating to the question, "Did the patient have difficulty focusing attention, for example, being easily distractible, or having difficulty keeping track of what was being said?" Feature 3 is assessed by the tester giving a positive or negative rating to the question, "Was the patient's thinking disorganized or incoherent, such as rambling or irrelevant conversation, unclear or illogical flow of ideas, or unpredictable switching from subject to subject?" Feature 4 is similar to item 1 in the 4AT. In this study, for the pre-CAM cognitive assessment we will use a set of questions covering the cognitive domains represented in the suggested tests in the CAM manual, including Days of the Week Backwards, counting from 20 down to 1, orientation questions, three-word recall, and clockdrawing, as well as simple orientation questions. All of these questions are used in routine clinical practice at the bedside.

Reference Standard Assessment

Reference standard assessment: This will be centred on the Delirium Rating Scale-Revised-98 (DRSR98)[45], requiring inspection of case notes, speaking to staff who know the patient, or speaking to the patient's relatives or others who know them (with patient consent). As per the instruction manual, the DRS-R98 will be supplemented by short neuropsychological tests

of attention and other domains, including Digit Span[27], the Observational Scale for Level of Arousal[46], the Richmond Agitation Sedation Scale[47] and the DelApp objective attentional assessment[48]. We will also perform simple orientation questions, and record any formal prior diagnosis of dementia and Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE)[49] scores. The DRS-R98 and supporting tests will be used to inform a binary ascertainment of delirium based on DSM-IV criteria. The final DSM-IV ascertainment of delirium will be based on a standardised process with final verification by the Chief Investigator, blind to the 4AT or CAM results. The panel of supporting tests, and the way the data are coded will be designed such that the performance of the 4AT can also be evaluated against the DSM-5 criteria[2]. The reference standard assessment will take approximately 15-20 minutes.

Ordering of assessments

All patients will undergo a reference standard assessment for delirium by the researcher who conducted the capacity assessment and consenting process. A different researcher will also ask each patient to undergo either the 4AT or the CAM. The reason that the researcher doing the capacity assessment and consenting process must also do the reference standard assessment is that the capacity and consenting process provides information to the tester over and above the normal 4AT or CAM testing process. This is not a concern for the reference standard assessment, which is aimed at providing a thorough assessment so as to optimise diagnostic accuracy. The order of these two assessments ([4AT or CAM assessment] and reference standard assessment) will be randomly allocated immediately after consenting, as will the assignment to either the 4AT or CAM. Each patient will receive the reference standard assessment by the same researcher who did the capacity and consenting process. The 4AT or CAM will be performed by a different researcher. When possible the IQCODE will

then be administered to a person who knows the patient well (within 4 weeks of the patient joining the study).

Randomisation procedure

The allocation sequence will be created using computer-generated random numbers. Participants will be randomised in a 1:1 ratio to be assessed using the 4AT or CAM experimental assessment. The order in which they receive the reference standard and experimental assessment will also be randomised in a 1:1 ratio. Randomisation will be stratified by study site with block allocation. The randomised allocations will be concealed until they are assigned as the randomisation system will be web-based and require a personal log-in and password. Once randomisation has been performed neither the researchers nor the participant will be blinded to the allocation as both will be aware of the assessments conducted and the order in which they are performed.

Outcome measurements (what, when, how)

Note that the outcome measurements for the primary study (the reference standard) are performed at almost the same time as the 4AT and CAM. The only subsequent data collection is capturing clinical outcomes at 12 weeks. This will be achieved through searching electronic patient records, telephone calls with participants, or face-to-face-interviews if still in hospital. The information gleaned at the 12 week point will at times be generalised due to participant recall or availability of full electronic records.

1) Primary outcome measure:

Diagnostic accuracy of the 4AT versus the reference standard delirium diagnosis

2) Secondary outcome measures:

(a) 4AT versus CAM in relation to reference standard delirium diagnosis

(b) Performance of 4AT cognitive test items (AMT4 and Months Backwards) in detecting longer-term cognitive impairment as detected by the IQCODE

(c) 4AT total scores as a predictor of the following clinical outcomes as determined at 12 weeks post-test: length of stay, falls, institutionalisation (as assessed by proportion of patients newly admitted to care homes or awaiting care homes at that time) and mortality

(d) Performance of individual items of the 4AT in relation to reference standard delirium diagnosis

(e) We will assess the 4AT total score as a measure of delirium severity.

(f) The primary output from the health economic analysis will be a comparison of the service delivery costs associated with the diagnostic accuracy of alternative (4AT vs. CAM vs. reference standard) triage tools for delirium.

Coding and recording assessments

The experimental assessments of delirium will be the 4AT and the CAM. The 4AT has a total possible score of 12: items (1) and (4) can score 0 or 4; items (2) and (3) can score 0, 1, or 2[12]. 4AT data will be used for the primary objective as a binary outcomes, with 0-3 scores giving a 'no delirium' classification, and 4-12 scores giving a 'delirium' classification; for the secondary objectives, continuous scoring, from 0-12, will be studied as a possible severity indicator, and scores of 1-3 (indicating cognitive impairment but not delirium) can be studied against other assessments of chronic cognitive impairment). The CAM will scored as delirium present or absent according to the algorithm. The 4AT, CAM scoring and reference standard scoring will be recorded on a paper Case Report Form.

Patient resource-use will be derived from medical records, including the 'TrakCare' (InterSystems Corporation, Cambridge, MA, USA) electronic patient record system, where available, as well as via patient or carer self-report. The self-report resource-use questionnaire

will include questions regarding inpatient health and social care utilisation with a maximum recall period of 16 weeks. The self-report resource use questionnaire will be developed specifically for the study for use by patient or proxy respondent using guidance from the Database of Instruments for Resource Use Measurement[50]. Administration of the questionnaire will be conducted at 12 weeks by one of the researchers in the study team, face-to-face where patients are still hospitalised, or via telephone. Data from the questionnaire will be recorded on a paper Case Report Form.

The data on all the Case Report Forms will be transcribed into a secure database by the researchers or a suitably qualified member of the research team. This will be conducted using Edinburgh Clinical Trials Unit Standard Operating Procedures. Quality checking will be performed in 10% of Case Report Forms.

Sample size calculation

450 patients will be randomised to assessment by 4AT and 450 to CAM. We will recruit sufficient patients to account for attrition; though we do not expect significant attrition because the recruitment, consenting and assessment process takes place over a small number of hours, in a single episode. Of the 450 patients within each assessment arm, 15% (67) would be expected to have delirium. The specificity of the triage tool would be estimated based on the 85% (383) without delirium, while the sensitivity would be estimated from the 67 with delirium. Based on analysis using the normal approximation to the binomial distribution, the two-sided 95% confidence interval widths for the specificity and sensitivity would be as shown in Table 2 for a range of levels of diagnostic test performance.

Table 2 Precision of specificity, sensitivity estimation

Parameter	True level of parameter	95% confidence interval width
Specificity	0.5	±0.050
Specificity	0.7	±0.046

Specificity	0.9	±0.030
Sensitivity	0.5	±0.120
Sensitivity	0.7	±0.110
Sensitivity	0.9	±0.072

It will therefore be possible to estimate the specificity precisely and the sensitivity with moderate precision. The precision in estimating negative predictive value would be expected to be similar to that for specificity; for positive predictive value it would be expected to be similar to that for sensitivity. For the secondary objective of comparing 4AT and CAM, based on analysis by continuity corrected chi-squared test, we have 83% power to detect a difference in specificity of 0.1, assuming a null hypothesis of specificity=0.70 for both tests and a two-sided 5% significance level. The corresponding difference detectable for sensitivity (null hypothesis sensitivity=0.7) would be 0.224 with 80% power.

Data analysis plan

The detailed statistical analysis plan will be agreed prior to database lock and will be prepared by individuals blinded to the randomised allocations.

Primary objective:

(a) 4AT vs reference standard: the diagnostic accuracy of 4AT versus the reference standard will be assessed using positive and negative predictive values, sensitivity and specificity. The exact binomial 95% confidence interval will be reported for each measure. A receiver operating characteristic (ROC) curve will be constructed to verify that the proposed cut point of greater than 3 on the 4AT score is appropriate. The area under the ROC curve and its 95% confidence interval will be reported.

Secondary objectives:

(a) 4AT vs CAM: differences in each of sensitivity, specificity, positive and negative predictive values between 4AT and CAM will be tested by Fisher's exact test and quantified by the difference in the two proportions (4AT-CAM) and its 95% confidence interval. To aid

comparison of 4AT and CAM, the overall performance of each will also be summarised using Youden's Index (sensitivity minus false positive rate) and the odds ratio of sensitivity to specificity.

(b) Performance of the 4AT cognitive screening items: is the 4AT an adequately sensitive tool for detecting general cognitive impairment as judged against a documented history of dementia and/or the Informant Questionnaire for Cognitive Decline in the Elderly? This objective will be addressed using the same methods as for the primary objective.

(c) 4AT vs clinical outcomes: as assessment of criterion validity, we will assess the performance of the 4AT in predicting length of stay, institutionalisation, and mortality, up to 12 weeks. Descriptive statistics of clinical outcomes will be presented for the groups with and without 4AT scores above the cut point of 3. The relationship between 4AT and each of mortality and institutionalisation will be analysed via logistic regression modelling; Kaplan-Meier curves and the Cox proportional hazards model will be used to assess 4AT as a predictor of hospital length of stay. The logistic regression and Cox models will adjust for age, gender and presence of dementia.

(d) Individual items: we will conduct analyses examining performance of individual items of the 4AT, e.g. is altered level of alertness alone a good predictor of delirium diagnosis? (Methods as per primary objective);

(e) Delirium severity: we will assess the 4AT total score as a measure of delirium severity by calculating the Spearman correlation between 4AT and DRS-R98 scores and its 95% confidence interval.

Full details of the proposed statistical analyses for the primary objective and secondary objectives (a) to (e) will be documented in a statistical analysis plan (SAP) which will include details of methods for calculating derived variables, any sensitivity and subgroup analyses,

and approaches to testing the assumptions in the statistical analyses. The SAP will outline the plan for validation of the statistical analysis.

Individuals with missing data for the reference diagnostic test will be removed from formal statistical analysis. Where any items of the CAM or 4AT were not able to be assessed, an overall delirium diagnosis will still be derived where possible based on the items which have been recorded. There will be no other imputation of missing delirium diagnoses.

(f) Delivery costs of the triage tools: We will estimate the delivery costs and subsequent resource consequences associated with the triage tools as a function of sensitivity and specificity from the perspective of the UK National Health Service. Potential resource consequences may include additional diagnostic procedures (e.g. more detailed cognitive screening and brain imaging), altered management as well as re-admissions. Healthcare resource use will be derived from medical records, including the 'TrakCare' (InterSystems Corporation, Cambridge, MA, USA) electronic patient record system, where available, as well as via patient or carer self-report. Monetary values will be attached to resource use, training and labour costs as well as the indirect costs of delivering each diagnostic tool using standard NHS pay and price estimates. Generalised linear models will be used to analyse 12 week cumulative costs which will inform longer-term resource consequences within a decision analytic model.

ETHICS

This study was granted ethical approval in Scotland (Scotland A NHS Research Ethics Committee REC 15/SS/0071) and England (Yorkshire and The Humber – Bradford Leeds NHS Research Ethics Committee REC 15/YH/0317).

Study oversight

Study oversight is through the Trial Steering Committee, which will meet every four months during the study. The Trial Steering Committee comprises two independent lay representatives, three independent experts (one of whom is the Chair of the Committee), the PI, the study statistician, and representatives from the Edinburgh Clinical Trials Unit.

Data Protection

Data will be collected and handled in line with sponsor and Edinburgh Clinical Trials Unit Standard Operating Procedures and NHS Trust policies. All electronic data will be link-anonymised.

DISCUSSION

This study was designed to validate the 4AT against a reference standard assessment, as well as compare it to another commonly used test for assessment of delirium. Since the initial study design, the 4AT has been widely adopted nationally and internationally. The 4AT has been incorporated into routine practice in multiple international centres, both in paper and electronic format, with many centres reporting 10,000 uses of the tool. The website www.the4AT.com has had an increasing degree of traffic, and the 4AT has been translated into multiple languages. The 4AT is also included in several national guidelines and position statements internationally as a recommended tool and it has been validated in other studies[38-42] but it is still essential that it is further tested in a large study. It is also essential to consider how well it identifies other types of cognitive impairment, relates to future outcomes, and its health economic impact.

In the initial design and implementation of the study, the main challenging aspects have been: 1) considering both Scottish and English legal and ethical framework to ensure that patients without capacity are included. Ethical approval for inclusion of these patients was granted though recruitment of this patient group proved difficult from the outset for several reasons.

Firstly, the narrow boundaries for the screening and identification strategy. This was addressed in a subsequent protocol amendment to aim for oversampling of patients at risk of developing delirium. Secondly, the availability of an appropriate individual to provide consent on behalf of the participant (i.e. a personal or nominated consultee, guardian, welfare attorney or nearest relative). Thirdly, a reluctance to consent due to perceived burden on participant. Persuading relatives of the value, importance and necessity of research even in clinically unwell patients demands a particular skillset from researchers and involved perseverance and excellent communication in order to achieve recruitment targets.

2) Recruitment and training of staff, with some staff moving to different posts, and new staff being recruited and requiring training; in each case detailed training supported by reading materials and practice sessions was provided.

We also acknowledge that it is possible that researcher bias may influence how the different index assessments (4AT or CAM) were scored. We also acknowledge that given the fluctuating nature of delirium, the gap between assessments potentially reaching two hours means that assessments could have different findings. We will conduct sensitivity analyses to analyse the impact of variations in the time gap between assessments.

CONCLUSION

The 4AT study aims to assess the validity of this rapid delirium screening tool that can be administered by a range of professional level healthcare staff to patients with sensory or functional impairments in a busy clinical environment, which also incorporates general cognitive assessment. We will also assess the later functional outcomes of people with and without delirium, and the health economic implications. The overall aim is to improve detection, and therefore management and outcomes, of this important and devastating condition.

REFERENCE LIST

1. American Psychiatric Association. Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* 4th ed. Washington, DC: American Psychiatric Association; 2000.

2. American Psychiatric Association, American Psychiatric Association. *Diagnostic and statistical manual of mental disorders : DSM-V.* 5th ed. Washington, DC: American Psychiatric Association; 2013

3. Siddiqi N, House AO, Holmes JD. Occurrence and outcome of delirium in medical in-patients: a systematic literature review. *Age and Ageing.* 2006;35(4):350-364.

4. Goldberg SE, Whittamore KH, Harwood RH, et al. The prevalence of mental health problems among older adults admitted as an emergency to a general hospital. *Age Ageing.* 2012;41(1):80-86.

5. Collins N, Blanchard MR, Tookman A, et al.. Detection of delirium in the acute hospital. *Age and Ageing.* 2010;39(1):131-135.

6. Delirium: diagnosis, prevention and management. 2010; www.nice.org.uk/CG103.

7. Davis DH, Muniz Terrera G, Keage H, et al. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. *Brain.* 2012;135(Pt 9):2809-2816.

8. MacLullich AM, Beaglehole A, Hall RJ, et al. Delirium and long-term cognitive impairment. *Int Rev Psychiatry.* 2009;21(1):30-42.

9. Witlox J, Eurelings LSM, de Jonghe JFM, et al. Delirium in Elderly Patients and the Risk of Postdischarge Mortality, Institutionalization, and Dementia: A Meta-analysis. *JAMA.* 2010;304(4):443-451.

10. 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 e241.

11. Anderson D. *Who Cares Wins.* London: Royal College of Psychiatrists; 2005.

12. Fong TG, Jones RN, Shi P, et al. Delirium accelerates cognitive decline in Alzheimer disease. *Neurology.* 2009;72(18):1570-1575.

13. Leslie DL, Marcantonio ER, Zhang Y, et al.. One-year health care costs associated with delirium in the elderly population. *Archives of Internal Medicine.* 2008;168(1):27-32.

14. Kakuma R, du Fort GG, Arsenault L, et al. Delirium in older emergency department patients discharged home: Effect on survival. *Journal of the American Geriatrics Society.* 2003;51:443-450.

15. Partridge JS, Martin FC, Harari D, et al. The delirium experience: what is the effect on patients, relatives and staff and what can be done to modify this? *Int J Geriatr Psychiatry.* 2012. 2013 28(8):804-12. doi: 10.1002/gps.3900 [Epub 2012 Oct 30]

16. Adunsky A, Levy R, Mizrahi E, et al. Exposure to opioid analgesia in cognitively impaired and delirious elderly hip fracture patients. *Arch Gerontol Geriatr.* 2002;35(3):245-251.

17. Department of Health Using the Commissioning for Quality and Innovation (CQUIN) payment framework: Guidance on new national goals for 2012-13. Department of Health, 2012.

18. Elie M, Rousseau F, Cole M, et al. Prevalence and detection of delirium in elderly emergency department patients. *Canadian Medical Association Journal.* 2000;163:977-981.

19. 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.

20. Davis D, MacLullich A. Understanding barriers to delirium care: a multicentre survey of knowledge and attitudes amongst UK junior doctors. *Age Ageing.* 2009;38(5):559-563.

21. Teodorczuk A, Reynish E, Milisen K. Improving recognition of delirium in clinical practice: a call for action. *BMC Geriatr.* 2012;12:55.

22. Kennelly SP, Morley D, Coughlan T, et al. Knowledge, skills and attitudes of doctors towards assessing cognition in older patients in the emergency department. *Postgrad Med J.* 2012.

23. Steis MR, Fick DM. Delirium superimposed on dementia: accuracy of nurse documentation. *J Gerontol Nurs.* 2012;38(1):32-42.

24. Steis MR, Fick DM. Are nurses recognizing delirium? A systematic review. *J Gerontol Nurs.*2008;34(9):40-48.

25. Inouye SK, van Dyck CH, Alessi CA, et al. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med.* 1990;113(12):941-8.

26. Borson S, Scanlan J, Brush M, et al. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry.* 2000;15(11):1021-1027.

27. Simon SE, Bergmann MA, Jones RN, et al. Reliability of a structured assessment for nonclinicians to detect delirium among new admissions to postacute care. *Journal of the American Medical Directors Association.* 2006;7(7):412-415.

28. Chester JG, Beth Harrington M, Rudolph JL. Serial administration of a modified Richmond Agitation and Sedation Scale for delirium screening. *J Hosp Med.* 2012;7(5):450-453.

29. Yates C, Stanley N, Cerejeira JM, et al. Screening instruments for delirium in older people with an acute medical illness. *Age Ageing.* 2009;38(2):235-237.

30. Harwood DM, Hope T, Jacoby R. Cognitive impairment in medical inpatients. I: Screening for dementia--is history better than mental state? *Age Ageing.* 1997;26(1):31-35.

31. Eeles EM, White S, Bayer T. Inability to complete cognitive assessment is a marker of adverse risk. *Age Ageing.* 2009;38(5):633-634; author reply 634-635.

32. Hall RJ, Meagher DJ, MacLullich AMJ. Delirium detection and monitoring outside the ICU. *Best Pract Res Clin Anaesthesiol.* 2012;26(3):367-383.

33. Kean J, Ryan K. Delirium detection in clinical practice and research: critique of current tools and suggestions for future development. *J Psychosom Res.* 2008;65(3):255-259.

34. Adamis D, Sharma N, Whelan PJ, et al. Delirium scales: A review of current evidence. *Aging Ment Health.* 2010;14(5):543-555.

35. Wong CL, Holroyd-Leduc J, Simel DL, et al. Does this patient have delirium?: value of bedside instruments. *JAMA.* 2010;304(7):779-786.

36. Ringdal GI, Ringdal K, Juliebo V, et al. Using the Mini-Mental State Examination to screen for delirium in elderly patients with hip fracture. *Dement Geriatr Cogn Disord.* 2011;32(6):394-400.

37. O'Keeffe ST, Mulkerrin EC, et al. Use of serial Mini-Mental State Examinations to diagnose and monitor delirium in elderly hospital patients. *J Am Geriatr Soc.*2005;53(5):867-870.

38. Bellelli G, Morandi A, Davis DH, et al. Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age and ageing.* 2014;43(4):496-502.

39. Lees R, Corbet S, Johnston C, et al. Test accuracy of short screening tests for diagnosis of delirium or cognitive impairment in an acute stroke unit setting. *Stroke*. 2013;44:3078-83

40. Kuladee S, Prachason T. Development and validation of the Thai version of the 4 'A's Test for delirium screening in hospitalized elderly patients with acute medical illnesses. *Neuropsychiatr Dis Treat.* 2016; 12:437-43

41. De J, Wand AP, Smerdely PI, et al. Validating the 4A's test in screening for delirium in a culturally diverse geriatric inpatient population. *Int J Geriatr Psychiatry.* 2016 Oct 20. [Epub ahead of print]

42. Hendry K, Quinn TJ, Evans J, et al. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. *Age and Ageing.* 2016 Aug 8. [Epub ahead of print]

43. McNarry AF, Goldhill DR. Simple bedside assessment of level of consciousness: comparison of two simple assessment scales with the Glasgow Coma scale. *Anaesthesia.* 2004;59(1):34-37.

44. Adamis D, Martin FC, Treloar A, Macdonald AJD. Capacity, consent, and selection bias in a study of delirium. *Journal of Medical Ethics.* 2005;31(3):137-143.

45. Trzepacz PT, Mittal D, Torres R, et al. Validation of the Delirium Rating Scale-revised-98: Comparison with the Delirium Rating Scale and the Cognitive Test for Delirium. *Journal of Neuropsychiatry and Clinical Neurosciences.* 2001;13(2):229-242.

46. Tieges Z, McGrath A, Hall RJ, et al. Abnormal level of arousal as a predictor of delirium and inattention: an exploratory study. *The American journal of geriatric psychiatry:official journal of the American Association for Geriatric Psychiatry.* 2013;21(12):1244-1253.

47. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med.* 2002;166(10):1338-1344.

48. Tieges ZS, Scott, K, Suchorab K et al. Development of a smartphone application for the objective detection of attentional deficits in delirium *International Psychogeriatrics.* 2015 Aug;27(8):1251-62. doi: 10.1017/S1041610215000186. [Epub 2015 Mar 6]

49. Jorm AF. A Short-Form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) - Development and Cross-Validation. *Psychol. Med.* 1994;24(1):145-153.

50. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials:a systematic review of studies funded by the UK Health Technology Assessment program.*Value Health.* 2010;13(8):867-872.

AUTHORS' CONTRIBUTIONS: All authors were involved in the design of the protocol for the study. AMJM initiated the conception of the study. AM, JSteven, PLB were involved in protocol design and acquisition of data; CJW initiated the design of the statistical analysis plan. JSmith initiated the design of the economic analysis. SDS and AMJM drafted the paper and all authors revised it critically for important intellectual content. All authors have approved the final approval of the version to be published, and agree to be accountable for all aspects of the work. AMJM is the guarantor.

ACKNOWLEDGMENTS

Thanks to the Trial Steering Committee.

FUNDING

This work was funded by National Institute of Health Research Technology Assessment Programme (NIHR DTA) grant number 11/143/01 (PI: AMJM; CO-I: SG, JSmith, AA, MG, SDS, TR, NS, CF, JH, CJW, AG). JB and CW were supported in this work by NHS Lothian via the Edinburgh Clinical Trials Unit. AA has received personal fees from Abbott Diagnostics, outside the submitted work. SG is funded by NIHR to undertake his role as NIHR HTA Clinical Evaluation and Trials Board Chair. SDS is funded by NRS (NHS Research Scotland) to undertake her role as NRS Ageing Specialty Group Lead (Scotland).

CONFLICT OF INTEREST

This work was funded by National Institute of Health Research Technology Assessment Programme (NIHR DTA) grant number 11/143/01 (PI: AMJM; CO-I: SG, JSmith, AA, MG, SDS, TR, NS, CF, JH, CJW, AG). JB and CW were supporrted in this work by NHS Lothian via the Edinburgh Clinical Trials Unit. AA has received personal fees from Abbott Diagnostics, outside the submitted work. SG is funded by NIHR to undertake his role as NIHR HTA Clinical Evaluation and Trials Board Chair. SDS is funded by NRS (NHS Research Scotland) to undertake her role as NRS Ageing Specialty Group Lead (Scotland).

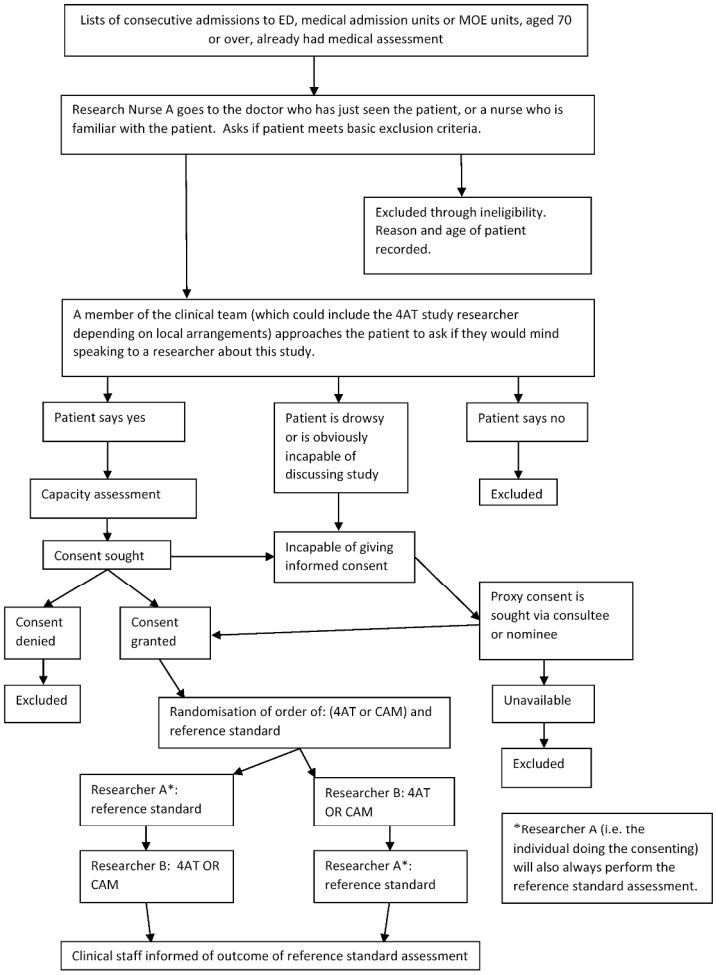


Figure 1: Study overview flowchart

210x297mm (300 x 300 DPI)

Appendix 1: Detailed consent procedures for incapacity in Scotland and England

Scotland

An appropriate legal proxy (that is, a guardian, welfare attorney, nearest relative, but not a member of the clinical team) will be approached by a member of the clinical team (potentially including researchers who are part of the clinical team) to be asked if they would be willing to consider hearing about a study involving the patient, and to potentially give consent on their behalf. If the proxy assents to hearing more about the study, the study team member responsible for consent will provide the proxy with information about: why they are being approached; the role of a proxy, explanation that acting as a proxy is voluntary; details of the study (as would be given to a participant with capacity). The proxy will be asked for advice on whether the participant should take part in the study and what, in their opinion, the participant's views and feelings would have been on taking part in the project had they retained capacity. Consent forms will be signed when the proxy is physically present. If no appropriate legal proxy can be identified within 96 hours, the patient will not be recruited to the study. This is because in Scotland patients with incapacity cannot be included in studies of nonmedicinal treatments unless there is a guardian, welfare attorney or nearest relative available to give consent.

England

If the patient is incapacitated at study entry then a personal consultee (usually a friend or relative) will be consulted and their opinion sought. The approach used will be similar to that detailed in the previous section when consulting legal proxies in Scotland. If the personal consultee agrees that their friend/relative can enter the study then we would ask them to sign a declaration form. If a personal consultee is not available for consultation then the treating doctor (who will be independent of the research team and of appropriate seniority),

will be asked to act as the nominated consultee and advise on inclusion in the study. If agreement is given it will be recorded on the declaration form.

All Trial Participants (England and Scotland)

All patients who lack mental capacity at the time of enrolment will be approached for consent to remain in the trial at the earliest opportunity once they regain capacity. Research staff have planned contact with study patients on the day of enrolment and only on one further occasion at 12 weeks when they will collect questionnaire data from the patient. If research staff become aware the patient has regained capacity while in hospital then written consent from the patient will be sought at this time. It is likely that in many cases the first contact by the research team will be at 12 weeks either in person or by phone.

The patient will be given the opportunity to either withdraw or remain in the study at this time. If the patient chooses to withdraw from the study they will be given the option of allowing/not allowing the use of data already collected. A patient information sheet will be posted to participants who wish to remain on the study and other patients on request.

If patients have not regained capacity at 12 weeks they will remain in the study based on the advice of the consultee or legal proxy.