



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/150019/>

Version: Published Version

Article:

MacLulich, A.M.J., Shenkin, S.D., Goodacre, S. et al. (2019) The 4 'A's test for detecting delirium in acute medical patients : a diagnostic accuracy study. *Health Technology Assessment*, 23 (40). pp. 1-193. ISSN: 1366-5278

<https://doi.org/10.3310/hta23400>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Health Technology Assessment

Volume 23 • Issue 40 • August 2019

ISSN 1366-5278

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

Alasdair MJ MacLulich, Susan D Shenkin, Steve Goodacre, Mary Godfrey, Janet Hanley, Antaine Stíobhairt, Elizabeth Lavender, Julia Boyd, Jacqueline Stephen, Christopher Weir, Allan MacRaid, Jill Steven, Polly Black, Katharina Diernberger, Peter Hall, Zoë Tiegies, Christopher Fox, Atul Anand, John Young, Najma Siddiqi and Alasdair Gray



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

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

¹Geriatric Medicine, Division of Health Sciences, University of Edinburgh, Edinburgh, UK

²Emergency Medicine, University of Sheffield, Sheffield, UK

³Health and Social Care, Leeds Institute of Health Sciences, School of Medicine, Faculty of Medicine and Health, University of Leeds, Leeds, UK

⁴School of Health and Social Care, Edinburgh Napier University, Edinburgh, UK

⁵Edinburgh Clinical Trials Unit, University of Edinburgh, Edinburgh, UK

⁶Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

⁷Emergency Medicine Research Group (EMERGE), NHS Lothian, Edinburgh, UK

⁸Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

⁹Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK

¹⁰Psychiatry, University of York, York

¹¹Hull York Medical School, York, UK

¹²Bradford District Care NHS Foundation Trust, Bradford, UK

*Corresponding author

Declared competing interests of authors: none

Published August 2019

DOI: 10.3310/hta23400

This report should be referenced as follows:

MacLulich AMJ, Shenkin SD, Goodacre S, Godfrey M, Hanley J, Stíobhairt A, *et al.* The 4 'A's test for detecting delirium in acute medical patients: a diagnostic accuracy study. *Health Technol Assess* 2019;**23**(40).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.819

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

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

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

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

Criteria for inclusion in the *Health Technology Assessment* journal

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

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

HTA programme

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

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

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

This report

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

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2019. This work was produced by MacLulich *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

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

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Honorary Professor, University of Manchester, and Senior Clinical Researcher and Associate Professor, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Director, NIHR Dissemination Centre, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

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

Abstract

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

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

¹Geriatric Medicine, Division of Health Sciences, University of Edinburgh, Edinburgh, UK

²Emergency Medicine, University of Sheffield, Sheffield, UK

³Health and Social Care, Leeds Institute of Health Sciences, School of Medicine, Faculty of Medicine and Health, University of Leeds, Leeds, UK

⁴School of Health and Social Care, Edinburgh Napier University, Edinburgh, UK

⁵Edinburgh Clinical Trials Unit, University of Edinburgh, Edinburgh, UK

⁶Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

⁷Emergency Medicine Research Group (EMERGE), NHS Lothian, Edinburgh, UK

⁸Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

⁹Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK

¹⁰Psychiatry, University of York, York

¹¹Hull York Medical School, York, UK

¹²Bradford District Care NHS Foundation Trust, Bradford, UK

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

Background: Delirium is a common and serious neuropsychiatric syndrome, usually triggered by illness or drugs. It remains underdetected. One reason for this is a lack of brief, pragmatic assessment tools. The 4 'A's test (Arousal, Attention, Abbreviated Mental Test – 4, Acute change) (4AT) is a screening tool designed for routine use. This project evaluated its usability, diagnostic accuracy and cost.

Methods: Phase 1 – the usability of the 4AT in routine practice was measured with two surveys and two qualitative studies of health-care professionals, and a review of current clinical use of the 4AT as well as its presence in guidelines and reports. Phase 2 – the 4AT's diagnostic accuracy was assessed in newly admitted acute medical patients aged ≥ 70 years. Its performance was compared with that of the Confusion Assessment Method (CAM; a longer screening tool). The performance of individual 4AT test items was related to cognitive status, length of stay, new institutionalisation, mortality at 12 weeks and outcomes. The method used was a prospective, double-blind diagnostic test accuracy study in emergency departments or in acute general medical wards in three UK sites. Each patient underwent a reference standard delirium assessment and was also randomised to receive an assessment with either the 4AT ($n = 421$) or the CAM ($n = 420$). A health economics analysis was also conducted.

Results: Phase 1 found evidence that delirium awareness is increasing, but also that there is a need for education on delirium in general and on the 4AT in particular. Most users reported that the 4AT was useful, and it was in widespread use both in the UK and beyond. No changes to the 4AT were considered necessary. Phase 2 involved 785 individuals who had data for analysis; their mean age was 81.4 (standard deviation 6.4) years, 45% were male, 99% were white and 9% had a known dementia diagnosis. The 4AT ($n = 392$) had an area under the receiver operating characteristic curve of 0.90. A positive 4AT score (> 3) had a specificity of 95% [95% confidence interval (CI) 92% to 97%] and a sensitivity of 76% (95% CI 61% to 87%) for reference standard delirium. The CAM ($n = 382$) had a specificity of 100% (95% CI 98% to 100%) and a sensitivity of 40% (95% CI 26% to 57%) in the subset of participants whom it was possible to assess using this. Patients with positive 4AT scores had longer lengths of stay (median 5 days, interquartile range 2.0–14.0 days) than did those with negative 4AT scores (median 2 days, interquartile range 1.0–6.0 days), and they had a higher 12-week mortality rate (16.1% and 9.2%, respectively). The estimated 12-week costs of an initial inpatient stay for patients with delirium were more than double the costs of an inpatient stay for patients without delirium (e.g. in Scotland, £7559, 95% CI £7362 to £7755, vs. £4215, 95% CI £4175 to £4254). The estimated cost of false-positive cases was £4653, of false-negative cases was £8956, and of a missed diagnosis was £2067.

Limitations: Patients were aged ≥ 70 years and were assessed soon after they were admitted, limiting generalisability. The treatment of patients in accordance with reference standard diagnosis limited the ability to assess comparative cost-effectiveness.

Conclusions: These findings support the use of the 4AT as a rapid delirium assessment instrument. The 4AT has acceptable diagnostic accuracy for acute older patients aged > 70 years.

Future work: Further research should address the real-world implementation of delirium assessment. The 4AT should be tested in other populations.

Trial registration: Current Controlled Trials ISRCTN53388093.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 23, No. 40. See the NIHR Journals Library website for further project information. The funder specified that any new delirium assessment tool should be compared against the CAM, but had no other role in the study design or conduct of the study.

Contents

List of tables	xiii
List of figures	xv
List of boxes	xvii
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 General introduction	1
Delirium: background	1
Underdetection of delirium	1
The need for a new assessment tool	2
<i>Health economics</i>	3
Overview of project	4
Chapter 2 Surveys of current practice	5
General introduction	5
Methods	5
<i>Survey development</i>	5
<i>Participants</i>	6
<i>Data analysis</i>	7
Results	7
<i>Survey A</i>	7
<i>Survey B</i>	11
Discussion	15
<i>Survey A</i>	15
<i>Survey B</i>	16
Conclusions and implications	18
Chapter 3 Clinical use and examples of clinical service evaluations	19
Introduction	19
Inclusion of the 4AT in policy, guidelines, advisory documents and websites	19
<i>Policy websites and documents that include the 4AT</i>	19
Use of the 4AT in clinical practice	20
<i>Examples of websites and documents documenting 4AT use in clinical practice</i>	21
<i>Case study: City Hospitals Sunderland NHS Trust</i>	21
<i>Case study: King's College Hospital NHS Foundation Trust, London</i>	21
Field testing of 4AT item 1 (level of alertness)	21
Conclusions	22
Chapter 4 Implementing delirium screening	23
Introduction	23
Research design	23
<i>Research objectives</i>	23

Methods	24
<i>Sampling strategy</i>	24
<i>Data collection</i>	24
<i>Data analysis</i>	26
<i>Research team</i>	27
<i>Ethics</i>	27
Findings	27
<i>The emergency department</i>	28
<i>Assessment units: medical assessment unit; elderly/frailty assessment unit; surgical assessment unit</i>	35
<i>Acute wards</i>	40
Discussion	46
<i>Multilevel knowledge of delirium</i>	46
<i>Service delivery context, purpose and value attached to delirium screening</i>	47
<i>Limitations</i>	48
<i>Conclusions and implications</i>	48
Chapter 5 Diagnostic accuracy of the 4AT	49
Introduction	49
Study objectives and hypotheses	50
Methods	50
<i>Ethics</i>	51
<i>Study design: overview</i>	51
<i>Inclusion and exclusion criteria</i>	51
<i>Sample size</i>	51
<i>Recruitment</i>	52
<i>Consent</i>	52
<i>Test methods</i>	52
<i>Research staff and training</i>	57
<i>Analysis</i>	57
<i>Methods: measures of diagnostic accuracy</i>	57
Results	58
<i>Descriptive statistics</i>	58
<i>Primary objective</i>	58
<i>Secondary objectives</i>	60
Discussion	64
Conclusions	67
Chapter 6 Health economics	69
Introduction	69
Methods	69
<i>Resource use, costs and data collection</i>	69
<i>Methods for analysing and handling missing data</i>	70
<i>Unit cost assignment</i>	70
<i>Cost calculation</i>	70
<i>Sensitivity analysis</i>	70
<i>Scottish hospital costs</i>	70
<i>English hospital costs</i>	72
<i>Costs of tests in hospital</i>	72
<i>Community health service costs</i>	72
<i>Costs of additional activities</i>	72

Results	72
<i>Cost analysis using original researcher delirium ascertainment</i>	72
<i>Cost analysis using final reference standard ascertainment</i>	74
Health economic model: cost-effectiveness analysis	75
<i>Model structure</i>	75
<i>Model parameters</i>	76
<i>Reporting</i>	76
<i>Methods for the sensitivity analysis</i>	76
<i>Results from the health economic model</i>	77
<i>One-way sensitivity analysis</i>	77
<i>Probabilistic sensitivity analysis</i>	77
Discussion	78
<i>Discussion of the cost analysis</i>	78
<i>Discussion of the cost-effectiveness analysis</i>	78
Chapter 7 General discussion	81
Research implications	83
Clinical implications	83
Acknowledgements	85
References	89
Appendix 1 Survey A	103
Appendix 2 Survey B	121
Appendix 3 Survey B: respondent characteristics	141
Appendix 4 Survey B: assessment tools used and the frequency of use among respondents	143
Appendix 5 User opinions of the 4AT from survey B	145
Appendix 6 Interviewees in qualitative study by hospital and department	149
Appendix 7 Interviewees in qualitative study by profession and seniority	151
Appendix 8 Topic guide for qualitative staff interviews: identifying delirium: knowledge, screening practices and barriers to identifying delirium in acute hospitals	153
Appendix 9 Comments and examples of suggested changes to the 4AT	157
Appendix 10 The STARD flow diagrams of recruitment at each site	159
Appendix 11 Receiver operating characteristic curve for 4AT diagnostic accuracy	163
Appendix 12 Diagnostic test accuracy of 4AT versus CAM for diagnosis of delirium	165
Appendix 13 Diagnostic test accuracy of individual items of 4AT for diagnosis of delirium	167

Appendix 14 Scatterplot of Delirium Rating Scale-Revised-98 scores and 4AT scores	169
Appendix 15 Kaplan–Meier plot for 4AT and length of stay	171
Appendix 16 Costs of tests: health economics analysis	173
Appendix 17 Community health service costs	175
Appendix 18 Costs of additional activities	177
Appendix 19 Health economics survey	179
Appendix 20 Model parameters	187
Appendix 21 Results from the health economics model	189
Appendix 22 Cost-effectiveness plane	191
Appendix 23 Cost-effectiveness-acceptability curve	193

List of tables

TABLE 1 Survey A: respondent characteristics	8
TABLE 2 Survey A: respondents' answers when asked to indicate which of the three tasks different professional groups should consider their own duty with regard to delirium detection	9
TABLE 3 Survey B: the extent to which respondents thought that various barriers were preventing the 4AT from being used more regularly in their units	12
TABLE 4 The 4AT	54
TABLE 5 Baseline characteristics stratified by delirium status	60
TABLE 6 Cross-tabulation of index test results (4AT and CAM) by results of reference standard	61
TABLE 7 Diagnostic test accuracy of the 4AT for diagnosis of delirium (defined by reference standard assessment)	61
TABLE 8 Components of the 4AT cognitive items stratified by dementia status, defined as a documented history of dementia and/or a score of ≥ 3.44 on the IQCODE	62
TABLE 9 Diagnostic test accuracy of the 4AT cognitive components for diagnosis of dementia, defined as a documented history of dementia and/or a score of ≥ 3.44 on the IQCODE	62
TABLE 10 The 4AT as a predictor of clinical outcomes	63
TABLE 11 Scottish hospital cost estimates	71
TABLE 12 English hospital cost estimates	73
TABLE 13 Costs (£) at 12 weeks for delirium vs. no delirium from the original researcher assessments	74
TABLE 14 Costs (£) at 12 weeks for delirium vs. no delirium by the final reference standard ascertainment	74
TABLE 15 Survey B: respondent characteristics	142
TABLE 16 Survey B: assessment tools used and the frequency of use among respondents	143
TABLE 17 Interviewees in qualitative study by hospital and department	149
TABLE 18 Interviewees in qualitative study by profession and seniority	151
TABLE 19 Diagnostic test accuracy of 4AT vs. CAM for diagnosis of delirium	166

TABLE 20 Diagnostic test accuracy of individual items of the 4AT for diagnosis of delirium	167
TABLE 21 Costs of tests: health economics analysis	173
TABLE 22 Community health service costs	175
TABLE 23 Costs of additional activities	177
TABLE 24 Model parameters	187
TABLE 25 Differences in costs and effectiveness and ICERs	189

List of figures

FIGURE 1 Page views per year on the www.the4AT.com website between 2011 and 2017	20
FIGURE 2 Diagnostic accuracy study: overview flow chart	56
FIGURE 3 The STARD diagram of flow of participants through the study (total across all three)	59
FIGURE 4 Decision tree	75
FIGURE 5 The STARD diagram of flow of participants through the Edinburgh site	160
FIGURE 6 The STARD diagram of flow of participants through the Sheffield site	161
FIGURE 7 The STARD diagram of flow of participants through the Bradford site	162
FIGURE 8 The ROC curve for 4AT diagnostic accuracy	163
FIGURE 9 Scatterplot of DRS-R98 scores and 4AT scores	169
FIGURE 10 Kaplan–Meier plot for 4AT and length of stay	171
FIGURE 11 Cost-effectiveness plane	191
FIGURE 12 Cost-effectiveness-acceptability curve	193

List of boxes

BOX 1 Examples of suggested changes to the 4AT from free-text comments in survey B	13
BOX 2 Reference standard assessment for delirium	53
BOX 3 The CAM: pre-scoring cognitive testing procedure	55

List of abbreviations

4AT	4 'A's Test	ICU	intensive care unit
A&E	accident and emergency	IQCODE	Informant Questionnaire for Cognitive Decline in the Elderly
AMT	Abbreviated Mental Test	IQR	interquartile range
AMT4	Abbreviated Mental Test – 4	ISD	Information Services Division
AMT10	Abbreviated Mental Test (different name for AMT)	MAU	medical assessment unit
CAM	Confusion Assessment Method	MDT	multidisciplinary team
CI	confidence interval	MMSE	Mini-Mental State Examination
CQUIN	Commissioning for Quality and Innovation	MoE	medicine of the elderly
CRF	case record form	NICE	National Institute for Health and Care Excellence
DRS-R98	Delirium Rating Scale-Revised-98	NPV	negative predictive value
DSM	<i>Diagnostic and Statistical Manual of Mental Disorders</i>	PPV	positive predictive value
DSM-3-R	<i>Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition – Revised</i>	PSSRU	Personal Social Services Research Unit
DSM-4	<i>Diagnostic and Statistical Manual of Mental Disorders, 4th Edition</i>	QALY	quality-adjusted life-year
DSM-5	<i>Diagnostic and Statistical Manual of Mental Disorders, 5th Edition</i>	ROC	receiver operating characteristic
EAU	elderly assessment unit	SAU	surgical assessment unit
ED	emergency department	SD	standard deviation
GP	general practitioner	ST3+	speciality trainee level 3 or above
ICER	incremental cost-effectiveness ratio	STARD	Standards for Reporting of Diagnostic Accuracy Studies
		UPR	unitary patient record

Plain English summary

Delirium is the sudden onset of confusion that can happen when someone is unwell. It is common in older people who go into hospital, and it is upsetting for both the patients and their families. Delirium is important to diagnose, because people with delirium do less well than those without, and it is often treatable.

The ideal way to diagnose delirium is with a full assessment by a specialist, but this is expensive and time-consuming. We therefore developed a short test called the 4 'A's Test (4AT). The four 'A's stand for Arousal, Attention, Abbreviated Mental Test – 4, and Acute change.

First, we interviewed hospital staff about delirium and the 4AT. We found that the 4AT was already widely used and that people found it easy to use.

We then tested how the 4AT performed in practice. A total of 785 recently admitted patients aged ≥ 70 years participated, of whom around one in eight had delirium. A researcher carried out the full standard delirium assessment on each patient and then a different researcher carried out the 4AT.

A normal 4AT score reliably ruled out delirium. An abnormal score was also reasonably effective in detecting delirium, but staff still needed to follow up such patients with a full assessment. People with higher 4AT scores stayed in hospital longer and were more likely to die, and their treatment was more expensive.

We conclude that the 4AT is a useful test to rule out delirium or to see if more detailed testing is required. It could help treat patients correctly and quickly. This would save money and improve outcomes.

Scientific summary

Background

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

Objectives

Phase 1

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

Phase 2

The primary objective was:

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

The secondary objectives were to:

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

Methods

Phase 1

Surveys

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

Qualitative study

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

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

Current use in clinical practice

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

Phase 2

Diagnostic test accuracy study

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

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

Health economic analysis

Cost analysis using data collected within the study

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

Cost-effectiveness analysis generated through a health economic model

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

Results

Phase 1

Surveys

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

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

Qualitative study

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

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

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

Current use in clinical practice

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

Phase 2

Diagnostic test accuracy study

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

Primary objective

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

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

Secondary objectives

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

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

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

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

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

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

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

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

Health economic analysis

To model the costs of delirium in the study population

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

Cost-effectiveness analysis generated through a health economic model

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

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

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

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

Conclusions

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

Trial registration

This trial is registered as ISRCTN53388093.

Funding

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

Chapter 1 General introduction

Delirium: background

Delirium is a severe and distressing neuropsychiatric syndrome that is characterised by acute deterioration in attention and other mental functions. The American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-4),¹ criteria for delirium are, in summary, a disturbance of consciousness (i.e. a reduced ability to focus, sustain or shift attention), and a change in cognition. This mental status deterioration develops over short periods of time (usually hours to days) and it tends to fluctuate.¹ The newer *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5),² criteria, published in 2013, are similar but have some terminological differences. Delirium is commonly precipitated by acute illness, trauma, or from the side effects of drugs. The presence of a 'general medical condition' is also one of the DSM-4 and DSM-5 criteria, although in practice this presence is usually assumed rather than specified in each case. Delirium is extremely common, affecting at least 15% of patients in acute hospitals,³⁻⁷ and it is associated with many complications and poor outcomes.^{4,8-15} Delirium is both a marker of current dementia^{4,16-18} and associated with the acceleration of existing dementia.¹⁹ In older patients without dementia, an episode of delirium strongly predicts future dementia risk.^{9,20} The economic burden of delirium, derived from 2008 US data, estimates the 1-year health-care costs to be US\$38B–152B.²¹ The detection of delirium is vital because its presence indicates acute systemic or central nervous system illness, physiological disturbance and drug intoxication or withdrawal; indeed, failure to detect delirium in the acute setting is associated with worse outcomes.²²

The specific management of delirium is of obvious and immediate benefit to patients in many clinical situations, for example in reversing opioid toxicity, enabling the treatment of peripheral infections that have presented with delirium, alleviating distress caused by anxiety and disorientation as well as directly by frightening delusions and hallucinations,^{13,23-25} and prompting more thorough assessment and treatment of symptoms.²⁶ For example, some studies²⁷ have found that surgical patients with delirium receive less analgesia than those with normal cognition; this matters not only because pain treatment is an end in itself but also because pain is itself a cause of delirium. Delirium management also includes communicating the diagnosis to patients and carers; for the latter, it is well known that delirium causes substantial distress^{28,29} and that, to some extent, clear communication of the diagnosis can ameliorate this. More broadly, detecting cognitive impairment in general (delirium, dementia, 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 the patient and their families, carrying out careful assessment of the 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.⁸ Detection of dementia has recently been highlighted in the Dementia Commissioning for Quality and Innovation (CQUIN) framework;³⁰ crucially, establishing if the patient has a 'clinical diagnosis of delirium' is a central element of the FAIR (Find, Assess, Investigate, Refer) algorithm at the heart of this framework.

Underdetection of delirium

Ample evidence^{7,31-33} shows that in medical, surgical, emergency department (ED) and palliative care settings, the majority of delirium is not diagnosed. It is unclear why detection rates are so low. A variety of types of study, including surveys of practitioners' attitude and knowledge, studies of the content of practitioners' education and training, and qualitative studies, have attempted to understand why such a common and serious condition remains so neglected in modern health care. One of the commonly identified issues is that staff are not certain about the diagnostic criteria for delirium, and, related to this,

what screening and/or diagnostic tools should be used. In a 2009 survey³⁴ of 784 UK trainee physicians, only 21% stated that they had a good knowledge of the diagnostic criteria for delirium, and only 8% reported using specific screening tools for delirium. A similar survey³⁵ recently repeated showed some improvements, but gave broadly similar results. A survey³⁶ in 2013–14 of undergraduate medical education found that, although delirium was included in teaching, the approach was highly heterogeneous and likely to lead to inconsistent knowledge. Qualitative analyses^{37,38} of practitioners in organisations have identified significant issues of attitudes towards ownership of the patient as well as other important factors in whether or not a patient with possible delirium undergoes an assessment for the condition.

Previous work shows that there are many reasons for delirium underdetection that act at many levels, from undergraduate education training to whole systems. These findings have pointed to several remediable factors. However, there is ongoing uncertainty around the present state of delirium assessment practice in the UK, in particular how different disciplines' approaches to detecting delirium are underexplored.

One obvious conclusion from the existing literature is that, to achieve consistently high rates of delirium detection in routine care, many practitioner and systems elements, or building blocks, need to be present and operating effectively. One of the essential elements is having assessment tools available that are validated and practical for the context in which they are to be used, and that are implemented consistently. Many diagnostic instruments make use of the *Diagnostic and Statistical Manual of Mental Disorders (DSM)/International Classification of Diseases*³⁹ criteria, but these have largely remained as research tools. For example, the short Confusion Assessment Method (CAM)⁴⁰ is commonly advocated in pathways and guidelines, but it requires specific training and takes 5–10 minutes to complete because it is preceded by a cognitive assessment and a brief interview.⁴¹ In many clinical units, the completion of such a test takes too long for practical use.

Rapid assessment of delirium presents several challenges, one of which is that patients show a wide range of levels of severity; many patients in acute settings are too unwell, too sleepy or too agitated to undergo cognitive testing or an interview.^{42–47} This problem of 'untestability' is likely to be another important factor in delirium underdetection. Most cognitive or 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 of these patients, and so no diagnosis, and often no specific treatment, is applied.

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. Therefore, the absence of a combined instrument that allows screening for both general cognitive impairment and delirium may contribute to the lack of specific delirium detection.

The need for a new assessment tool

Given the multiple constraints of the acute environment, the range of staff who might be expected to screen for delirium, the common co-existence of delirium and dementia, and the heterogeneity of patients, we judged that a screening tool should have these features:

1. be short (< 2 minutes)
2. be easy to learn
3. be easy to administer and score
4. be able to be used by professional-level health-care staff from a variety of disciplines
5. allows scoring of patients who are too drowsy or agitated to undergo cognitive testing or clinical interview
6. take account of informant history
7. be able to be administered through written questions to people with severe hearing impairment
8. be able to be administered to patients with visual impairments
9. not require subjective judgements based on interview

10. combine delirium screening with general cognitive screening
11. not need a quiet environment for administration
12. not require physical responses, such as drawing figures or clocks.

There are multiple instruments for delirium screening, diagnosis, severity assessment and monitoring.⁴⁸⁻⁵¹ Therefore, before deciding to design a new screening tool, we 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 duration alone. The remaining scales lacked important features such as general cognitive screening. Thus, we 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 National Institute for Health and Care Excellence (NICE) guidelines on delirium,⁴ which emphasised the need for research on a screening tool for delirium that would be suitable for routine use.

The subsequent design process involved scrutiny of each of the nearly 30 delirium assessment tools published in 2010, 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.^{46,52-54} 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 it is 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 the light of this, we decided that the new test should include cognitive screening sensitive to general cognitive impairment, but also include items on altered level of alertness and change in mental status, both of which are strong indicators of delirium. The new test was named the 4 'A's Test, or 4AT, with the 'A's standing for Alertness, Abbreviated Mental Test – 4, Attention (Months Backwards test), and determination of Acute Onset (see *Table 4* and www.the4AT.com). The first version of the 4AT was drafted and tested informally by colleagues; changes were made based on feedback, and updated versions were tested again. After several iterations, involving 20 doctors and nurses with varying levels of experience, the final version was produced. An initial audit (unpublished) in 30 inpatients found encouraging results.

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, with generally positive feedback from users (see *Chapter 3*). Several validation studies have been published, showing supportive results, with satisfactory diagnostic accuracy (see *Chapter 5*).⁵⁵⁻⁶⁰ An additional study also assessed the usability of the 4AT in the palliative care setting, finding positive results.⁶¹ However, there is still a need for further diagnostic accuracy studies to inform clinicians about which tools to use in which contexts, especially given the large number of tools and validation studies of such tools available.

Health economics

This study also wanted to compare the implications of any differences in diagnostic accuracy between the CAM and the 4AT with respect to health-care costs. Delirium is known to be economically costly,^{21,62} and the prevention of delirium is cost-effective.^{63,64}

Overview of project

This project has the overarching goal of assessing the need for further development and then testing the diagnostic accuracy of the 4AT delirium assessment tool. The work is divided into two phases, and there is a health economics project that ran in parallel with the second phase.

Phase 1 uses practitioner surveys, qualitative studies and information from the existing use of the 4AT in clinical practice to assess usability and to determine whether or not any changes are needed to the current structure and scoring of the tool. The aims of phase 1 were, specifically, to understand more about health-care professionals' knowledge of and attitudes to delirium assessment, using survey and qualitative study methodology. We also specifically surveyed practitioners who had experience of using the 4AT. The main aim of phase 1 was to determine if changes needed to be made to the 4AT in advance of phase 2.

Phase 2 is a diagnostic accuracy study with the primary objective of determining the diagnostic accuracy of the 4AT for delirium detection. The secondary objectives of phase 2 are to compare the performance of the 4AT with that of the CAM (because the CAM is recommended in some guidelines, and it is of value to know if the 4AT performs similarly despite being shorter); to examine the performance of the cognitive test items in the 4AT in detecting general cognitive impairment; to determine if the 4AT predicts length of stay, new institutionalisation, and mortality at 12 weeks; and to examine the contribution of individual items of the 4AT to overall delirium diagnosis.

In the health economics study, conducted in parallel with phase 2, we wanted to estimate the economic implications of differences in diagnostic accuracy between the 4AT and the CAM. This was based on first estimating the costs of delirium care over 12 weeks. From this, the costs of true positives, true negatives, false positives and false negatives were estimated, and these were applied in the context of a model of the diagnostic test accuracy results of the 4AT.

Chapter 2 Surveys of current practice

General introduction

Although the severe impacts of delirium are becoming clear,^{8,10,15,65,66} it has been amply documented that delirium remains substantially underdetected both in general settings and in the intensive care unit (ICU).^{7,16,31–33,67–72} There are many possible reasons for this, which surveys on practice and knowledge among professionals have helped to shed some light on.^{34,36,73–93} The surveys vary considerably in the mix of professionals studied, the focus of the questions, the health-care setting, and so on. The majority of published surveys relate to ICUs alone, with relatively small numbers involving health-care practitioners from other settings. Most surveys have primarily focused on the attitudes, clinical practice and knowledge of various health-care practitioners with respect to delirium management, rather than attention on delirium detection and assessment.

Some notable findings from the surveys relevant to detection assessment follow. Most respondents across disciplines believe that delirium is underdiagnosed^{34,78,86,89} and that the treatment of delirium is important.^{74,78,83,89} Yet surveys frequently indicate that basic knowledge of delirium is inadequate.^{34–36,77,90,91,93,94} In one study,⁷⁷ 42% of nursing and medical staff working in one of three Scottish ICUs did not know that delirium was associated with an increased 6-month mortality rate and 44% had never received training on ICU delirium. In a survey of UK surgical training doctors,⁹⁰ only 2% were familiar with the diagnostic criteria for delirium. Published guidelines in the UK⁴ recommend daily screening in ICUs using validated delirium screening tools. Despite this, surveys generally indicate that a small minority of practitioners routinely screen for delirium using tools.^{34,77,78,82,83,85–87,89,93,94}

Greater knowledge of attitudes and clinical practice with respect to detection and assessment is essential in identifying factors that contribute to the underdetection of delirium and could help in the development of effective screening tools, education and clinical implementation strategies. Currently, although the existing surveys provide useful information, there remains a lack of understanding of the specifics of delirium assessment in general settings and in different disciplines.

In this chapter we describe two survey studies. The aim of the first, survey A, was to gain more understanding about the knowledge of and attitudes to delirium in general. The survey was in four practitioner groups (medical practitioners, nurses, occupational therapists and physiotherapists) working in a variety of inpatient and outpatient settings in the UK.

The second survey, survey B, was distinct from survey A in that it had a specific aim of gauging attitudes to and potential issues around the use of the 4AT. This addressed both the practical use and other aspects. The aim of the survey was to identify any potential changes to the instrument, or to its guidance notes, that could improve its usability before its evaluation in the diagnostic accuracy study. There was some conceptual overlap with survey A, but survey B was specifically aimed at understanding more about the use of the 4AT.

Methods

Survey development

The surveys (survey A is in *Appendix 1* and survey B is in *Appendix 2*) were developed in multiple stages. They were initially drafted in web form and revised by Antaine Stiobhairt, Alasdair MJ MacLulich and Susan D Shenkin based on previous literature and personal clinical experience of delirium. The revised surveys were reviewed by members of the study team, who assessed face validity, structure and clarity,

and further revisions were made. Subsequently, survey A was piloted with 19 additional health-care practitioners outside the study and survey B was piloted with five. In each case, participants were asked to comment on content and any technical problems. Minor amendments were made to both surveys.

In both surveys, the items were presented in a fixed order (with no randomisation) and included multiple choice, ranking, five-point Likert scale and open-comment response formats. Response options including 'N/A' (not applicable), 'don't know' and 'other (please specify)' were provided throughout the survey, and the majority of questions were optional in order to minimise response coercion and attrition. The surveys began with six demographic items, two of which (career stage and specialty) were presented only to respondents who had a primary qualification in medicine. Some additional questions were presented to those with this qualification (see *Appendix 1*). Both surveys finished with an open-comment box in which respondents were invited to 'comment on any of the issues raised in this survey or additional issues surrounding the detection and assessment of delirium that have not been addressed'. No incentives were offered for survey completion.

The surveys were considered to be service evaluations, as the participants were anonymous health-care practitioners and the surveys were of current practice; thus, formal research ethics approval was not required. The surveys were hosted on the internet using Survey Monkey (www.surveymonkey.com). Invited recipients who clicked on the hyperlinked URL in the e-mail were initially presented with a web page that described the study in greater detail and explained that participation was anonymous, that no computer location information or cookies would be collected, and that the results would be published and may involve direct quotations. Potential respondents were then presented with the question 'Do you agree to participate in this survey and consent to the potential use of your anonymised responses as described above?'. Those who chose 'I agree' proceeded to the survey, whereas those who chose 'I do not agree' were directed to a page that explained that they had to agree if they wanted to participate. As it was anonymised, the survey did not allow for checking for multiple responses from single users. No time-stamp data for individuals were analysed.

Participants

For Survey A, the health-care practitioners of interest were medical practitioners, nurses, occupational therapists and physiotherapists working in the UK who came into contact with delirium as part of their daily routine. These professional groups were selected as they were numerically large and came into frequent contact with patients with delirium. A list of e-mail addresses of potential respondents was generated through networks of study authors and through internet searches focused on, but not limited to, EDs, ICUs, acute assessment units/medical assessment units (MAUs), elderly care, orthopaedics, oncology, stroke units and palliative care, as delirium is common in these settings. Contact details of relevant associations/societies and trusts/health boards were also obtained through internet search engines.

A standard invitation e-mail was sent to individual practitioners on 30 September 2014. Amended versions of this e-mail were sent to trusts/health boards across the UK, and to relevant societies/associations (geriatrics, psychiatry, acute medicine, nursing, palliative care, etc.). All recipients were asked to forward the e-mail to relevant practitioners in their own contacts list and to consider displaying an A4 poster highlighting the survey in their staff areas. Reminder e-mails were sent on 21 October 2014 and 18 November 2014. Each of the study collaborators forwarded the e-mail to their contacts informally during this time. The survey was closed on 3 January 2015, by which time participation had tapered off.

Survey B was intended for any health-care practitioners worldwide who had used the 4AT to screen patients for delirium. A convenience sample of suitable practitioners was obtained through multiple methods. A standard invitation e-mail was distributed to practitioners working in units where the 4AT was known to be in use on 4 November 2014. Practitioners who completed survey A (Delirium Detection and Assessment in Clinical Practice), who had confirmed that they had used the 4AT and who had provided their e-mail address were e-mailed on 6 November 2014. All recipients were asked to forward the e-mail to 4AT users in their own contacts list and to consider displaying an A4 poster highlighting the survey in

their staff areas. Reminder e-mails were sent on 2 December 2014. Each of the study collaborators forwarded the e-mail to their contacts informally during this time. Posters were also displayed in staff areas of units in the Royal Infirmary of Edinburgh that use the 4AT, and paper slips containing the link were handed to practitioners informally. The survey was closed on 22 January 2015, by which time participation had tapered off.

Data analysis

Quantitative data were analysed using R version 3.0.2 (2013; R Core Team, The R Foundation for Statistical Computing, Vienna, Austria). The majority of the data had non-normal distributions and heterogeneous variance across groups; therefore, medians and interquartile ranges (IQRs) are reported throughout and non-parametric statistical tests were used.

For comparative data in survey A, the threshold for statistical significance was a p -value of < 0.05 . Between-group analyses were carried out using chi-squared tests where both variables were categorical, using Kruskal–Wallis tests with Holm–Bonferroni method followed by Mann–Whitney U -tests. For items whose responses were given on five-point Likert scales, verbal responses were converted to the corresponding numeric responses 1–5 to facilitate between-group analyses, and analysed as continuous variables. Mean scores and standard deviations (SDs) were reported for these converted responses, as the median and IQR rarely varied from 3 and 2–4, respectively, owing to the restricted scoring range of 1–5. Between-group analyses involving work settings were carried out on participants who worked in single inpatient settings only, in order to avoid the possible confounding effects of working in multiple settings. Owing to the large number of significant pairwise comparisons, details of significant effects were not reported in text when the effect size $r < 0.15$, but were included in tables only. Unanswered items and sample attrition resulted in missing data and fluctuating sample sizes across and within items; therefore, the results are based on the total number of responses for each item, excluding cases of missing data and when respondents chose ‘N/A’ or ‘don’t know’, unless stated otherwise. No statistical corrections were applied.

Qualitative data submitted through open-comment fields were reviewed informally and the decision on which quotations to report was based on what was judged to contribute to improve the implementation or refinement of the 4AT.

Results

Survey A

A total of 2671 practitioners agreed to participate in the survey. Of those, 172 stopped before completing the core demographic questions; 137 completed these questions but did not continue; 12 worked outside the UK; 41 had a primary professional qualification in an area other than medicine, nursing, occupational therapy or physiotherapy; for two their primary professional qualification could not be verified; and one was retired. These cases were excluded. Data from 2306 (86%) respondents were retained. A summary of the respondent characteristics is shown in *Table 1*.

All estimates relate to the frequency of respondents except where stated otherwise. Numbers in cells pertaining to settings are not equal to the total sample size, as respondents could choose multiple settings.

Awareness of delirium

Respondents were asked whether or not they thought that there had been an increase in awareness of delirium among colleagues in their specialty in the previous 3 years: a large majority of 83% (1392/1680) said ‘yes’. Post hoc analysis revealed significant differences across inpatient settings ($\chi^2(6) = 84.9$; $p < 0.001$); increased awareness was reported most frequently by respondents working in intensive care (95%, 149/157) and acute inpatient medical settings (90%, 534/593), and least frequently by those working in EDs (68%, 141/208) and rehabilitation units (67%, 28/42). The majority perceived increased mentioning of delirium in professional domains such as professional conferences (77%, 1076/1391), clinical journals (77%, 1101/1426),

TABLE 1 Survey A: respondent characteristics

Characteristic	Medicine (tertiles of years of experience)				Overall	Specialty			Total
	Unknown	Lower	Mid	Upper		Nursing	Occupational therapy	Physiotherapy	
Frequency (% of sample)	5	145	271	680	1101 (47.75)	668 (28.97)	271 (11.75)	266 (11.54)	2306 (100)
Years qualified, median (IQR)	16 (8–26)	2 (1–3)	8 (6–10)	19.50 (14–25)	14 (7–22)	21 (12.75–30)	10 (5–16)	12 (6–20)	15 (8–24)
Setting (n)									
Acute inpatient medical	1	83	124	253	461	224	162	155	1002
ED	1	30	96	183	310	27	14	20	371
Intensive care	1	10	18	81	110	90	2	25	227
Rehabilitation	0	3	12	44	59	35	46	62	202
Surgical ward (exc. ortho.)	2	6	8	24	40	55	11	25	131
Palliative care	0	1	3	20	24	28	4	3	59
Psychiatry	1	14	16	60	91	83	20	12	206
Other ^a	2	19	24	139	184	214	75	75	548

a A total of 436 out of 548 chose 'other' alone. Of these, 202 worked in the community (general practitioner, $n = 48$; mental health, $n = 58$; rehabilitation, $n = 12$; and various, $n = 84$), 50 worked in inpatient settings (mental health, $n = 13$; anaesthetics, $n = 10$; high-dependency unit, $n = 5$; various, $n = 22$) and 24 worked in outpatient settings (rehabilitation, $n = 3$; various, $n = 21$). For the remaining 158 respondents, it was unclear whether they worked in inpatient, outpatient or community settings (mental health, $n = 49$; research, $n = 10$; intermediate care, $n = 10$; various, $n = 89$).

Acute inpatient medical comprises acute assessment unit/MAU, medicine of the elderly, orthopaedics, oncology, stroke, internal medicine specialist ward (e.g. cardiology, respiratory, gastroenterology, nephrology, endocrinology, neurology). Surgical ward (exc. ortho.) refers to surgical ward not including orthopaedics.

clinically related websites (77%, 1043/1351), training events (76%, 1221/1603) and training curriculums (71%, 1010/1406). By contrast, much lower proportions perceived an increase in general media coverage (e.g. BBC News, newspapers; 30%, 440/1489) and coverage on social media (for respondents who were users), such as Twitter (www.twitter.com; Twitter, Inc., San Francisco, CA, USA) (35%, 203/575) and Facebook (www.facebook.com; Facebook, Inc., Menlo Park, CA, USA) (13%, 88/675). Responses given in an open-comment field following this question focused on increased awareness of delirium in the workplace, with references made to the promotional work of individuals or teams, internal teaching sessions, in-house/ on-the-job training, information displayed in clinical areas, and awareness raised indirectly through educational drives for sepsis and frailty. The remaining comments focused on increased awareness gained through NICE guidelines and e-learning modules.

Detection of delirium

When respondents were asked their opinion on what percentage of patients with delirium in their unit had their delirium diagnosed and documented, using the bandings '0–20%', '21–40%', '41–60%', '61–80%' and '81–100%', the responses for each banding were 14% (275/2010), 18% (352/2010), 22% (449/2010), 26% (526/2010) and 20% (408/2010), respectively. This suggests that those surveyed feel that there is substantial underdiagnosis of delirium, with a minority of units perceived to be achieving at least 80% (estimated) diagnosis rates. Estimates of diagnosis rates differed significantly by clinical area, with EDs showing the lowest estimates (mean 2.6, SD 1.12) and acute inpatient medical and rehabilitation settings showing higher estimates (mean 3.45, SD 1.22, and mean 3.6, SD 1.3, respectively). Overall, respondents believed that the largest contributor to the underdiagnosis of delirium in their units was 'difficulty discriminating delirium from dementia', with 44% (907/2072) stating that this made a 'large' or 'very large' contribution. This was followed by 'lack of staff confidence in assessment' (39%, 809/2050) and 'lack of staff knowledge of delirium' (39%, 817/2083). Notably, the difference between all of these contributors is limited to a range of 11%, and more than one-third of respondents considered each of the listed contributors to have a 'large' or 'very large' effect, indicating that respondents consider a complex combination of factors to be responsible for the underdiagnosis of delirium.

Regarding the role of different disciplines in 'flagging potential cases', 'screening high risk patients' and 'making a formal diagnosis', responses indicated that virtually the whole sample considered doctors to be responsible for diagnosis, with other disciplines considered to have a much smaller role in this. Nurses were considered to have the most important role in screening high-risk patients, whereas physiotherapists and occupational therapists were considered to have the main responsibility for 'flagging potential cases' (Table 2).

Attitudes to the importance of making a formal diagnosis of delirium

Respondents were asked to indicate their agreement with three statements concerning delirium care. When asked to indicate their agreement with the statement 'making a formal diagnosis of delirium is important to provide good delirium care', 5% (92/1999) of respondents chose 'strongly disagree' or

TABLE 2 Survey A: respondents' answers when asked to indicate which of the three tasks different professional groups should consider their own duty with regard to delirium detection

Professional group	Task		
	Flagging potential cases	Screening high-risk patients	Making a formal diagnosis
Doctors	43% (868/2004)	56% (1113/2004)	96% (1926/2004)
Nurses	67% (1322/1983)	78% (1553/1983)	23% (451/1983)
Occupational therapists	83% (1619/1947)	46% (887/1947)	10% (191/1947)
Physiotherapists	88% (1738/1984)	32% (633/1984)	7% (130/1984)

All estimates are shown as percentage (proportion) of the total responses for each practitioner type.

'disagree' and 5% (98/1999) chose 'neither agree nor disagree', whereas 91% (1809/1999) chose 'agree' or 'strongly agree'. When asked to indicate their agreement with the statement 'distinguishing between delirium and dementia is important in providing good care', 4% (73/2001) chose 'strongly disagree' or 'disagree' and 3% (69/2001) chose 'neither disagree nor agree', whereas 93% (1859/2001) chose 'agree' or 'strongly agree'. Finally, when asked to indicate their agreement with the statement 'delirium treatment improves patient outcomes', 4% (76/2000) chose 'strongly disagree' or 'disagree' and 7% (146/2000) chose 'neither disagree nor agree', whereas a vast majority of 89% (1778/2000) chose 'agree' or 'strongly agree'.

Personal practice

When respondents were asked to rate their level of confidence in their own ability to detect delirium, 15% (303/2064) stated that this was 'very low' or 'low', 78% (1600/2064) stated that it was 'moderate' or 'high' and 8% (161/2064) stated that it was 'very high'. Post hoc analysis revealed significant differences in mean scores (very low = 1; low = 2; moderate = 3; high = 4; very high = 5) between inpatient settings [$\chi^2(6) = 102.6; p < 0.001$]. Respondents working in psychiatry ($n = 165$; mean 3.8, SD 0.84) and those working in intensive care ($n = 163$; mean 3.7, SD 0.76) were more confident than those working in surgical wards ($n = 66$; mean 3.0, SD 0.93; $p < 0.001$ for both) and those working in rehabilitation units ($n = 49$; mean 2.8, SD 0.9; $p < 0.001$ for both).

When respondents were asked whether or not they had ever used a tool (i.e. a specific tool or cognitive test) to detect delirium, 54% (1103/2061) stated 'yes'. Significant differences were also revealed between inpatient settings [$\chi^2(6) = 116.3; p < 0.001$]: 90% (147/163) of respondents working in intensive care, 65% (107/164) working in psychiatry and 58% (425/729) working in acute inpatient medical settings reported having used a tool to detect delirium. Rates of tool use ranged from 33% to 47% among respondents working in other settings. The most frequently used delirium screening tools used were the CAM (61%, 630/1041), the 4AT (60%, 625/1043) and the CAM for the ICU (42%, 430/1044). The most widely used cognitive test was 'orientation to time, place, person' (91%, 940/1028), followed by the AMT10 (Abbreviated Mental Test) (75%, 770/1026) and the Mini-Mental State Examination (MMSE) (74%, 762/1028), respectively. When medical practitioners were asked whether or not they thought that patients needed to be sufficiently conscious to produce verbal responses so that a bedside assessment for delirium could be undertaken, 72% (769/1073) stated 'yes'. A majority of medical practitioners (88%, 947/1077) reported that they 'frequently' or 'almost always/always' sought a history of mental status from collateral sources [e.g. family or general practitioner (GP)] for patients with cognitive impairment. Both individual and organisational practice in recording delirium varied, with a range of terms used for patients likely to have delirium.

When asked whether or not guidelines for delirium detection existed in their units, 64% (1021/1605) reported that these did exist. Of those who said that guidelines existed, 22% (211/953) thought that these were 'never/rarely' or 'sometimes' followed, 58% (554/953) thought that they were followed 'about half of the time' or 'frequently', and 20% (188/953) thought that they were 'almost always/always' followed.

Terminology

Medical practitioners were asked to indicate the term they would be most likely to apply in practice to a patient who 'presents with recent onset drowsiness and is not producing verbal responses but is responding intermittently to one-stage commands'. The term 'delirium' was chosen by one-third of respondents (35%, 386/1090). This was followed by 'obtundation' (23%, 248/1090), 'stupor' (15%, 166/1090), 'encephalopathy' (4%, 41/1090) and 'coma' (3%, 36/1090). Among the 20% (213/1090) of medical practitioners who chose to specify their own term, a majority ($n = 136$) said altered, decreased or fluctuating 'consciousness' or said that they would refer to the patient's score on the Glasgow Coma Scale. Other repeated terms included 'confusion', 'acute confusion' and 'acute confusional state' ($n = 12$). Twenty-seven respondents stated that they would need further information on the wider clinical context to be able to make a judgement, and nine

stated that they would not apply a label at all, but rather would describe the patient's symptoms. Although almost two-thirds of medical practitioners chose a term other than delirium, in a follow-up question two-thirds (66%, 732/1101) stated that the same patient was 'likely' or 'very likely' to have delirium, indicating that terms are considered somewhat interchangeable by many medical practitioners. By contrast, 6% (67/1101) stated that it was 'very unlikely' or 'unlikely', and 26% (288/1101) stated that it was 'neither likely nor unlikely', that this patient had delirium.

The term used most often in respondents' units to describe patients with 'acute deterioration in cognition or other mental functions caused by an acute medical problem, drug side-effects or other acute causes' was 'confusion', with 61% (1250/2046) stating that this term is 'frequently' or 'almost always/always' used in their unit. This was followed by 'acute confusional state/acute confusion' (45%, 1032/2306), 'delirium' (33%, 682/2058) and 'septic encephalopathy' (2%, 48/1951), respectively. The majority of the 129 participants who chose to report alternatives recorded informal descriptions such as 'agitated', 'knocked-off', 'muddled', 'disorientated', 'withdrawal' and 'drowsy'. Specific terms reported included 'acute on chronic confusion', 'cognitive impairment', 'acute cognitive impairment' and 'dementia', with some practitioners stating that these terms were often misused in their units.

Survey B

A total of 117 practitioners agreed to participate in the survey. Fourteen stopped before completing the core demographic questions (items 1–4) and three stated that they had never used the 4AT in clinical practice. These cases were excluded, giving a final sample of 100 (88%) respondents. The geographical distribution was Scotland ($n = 62$), England ($n = 28$), Italy ($n = 5$), the USA ($n = 3$) and Australia ($n = 2$). *Appendix 3* gives a summary of respondent characteristics.

In addition to the 4AT, a large proportion of respondents reported that they had used the CAM (68%, 61/90). Among the 28 respondents who named additional tools they had used, the Abbreviated Mental Test ($n = 15$) and the MMSE ($n = 8$) were the two most commonly mentioned (see *Appendix 4*).

All respondents reported having at least moderate confidence in their ability to detect delirium (moderate 28%, 28/100; high 52%, 52/100; and very high 20%, 20/100). The interval from which the 4AT was first used by the individual to the time of completing the survey varied across the sample ['< 1 month', 10% (9/90); '1–6 months', 21% (19/90); '7–12 months', 18% (16/90); and '> 1 year', 51% (46/90)].

A detailed summary of user opinions on the 4AT, collected as part of survey B, is provided in *Appendix 5*. Regarding general opinions on the 4AT, 84% (59/70) had positive views, 14% (10/70) had neutral or mixed views and one respondent (1/70) had a negative view. When respondents were asked how often they used the 4AT with patients at risk of delirium, 33% (30/90) stated 'never/rarely' or 'sometimes', 14% (13/90) said 'about half of the time' and 52% said 'frequently' or 'almost always/always'. When respondents were asked if they thought that using the 4AT as part of routine assessment was feasible in their unit, most (95%, 81/85) said 'yes', with many referring to the ease (39%, 15/38) and speed (29%, 11/38) of use. One question addressed the extent of knowledge of delirium that respondents thought necessary for health-care practitioners to have in order to use the 4AT effectively: 7% (6/84) stated 'none/very little', 58% (49/84) stated 'some' or 'a moderate amount' and 35% (29/84) stated 'quite a bit' or 'an extensive amount'. With respect to specific training in use of the 4AT for health-care practitioners to be able use the tool effectively, 17% (14/84) stated 'none/very little', 60% (50/84) stated 'some' or 'a moderate amount' and 24% (20/84) stated 'quite a bit' or 'an extensive amount'.

When respondents were asked whether or not the 4AT was used as part of routine assessment by them or others in their unit, 69% (57/83) said 'yes', with two respondents commenting that they were currently carrying out audits of 4AT use in their units. Regarding barriers to using the 4AT in respondents' units, several were identified (*Table 3*). The use of an alternative tool was not considered by most to be a

TABLE 3 Survey B: the extent to which respondents thought that various barriers were preventing the 4AT from being used more regularly in their units

Barrier	Extent, % (frequency)				
	Very small	Small	Moderate	Large	Very large
Time constraints	9.64 (8/83)	36.15 (30/83)	27.71 (23/83)	15.66 (13/83)	10.84 (9/83)
Existing use of/familiarity with an alternative tool	37.35 (31/83)	24.10 (20/83)	24.10 (20/83)	12.05 (10/83)	2.41 (2/83)
Lack of staff confidence in using the tool	8.43 (7/83)	28.92 (24/83)	34.94 (29/83)	19.28 (16/83)	8.43 (7/83)
Lack of perceived need to use a delirium screening tool	7.23 (6/83)	16.87 (14/83)	32.53 (27/83)	37.35 (31/83)	6.02 (5/83)
Lack of staff knowledge of delirium	6.02 (5/83)	28.92 (24/83)	40.96 (34/83)	21.69 (18/83)	2.41 (2/83)

significant barrier. Opinions on other barriers showed a broad distribution across the sample, suggesting that implementation of delirium assessment tools is complex. Two particular comments from the free-text comments emphasised that the extent of 4AT use depends on broad systemic and cultural factors and not simply on the merits of its utility:

We tried it but then moved back to AMT + CAM as AMT already performed by Nursing staff as part of basic assessment.

Consultant, ED/MAU

We have done some improvement work with it but there was some resistance from colleagues about using it as a screening tool.

Speciality trainee level 3 or above (ST3+), MAU/Medicine of the elderly (MoE)

To gather information about ease of use of the 4AT, respondents were presented with descriptions of three patient groups and asked to indicate their typical experience of using the 4AT with each group (see *Appendix 2*). For 'drowsy patients who cannot produce verbal responses', 51% (36/70) reported that using the 4AT was 'very easy' or 'easy', 17% (12/70) reported that this was 'neither easy nor difficult' and 31% (22/70) reported that this was 'difficult' or 'very difficult'. For 'patients with dementia who are alert and able to converse', 77% (60/78) reported that this was 'very easy' or 'easy', 17% (13/78) reported that this was 'neither easy nor difficult' and 6% (5/78) reported that this was 'difficult' or 'very difficult'. For 'patients who are agitated and distressed', 44% (33/75) reported that this was 'very easy' or 'easy', 20% (15/75) reported that this was 'neither easy nor difficult' and 36% (27/75) reported that this was 'difficult' or 'very difficult'.

Respondents were also asked how long items 1–3 (alertness, abbreviated mental test-4, attention) of the 4AT typically take to complete: 12% (10/81) stated '< 1 minute', 54% (44/81) stated '1–2 minutes' and 33% (27/81) stated '3+ minutes'. Some respondents added that the time taken is 'affected by deafness', that they usually have a 'conversation at the same time' and that 'longer time usually hints at worse performance'. Most of those who commented explained that the time taken to obtain collateral history from patients' family, carers, GP or medical records was highly variable. These findings confirm that the bedside components are brief in most patients but that item 4 (determination of acute onset) can be time-consuming.

A detailed account of responses to survey B questions along with examples of free-text responses is reported in *Appendix 3*. In addition to giving feedback on the 4AT itself, respondents were asked whether or not they would suggest making any changes to it. The majority did not propose any changes, with 6–10 comments on each item and the guidance notes. Respondents' comments could be grouped into the following themes: changes to item content, clarity, visual presentation, the scoring system and general comments.

Examples are shown in *Box 1*. There were also some queries about validity and diagnostic accuracy of the individual items. Similarly, a small number of respondents suggested including 'time' alongside the four items of the AMT4.

The findings of survey B were discussed with members of the study team, and considered in the light of other information, including additional external validation studies that had been published, the clinical service evaluations that had been collected as part of the study process, and other feedback about the 4AT from outside the study. The team decided that, in the absence of consistent feedback concerning a potential specific change to the 4AT, the study process had determined that no change to the 4AT was required in advance of the diagnostic accuracy study.

BOX 1 Examples of suggested changes to the 4AT from free-text comments in survey B

Changes to item content

A number of respondents suggested making changes to the content of the items. One respondent proposed using the AVPU scale for 'alertness'. Others suggested greater specificity:

'Clearly abnormal' automatically scores 'possible delirium'. Perhaps break this down to, for example, 'comatose, drowsy, agitated' and provide relevant scores.

FY1–2, surgical ward

There should be a box for hyperactive/over stimulated behaviour.

Nurse (30 years), MoE

Some respondents reported that they had experienced difficulties when asking patients to state the months of the year backwards, and suggested including an alternative measure of attention:

Make other suggestions for different types of attention questions as some patient are illiterate and can't do the months of the year backwards.

Nurse (10 years), MoE

[A]dvice on test to use in non verbal or aphasic patients or in non-english-speaking patients.

Consultant, MoE

[P]eople sometimes refuse to do, so I resort to using 20-1.

Consultant, MAU/immediate care

Clarity

In reference to item 1, multiple respondents pointed out that there were no examples of what constituted 'clearly abnormal'. Some respondents felt that item 4 needed to be clearer about what exactly is being assessed and how it should be assessed:

More emphasis on 'is this their normal behaviour or has there been a change'.

Consultant, MoE/stroke

A lot of staff think its an acute and fluctuating change in their physical condition not the cognition! Maybe it needs to be explained more or maybe its a teaching problem we need to take on board.

Nurse (10 years), MoE

BOX 1 Examples of suggested changes to the 4AT from free-text comments in survey B (continued)

I would change the description from over the 'last 2 weeks' to 'in the last few days or weeks' . . . to take account of those with a delirium who have been treated in the community but have not improved.

Nurse (30 years), MoE

There is a need to be explicit about what to do if there is no one to give a history.

Nurse (30 years), MoE

Visual presentation

A number of comments focused specifically on the visual presentation of information on the document. For example:

I'd put [item 1] as Item 3 or 4 - I assess it throughout rather than as a stand-alone item at the start. Rarely am I just doing the 4AT, usually it's part of a more general assessment.

Consultant, MoE

Scoring system

Some respondents suggested making changes to the current scoring system. For example:

I wonder if the timeline of 2 weeks would capture most/all cases of delirium esp if there has been a longer duration of the index episode. Could a breakdown of the absolute point score of '4' currently given for a 'Yes' response, be subdivided into a score of '2' for between 2–4 weeks; and perhaps as score of '4' for < 2 weeks?

Consultant, MoE/rehab

Alertness - 0,0,1,3. Feels more realistic and reflective of variation. AMT4 - 0,0,1,2,3. I do not believe that 1 mistake deserves a score at all. It would be fairly easy with normal brain function even in young patients to make a single mistake. A score of 2 would be from my AMT5 (including 'time' to nearest hour).

ST3+, MAU/MoE

General comments

One respondent suggested the need to provide more information about the clinical significance of the results in the guidance notes:

There is some confusion about repeated testing and if this means the delirium is resolving. Capturing that would be very helpful, like a MEWS [Modified Early Warning Score].

Consultant, MoE/stroke

Another respondent highlighted an issue that might be of growing concern as hospitals gradually go 'paper light'. In reference to item 4, they stated:

This part of the test is frequently the most inaccurately completed part of the test. However this is due to the trust version of the test has been created on a computer and the guidance sentence underneath is not present.

Nurse (5 years), MoE

Discussion

Survey A

Survey A had a substantial number of respondents and is, to our knowledge, the largest survey on delirium to date. Because of the nature of the recruitment process, respondents were likely to have had an interest in delirium and/or to have been interested in the management of cognitive impairment more generally. Nevertheless, given the sample size, the results do suggest that there is increasing awareness of delirium among hospital staff in the UK. Additionally, respondents indicated that the majority of units in which they worked had guidelines for delirium detection. These findings indicate potentially positive trends in improving delirium detection and, therefore, care in the UK. However, other findings suggest that substantial challenges remain.

Nine out of 10 of respondents agreed or strongly agreed that formal diagnosis of delirium is important, that distinguishing between delirium and dementia is important, and that delirium treatment improves patient care. Although these data are from a potentially biased sample of health-care professionals, they support the notion that a substantial proportion of practitioners believe that delirium is worth diagnosing. This view in the context of the well-documented poor rates of delirium detection demonstrates a challenging paradox in mainstream clinical practice.

Some of the findings of this study help to address the question of why this paradox exists. One of the striking results is the range of terminology used to describe a patient with 'recent onset drowsiness' who is 'not producing responses but is responding intermittently to one stage commands'. This clinical scenario is clearly consistent with a diagnosis of delirium according to DSM-5 criteria and the accompanying guidance notes, especially considering that delirium is much more common than alternative diagnoses. Yet only one-third of respondents favoured the formal term 'delirium' being applied to this case, with several non-diagnostic, ill-defined terms, such as 'obtundation', collectively being suggested more often. Notably, a follow-up question asked respondents if the patient described at the beginning of this paragraph had delirium, to which two-thirds of respondents responded that this was 'likely' or 'very likely'. Additionally, a different question directed at medical practitioners found that 72% believed that the patient needed to be able to produce verbal responses to allow bedside assessment of delirium; this view is not aligned with guidance in the DSM-5, which deems it possible to assess delirium in non-comatose patients with acute mental status disturbance who are incapable of speech. These findings indicate that there is remarkable inconsistency in the terms used to describe the clinical states most compatible with an initial diagnosis of delirium. The results in relation to the variable terminology used in the organisations of respondents parallel the findings from individual practitioners. The consequences of using inconsistent terminology are potentially serious, including failure to apply agreed treatment pathways, impaired communication among staff and an absence of adequate communication of the diagnosis to patients and carers. Moreover, an incoherent approach to diagnosis among senior staff makes attempts to provide effective training to junior practitioners and students much more challenging. Thus, these findings have broad implications for education and training, including continuing professional development.

Practitioners showed a range of levels of confidence in their own ability to detect delirium. Given that the sample is likely to be biased towards those with an interest in the condition, it is of concern that at least a substantial minority of respondents, including medical practitioners, have very low to moderate confidence. Given that delirium is very common, affecting at least one in eight hospital patients, this lack of confidence does not stem from unfamiliarity with the condition; rather, it is likely to result from insufficient education and training. The general incoherence around approaches to delirium detection may also contribute to the lack of confidence in many practitioners. The findings showed that just over half of respondents had ever used a specific delirium detection tool; non-specific cognitive tests were more widely used. Related to this, although around two-thirds of units had delirium guidelines, respondents reported that in only a minority of these were the guidelines followed.

Another informative finding concerns the perceptions of the responsibilities of different disciplines with respect to delirium. In particular, respondents mostly felt that nurses were responsible for screening for delirium but not for diagnosing it. It is not rational or pragmatic to restrict diagnosis of such a common condition to medical staff alone, when the diagnosis can usually be made readily with bedside assessment and informant history (or personal knowledge of the patient). On the contrary, it can be argued that nurses who have appropriate training (including an awareness of the DSM-5 criteria) are in a strong position to make a diagnosis, because they have the most direct contact with the patient of any health-care practitioner. Notably, the prompt diagnosis of delirium is advocated by many policy-makers, such as Healthcare Improvement Scotland. This finding has implications for both policy and clear decision-making around the explicit roles of doctors and nurses in detecting delirium and initiating early care. With respect to physiotherapists and occupational therapists, an awareness of the main features of delirium, perhaps coupled with the use of a screening tool such as the 4AT, could lead them to report potential delirium to colleagues qualified to make a diagnosis.

Limitations

Some limitations of survey A study should be acknowledged. This study did not have a population sample. The nature of the survey means that the proportion of responding practitioners who had an active interest in delirium was probably higher than that among the whole population of practitioners in the UK. E-mails were initially sent only to practitioners for whom an e-mail address was available, and recipients then decided whether or not to participate, and also whether or not to forward the e-mail to colleagues. Although the core distribution list contained practitioners from 165 trusts/boards, 112 palliative care services and each of the authors and multiple organisations were asked to forward the invitation e-mail to relevant practitioners in their contacts list, and this snowball distribution method is likely to have resulted in clusters of respondents in particular locations. The proportion of practitioners in the sample from each UK nation does not reflect the true ratio of practitioners distributed across the UK. In particular, Scotland is over-represented, which is relevant because promotional efforts in NHS Scotland around delirium care by the government body Healthcare Improvement Scotland began in 2012. Although efforts were made not to lead respondents, the use of the term 'delirium' in the invitation e-mail and study title may have primed respondents and introduced bias when asking about the terms used by them and others in their units. As a result, the term 'delirium' was potentially over-reported in this survey in comparison with practice in general. Although the survey was reviewed by all of the authors on two occasions for face validity, clarity and structure, and then piloted with 19 health-care practitioners, neither test-retest nor inter-rater reliability was assessed. Anonymous participation precluded calculations of reliability for the study sample; however, rough estimates could have been generated from pilot participants. The analysis was mainly purely descriptive. Group comparisons among professions were conducted post hoc. Given this, and also that the sample was not representative, no strong conclusions can be drawn from the results about the differences among professional groups. Nevertheless, the findings provide some evidence that practitioner expectations about delirium diagnosis are different.

Survey B

Survey B was directed at respondents who had experience of using the 4AT. The final sample size was 100 and comprised mostly doctors and nurses. A large majority of respondents reported carrying out delirium assessments 'frequently' or 'almost always/always', meaning that this is a sample that is likely to include a substantial proportion of professionals experienced in delirium assessment. In addition to responses to the structured survey questions, multiple free-text comments were provided.

Taking the findings as a whole, respondents generally viewed the 4AT as a useful, rapid and practical tool. Several comments were made about potential changes. However, no clear problem with the 4AT emerged consistently. Many of the issues raised relate not to the 4AT specifically but to general challenges in the assessment of delirium, such as the availability of informant history and the time it can sometimes take to get this history, and basic knowledge of the features of the syndrome itself. Other issues relate to possible modifications of the cognitive tests used, the time frame over which altered mental status is considered to indicate delirium, and so on. Many of these suggestions are reasonable and reflective of variations in

accepted, mainstream practice. However, given that the current version of the 4AT is in wide use, supported by several validation studies and mentioned in several policy statements and guidances (see *Chapter 3*), in the absence of strong positive evidence in favour of these modifications being introduced, retaining the current version is the most pragmatic option.

Around one-third of practitioners reported that it was 'difficult or 'very difficult' to complete the 4AT for 'drowsy patients who cannot produce verbal responses'. Items 2 and 3 of the 4AT allow for scoring such 'untestable' patients, and item 1 allows for scoring drowsy patients. This finding suggests that education about the features of delirium with respect to reduced arousal would be helpful, and specific training on the 4AT concerning this issue would reduce the uncertainty reported by some practitioners. With respect to training, most respondents stated that at least some training in using the 4AT would be needed for the tool to be used effectively. The proportions of respondents stating this were similar to the proportions responding to the question about whether or not training in delirium in general would be required to use the tool. Although the 4AT was designed to be simple and practical, and usable without specific training, the survey findings suggest that users must have a basic understanding of delirium if they are to use the 4AT. This should include understanding that a reduced level of arousal is commonly seen in delirium and contributes to that diagnosis. A number of comments suggested that more guidance is needed about what to ask carers or contacts in terms of changes in patient behaviour. This knowledge is best classed as part of a practitioner's general information about delirium detection; nevertheless, given the lack of other approaches to delirium detection in routine care, more information on how this should be done in the context of using the 4AT would be valuable. Additionally, in conjunction with this, some training in the use of the 4AT would appear to be beneficial in promoting its effective use. Given the lack of space on the one-page 4AT form, such additional education and training on delirium in general and on the 4AT would best be provided on easily accessible websites, including www.the4at.com.

Some queries arose about the scoring of the cognitive test items. Cognitive tests used in hospital practice show variable diagnostic accuracy for dementia and delirium,^{46,52,95,96} and no single test provides sufficiently good performance for it to be used in all contexts. In addition, cognitive tests performed in isolation inform clinical judgement, but cannot be performed as diagnostic tests alone. Therefore, practitioners using the AMT4 and Months Backwards tests need to be aware of both the value and the limitations of the information provided by the test results. For practitioners seeking further information about the cognitive test items in the 4AT, the URL of the 4AT website is given on the form (and, indeed, is easily found through an internet search), and an up-to-date list of specific 4AT validation studies, as well as relevant studies relating to the cognitive test items, is provided on that website.

A need for both the scoring and the guidance to be on the paper or electronic documentation was also identified. This sometimes seems to have been lost when the 4AT has been incorporated into a larger assessment such as an electronic patient assessment form. This could be addressed by providing clear links to the 4AT guidance notes or the website.

The current findings emphasise that the use of the 4AT is not based solely on the utility of the tool itself, but also depends on broader systemic and cultural factors. All of those participating in survey B had some experience of delirium assessment, and yet there was some evidence of lack of knowledge about some general aspects of delirium, which had an adverse effect on participants' understanding of how to use the 4AT. Furthermore, respondents identified external factors that hindered their use of the 4AT, and presumably their use of any other delirium assessment tool.

Limitations

Survey B had some limitations that must be acknowledged. The sample size was relatively small, and most participants were based in the UK, with a disproportionate number from Scotland. This limits the generalisability of the findings. The sample was not representative of all users of the 4AT; rather, it probably was biased towards those with more experience of its use, and also possibly those with a more favourable opinion of the tool than users of the 4AT as a whole. Most respondents were medical

practitioners, which meant that there was limited information from nursing staff and staff from other disciplines. The free-text comments were not subjected to a formal analysis such as content analysis. Nevertheless, the study yielded considerable useful information regarding opinions on the use of the 4AT from both the structured questions and the free-text comments.

Conclusions and implications

Two major issues were addressed in these surveys. The first was to develop a deeper understanding of broader issues around delirium assessment in the UK health-care practitioners. The results from survey A provided valuable insights into the many barriers that prevent the effective detection of delirium in mainstream care. These include variable knowledge among practitioners of delirium and its features, a lack of confidence among many practitioners in their ability to detect delirium, inconsistent use of terminology, lack of compliance with guidelines, and varying opinions on the roles of different health-care staff in detecting delirium. These findings are relevant to education policy in relation to undergraduate and professional levels, as well as the design of effective systems, and for creating an organisational culture that facilitates delirium detection. The organisational issues are explored in more detail (see *Chapter 4*) in the qualitative studies that also form part of phase 1 of this project. The second major issue addressed by this part of the project was a specific survey examining opinions on the 4AT, looking in particular at the issue of whether or not there was a need to modify the 4AT in advance of the diagnostic accuracy studies in the second part of this project. Respondents were generally positive about the 4AT. There were some suggestions about how it could be modified, but in the absence of a consistent message about these the team decided that no changes were necessary. However, survey B did yield important information about the need for general delirium education and pointed towards the benefits of providing education on use of the 4AT to enhance its effectiveness.

Chapter 3 Clinical use and examples of clinical service evaluations

Introduction

The 4AT was initially developed as a practical clinical tool, as described in the introductions to *Chapters 1* and *5*. At the time that the main grant application funding for the present project was submitted, the 4AT was already being used as a clinical tool in some settings, and two diagnostic test accuracy studies had been published. Since then, clinical use of the 4AT, both nationally and internationally, has expanded rapidly. It also now has a substantial presence in multiple policy, guidelines and advisory documents and websites. It has been translated into Arabic, Thai, Russian, German, Spanish, French, Italian, Dutch and Norwegian (with active plans for it to be translated into Japanese, Serbo-Croat and Hebrew). In addition to the 4AT being the subject of several diagnostic accuracy studies, it is in use as a tool in some other studies of delirium.^{97–99}

This brief chapter (1) provides an overview of current knowledge about the clinical use of the 4AT, including its presence in policy and guidelines documents, and (2) provides examples of clinical service evaluations, including field testing item 1 of the 4AT.

Issues around item changes, for example item 4 (acute onset and fluctuating course), were explored in the qualitative studies and surveys, and the study team concluded that no changes to the 4AT were required.

Inclusion of the 4AT in policy, guidelines, advisory documents and websites

The 4AT is included in multiple local, national and international policies, guidelines, advisory documents and websites, and reports of use in clinical practice. Some examples are provided here. Websites were accessed on 5 December 2017.

Policy websites and documents that include the 4AT

- *Scottish Standards of Care for Hip Fracture Patients 2016*: the 4AT is the recommended delirium assessment tool.¹⁰⁰
- Healthcare Improvement Scotland's *Delirium Toolkit*: the 4AT is the recommended delirium assessment tool.¹⁰¹
- *Managing Falls and Fractures in Care Homes for Older People – Good Practice Resource*: the 4AT is the recommended delirium assessment tool.¹⁰²
- *Dementia Revealed: What Primary Care Needs to Know*: the 4AT is included as one of the recommended tools.¹⁰³
- *Assessing for Cognitive Impairment in Older People: Clinical Audit 2014–15*: the 4AT is one of the recommended tools.¹⁰⁴
- *Falls and Fracture Consensus Statement*: the 4AT is now included in the list of 'quality metrics required to pass best practice tariff'.¹⁰⁵
- London Major Trauma Systems' *Management of Elderly Major Trauma Patients*: the 4AT is one of two recommended delirium assessment tools.¹⁰⁶
- US Department of Veterans Affairs' *Delirium Information for Professionals*: the 4AT is among the recommended tools for delirium assessment.¹⁰⁷
- American Nurses Association: the 4AT is one of the included tools for delirium assessment.¹⁰⁸

- *Early Identification and Initial Management of Delirium in the Emergency Department/Acute Medical Assessment Unit*: the 4AT is included as the main delirium assessment tool.¹⁰⁹
- Australian Commission on Safety and Quality in Health Care's *Delirium Clinical Care Standard*: the 4AT is included as a recommended tool.¹¹⁰
- Health Navigator New Zealand (website providing clinical information for clinicians in New Zealand): the 4AT is the recommended delirium assessment tool.¹¹¹

Use of the 4AT in clinical practice

There is no central record of which screening tool is used in which hospitals, and so it is impossible to determine everywhere that the 4AT is currently used, but through requests on social media, and contacts with members of the Scottish Delirium Association and European Delirium Association, we have determined that it has been rapidly adopted throughout the UK. In England, it has been widely adopted as a screening tool to fulfil the CQUIN target, for example at Guy's and St Thomas's Hospital, King's College Hospital, York Teaching Hospital NHS Foundation Trust, Sunderland NHS Trust and Northampton General Hospital (see case studies below). It is also widely used throughout Scotland to screen for cognition on admission. The Older People in Acute Care programme from Healthcare Improvement Scotland reports that there is no nationwide collection of data on the 4AT, but it is being used in the majority of NHS boards (e.g. Royal Infirmary of Edinburgh and Western General Hospital NHS Lothian; Royal Victoria Hospital, Dundee), and options are being explored for testing its use in care homes and community hospitals. Additional information regarding use of the 4AT in clinical practice includes findings from survey A in this report (see *Chapter 2*). Indirect evidence of the use of the 4AT in clinical practice also comes from the number of hits on the 4AT website. This website was launched in February 2011. The number of page views increased from 1254 in the year beginning June 2011 (the point at which analytical data are first available) to 61,269 in the year beginning June 2016 (Figure 1). The top five countries recorded are the UK, Australia, the USA, Ireland and Canada.

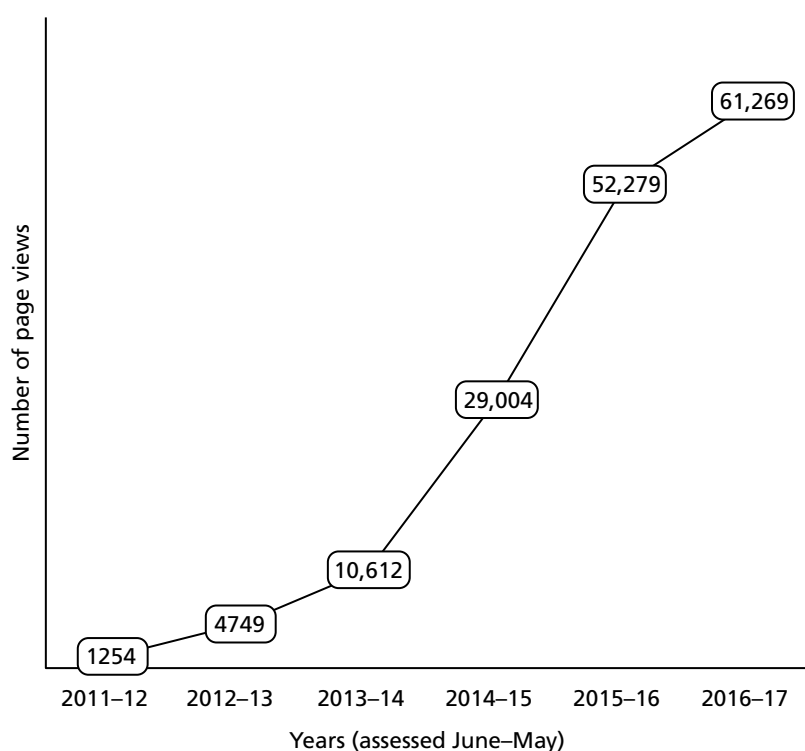


FIGURE 1 Page views per year on the www.the4AT.com website between 2011 and 2017. Source: Google Analytics. (Google Inc., Mountain View, CA, USA).¹¹²

In the following sections, we provide some examples of websites and documents relevant to the clinical use of the 4AT, and then we provide two case studies of clinical use. The case studies were selected based on study team members' knowledge about centres that have established 4AT use.

Examples of websites and documents documenting 4AT use in clinical practice

- *Identifying Critical Success Factors for Improved Outcomes for People With Dementia and Their Carers in Acute Care.*¹¹³
- *Improving Older People's Acute Care: Impact Report:* summary report of multiple quality improvement projects including delirium in several hospitals in Scotland, involving use of the 4AT.¹¹⁴
- National Hip Fracture Database Annual Report 2016: the 4AT is documented as having been introduced as a standard clinical tool for delirium detection in hip fracture patients.¹¹⁵
- *Think Delirium: Recognition of Delirium in an Acute Medical Unit:* clinical audit, Ninewells Hospital, Dundee, showing improvement in delirium detection following introduction of the 4AT.¹¹⁶
- *Improving Identification and Management of Delirium in the Orthopaedic Trauma Unit:* quality improvement project on increasing the use of the 4AT in an acute hip fracture ward.¹¹⁷
- Derbyshire Healthcare NHS Foundation Trust's annual report and accounts 2015/16: the report documents that the 4AT was introduced to improve delirium detection.¹¹⁸

Case study: City Hospitals Sunderland NHS Trust

City Hospitals Sunderland NHS Trust is a large district general hospital with ≈ 900 beds, of which ≈ 200 are for acute care of the elderly. The 4AT has been embedded in the trust since 2013. It is used for all acute admissions, and in the ED it automatically appears in the nursing admission documentation to be completed for all admitted patients who are aged ≥ 65 years. Score distribution is approximately 65% of patients with a score of 0, 20% of patients with a score of 1–3 and 15% of patients with a score of > 3 . All patients scoring ≥ 1 (and patients about whom there is a clinical concern) are reviewed by the specialist delirium dementia outreach team, comprising four band 7 practitioners (three nurses and one physiotherapist), a 0.5 whole-time equivalent senior pharmacist and a 0.5 whole-time equivalent band 3 administrator, with supervision from 1.5 whole-time equivalent consultant geriatricians. The team assesses 300–400 new and 100–200 review patients each month (total to June 2017, $n = 1218$; mean age 78.1 years, range 61–102 years). This has resulted in improvements in achievement of the CQUIN target and in delirium diagnosis recording, and a reduction in referrals to the mental health liaison team. Of note is that approximately 75% of patients seen by the delirium dementia outreach team are identified after a positive score on the 4AT, but 25% are also referred by clinical staff as a result of concerns. The majority of these patients score 0 on the 4AT, but, on review, are confused and/or drowsy, and so the score is likely to have been inaccurate. Written feedback from the team was provided: in summary, the 4AT has been widely adopted and is considered easy to use and quick to complete. Patients with scores of 0 may be 'underscored'; and local review has concluded that this is a staff training issue, rather than an issue requiring a change to the tool.

Case study: King's College Hospital NHS Foundation Trust, London

King's College Hospital NHS Foundation Trust is a large acute care teaching hospital with 950 beds. The 4AT is embedded as the screening test for all acute admissions. A delirium and dementia team was established to review patients who screened positively on the 4AT.¹¹⁹ It has been difficult to validate because of issues with retrieving accurate information from primary care. However, there is a good correlation between higher 4AT scores and delirium features. Just over 80% of individuals screened go on to receive a diagnosis of dementia or delirium at follow-up. Written feedback from the team is that the 4AT is easy to use and score.

Field testing of 4AT item 1 (level of alertness)

The main purpose of this subproject was to assess the agreement between raters for item 1 (level of alertness). The reason for assessing this item alone was that the study team considered the brief cognitive test items (items 2 and 3) to be well studied and relatively simple tests with less need for assessment of

agreement, whereas a binary judgement of level of alertness was more likely to be prone to disagreement and, therefore, might need to be changed in advance of phase 2.

This subproject was conducted using paired assessors using the 4AT in routine clinical practice in the Royal Infirmary of Edinburgh. In the acute hip fracture ward, it is common for patients to be reviewed by a pair of practitioners (either two doctors, or a doctor and a nurse).

For this clinical evaluation, when practitioners were assessing patients in pairs, one would perform the 4AT and the other would record their responses to each item, blind to the other rater's scores. We performed 50 paired observations, scored as either negative or positive for the presence of abnormal level of alertness. When there was disagreement, raters discussed this (without changing the ratings), recording their opinions about why there had been disagreement.

The number of positive observations was 16 out of the total of 100. There were four instances of disagreement across the 50 paired assessments. This yielded a Cohen's kappa statistic of 0.70, which indicates substantial agreement. Raters disagreed on two occasions about whether or not a patient had agitation, and on two occasions about whether or not a patient's reduced level of arousal was low enough to trigger a positive score. On all four occasions, both raters noted that the patient's altered arousal was relatively mild.

In summary, the field testing showed that there would be some differences of interpretation regarding whether or not mildly altered alertness (whether this is restlessness or sleepiness) is considered to be abnormal. This is an inherent problem of using a binary scoring system. An item involving more gradations would be likely to lead to greater agreement. However, the additional time cost and complexity of using an item with multiple levels conflicts with the need for a tool that is not only rapid but also easy to understand and use in a busy clinical context. On balance, the evidence suggests that item 1 in the 4AT performs at an acceptable level.

Conclusions

Since its release in 2011, the 4AT has become widely used for detecting delirium in clinical practice nationally and internationally. This has been driven by the apparent need for such a tool, as identified in the 2010 NICE guidelines on delirium,⁴ in combination with the simplicity of the 4AT and the initial evidence that it has sufficient validity. Although it is difficult to get precise estimates of use of different tools in clinical practice, the 4AT is likely to now be one of the leading tools for delirium assessment in current use in the NHS in the UK. Information from clinical service evaluations and uptake by major policy-makers demonstrates that, overall, the 4AT is considered practical and valuable.

Chapter 4 Implementing delirium screening

Introduction

This chapter reports on the qualitative substudy part of phase 1 of the research. The aim of the exploratory study was to provide an insight into the practice of delirium screening and assessment in acute NHS settings from the perspectives of different health professionals.

As discussed in *Chapter 1*, although delirium is a common complication of acute illness, particularly in older people, and incurs considerable costs at human, clinical, service and societal levels, the problem of underdetection persists. Although surveys have shed important light on the knowledge of, and attitudes to, delirium among health professionals in different health systems (see *Chapter 2*), there is limited research on the practice of delirium recognition and assessment in the real-life context of health-care delivery across different service delivery locations within the acute hospital.^{120–128} Most of these published studies relate to the work of nurses only;^{120,121,123–127} take place in various inpatient settings (palliative care,^{121,124} orthopaedic,^{125,126} medical/surgical¹²⁷ and older people^{122,123,128} wards); and take place in different countries. The main conclusions are that nurses experience discomfort when, and lack knowledge and understanding of, working with and responding to patients with delirium; and that the focus is often on managing disruptive behaviours rather than on investigating and treating a medical syndrome. Negative attitudes and knowledge and skills gaps at individual and professional levels have been suggested as reasons for the underdetection of delirium and poor treatment. However, knowledge, skill gaps and work practices are not separate from, but are located in, cultural, organisational and interprofessional contexts. Despite this, few studies^{37,38,41,122,129,130} have examined barriers to delirium screening, assessment and management at these levels. In this study, we sought to provide insight into delirium knowledge and the value attached to delirium identification, and how both are shaped by professional role, organisational purpose and routines in different settings along the patient journey, into and through hospital, in the UK NHS. This was to inform understanding of the factors implicated in the implementation of a screening tool for delirium in practice.

Research design

Research objectives

The study objectives were to:

1. examine screening practices for delirium at particular locations in the patient journey into and through hospital
2. explore staff knowledge of delirium, the value attached to identifying delirium, and the strategies called forth to manage it, at these same service delivery locations
3. determine the organisational, environmental and system-level barriers at locations along the patient journey that have an impact on the systematic identification of patients with possible delirium
4. consider general staff use of screening tools and the 4AT specifically, the contexts in which the 4AT is used, and the speed, ease and value attached to its use.

We performed a multimethod, qualitative, mixed-methods study comprising in-depth interviews with health professionals and observation of practice in different locations along the patient journey: the ED, the MAU, the surgical assessment unit (SAU), the elderly assessment unit (EAU), and selected acute wards. The rationale for combining interviews and observation was that whereas interviews can explore staff understanding of delirium, and the meaning of, and value attached to, delirium screening in particular settings, observation provides a picture of what staff actually do in the real-life context of health-care delivery in time and space. This includes what people take for granted or are unable to articulate.

The overall aim of the study, therefore, was to provide a descriptive and explanatory account of how hospital staff from different disciplines and at different levels of seniority understood, made sense of and attached value to identifying patients with delirium, in the context of the organisation of work within settings intended to achieve different purposes.

Methods

Sampling strategy

Hospitals

We sought to include hospitals in which the 4AT was routinely used as a screening tool (4AT 'experienced' sites) and hospitals in which it was not (4AT 'virgin' sites). Initial findings from the survey of health professionals in the four countries of the UK (England, Scotland, Wales and Northern Ireland) (see *Chapter 2*) had indicated that, apart from a small number of hospitals in England, the 4AT was primarily in routine use in Scottish hospitals at that time. Furthermore, among the majority of respondents from other UK countries, there was no consistent use of any specific screening tool for delirium.

We purposively selected three general hospitals (two in England and one in Scotland). Two (Avonfield and Cranford) were part of large, acute foundation trusts located in urban centres in the north of England and were 4AT 'virgin' sites. Neither had adopted specific guidelines for detecting and managing delirium at the whole-hospital level, although in both hospitals initiatives had occurred at the elderly care directorate and ward levels to improve the detection, management and prevention of delirium, drawing on NICE guidelines.⁴ Neither hospital had adopted a single delirium screening tool, although Cranford was considering implementing the Single Question in Delirium, a single question prompt that asks 'Is this patient more confused than before?'. The third hospital, Denbury, was in an urban area of Scotland, within a large health board spanning metropolitan areas and a rural hinterland. The 4AT had been introduced throughout the hospital, along with local guidelines on the detection and management of delirium. All three hospitals were attached to medical schools. A pseudonym is given for each site.

Participants

Within each hospital, we adopted a multilayered approach to secure engagement in the study. Initially, we arranged meetings with departmental managers, clinical leads for delirium and dementia and senior medical and nursing staff in order to introduce the study, gauge interest in participation and secure access. Staff provided orienting guidance on the process of screening and assessment, identified key informants and offered a point of contact for each department.

Within each department/service location, we approached the senior clinician or manager to discuss the study, seek permission to conduct observation and interview staff from different disciplines and at different levels of seniority. We provided information leaflets and posters to advertise the study. In some settings, we were able to attend team meetings and explain the research; in others, information leaflets were distributed to staff. Our interest was in identifying and approaching individuals who played a specific role in the process of assessment, information gathering and diagnosis. In all settings, a senior clinician was invited to take part in an interview and then suggested others in different roles to contact. Observation also served to identify key practitioners to approach; for example, an informant conversation with an individual about their work on occasion resulted in an invitation to take part in a formal interview. All of those approached agreed to take part in the study.

Data collection

Interviews

We conducted qualitative interviews with a purposive sample of staff from each service location (ED), different types of assessment unit and acute wards, from different disciplines, and with varying levels of

experience and seniority. Staff provided orienting guidance on the process of screening and assessment, identified key informants and offered a point of contact for each department.

A total of 19 health professionals in the 4AT 'virgin sites' took part in a formal interview: 11 based at Avonfield and eight based at Cranford. Another 10 staff participated in lengthy informant discussions in the course of observation; detailed notes were made and subsequently written up. In the 4AT 'experienced' site, 24 interviews were conducted. In total, 43 interviews and 10 informant discussions were carried out. The number of interviewees in each organisational setting and their seniority and discipline within the 'virgin' and 'experienced' sites are shown in *Appendices 6 and 7*.

Interviews were conducted using a topic guide (see *Appendix 8*), which was used as an aide-memoire; interviewees were invited to discuss issues relevant to their professional role, task and work setting. The topic guide was developed from a review of the literature, prior experience of research on delirium practice in acute care¹²² and discussion with clinical colleagues in the wider research team, including the chief investigator and co-applicants working in varied clinical settings. The focus was on how staff perceived their role within the setting and the nature and pattern of their work routine; the context in which they were involved in identifying patients with delirium; their current strategies for detecting delirium, including triggers; what actions or investigations might be prompted following screening and assessment of probable delirium; and issues pertaining to the treatment and management of patients with delirium. This line of questioning permitted further exploration of an individuals' knowledge and understanding of delirium; strategies for dealing with it specific to their role or department; and features of the setting that facilitated or inhibited understanding of, identification of, and action on delirium. This set the context for asking respondents for their general views on delirium screening tools and their specific views about the 4AT. The mean duration of interviews was 50 minutes (range 35–65 minutes). The interviews were usually conducted in a private space (an office or a meeting room), although they were frequently subject to interruptions and then resumed.

Observations

Ethnographic observation, including shadowing members of staff, was employed to develop understanding of the rhythm and pattern of routine activity in each organisational setting; how the work of assessment and care delivery was accomplished; and how delirium screening, assessment and management fitted into this. During observation, researchers engaged staff in informal conversations in order to clarify aspects of practice and work organisation. These ranged from chats to clarify an event, activity or professional role, to lengthy exchanges to provide insight into sequences of action and work processes. Locating respondents' comments in the contexts in which they were made facilitated deeper understanding of their meaning. Additionally, we collected documents relating to procedures (e.g. care pathways, use of structured assessment tools) and processes (assessment, diagnostics and action-planning).

A total of 45 hours' observation took place across the three hospitals, each session lasting around 3 hours at a time. Observation initially centered on the ED, intentionally replicating the starting point of the patient journey, as the ED might be considered a key location for early detection of delirium. At the same time, the pace and complexity of the practices and processes in ED required a longer period of observation to better understand staff roles and routines. Then, we moved on to the next point in the patient journey, the assessment units, which comprised three MAUs, two EAUs/frailty units (dedicated assessment units for older people) and one SAU. Finally, we moved on to acute wards; these were mainly care of older people wards, but also included stroke, cardiothoracic, orthopaedic trauma and post-surgical wards. Not all patients follow such a linear path; some bypass the assessment unit and are directly admitted to a ward, and others go straight to an ICU. However, the assessment unit is an important staging point for a more detailed assessment of patients and for decision-making about acute admission or diversion elsewhere. We selected a range of types of assessment units and wards to ensure that we included settings where staff were likely to vary in their knowledge of delirium (and dementia). We did not include ICUs in the

study; observation in this setting was considered intrusive, and the research timescale did not permit the detailed negotiation that this participation would require.

Data analysis

All audio-recorded interviews were anonymised and transcribed verbatim. As our focus was on interview content, transcription notations were used only to indicate pauses, cross-talking and hesitations. Researchers conducting the interviews listened to the audio-recording alongside the transcribed text for accuracy of transcription. They also completed a pro forma for each interviewee, which included the interviewee's salient characteristics (profession, grade/seniority, gender and organisational setting/ward type). A summary was produced as soon as possible after the interview; this included the main issues raised, the research questions on which the contact was most focused, new ideas generated to pursue in the next interview, and/or areas that were insufficiently explored. Audio-recordings, transcriptions and summaries were shared and discussed within each local team; transcriptions and summaries were shared between the two local teams. We did not send transcriptions to interviewees for comment; although this offers potential for further discussion and elaboration, its use as a method of validation is contested, reflecting varied methodological and epistemological stances.

Contemporaneous notes from informant interviews and field notes were written up as expanded accounts as soon as possible after these events.^{131–133} The researchers' impressions of, and reactions to, what they had seen and heard were recorded separately in a chronological journal. All data sets were stored electronically and organised by site and setting. Research teams involved in the English and Scottish sites met regularly to discuss the emerging data and what to make of them, and to compare and contrast findings within and across sites; team members also communicated regularly by e-mail and telephone.

Data from all sources were analysed using grounded theory methods,¹³⁴ including simultaneous data collection and analysis, constant comparison, search for negative cases and memo-writing. Multiple readings of all interview transcripts, summaries and field notes by team members individually familiarised researchers with the data, including their variability and range within and across settings and at individual and organisational levels. Through discussion, we developed tentative ideas, captured in memos, of what was going on at both levels. For example, what were the distinctive features of clinical work undertaken in different organisational settings; how did these inform the delirium-related work carried out by individual staff; and how did delirium knowledge across sites and settings influence the value attached to the work of delirium identification, assessment and action?

Data sets (transcripts and field notes) relating to each organisational setting were subject to initial open coding by local research team members, individually and together. This first phase involved synthesising, integrating and organising data segments into named meaning units on hard data copies, with each data source examined on its own terms. In this way, we sought to retain the coded segments in the context of the whole interview or field note. Initial codes were provisional, primarily descriptive and close to the data, and aimed at understanding participants' meanings, accounts and actions. Coded transcripts were shared, compared and discussed among the research team, differences in interpretation often giving rise to new insights. For example, participants in Denbury often prefaced their accounts of action on delirium in terms of what they should do, suggesting the need to examine how normative practices relating to delirium identification were introduced, and what was shared, subverted or sustained over time.

From the initial codes, we agreed which to pursue for more in-depth focused coding. The open code 'knowledge of delirium', for instance, was developed as an analytic category during focused coding comprising several components (formal knowledge of diagnostic criteria; practice knowledge based on exposure to patients with delirium and its effects on them; and tacit knowledge from practice experience, emotions, insights and observation). Analytic categories, components and descriptors, and the research objectives they related to, were displayed and managed in Microsoft Word (Microsoft Corporation, Redmond, VA, USA) files organised by site and setting. With regard to knowledge, for example, we compared and contrasted different forms of knowledge to examine how they variously shaped meaning

and action on delirium identification between settings, between professionals and across sites, and created diagrams of these relationships. These were captured in memos and further tested out with the data.

In addition to coding and categorising data, we explored the data narratively to examine processes and context systematically. For example, regarding delirium identification, we perused larger sections of transcripts and notes to examine the narrative structure of accounts: what preceded a particular account/incident and what followed from it to capture the processes involved. Comparing and contrasting narratives between settings cast light on how features of organisational context affected these processes. Each stage of analysis involved discussion within the research team: codes and categories were reviewed; memos and diagrams were shared; and explanations developed were refined and verified, moving iteratively between the empirical data and sense-making in relation to them.

Research team

Two investigators (EL and AS) conducted the interviews and observations, with academic support from Mary Godfrey and Janet Hanley (both experienced qualitative researchers). Organised as two local teams, Elizabeth Lavender and Mary Godfrey pursued the research in the English sites and Antaine Stíobhairt and Janet Hanley did this in the Scottish site. They came from different disciplinary backgrounds – psychology (EL and AS), sociology (MG) and nursing (JH) – and from varied research areas: applied health and social care research relating to older people, including with cognitive impairment (MG), adult mental health (EL), learning disability and delirium (AS), and the introduction of new health technologies particularly in primary and community care (JH). None was a medical practitioner. Their varied backgrounds and perspectives generated openness to exploring different ways of considering issues and problems.

The whole team met together for a day at critical points: at the beginning of the research to develop the fieldwork plan; during fieldwork to reflect on emerging data, consider variation within and between sites and what shaped it, and identify new lines of enquiry; and during analysis to discuss and refine codes, categories, analytic concepts and narrative content. Regular telephone conversations, e-mail discussions and sharing of memos, diagrams and matrices occurred between investigators in an ongoing dialogue and to ensure consistency of analytic approach in the three sites. Within local teams, formal meetings and informal discussions took place approximately weekly.

Ethics

Ethics approval for the study in the two English sites was obtained from the National Research Ethics Committee (reference 14/SW/1095). The main ethical issue was not seeking formal consent for general observations of routine practice. As observations were unobtrusive and carried out in public or semi-public spaces, and did not identify individual staff, patients or visitors by name, we considered that it would be impractical, and probably more intrusive, to seek formal written consent from all those present. Instead, we sought informal verbal consent. At the beginning and end of each observation, the researcher 'signed in' and 'signed out' with a senior staff member. No difficulties were encountered in any of the sites.

Advice from the National Research Ethics Committee in Scotland was that as the focus was on examining the 4AT in use, the study was more appropriately characterised as audit and, therefore, required only local clinical governance approvals; these were obtained. The conduct of the study in this site conformed to the same ethical principles as informed research in the English sites.

Findings

The findings are organised as follows. We examine each setting separately, their organisational and care environments and the views and experiences of different professionals within them pertaining to their knowledge about, and practice in, delirium detection. We consider how the context and purpose of each setting impacts the value attached to delirium screening. Then we compare and contrast findings across settings and between professionals to consider the barriers to delirium screening, with a focus on the

pattern of variation between the 4AT 'virgin' and the 4AT 'experienced' hospital sites. Finally, we draw out the implications for understanding delirium screening implementation and practice. In *Appendix 9* we report some views of staff on the 4AT screening tool within the 'virgin' sites as they 'thought aloud' on its use in context of their particular setting, and similarly for the 4AT 'experienced' sites.

Each setting is introduced with a snapshot drawn from observation of the organisational context in which the work of delirium screening and assessment occurred. This is to convey the type and pace of activity within the department, the speed of routine processes, and how staff worked together to accomplish their role and purpose.

The emergency department

Studies examining recognition, detection and delirium occurrence in EDs have been a particular focus of research interest on delirium. EDs are both a patients' first point of entry into general hospitals and a location in which decisions about appropriateness of acute care are initially made. From systematic reviews^{135,136} of observational and prospective studies over two decades, a yawning gap exists between the occurrence of delirium among older patients in EDs as assessed by researchers using various measurement tools, and what is identified by ED physicians, typically based on a review of charts/documents. Primarily conducted in the USA and Canada, study findings suggested the presence of delirium in approximately 7–10% of older people admitted to the ED; this is lower in some studies¹³⁷ and considerably higher among subgroups of older people, such as those who are older, in long-term care or living with dementia.³² Furthermore, delirium is missed in between half and three-quarters of cases comparing rates of delirium as measured by researchers with detection rates based on chart review.^{22,31,32} Although some of this variation may be attributed to differences in methodology and heterogeneity in the measures employed,^{135,136} it is also likely that contextual factors are implicated.

Emergency departments: an observational snapshot

Emergency departments were variously calm and chaotic: either eerily empty and anticipatory, or alarmingly active and noisy. The general atmosphere in reception was steadily busy yet reassuringly calm. Walk-in patients checked in at reception, and were directed to minor injuries or 'see and treat' waiting. Time passed slowly here. The low hum of a television screen appeared to offer people some distraction as they waited for attention.

Passing from reception into the main treatment area, sound and commotion intensified. There was an impression of constant movement. Staff walked quickly between treatment areas; patients were wheeled in and out of assessment cubicles; family members entered, looking lost and anxious; porters, therapists and non-medical staff searched for patients; other visitors to the department – paramedics, police, social care workers – attempted to attract the attention of relevant practitioners and sometimes patients.

This noisy, chaotic, disorientating atmosphere concealed the composed co-ordination of the operation beneath. ED staff faced an unremitting stream of incoming patients and tasks to be carried out, yet they appeared to work steadily, each playing their part in the overall purpose, namely accomplishing the timely and appropriate throughput of patients. In the centre of observed departments were large computer tracking screens on which each patient's location, assessments, test results and time since arrival served as a constant reminder and additional pressure.

Staff greeted incomers, taking each case in succession, filtering and processing patients onwards, and drawing on different sources of information to effect appropriate movement, all the while conscious of the clock ticking and the steady accumulation of incoming patients. Porters transported sometimes confused, upset or unwell patients to radiography or to bays for further investigation; technicians took blood samples; nurses and junior doctors traced patients' progression through investigatory staging posts, picked up observations and carried out detailed assessments. Amid the *mêlée* were the patients, sometimes accompanied, often alone as they were guided through and onwards. Staff – primarily medical and

nursing – carried out their tasks systematically, concurrently and collaboratively, constructing a picture of the patient, a clinical case.

Flow and processing were paramount, efficiency and accuracy crucial. Sporadically this pattern was interrupted by an incoming emergency. Alerted by an alarm, senior doctors and nurses gathered, ready to respond, primed to act, and then melting away back to their patients and paperwork when no longer needed.

Patients arriving by ambulance came in through the back door, accompanied by paramedics and sometimes by a relative or friend. Paramedics waited to hand over their charge, usually to a senior doctor or nurse, who took the patient's information, including a brief history of the precipitating medical event, biographical details and medication taken; conducted a rapid assessment; and ordered appropriate tests and further investigations. With an experienced clinician at handover, the process took about 10 minutes per patient. Sites varied in the type of professional undertaking this task (e.g. in Avonfield this was a senior doctor, and in Cranford a senior nurse practitioner). Their early involvement appeared to aid efficiency as long as the staff in the bays could respond to the pace.

The specific aim at handover was the rapid, accurate movement of patients to the next appropriate location, whether for immediate treatment in resuscitation, for those who were very unwell, or for further assessment and information gathering by staff in the bays. Here, nursing staff conducted clinical observations and gathered information about the person, including their social circumstances, functional abilities, and health and well-being; while medical staff, typically junior doctors, carried out clinical assessments, interpreted test results, and, drawing on the multiple sources of information and discussion with nurses, developed a possible differential diagnosis and plan of action for review with senior staff. Patients were discharged or admitted, the ultimate aim of ED being the effective direction of the patient either onward through acute care or discharged home. A marked feature of the process of decision-making regarding movement of patients between one assessment space and another and within each space was vertical and horizontal collaboration between staff. Even so, there were times when the sheer numbers of patients coming in constrained the smooth flow.

Location of resources at the 'front' door of the hospital had the potential to facilitate further observation and assessment of those deemed medically fit for discharge but who might need support to ensure a 'safe discharge'.

Distinctive and common features of ED organisation to achieve its clinical and service purposes were the temporal frame within which the work was organised; the division of labour in designated spaces where particular types of action and interaction occurred (e.g. triage, pitstop, resuscitation); systems and processes to achieve collaborative and co-ordinated team working to appropriately respond to patients' immediate clinical needs, pursue investigations and information gathering to determine where next for the patient; and maintain pace and flow.

Emergency department: knowledge and understanding of delirium and delirium detection

Respondents to survey A (see *Chapter 2*) generally reported an increased awareness of delirium, primarily attributing this to greater exposure through professional journals and conferences, prompted by NICE guidance⁴ and initiatives such as 'Think Delirium' in Scotland. Nevertheless, also evident from the survey was wide variation in awareness of, knowledge about and attitudes to delirium among staff in different organisational settings and across different disciplines. In particular, the perception that staff awareness of delirium had generally increased was not shared by survey respondents working in EDs. However, interviews with ED staff convey a more nuanced picture, as considered below.

Types of knowledge

Senior ED doctors in 'virgin' sites suggested that they had heightened knowledge of delirium and of the consequences of undetected delirium, meaning that they were more attuned to the importance of identifying it. Furthermore, understanding of the delirium trajectory, including its deleterious consequences for patients in both the immediate and the longer term, increased the salience of work to detect it. The combination of clinical knowledge and experience meant that senior staff were able to quickly recognise a particular constellation of factors in the patient that made them think of delirium. This tacit knowledge of delirium, aligned with the clinical presentation, sensitised these professionals to noting and interpreting observational cues to construct a clinical picture:

There comes a point in a senior doctor's life where the database of thousands and thousands of patients that I have seen is sufficient for me to be able to recognise certain clinical patterns both from history and examination and that's often brief . . . but it's quite common to get the key points of the history that will point towards the diagnosis . . . so recommending a comprehensive history and a full physical examination . . . in the first place, and then targeting investigations to those most relevant . . . there's a number of different things that can cause it.

Consultant, ED, Avonfield 04

Importantly, these different types of knowledge were not shared by all staff in the ED. A common view expressed by senior doctors was that although staff generally were 'more willing to consider delirium as part of the clinical picture' they lacked knowledge of the profound negative effects of undetected delirium:

Do they understand it? No, I don't think they appreciate it. I think that people are getting better at picking it up but do I think that they recognise the significant mortality rate . . . that's associated with it? No I don't think so. I think they might vaguely be aware but . . . I think the numbers would startle them.

Consultant, ED, Avonfield 01

Consequently, it was suggested that these staff might be less focused on delirium detection, according to its lesser significance in the overall assessment process and in the construction of the differential diagnosis. Lack of clinical experience with patients with delirium was also seen as a barrier to the development of tacit knowledge.

The enhanced understanding of delirium that accompanied knowledge and experience meant that senior staff were often able to quickly recognise a particular constellation of factors in the patient that made them think of delirium. By incorporating knowledge of delirium with the clinical presentation, these professionals noted and interpreted observational cues to construct a clinical picture; furthermore, understanding of the delirium trajectory increased the salience of detecting it.

Less experienced staff, or those with different professional roles, might have only *some* elements of this delirium knowledge, leaving them less able to recognise its presentation and its meaning for the patient. Introduction of a screening tool might, some suggested, improve recognition of delirium, but only if the pattern of signs presented by the patient offered a signal or trigger to the practitioner to consider delirium as a possibility. This in turn was seen to require at least some knowledge of the circumstances in which delirium typically developed. The corollary was that senior staff might perceive use of a screening tool as less relevant, focusing instead on assessment using a range of clinical tools among those patients who exhibited signs of probable delirium.

Inconsistency of knowledge

Among junior staff, the pattern of awareness and knowledge of delirium was inconsistent, shaped partly by training and partly by previous experience in a setting in which delirium was prevalent. One interviewee described how knowledge of delirium gained through geriatric experience in a care of older people ward, in addition to taught sessions, had helped him to understand delirium better, and to develop a proactive

approach towards cognitive screening which carried over into his practice in potentially less delirium-prevalent settings:

I imagine that if I'd not come from a ward, a base ward, where there was very much a focus on cognitive screening, I think I would have had much more of a hit and miss approach to the deployment of cognitive screening and I wouldn't have easily recognised its importance and prioritised doing it above other things for the patient.

Junior doctor, ED, Avonfield 06

Several others indicated that their knowledge of delirium developed into deeper understanding only when they connected their theoretical learning with practical experience of patients with delirium and its effects. Thus:

At medical school you are taught that delirium is a medical emergency, that it's an acute medical problem that needs treatment straightaway. But as much as I understand the seriousness of it, I generally appreciate finding out the source of it . . . It's more that this patient has delirium because they've got sepsis. I need to treat the sepsis, so that's what worries me more to be honest . . . So if I'm worried that . . . the patient is kind of disoriented to time and place, confused, then I'll go down that avenue in terms of what I should be looking for in the collateral history.

Junior doctor, ED, Avonfield 11

Awareness of the significance of delirium therefore enhanced sensitivity to observational cues that might indicate its presence. Furthermore, it translated into a line of exploration in history-taking and information-gathering as part of developing a differential diagnosis.

The following account further shows how important it is to understand the significance of delirium so that it can be prioritised when making a differential diagnosis. Here, a junior doctor with interests in psychiatry and emergency medicine and a recent rotation in psychiatry described how he would elicit information from an older patient, among those 'presenting an extra challenge':

I tend to ask open questions to start off with because you can just gauge whether the patient is completely oriented and knows what's going on and if I'm worried that I don't think that patient is kind of oriented to time and places, confused, then I'll go down a completely different avenue in terms of what I'll be looking for in the collateral history; if anybody's with them at the time, and chatting to them about what's happened. And I also tend to go through the AMT [cognitive assessment tool] with patients like that as well . . . it's very difficult to know what's new for them because sometimes you'll turn around to the relatives and you'll say, you know, 'Is this a new amount of confusion do you think?' and they'll be like, 'I'm not really sure to be honest,' and . . . you have to treat it as it's new . . . but trying to figure out a source of that confusion as well . . . And then depending on how well that patient is and whether there's an obvious source and what their home circumstances are like, you know, sometimes you can get them back home with a course of antibiotics, other times you need to admit them because it's not safe to send them home, given their kind of social support, and the fact that they're just very confused . . . More often than not I've ended up admitting people that have come in acutely confused, because they've just not been well enough to go home . . . and there's been other issues like they've been tachycardic or hypotensive, and so they've needed to stay in clinically.

Junior doctor, ED, Avonfield 07

It is notable here that the use of the terms 'confusion' or 'acute confusion' did not necessarily denote a lack of knowledge of delirium. The terms in this context were employed as descriptive labels drawn from the patient's history and observational cues that opened up lines of investigation to probe the source,

and that might conclude with a diagnosis of 'probable delirium'. When such a cause was not evident, this was generally viewed as a reason for admission in both the 4AT 'virgin sites':

I think anyone who's acutely confused and that you haven't . . . got an easily reversible cause . . . I think they would need admission really . . .

Junior doctor, ED, Cranford 01

And:

I would imagine that a lot of patients with confusion get admitted and it's left at a later point for people to work out if their confusion is new, the extent to which it's new . . . I would think it's difficult to arrange the speedy discharge of an A&E [accident and emergency] attender when you can't clarify the history of confusion.

Junior doctor, ED, Avonfield 06

A key issue emerging from the findings from these 4AT 'virgin' sites (i.e. where there was no systematic, routine delirium screening tool in use, including the 4AT) is that assigning value to delirium detection is at least partly contingent on experiential knowledge of delirium and its impact on patients, as well as learned knowledge. We have characterised professionals with such multilevel knowledge of delirium as 'delirium knowledgeable'.

The routine use of a delirium screening tool might be expected to enhance delirium awareness, if not knowledge of delirium, which in turn would result in its more consistent detection. The relationship appeared to be less straightforward, as suggested by interviewees from an ED within a 4AT 'experienced' site. Here, although the 4AT had been introduced into the ED more than a year previously, it was not currently in routine use. Delirium knowledge was described as patchy:

So I think it's still under-recognised . . . and you'll maybe get a handover from triage that the patient's agitated, or . . . they could be septic, that's another term. So they might have a high temperature, and it's only really picked up that they're delirious I think if someone is familiar with the concept . . . and that's usually after they've seen a doctor or an experienced nurse who's maybe done some geriatric training. But even then it can slip through the net.

Consultant, ED, Denbury 17

'Delirium knowledgeable' staff, however, were likely to 'see' behaviours as possible indicators of delirium, requiring further investigation. Similar to 'delirium knowledgeable' participants in the 4AT 'virgin' sites, it appeared that knowledge of delirium among some ED staff made them more sensitive to indicators of probable delirium when taking a patient's history. From the account below, it is evident that eliciting understanding involved a combination of strategic questioning, interpretation based on tacit knowledge of risk factors for delirium, and experience of and sensitivity to the meaning of observational cues:

Within a few minutes of speaking to them, I'll appreciate whether they've got an understanding as to where they are, why they're here . . . their presenting complaint would be one for me . . . the story of their presentation would make me think. Before I go in to see somebody I often have a very quick look in their medical history. So anything there . . . that they've been in before with confusional states, they've had UTIs [urinary tract infections], they may already have a diagnosis of cognitive impairment. So I may have that information before even going to the room. And also just how I see them. What they look like as I'm approaching them . . . all of those little things would trigger in my head . . . what's going on here . . . And I would note on the history 'have you considered . . . [Delirium]; 'does the patient need this'. But it would be the area the patient goes to for their management that should be the area that does the 4AT.

Senior nurse, ED, Denbury 16

Among 'delirium knowledgeable' staff, although the type of work involved in eliciting and noting the presence of delirium in the ED is described in this account, systematic screening of delirium was viewed as problematic in this setting. To understand why, we next consider the specific ED context and purpose.

Emergency departments: context, purpose and their impact on the value attached to delirium screening

Contextual factors pertaining to the nature of the care environment in the ED were viewed as impinging on both the capacity and the value attached to delirium screening and detection. These were of three broad types and operated across the 4AT 'virgin' and 'experienced' sites: the fast pace of work (and rapid turnover of patients) within a short, delimited time; the impact of the admission process on patients; and the purpose of the work in the ED.

Pace of work in the emergency department

Across all sites, the pace of work was uniformly perceived as a barrier to delirium screening:

The problem is that it's such a rapid turnover there's probably a lot of concern from staff that they don't want to fill in another score sheet or another tool . . . I've had similar problems trying to institute scores for frailty . . . no one really wants to have more work to take on . . . it's probably quite a short-sighted view . . . the scores we have on the triage sheet at the moment are vastly undercompleted even for such questions as 'what's the patient's early warning score' and 'do you think the patient's septic?'

Senior doctor, ED, Cranford 02

Two specific features were raised that highlighted how critical time was in the ED: (1) the short time available for working with each patient to arrive at a decision about the immediate action to be taken; and (2) the access to information relevant to addressing the key features of probable delirium, namely whether or not the 'confusion' was of rapid onset and was fluctuating.

Time in the emergency department to achieve purpose

Decisions about the immediate action to be taken for the patient were made quickly, usually within 4 hours, to meet the performance target and manage demand. Even a screening tool described as brief and concise, such as the 4AT, was considered impractical, or at least it was not seen as a priority in this context, as a nurse in the 4AT 'experienced' site indicated:

It's not consistently completed particularly on days like today, when you are extremely busy. It's probably not one of the first things that you would prioritise . . . because of staffing more than anything . . . it's not something that people automatically think about.

Staff nurse, ED, Denbury 18

This pattern was corroborated by interviewees working on the MAU and acute wards in this site, who reported that the 4AT was rarely completed for patients who came to them. Typically, documentation from the ED comprised a patient history, and information on acute presentation and differential diagnosis. Even so, staff acknowledged that it was difficult to accurately assess cognitive impairment in the ED environment because of the impact of the admission and the event that gave rise to it (see *Impact of the admission process on patients*).

A key feature of delirium is its fluctuating course and recent onset. In the ED, time and context were perceived as barriers to accessing relevant information for addressing what was an apparently simple question. Thus:

The acute . . . and fluctuating course . . . that's so difficult in ED . . . when this is done, within 4 hours of them coming in. Unless we've got the family there or we know them well, I think that it'd be really difficult to get a judgement on that . . . unless you've got a really good history from someone else I don't think you'd be able to . . . score that very accurately.

Junior doctor, Care of older people ward, Denbury 13

Impact of the admission process on patients

Features of delirium, such as 'alertness', were viewed as affected by time and context, with the busy, dynamic environment making it difficult for patients to concentrate. It was also argued that the trauma of the event giving rise to admission meant that patients required reassurance and support, and 'tests' that were aimed at identifying deficits might add to a patient's anxiety. For 'delirium knowledgeable' practitioners, having reassuring conversations with patients, and employing 'comforting' strategies to alleviate distress, could enable understanding of cognitive impairment as part of a whole picture:

I think when people come in and they're either in pain or distressed or, you know, that they've had a journey by ambulance, they're excited, they're you know, upset by that . . . It's a busy noisy department and so concentration possibly is an issue for patients. It's probably not the best time . . . to be asking them to count backwards, or you know 'tell me the numbers from' . . . We would all struggle with no cognitive impairment under those circumstances . . . So I do get that . . . just informally, chatting to them can give you quite a lot of information without putting them through the added sort of almost 'examination' stress, of doing a formalised test.

Senior nurse, ED, Denbury 16

The approach conveyed here offers a partial explanation of how 'knowledgeable delirium practitioners' are sensitive to, and take account of, delirium risk as part of their overall assessment, yet eschew the use of a screening tool.

Purpose of emergency department work

Several participants described the ED as a 'blank canvas' insofar as the work was unpredictable. The primary purpose of the work was viewed as obtaining a differential diagnosis to inform decision-making about an immediate course of action within a designated time. The process was conceived of as 'layers of decision-making': first, deciding if the patient was seriously ill; second, deciding whether or not to admit the patient, cognoscente of the gatekeeping function of the ED; and third, deciding whether or not a probable diagnosis fitted the overall patient picture. Prioritisation of delirium screening in this setting was influenced by its perceived value in contributing to such decision-making. Even when delirium detection in itself was valued, unless it influenced decision-making it might not be prioritised:

The icing on the cake almost is where a diagnosis fits into that because you may not know precisely what the diagnosis is . . . you might have a list of differential diagnoses . . . so that for the person in front of you, you don't actually know what the cause of their problem is but you know that they can't go home . . .

Consultant, ED, Avonfield 01

Across all sites, it was suggested that a screening tool (whether the 4AT or another tool) would be used more consistently if staff understood why it was important to use it, if they could see tangible benefits for patients and if screening was seen to make a difference to patient outcomes in this particular setting:

It's a good concept. The problem is that trying to change people's behaviour is very difficult if they don't understand why they are doing something . . . What we kind of want to know, in the emergency department, is . . . if by doing this score are we making a difference? And, you know, how many, what effect is this having on patients dying or on saving patients from getting unnecessary treatment.

Consultant, ED, Denbury 17

Even so, an assessment of benefit would probably involve consideration of the nature of the ED environment and the challenges of context and time in that setting. Additionally, it was noted that experienced and 'delirium aware' ED staff considered delirium risk as a factor in their overall assessment, although this drew on pattern recognition through multiple sources of information, including interpretation of observational cues. Non-'delirium aware' staff, by contrast, were unlikely to 'see' and interpret this information. Paradoxically, all other things being equal, knowledge of delirium in the ED context might

obviate the value attached to the use of a screening tool, although, similarly, a lack of delirium knowledge contributed to non-use of such a tool.

Assessment units: medical assessment unit; elderly/frailty assessment unit; surgical assessment unit

In contrast to the ED, there is a paucity of research on delirium prevalence and detection in assessment units. Generic assessment units are similar to EDs in that they manage demand and flow near the point of entry to the hospital within a relatively short time and have a diverse patient profile in terms of age and clinical presentation. However, dedicated assessment units for older people, although subject to the same time and pace pressures, are more likely to be staffed by those with interest and expertise in the care of older people.

Assessment unit: an observational snapshot

These varied types of assessment unit operate at the next point of the patient journey from the ED. These are short-stay facilities (usually for between 24 and 48 hours or up to 72 hours) to enable further observation, investigations and decision-making in order to develop a management plan for the patient. The work here includes determining the most appropriate course of action and destination for patients [acute ward admission, discharge to an alternative setting (intermediate care, community hospital, hospital at home) or discharge home], and mobilising the necessary support and resources to achieve the intended outcome. Whereas assessment units within two hospital sites included dedicated units for older people or those with frailty (EAU), the third operated as a generic unit (MAU) within which patients with medical needs were accommodated irrespective of their age. In one site, we included a SAU.

The physical environments of assessment units in the study had a similar layout to acute wards with varied combinations of male and female bays and single rooms; and a work station in which medical and nursing staff congregated, checking computers, working the phones, examining case notes. The rhythm of the daily routine was shaped by purpose: engagement in clinical observations and multiple investigations to determine destination outcome; and provision of medical, personal and emotional care for patients with variable levels of acute need. Both lines of work contributed to an intense, noisy, busy, and fast-paced care environment, albeit one that was typically somewhat less frenetic than that in the ED.

An area of high patient flow, the impression of the MAU/EAU was a place of continuous movement – of patients and staff – and a cacophony of sound. New patients were brought in on trolleys by porters; others sat, dressed, by their beds waiting to be discharged; yet others were wheeled away to acute wards. Medical and nursing staff moved between patients and clinical stations, variously carrying out examinations, ordering investigations, perusing results, conducting observations and eliciting relevant clinical and biographical information from patients. Nurses in particular spent considerable time on the telephone checking out information on medical history, service use and patient background with primary and community care professionals and with family members. Ward rounds, several times daily, drew on these multiple sources of information to achieve pace and flow.

The orderliness of the process could be deceptive; conversations with individual patients were disrupted as staff were called away to respond to the sound of a buzzer, a patient calling out or a newly admitted patient needing to be settled in. The flow was such that being a staff member down or having a very ill patient could transform what might be construed as organised chaos into a chaotic care environment. In contrast to typical acute wards, patient movement in and through the setting continued through the night.

The clinical rhythm was punctuated by the care regimen: mealtimes, physical care (washing, assistance with toileting, mobilising) and housekeeping tasks (bed-making, distributing drinks, cleaning). Here, health-care assistants moved quickly through the bays and rooms, cleaning, turning, washing, toileting, serving meals and conversing with patients, creating multiple opportunities to notice slight changes in the patient's condition, the communication of which relied on a verbal exchange with nursing or medical colleagues.

There were similarities and differences between the MAU's organisation to achieve its clinical and service purpose to that of the ED. One similarity was the timeframe within which the work was organised, although in the MAU this was longer. Extending the period from 12 to 48 hours meant that the work encompassed tasks relating to personal care, bodywork and nurture (nutrition and hydration), as well as clinical observations, investigations and more detailed information-gathering from diverse sources. Observation of patients over different times of the day and night, and ongoing direct patient contact through care routines aligned with enhanced knowledge of the person's usual state through the information-gathering process, made assessment of 'fluctuating' and 'new onset' easier.

Assessment unit: knowledge and understanding of delirium and delirium detection

Medical assessment unit

Like their ED colleagues, members of the multidisciplinary team (MDT) in assessment units were required to draw on their knowledge of a wide range of conditions to pursue an investigatory strategy to develop and refine their diagnosis and to implement a management plan for the patient. Importantly, and in contrast to the ED, the lengthier assessment period meant that ascertaining relevant knowledge of the patient's biography and medical history was more feasible, despite the fact that time was also regarded as a scarce resource.

Similar to in the ED, some senior doctors in this setting conveyed a breadth of clinical experience through which they had acquired knowledge and understanding of delirium: they were 'delirium knowledgeable'. Additionally, specific local initiatives directed at enhancing awareness of delirium among staff, as in Cranford, had, it was suggested, improved delirium knowledge among MAU professionals:

Nurses here are fairly switched on as far as delirium goes. We don't have so many patients here as we do on elderly care who have fluctuating confusion, but we do get it here and nurses deal with it. I trust them.

Senior doctor, MAU, Cranford 02

In Denbury, implementation of the 4AT in the MAU was regarded as having enhanced awareness of delirium. This was symbolised in the language employed, with the term 'delirium' explicitly used instead of 'cognitively impaired' or 'not quite right'. The 4AT's introduction had been supported by a short information session on its use as well as on the significance of delirium. Nursing staff were designated as responsible for screening for delirium and completing the tool, with the 4AT being incorporated into the nursing assessment. Even so, interviews showed that multilevel delirium knowledge among MAU staff in this site was variable.

'Delirium knowledgeable' interviewees in Denbury MAU described the process of identifying patients with probable delirium in a similar way to 'delirium knowledgeable' participants in 4AT 'virgin' sites, namely that it was seen as involving more than completing a score on the tool. It involved 'empathic connection' with a patient's distress and anxiety as a result of the event giving rise to the admission, history-taking in a 'comforting' conversational style to help the patient relax, and eliciting information 'almost like a chat' to ascertain whether or not the patient might have cognitive impairment. Discrepancy in the accounts from the patients were described as 'red flags', triggering questions such as 'What is going on here?' and 'Is this normal?'. These also prompted further information-gathering. Completion of the tool by these nurses in this context formalised what might in the past have been a comment to the doctor that the patient was 'really confused but it isn't normal for her'. They regarded screening for delirium as a necessary and legitimate part of the habitual assessment work in the MAU. A positive screen also directed and prescribed action:

I would check what bloods have been done in accident and emergency and probably add a few more tests because we're looking at a confusion screen . . . widening the screen . . . and should help get a diagnosis quicker [of possible infection] and work out the appropriate antibiotic.

Staff nurse, MAU, Denbury 04

Whereas delirium knowledge and its management enabled nursing staff to take the initiative and act on the results of screening, for others the perceived value of screening for patient care was not self-evident. This had a knowledge dimension: if staff did not understand the significance of delirium, they were less motivated to use the tool. This also had an action dimension: if the pathway from detection to action was unclear, sustaining engagement with screening could be compromised.

I think there's an assumption that we'll just carry on with this person who has got known cognitive impairment and we'll carry on observing it, which I suppose is what the 4AT is for . . . But what are we doing with it? . . . Are we doing anything about a known cognitive impairment when they're scoring one to three? I don't know.

Staff nurse, MAU, Denbury 08

Several nurses indicated that, when under time constraints, they would prioritise specific assessments over the 4AT; skin/pressure assessments were specifically mentioned as more visible indicators of performance and care quality. Although noting that they had used the delirium screening tool on only a few occasions, they also acknowledged that 'others do it more often'. Yet others prefaced their accounts of use by referring to what 'should be done': it was what was expected of them.

Variable knowledge of delirium and lack of clarity on the pathway from detection could affect consistency in use of the screening tool.

Surgical assessment unit

We conducted the study in only one SAU; therefore, the picture that emerged from comments of interviewees, of a low level of understanding and awareness of delirium, may not be typical of these units. Even so, this echoes the accounts of the care of older people liaison team on orthopaedic wards in Denbury.

In Cranford's SAU, senior nursing staff acknowledged the delirium knowledge deficit, and viewed this as problematic in the light of the high level of risk among one group of their patients, namely older people with hip fracture. Even though information-gathering did not include use of a screening tool for delirium, the nursing assessment pro forma contained four questions relating to risk factors for delirium, based on NICE guidelines (having current hip fracture, being aged ≥ 65 years, having cognitive impairment and having severe illness). For those at risk of delirium, a pro forma delirium prevention care plan was to be completed (e.g. action on hydration, nutrition, pain, medication review). Both were introduced on wards as part of a drive to increase awareness of delirium and delirium prevention (albeit without explanation, training or support in use). Interviewees here were unclear about how these risk factors were established and about the content and purpose of the delirium prevention plan. Unsurprisingly, they reported that the pro forma was not completed on the unit. A senior nurse, who was supportive of the idea of delirium knowledge as enabling staff to 'see' the value of identifying delirium, welcomed the introduction of a short, easy-to-use tool such as the 4AT, but went on to suggest that, to trigger use in this setting, staff needed to understand why they should invest in it and what action would be taken as a consequence:

It looks pretty user-friendly and that it won't take too long to do which is the key, you know, 'cos otherwise people just won't do it. But then obviously we need to know why we're doing it and what to do once they trigger a certain score as well.

Senior nurse practitioner, SAU, Cranford 06

Elderly assessment unit/frailty unit

In EAUs/frailty units, staff presented as experienced in and knowledgeable about the multiple needs of older people, including those arising from a cognitive impairment. Among their patients, delirium was common. Identifying, assessing and developing a management plan for those with delirium was seen as a routine part of their work, although how this was accomplished differed between the two 4AT 'virgin' sites.

In Cranford EAU, although the patient assessment form did not include a delirium screening tool, it was commented that experience in caring for older people meant that 'acute confusion' was at the forefront of thinking and that any indication of inattention would sensitise staff to look for 'reversible causes'. In this site, the staff's perceived skill in noting and making sense of changes in the patient contributed to the low value attached to screening tools; confidence in observing and reporting changes meant that medical staff could proceed to assessing and investigating factors contributing to delirium. Moreover, as the purpose of the setting was to develop a management plan, tools were indicative of a problem that would prompt formation of an action plan, but the view was that they did not in themselves inform such a plan:

As a geriatrician what you are doing is looking at the whole picture to inform the most appropriate management plan . . . you are using your observational skills as a doctor; you are also drawing on cues from staff from their knowledge of the patient. Nursing staff will also know – but they may not use the word 'delirium'. They've observed that the patient has become more muddled, more confused as they've seen them on the ward; or the relative has said they are more confused now than they were at home.

Consultant, EAU, Cranford 05

In Avonfield, the AMT10 was conducted with all patients as a case-finding tool for dementia, with a flow chart to delineate resulting action. Aligned with the collateral history, this was also used to identify new, acute confusion. The rationale for this approach was the co-occurrence of delirium and dementia and to engage staff in thinking about the link between them.

The pattern observed previously in ED – that understanding and knowledge of delirium meant that staff were more sensitive to picking up on observational cues that might, for example, indicate change in alertness and perceptual disturbances – was evident in both units. An understanding of the risk factors and a direct, routine involvement with patients meant that nursing and care staff were seen to play a crucial role in ascertaining if a patient did not seem to be 'quite right'. Both doctors and nurses interviewed emphasised their joint and collaborative endeavour in identifying delirium: 'they appreciate that it means something's going on that needs investigating'. Medical staff were seen to note and act on the information communicated by nursing and care staff; the latter in turn felt valued and confident about communicating their observations in MDT meetings, in ward rounds and informally with medical colleagues.

Medical assessment unit/elderly assessment unit: context and purpose – impact on the value attached to delirium screening

Two contextual features of the care environment that impacted on delirium screening and detection in MAU/EAU settings in the same way as in the ED were the pace of work and the perceived alignment of organisational purpose with delirium detection and assessment. Other features more specific to the MAU/EAU setting were the systems in place for integrating delirium detection into the care culture and the processes established to legitimise the contribution of different groups of staff.

Pace of work

A feature of the MAU/EAU was that, in comparison with the ED, staff had a longer and more sustained period in which to observe, collect collateral information and review a patients' progress. As patients were admitted at all times of the day and night, the constant busyness and noise of the unit disturbed those trying to rest or sleep. Among some staff, the continuous hum of activity and the merging of day and night meant that it was a 'bad environment' in which to screen for delirium. Furthermore, it was considered that patients were likely to be disoriented and distressed, particularly those admitted during the night, on account of both the medical event that precipitated hospital admission and the admission itself. Generally, although the nature of the care environment in the MAU might make it difficult to obtain an accurate baseline picture from which to determine delirium, the consensus was that it did not create an insurmountable barrier.

Systems for integrating delirium detection into routine practice

In the 4AT 'experienced' MAU setting, participants described the work undertaken to implement the 4AT into ward routines. The 4AT was incorporated into the unitary patient record (UPR), which comprised multiple screening and assessment tools and was intended to be completed by nursing staff for all new admissions along with documentation of 'care rounding'. The documentation involved daily interval review and documentation by nurses of skin/pressure areas, mobilisation, toileting, continence and catheter care hygiene, analgesia, pain, fluid balance and care plans. Direct, ongoing engagement of nursing (and care) staff with patients meant that they were seen as well placed to observe change and fluctuations. A division of labour was in place such that nurses were responsible for completing the delirium screening tool and junior doctors conducted the cognitive assessments (AMT and MMSE) as part of the clerking process.

Some aspects of the system to implement delirium screening in Denbury were not universally regarded as meaningful. A senior nurse questioned the value of screening everyone who met the specific age criteria; she felt that screening should be more flexible and draw on tacit knowledge about and expertise in delirium risk: 'use your head'. Thus, it was argued that observation of and conversations with patients should inform which patients were 'at risk' of delirium and, therefore, should be screened. Such an approach was, however, predicated on having an in-depth knowledge of delirium and delirium risk:

In our paperwork it says over 65 years . . . we're trying to make it 'use your head', basically. If they're 55 but they're ill enough that you're concerned that they've got some sort of delirium/sepsis, use the tool . . . But the lady in bed [bed number], who is over 65, so technically we should be doing the 4AT. She's all with it. She's in for cardiac reasons. She's no side-symptom of infection. And I just explained it to her by 'it's a screening tool we have for anyone over 65, do you mind?', 'no', but she laughed all through it . . . I was a bit embarrassed about her but . . . that's the first time I kind of thought 'why am I doing this?'

Senior nurse, MAU, Denbury 04

In the context of a MAU, responding to the heterogeneity of the needs of patients across age groups and with multiple and varied presentations posed particular issues for some staff in using the 4AT delirium screening tool. Highlighted was the way in which staff with knowledge of particular conditions employed that specific knowledge and skill as a lens to make sense of, or contextualise, observational cues and behaviours. The extended narrative account below about toxicology illustrates how the overall pattern of presentation confers meaning. It also suggests that practitioners may judge that routinely screening patients is unnecessary without considering the delirium risk, and that they may have an alternative explanation for symptoms because of their knowledge and skills in a different field of medicine:

For the younger age group you . . . look for the obvious first. If they've just drunk a bottle of vodka you're looking at intoxication. What's the chance of this person . . . having a delirious state before coming in? If there's no history of it look at the obvious and deal with it from there . . . Or . . . if there's a history of an ongoing affective disorder or something else going on from a psychiatry point of view I would think psychiatry rather than a delirium. And if you get the history then it's going to show up . . . So the history taking is going to give you the indication. The 4AT is a bit of a snapshot when they arrive; you need a bit more of a history. So regarding the criterion: 'has there been a kind of sudden onset?' Well your history taking is going to show you that. So it's part of it from there. But if you . . . get a 3 out of 4 on 4AT . . . is it going to change? And it's just another number where you think 'fine', but that could be explained by the fact that they've just drunk a huge amount of alcohol 4 hours ago. You're giving yourself a reason why there's a 4. You don't necessarily need to be think that . . . we're looking at delirium; no, we're looking at an intoxication.

Senior nurse, MAU, Denbury 08

Legitimising the contribution of different staff groups

Among some nurses, the value of the standardised tool was twofold. It gave credence to nursing observations and enabled nurses to exercise professional autonomy to pursue investigatory action, without having to defer to medical colleagues to endorse it.

The understanding of delirium in the context of a perceived valued professional role in detection and action contributed to ownership of the process and sustained ongoing use. From this perspective, screening was not simply a 'tick-box' exercise. It involved knowledge-informed interpretation and decision-making on the part of nursing staff.

Others questioned whether or not medical colleagues accorded value to the nurses' work in detecting delirium. If their work within the overall process of detection and action on delirium was neither understood nor accorded value, this could also inhibit sustained participation:

Doctors aren't coming to say '4AT, we're acting on it, we've done this . . .' And there's not enough of: 'what we are doing next, are we responding to a sepsis or are we responding to . . .' It's a multidisciplinary team thing at the end of the day . . .

Staff nurse, MAU, Denbury 07

Overall, delirium screening and assessment was viewed as consistent with the purpose of assessment units, namely to develop a patient management plan within a more extended timescale than was possible in the ED. Nevertheless, as evident in all sites, knowledge of delirium affected the significance attached to detecting and assessing it when formulating a plan of care.

From the accounts provided by the ward staff interviewed, the MAU had relatively high 4AT completion rates (reported as 60–70%). Factors contributing to variability included variable knowledge of delirium, which negatively affected the value attached to identifying it; and uncertainty about whether or not screening was followed by action and, therefore, accorded value by members of the MDT. Moreover, although the delirium screening tool appeared a simple and 'objective' indicator of probable delirium, the interpretation of some items, such as 'recent' and 'fluctuating', were not regarded as straightforward; rather, they made sense as part of an overall pattern in context. Similarly, presenting features that might indicate probable delirium could, depending on the context, be interpreted as meaning something else.

Acute wards

The occurrence of delirium on acute wards varies with patients' severity of illness and age profile. This pattern was evident in a study¹³⁸ that uniquely examined the prevalence and incidence of delirium across acute wards in a large, tertiary hospital using a common suite of screening and diagnostic tools. Whereas an overall prevalence of 20.7% represented the burden of delirium in the hospital, this varied between wards, being highest among patients on medical, neurosurgical and orthopaedic wards and lowest in general surgical wards. Furthermore, those in advanced older age (i.e. ≥ 80 years) had nearly 35% delirium prevalence, compared with around 5% among patients aged < 50 years; and over half of patients with delirium had a pre-existing cognitive impairment. These findings underscore the significance of delirium detection and management being a major focus of clinical and care work on particular acute wards; the importance of detecting and assessing both delirium and cognitive impairment; and consideration of their reciprocal impact on treatment and recovery.

Acute ward: an observational snapshot

The organisation of space in acute wards was not dissimilar to that in assessment units. There were varied combinations of four- and six-bed bays and single rooms, clinical areas, and staff working spaces of various sizes. Policy initiatives in the UK^{139–141} have directed attention to the need for environmental changes on acute hospital wards to create easier-to-navigate spaces for patients who have a cognitive impairment. This has included colour-coding of bays and using signs on toilets and bathrooms. Most care of older people wards in the study (but not other medical wards) had pursued such changes, although the

nature and extent of these varied considerably; for example, space for patients to engage in activities or interact socially with each other was very limited.

Similar to assessment units, the daily rhythm of acute wards was shaped by the purpose of the setting: making clinical observations and dispensing medication, holding ward rounds and MDT meetings to develop and review treatment plans, progress, discharge readiness and destination; and providing medical, personal and emotional care for patients with varying levels of acute need. The daily routine and the tasks that constituted it had a similar shape and content on all study acute wards, although the timing and degree of flexibility varied. The routine working day for ward staff began early: handover from the night to the day shift around 7 a.m., followed by patient-related tasks such as dispensing medication, assisting with personal care (such as washing and toileting) and organising and supporting patients with their breakfast. These tasks were often opportunities to converse with patients and develop knowledge about their biography and preferences.

The period following breakfast to late morning was a flurry of activity and cacophony of sounds as members of the ward team (therapists, doctors, nurses, care staff and ward clerk) and off-ward staff (phlebotomists, pharmacists and porters) went about their work. For nursing and care staff, work with individual patients was often interrupted as they responded to others pressing buzzers for assistance or calling out in distress.

At midday, the round of mealtimes, toileting and dispensing of medication resumed. On several wards, open visiting from late morning saw a trickle of visitors, who were often encouraged to assist their relative or friend with eating. On most wards, visiting times were scheduled for up to 2 hours from 2 p.m., and for up to 2 hours from 6.30 p.m. In practice, there was considerable flexibility: when patients were very ill, or when visitors had travelled a long distance or when they had simply arrived early. The steady stream of visitors in the afternoon altered the ward rhythm as movement and noise emanated from patients and visitors. Their departure heralded the repetition of routine tasks: meal, clinical observations, toileting and then visiting again. The exiting of visitors in the evening was the signal for commencing night work. Visiting times were often a period during which nurses sought out, or were approached by, relatives to give or elicit information.

Formal handover from day to night staff reversed the sequence of the early morning, and another busy period of putting patients to bed, toileting and medications began. For patients, the dimming of the central lights in the bays signalled the day's end, although movement within bays and between beds, and between bays and toilets, continued into the night. In summary, the rhythm of acute wards in comparison with that of the ED and the MAU appeared more orderly and less frenetic, although for staff there was no let-up in a round of activity that involved continuous movement. Nevertheless, staff had opportunities within their daily routines to observe and converse with patients and to develop knowledge about what was 'normal' for those patients in terms of presentation and behaviour.

Acute wards: knowledge and understanding of delirium

From the findings relating to the ED and the MAU, one would anticipate that awareness and knowledge of delirium and the value attached to identifying it would vary greatly depending on the ward patient profile. This was broadly substantiated, although the depth of knowledge of delirium among staff on care of older people wards was also variable.

Care of older people wards

As noted earlier, delirium and dementia often co-occur in older patients and the prevalence of delirium on care of older people wards is high. In the study, most of the wards were care of older people wards. Although medical and nursing staff on these wards had experience of 'seeing' and responding to patients with both delirium and delirium on dementia, knowledge of delirium embraced the spectrum from 'limited' or 'general awareness' to 'delirium knowledgeable' as characterised above, namely in-depth

understanding of its heterogeneous presentation, trajectory, distress caused to patients and anxiety of relatives.

The 4AT 'virgin' care of older people wards in this study was, in some respects, atypical of care of older people wards generally in their knowledge of delirium and in the value attached to identifying and managing it. The Avonfield ward was designated for patients with dementia and delirium; and the Cranford ward had been involved in research on delirium prevention.

On both wards, having knowledge of delirium and the significance attached to it was viewed as sensitising staff to observable signs of change. Thus:

The wards I work on staff understand the value of identifying delirium . . . they appreciate that it means something's going on that needs assessing . . . and even hypoactive delirium, you know, they notice if someone's not waking up for them and taking their medication . . . they do pick it up and I feel they do know to act on it . . .

Consultant, frailty assessment unit and care of older people ward, Avonfield 02

The close working environment and frequent interaction between staff, patients and caregivers were seen as enabling staff to develop a shared understanding of what was 'normal' for the patient and to work with the patient as a person:

I think perhaps in terms of a general experience point of view I suppose it is good to have because I don't know where else you'd get this experience because [delirium] is a common problem, especially in older people . . . you are treating the patient as a whole rather than just the illness that they've come in with, which is good.

Junior doctor, Care of older people ward, Avonfield 05

Health-care assistants, nurses, doctors and therapists conveyed a deep understanding of people with a cognitive impairment. There was an expectation on nursing and care staff to spend time with patients and acquire personally meaningful knowledge of those patients as individuals:

You get to know your patients when they're not right: 'oh Elsie's not right today . . . she's sleeping more than usual . . . not communicating as she would normally . . .' we spend a lot of time with patients so we notice changes when you're washing or bed-bathing the person . . . and because everybody's different . . . it's really important that you know the person as an individual and what they're like on a daily basis . . . we talk about those changes at handover.

Health support worker, Care of older people ward, Avonfield 09

Personally meaningful knowledge acquired by health-care assistants through their direct contact with patients was valued, communicated, listened to and shared in multidisciplinary fora.

In care of older people 4AT 'experienced' wards, nursing staff reported that the introduction of local guidelines for delirium, aligned with organisational impetus on screening with 4AT implementation, had brought delirium further to the front of awareness; and that this was in the context of a pre-existing knowledge gap. Thus:

Interviewer: . . . what sort of training have you had in the past on delirium?

Interviewee: [shakes head] . . . None [then goes on] . . . I have looked at information websites for myself . . . and there has been a big drive on delirium just now in the hospital as well as a sepsis drive, to increase detection of delirium and get us to distinguish between delirium and dementia . . .

Staff nurse, Care of older people ward, Denbury 02

Knowledge of the significance of delirium conferred priority on detection and action:

... the recognition of delirium is crucial 'cos it's like any acute medical condition it needs to be addressed as importantly as you would do the rest, the heart attack or anything else ... we're all getting better at identifying it earlier ... and there's quite a focus on it in care of older people ... now it should impact their medical care, their nursing care and whether they will have the ability to engage with physio[therapist] or OT [occupational therapist] so it has to be part of reviewing their progress or their medical condition ...

Clinical fellow, Care of older people ward, Denbury 09

Awareness of delirium, it was suggested, meant that 'problem' behaviour of, for example, patients 'kicking, spitting and aggressive' could be understood in context of their condition, facilitating empathic connection with the patient:

It makes it, in your head, more acceptable, because you know that [normally] they would never act that way, they wouldn't ...

Staff nurse, Care of older people ward, Denbury 21

The level of awareness and attention about delirium on care of older people wards was in marked contrast to that reported among staff in other medical and surgical wards. Junior doctors in particular, from their experience of multiple rotations in various surgical and medical specialties, offered a wide comparative perspective. The prevalence of delirium (or cognitive impairment) on specific surgical wards, such as cardiothoracic, was viewed as very low in comparison with that on care of older people wards. The profile of patients (younger) and the clinical purpose (mainly elective procedures) suggested low risk, which was seen to explain why neither cognitive nor delirium screening appeared on clerk-in sheets, also justified on the basis of very low prevalence. On orthopaedic wards, the high prevalence of delirium among subgroups of patients, particularly older people with a hip fracture, had been addressed by having a liaison care team for older people to provide support for these patients. Although working closely with nurses and junior doctors on these wards, a division of labour existed such that senior surgical staff were focused on surgical aspects of care and had little involvement with nurses and junior doctors in relation to other aspects of the work.

Therefore, from a comparative perspective, staff on care of older people wards had a heightened awareness of delirium, given their exposure to patients at risk of, and experiencing, delirium. Moreover, nursing and care staff were in a powerful position to note changes in patients because they had more direct and prolonged contact with them. However, the descriptions of practice revealed that some staff were uncertain about what to do if they identified delirium in this setting.

Acute wards: context and purpose: impact on the value attached to delirium screening

The features of context in the acute ward that operated as barriers to, or facilitators of, delirium detection, were systems for integrating delirium detection into routine care management; and interdisciplinary working practices legitimising the contribution of different groups of staff. These were, in turn, aligned with the primary purpose of work in acute ward settings, namely to provide treatment and care to effect recovery and appropriate discharge of patients with conditions that had a risk of cognitive impairment, including delirium.

Systems for integrating delirium identification into care management in routine practice

Our study was focused on delirium detection. What emerged from interviews with staff on care of older people acute wards was that there was an intimate relationship between knowledge of delirium, the value attached to detection and the processes in place for the subsequent management of patients with delirium.

In Denbury, study participants were from three different care of older people wards. Staff interviews in each ward showed some ambiguity about which professionals were responsible for detecting new cases of delirium and the use of the 4AT in that process; the impact that detection had on the management of

patients identified with delirium, including those for whom a positive screen had been completed in the MAU; and what systems were in place to review recovery from delirium. It is possible that the disparate findings reflect the fact that interviews took place while considerable movement of staff was happening: junior medical staff had just begun their new rotation, which also coincided with the recruitment and induction of nurses into the department. Thus, the findings may pertain to local, temporal features of care delivery. Even so, they pose organisational issues relevant to embedding new practices on delirium into work routines.

Delirium detection: division of labour or disconnection?

Among the junior doctors interviewed from Denbury care of older people wards, there was consensus that, as part of the clerking-in process, it was their responsibility to ensure that a cognitive assessment was conducted, typically using the AMT10. The 4AT, on the other hand, was for nurses to complete, consistent with the system operating on the MAU.

From the doctors' perspective, their cognitive assessments informed a plan of treatment that was then reviewed by a senior doctor in which 'it's the AMT we refer to'. The 4AT, on the other hand, was in the nursing profile (on their assumption that it was completed by nursing staff) and not something to which they would refer or look at:

Whereas the AMT because it's in our . . . booklet . . . that is what we look at . . . the medical team don't use the 4AT very much . . . I don't think many people would go into the nursing notes to find the 4AT score . . . I've never had one of the nurses flag the 4AT score to me. They might flag that they think the patient is confused . . . I've never had anybody say: 'oh we're worried about . . . and this was their score'.

Junior doctor, Care of older people ward, Denbury 03

Among doctors, the rationale for using the AMT10 was twofold: it was the preferred tool on the wards and on their clerking-in checklist; and it was familiar: 'I know it by heart; I don't need to refer to it'. The 4AT, on the other hand, was unfamiliar and the meaning of the score as a representation of the patient's presentation was not obvious.

Generally, the picture that nursing staff conveyed was that if the 4AT was completed on the MAU, it was not repeated on admission to the ward; only in its absence would it be done on ward admission.

'Delirium knowledgeable' nurses emphasised their reliance on behavioural cues that might indicate a change and a probable new episode of delirium; the formal system for delirium detection using the 4AT did not feature:

I think the nurses would go to the doctors . . . and say 'we think' or during a ward round . . . or one of the rapid rundowns we would say: 'overnight this . . . [happened] . . . 'what do you think, we've sent urine . . . blood'. A lot of the time we do take the initiative . . . if we think there's an infective source.

Staff nurse, Care of older people ward, Denbury 02

These nurses would take the initiative in investigating the source of the observed changes, drawing on tacit knowledge – 'deep down you can distinguish delirium from dementia' – and on learned and experience-based knowledge: they did not wait for medical endorsement of the decision before acting.

Among staff generally, it was reported that serial use of the 4AT did not occur, or even that it was appropriate:

It's almost by the time [the patient gets to the ward] I suspect we've already got all the information that a 4AT would capture and so we should be doing more than just a 4AT . . . it gives us a baseline . . . then it's not just screening . . . it becomes an investigation doesn't it? . . .

Occupational therapist, Care of older people ward, Denbury 22

Among some senior doctors, the interface between the detection and management of patients with delirium posed challenges that required further work to address:

I think everybody should be screened, but then it's knowing what to do after the screening tests, and how to do it . . . it's then the follow-up, and I think that's what we struggle with . . . that's a general problem [not unique to us]. We should reassess things . . . [some patients with delirium] might well have an undiagnosed dementia that we can't diagnose yet because they've got delirium . . . and for [new cases] is a different approach needed?

Consultant, Care of older people ward, Denbury 15

Integrating detection with the work of management

In the MAU, the purpose of delirium detection using the 4AT was broadly understood by staff, and there were systems in place to include the 4AT within the nursing UPR, thereby establishing clear lines of accountability for completing it as part of the nursing assessment. By contrast, ward nursing staff conveyed considerable uncertainty about the purpose of the 4AT in this setting. Although it was included in the nursing assessment, medical staff assumed the lead in assessing cognitive impairment and forming an action plan in relation to it and nursing staff seen to provide a supporting role in feeding back observational cues that might be indicative of delirium.

Senior staff on Denbury wards referred to routine MDT meetings as arenas in which the progress of recovery from an episode of delirium was discussed and new cases of delirium were identified. This included decision-making on whether a patient was safe enough to move to another ward or to engage with rehabilitation or for the team to assess the person's progress with recovery. They saw nursing staff as playing a critical role in this decision-making as a consequence of their more direct and ongoing interaction with patients. Even so, for nurses, the question posed was where the tool fitted into routine practice.

Aspects of the ways in which delirium management was accomplished on Avonfield's care of older people ward are also illustrative of integration of delirium detection and management and co-ordination of interdisciplinary work is practised. Similar to on Denbury wards, delirium was initially identified in the EAU, using the AMT10. Patients with delirium then underwent routine investigations to examine aetiology. As part of the admission to the ward, a joint treatment plan was discussed at the MDT meeting. The ward care culture prioritised spending time with patients, and also drawing on the families' own knowledge of the person and engaging in practices that were regarded as the 'good nursing aspects of delirium management'. These included non-pharmacological strategies, such as attention to hydration and nutrition, sleep, hygiene, mobilising patients and stimulation. Nursing and care staff reported that knowledge of patients, aligned with understanding of delirium, meant that they were sensitive to what was 'normal' for the person. Observations were communicated formally and informally through forums such as MDT meetings and handovers. Person-knowledge informed how staff worked with patients to effect a calm and supportive environment:

Some of our patients with delirium we have to provide one-to-one nursing which makes it difficult for the rest of the ward 'cos that takes one whole person off the ward team . . . so a patient we had recently, he just wanted you to walk around and hold his hand all the time, otherwise, you know, he'd get distressed or fall or something, so it was doing that. Or it's observing somebody from a bit of a distance 'cos some patients get overwhelmed if somebody's with them all the time. It's just like I say, it's just assessing each patient individually.

Senior nurse, Care of older people ward, Avonfield 08

The ongoing review of progress with treatment plans, supported through the systematic observation of patients' emotional, behavioural and relational action and interaction and involvement of families, contributed to the evaluation of progress. The work was conceived of as a joint and shared enterprise involving all members of the MDT. It was the consistency in the way in which joint work was pursued that reinforced the sense of value in each other's particular disciplinary expertise.

Discussion

The novel features of this study were, first, exploration of the relationship between knowledge of delirium among health professionals, the value attached to detecting it, and the practices engaged in to identify and manage it; and, second, examination of how knowledge and practice relating to delirium plays out at different points in the patient journey into and through the acute hospital. Although current evidence from systematic reviews^{135,136} report considerable underdetection of delirium in specific settings, such as the ED, and qualitative research^{121–127} reports attitudes to, and knowledge about, delirium among nurses in various acute wards, there is a paucity of studies that have addressed how knowledge is distributed between settings and professionals and how organisational context shapes the value attached to delirium screening and practices.

Multilevel knowledge of delirium

Knowledge and understanding of the significance of delirium, including its impact on patient distress and outcomes, were critical to the engagement of staff in delirium detection. Across all settings, delirium detection and the value attached to it were contingent on a particular type of knowledge held by professionals. This was more than 'awareness' of the syndrome, including its diagnostic features. Although important, such awareness was insufficient for staff to invest time and effort to systematically carry out the work of identification and action in relation to it. In addition, more in-depth knowledge of the impact that delirium had on patients, the distress that ensued as a consequence of the delirium and its immediate and longer-term effects on clinical and service outcomes was required. Acquisition of this multilevel knowledge meant that staff were able to recognise particular constellations of factors in the patient situation that made them think of delirium. This tacit knowledge, aligned with the clinical presentation, sensitised professionals to note and interpret observational cues to construct a clinical picture. This was built up collaboratively, employing a comprehensive investigatory approach combining observational cues, patient assessments, collateral history, medical notes and nursing observations, in a process referred to elsewhere as horizontal and not vertical use of expertise.³⁷

The significance of multilevel knowledge is reinforced from studies of the experiences of people with delirium and the staff caring for them. In a synthesis of qualitative studies of patients and nurses experiences of delirium,¹²⁰ the authors concluded that patients who had suffered delirium experienced safety and comfort from being understood, supported, believed and responded to with reassurance and care and were able to distinguish the quality of care delivery from how they were acted on, spoken to and touched. Nurses who lacked delirium knowledge were more likely to view patients' behaviour as strange, incomprehensible and a source of stress and irritation; and were unable to respond with support and empathy. Other studies focusing particularly on nursing staff^{120,121,123–128} have emphasised the importance of staff attitudes and responsiveness to the emotional needs of patients with delirium.

We found that the introduction of an easy-to-use screening tool had the potential to raise awareness of delirium when accompanied by learning and changes to, or modification of, existing systems for information gathering, for example including the tool as an integral component of the assessment process. However, insofar as delirium knowledge remained patchy, the outcomes poorly understood and the benefits of detection not visible to practitioners, the practice of delirium detection was inconsistent given the competing demands of the setting, findings that are echoed in other studies.^{37,38,121,123–128,130} Thus, van den Steeg *et al.*¹³⁰ concluded that poor delirium knowledge and skills among nurses acted as barriers to delirium guidance adherence, including systematic use of a risk screening instrument. Problems of adherence in turn undermined belief in the benefits and goals of screening; the consequence was that nurses viewed screening as simply another form, the completion of which took time away from patient care delivery.

Health professionals (irrespective of role) with extensive clinical experience and/or exposure to patients at risk of delirium, and those with expertise in the care of people with a cognitive impairment, were more likely to understand the significance of delirium detection, be sensitive to observational indicators of 'something not being quite right,' and pursue lines of information gathering, investigation and interpretation to make sense of symptom patterns. These 'delirium knowledgeable' clinicians spanned professional disciplines. They were

commonly found among those with extensive clinical experience and/or with specific skill and expertise in the care of older people, and had additionally made the connection between the diagnostic features of delirium and its significance for patient outcomes. 'Delirium knowledgeable' nurses, for example, would initiate diagnostic investigations to ascertain the source of delirium without deferring to medical staff. That 'delirium knowledgeable' practitioners were found across settings provides partial explanation of the variability in identifying and responding to delirium. The paradox was that for 'delirium knowledgeable' practitioners in some settings, a screening tool as an aide memoire for practice was seen as unnecessary, although among 'delirium knowledgeable' nurses in the 4AT 'experienced' site, the tool reinforced their expertise vis-à-vis medical staff.

Service delivery context, purpose and value attached to delirium screening

The second feature of the study findings relates to how delirium knowledge and practice played out at different points in the patient journey into and through the acute hospital.

The value and priority attached to delirium detection varied with the clinical purpose and organisational context. In an ED environment, the purpose of the setting – obtaining a differential diagnosis to inform decision-making on an immediate course of action within a very short time frame – did not make formal delirium detection salient at this point in the patient journey. The question posed by 'delirium knowledgeable' practitioners was whether or not the systematic detection of delirium in this context, with its attendant problems (difficulty of ascertaining a baseline picture; time constraints; and pace and flow pressures), would make a difference to patient outcomes. On the one hand, 'delirium knowledgeable' clinicians described their approach to practice as being sensitive to observational signs of 'acute confusion, and then engaging in a comprehensive assessment process, drawing on multiple sources of information to guide investigations and action, suggesting that there was value in this work. For example, action on identifying and responding to infection as a cause of delirium meant that treatment could be speedily initiated to benefit patients. More problematic was introducing a detection tool for all patients meeting specific criteria (e.g. age). It was notable that although the 4AT had been introduced into the ED in Denbury some time prior to this study, the consensus (not just from ED interviewees but from staff in the MAU and wards) was that it was no longer completed. Knowledge and purpose, then, were key factors in variability of delirium recognition in the ED.

Although some contextual features of MAU settings were thought of as barriers to delirium detection, it was generally agreed that identifying delirium was consistent with their organisational and clinical purpose: to enable further observation, investigations and decision-making to develop a management plan for the patient. Here, too, knowledge and understanding of delirium were critical to engaging staff in the work of delirium detection, creating the conditions for staff to invest in it. Even so, organisational factors including systems for integrating detection in routine care delivery, attention on mechanisms for legitimising and valuing the role and contributions of different professionals, and clarity about how delirium detection informed treatment and care were also necessary to embed delirium detection in care routines. Critical also was that the outcomes of investigations and the determination of actions arising out of them were clearly communicated to those charged with the task of carrying them through at the next point in the patient journey.

The ward was the main site both for ongoing management of patients with delirium and for identifying new cases. In high delirium prevalence settings, such as care of older people wards, staff were more aware of and knowledgeable about delirium. Junior doctors, nursing and care staff reported greater awareness of delirium than colleagues in other medical and surgical wards; they also conveyed greater consistency in their understanding of delirium and ability to distinguish it from dementia, a commonly cited reason for underdetection.

Even so, practice in managing delirium was variable. From this study, as from other research,^{37,38} critical to the integration of delirium detection and management on acute wards, in addition to multilevel knowledge of delirium, were working practices that placed value on person-centred approaches with patients and acknowledged the expertise of different groups of staff, and systems in place for co-ordinating, evaluating, communicating and sharing progress on recovery.

Our findings do not offer grounds for complacency, as evidence from other studies attests.^{122,123,128} Person-centred approaches to practice in respect of people on acute wards living with dementia (and delirium) are the exception and not the norm, as reported in national audits carried out in NHS hospitals in England.^{142,143} A recent study¹⁴⁴ of practice in respect of people with a cognitive impairment on 10 care of older people and orthogeriatric wards in hospitals in three English regions found that knowledge and understanding of delirium was variable and that consistent collection of data on delirium occurred on only a minority of wards where staff both had knowledge of delirium and had valued ongoing assessment and review as a means to effect and evaluate better management of it. It is perhaps surprising, given the co-occurrence of delirium and dementia, that policy emphasis on dementia in the UK,^{140–142} and reflected performance targets relating to it, has not also accorded more visibility to delirium.

Across all sites and settings, professional roles and expertise shaped knowledge and action, although the purpose of the setting and interdisciplinary practice were also key. Generally, medical staff assumed primary responsibility for diagnostics and overall treatment planning, but nursing and care staff played a pivotal role in noting and communicating change in patients and in pursuing work practices implicated in non-pharmacological delirium prevention and management interventions. In this respect, interdisciplinary working practices contributed to the establishment of a culture of care in which delirium detection and management were integrated into the broader work of supporting patients with a cognitive impairment in acute care.

Limitations

This was an exploratory study and its limitations should be acknowledged. It was undertaken in three sites only; the number of interviews conducted in each site was small, as was the number of interviewees drawn from different organisational settings on the patient journey from ED through to the ward. At ward level, the majority of those interviewed worked in care of older people wards. Thus, the findings are likely to overstate the engagement of staff in delirium screening compared with staff in wards in which knowledge and salience of delirium was low. At the assessment unit level, the study revealed different forms of service organisation from generic units to surgical units to designated units for older people and those with frailty. In the context of this study, although the findings are suggestive of how different organisational cultures shape the salience attached to delirium detection and practice, the full import of these different organisational forms on strategies to improve practice requires further systematic exploration.

Conclusions and implications

From this study, it is helpful to consider improvement in the identification and care of people with delirium as involving several interacting layers. At the individual level, multilevel knowledge of delirium creates the conditions in which delirium detection and action in relation to it is understood as meaningful and invested in, with regard to the time and skill to engage with it. The second layer refers to the existence of systems, mechanisms and processes to collectively engage and co-ordinate the work of different groups of staff to implement and sustain delirium detection and management in routine practice. This includes appropriate tools for identifying delirium. The third layer relates to the organisational purpose of the setting, namely to what extent can work practices and the division of labour through which these are carried out be modified or adapted to enhance delirium detection and achieve clinical and organisational outcomes. This poses a particular challenge for the ED. Considerable further research is needed to address these multiple challenges.

Chapter 5 Diagnostic accuracy of the 4AT

Introduction

It is essential that delirium, and any cognitive impairment (e.g. owing to dementia, depression or learning disability), is detected early to allow any underlying conditions to be identified and managed;²⁷ and to ensure that appropriate care pathways are followed, capacity is reviewed, and families and legal guardians are kept informed. At least two-thirds of cases are missed in the ED and general medical settings.^{7,31,32,145} There are several general reasons underlying this, including inadequate education and training, lack of a culture that facilitates delirium detection, and time constraints. However, another key reason may be the lack of a very rapid, simple, validated assessment tool that can be used without specific training.³⁴

The short CAM⁴⁰ is the most-studied short delirium assessment tool.¹⁴⁶ It comprises a four-feature diagnostic algorithm, based on the *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition – Revised (DSM-3-R), in which each feature is scored following a brief bedside interview and a cognitive assessment using a tool such as the Mini-Cog.¹⁴⁷ The CAM is included in multiple guidelines and policy documents, and has been translated into several languages.¹⁴⁸ It is the delirium assessment tool most used in research studies, and is the most cited in the literature, with almost 2000 citations since publication in 1990. It is the tool most commonly advocated in clinical pathways and guidelines globally, although the 4AT now has comparable representation in several countries, including the UK and Australia. Studies of the diagnostic accuracy of the CAM generally show good specificity, although the sensitivity is more variable, ranging from 0.13 to 1.0.^{4,49,149–154} The variability in sensitivity across studies appears partly to relate to the professional background of the rater, and the degree of training of the rater; in the studies showing lower sensitivities, raters have tended not to be medically qualified and/or have not had specific training in the use of the CAM. Additionally, validation studies of the CAM have been performed in various settings, with different populations and different exclusion criteria.

As a tool for mainstream routine use to detect delirium, the CAM may not be ideally suited to all settings. Because of the need for a bedside interview and brief cognitive test before scoring, it generally takes 5–10 minutes to complete. In addition, many acutely unwell patients have a reduced level of arousal^{42,47,155} and are unable to undergo cognitive testing or even a simple interview. The CAM has an algorithmic structure such that feature 1 (acute onset and fluctuating course) and feature 2 (inattention) must both be positive before proceeding to score feature 3 (disorganised thinking) and feature 4 (altered level of arousal). A patient with a severely reduced level of arousal (e.g. they are responding to verbal stimulation with only slight movement) cannot readily be assessed for inattention according to the guidance in the CAM manual,¹⁵⁶ and thus such patients cannot be scored positive for delirium. Additionally, early in the hospital stay, positive evidence for acute onset or fluctuating course is often not available.⁵⁹ Again, in these circumstances, patients cannot be scored positive. Another issue is the need for subjective, binary assessments of aspects of mental status, such as inattention. These assessments are less reliable and more challenging for staff, particularly non-specialists, than more objective tests. For example, one study found a kappa of 0.66 for the CAM inattention item when measured by trained assessors.¹⁵⁷ Another challenge is that, according to the CAM training manual, those using the tool should undergo specific training; this training takes several hours and it may not be feasible to provide this to all users in a given health-care setting.

The 4AT⁵⁶ was developed in 2010 as a screening tool for delirium, and to identify pre-existing cognitive impairment. The development was prompted by the need for improved detection in the first author's institution, a need identified as part of a larger planned programme of improved care for delirium. Although at that time many delirium assessment tools were already published, after extensive searching and trials of some available tools, the authors took the view that none of these had the required characteristics. These included being short; being easy to learn, administer and score; being usable in patients unable to

produce verbal responses; being as inclusive as possible; and being usable by health professionals from a range of disciplines. The tool also had to include general cognitive screening items to avoid the need for two separate instruments. The test was initially drafted and then assessed for practicality and usability in clinical use, and redrafted and reassessed, with several iterations of this process. An audit (unpublished) of the completed version in 30 inpatients comparing clinical use of the 4AT with the independent reference standard DSM-4¹ assessment found 100% sensitivity (95% CI 69% to 100%) and 90% specificity (95% CI 68% to 99%). The tool was then published on www.the4at.com to facilitate local institutional use. Following this, the 4AT began to be adopted in clinical units in several centres in the UK and internationally. A validation study in geriatrics inpatients ($n = 234$) found a sensitivity of 0.9 and a specificity 0.84 for delirium.⁵⁶ Another early study in stroke patients ($n = 110$) found a sensitivity for detecting delirium of 1.0 and specificity of 0.82.⁵⁵ However, given the limitations of previous tools, and the promising early diagnostic accuracy evaluations and increasing clinical adoption of the 4AT, it was considered necessary to proceed to a larger validation study. The present study was funded in 2013, and commenced recruitment in 2015. Since the funding award for the present study, other studies evaluating the 4AT have been published.^{57–59,61,158} Those reporting diagnostic accuracy data^{57,59,60} (total $n = 769$) found sensitivities ranging from 0.83 to 0.93 and specificities ranging from 0.7 to 0.91.

The main aim of the present study was to provide a prospective evaluation of the diagnostic accuracy of the 4AT. Additionally, because the CAM is widely used in research and clinical practice, the study also involved a corresponding evaluation of the CAM to provide a comparison under the same study conditions. Secondary objectives were to assess the performance of the cognitive testing items embedded in the 4AT, the relationship between the 4AT and clinical outcomes, the performance of individual items of the 4AT in detecting delirium (to determine the contribution of these items to the overall diagnostic performance), and the relationship between the 4AT and delirium severity as assessed with a standard delirium severity scale.

Study objectives and hypotheses

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-4 diagnosis.

The secondary objectives are:

- (a) to compare the performance of the 4AT with that of the 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 the 4AT scores predict 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.

Methods

We followed the STARD (Standards for Reporting of Diagnostic Accuracy Studies) 2015 guidelines¹⁵⁹ for reporting diagnostic accuracy studies. The study was registered as International Standard Randomised Controlled Trial Number (ISRCTN) 53388093 and UK Clinical Research Network ID 19502, and the protocol was accepted for publication in advance of any database locking and any statistical analyses.¹⁶⁰

Ethics

This study was granted ethics approval prior to data collection 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 design: overview

Patients aged ≥ 70 years in EDs or acute general medical wards were prospectively recruited in three sites (Edinburgh, Bradford and Sheffield). The total target number of patients was 900 (all sites). Each patient underwent (1) a reference standard delirium assessment lasting up to 20 minutes, and (2) either the 4AT or the CAM (lasting up to 10 minutes). Randomisation was used to allocate participants to the 4AT or the CAM and to determine the ordering of the reference standard and 4AT or CAM assessments. A questionnaire on pre-admission cognitive function was administered to an appropriate informant (if one was available). At 12 weeks, data on clinical outcomes, including length of stay, institutionalisation and mortality, and resource utilisation were collected from medical and social care records, or from patients if the records did not provide sufficient information.

Inclusion and exclusion criteria

Inclusion criteria

- Was aged ≥ 70 years.
- Had been acutely admitted to the ED (within 12 hours of attending) or acute general medical and geriatrics units (within 96 hours of admission to the ward). ED patients were recruited only from those who were brought in by ambulance as an emergency or through their GP, to ensure that recruited patients were representative of those likely to be admitted to hospital as well as having a similar profile of underlying cognitive impairment or other comorbidities.

Exclusion criteria

- Had acute life-threatening illness requiring time-critical intervention (e.g. ST-elevation myocardial infarction, septic shock or severe pulmonary oedema).
- Was in a coma ['unresponsive' on the AVPU (Alert, Responsive to Verbal Stimulation, Responsive to Pain, Unresponsive) scale].¹⁶¹
- Was unable to communicate in English, had severe dysphasia.
- Had previously enrolled in the 4AT study.

Sample size

It was planned that a total of 900 patients would be randomised: 450 to assessment by the 4AT and 450 to assessment by the CAM. Of the 450 patients within each group, 15% ($n = 67$) would be expected to have delirium.³ The specificity of each triage tool would be estimated based on the 85% ($n = 383$) without delirium; sensitivity would be estimated from the 15% ($n = 67$) patients with delirium. This sample size would allow precise estimation of specificity and moderately precise estimation of sensitivity. Based on analysis using the normal approximation to the binomial distribution, the two-sided 95% confidence interval (CI) width for specificity would range from ± 0.030 to ± 0.050 , and for true specificity would range from 0.90 to 0.50. The corresponding CI widths for sensitivity would range from ± 0.072 to ± 0.120 .

For the secondary objective comparing the 4AT and the CAM, based on analysis by a continuity-corrected chi-squared test, we have 83% power to detect a difference in specificity of 0.10, assuming a null hypothesis of specificity = 0.70 for both tests and a 5% significance level. The difference detectable for sensitivity (null hypothesis sensitivity = 0.70 on both tests) would be 0.224 (80% power).

Recruitment

Patients were recruited between 08:00 and 22:00. Eligibility screening was carried out by a member of the clinical team, yielding sets of names of potentially eligible patients. Patients were screened in alphabetical order from study opening in October 2015 until July 2016.

This recruitment strategy was changed from July 2016 because analyses suggested that patients at a lower risk of delirium (i.e. those not requiring capacity assessment) were more likely to be recruited than those with impaired capacity. The new recruitment strategy lasted 5 months. The strategy involved using a pragmatic approach to facilitate some oversampling of patients at higher risk of delirium. From the lists of patients identified as per the original strategy, patients considered at higher risk of delirium on clinical grounds (e.g. older age, likelihood of admission, a greater number of comorbidities) were approached first, rather than this being done in alphabetical order. This change was approved by the funder and the Trial Steering Committee.

Potentially eligible participants ($n = 4928$) were identified as those without critical illness or not in coma, in the ED within 12 hours of attendance, or on acute medical wards within 96 hours of attendance, in three sites (Edinburgh, Bradford and Sheffield: see *Appendix 10* for numbers screened, eligible and consented in each site). Study recruitment commenced on 19 October 2015 and was completed on 30 December 2016, with final follow-up data collection and database locking on 29 June 2017. In total, 843 patients were randomised.

Consent

Participants were eligible for this study whether or not they had capacity to consent. Informed consent was sought by trained researchers using a combined informal capacity assessment/consent process. Both verbal and written information was provided about the study, using a style and format suitable for the participant group (i.e. for varying levels of capacity). The researcher then asked the potential participant to recount the study information to check their understanding. This, together with the treating team views, was used to assess the person's capacity to consent. When the potential participant lacked capacity to consent, recruitment proceeded under the provisions of the Mental Capacity Act 2005¹⁶² in England or the Adults with Incapacity (Scotland) Act 2000,¹⁶³ using an appropriate personal or nominated consultee, guardian, welfare attorney or nearest relative (see the published protocol¹⁶⁰ for additional details, including different approaches in Scotland and England).

Test methods

Reference standard assessment for delirium

The reference standard assessment for delirium (*Box 2*) was based on a combination of the Delirium Rating Scale-Revised-98 (DRS-R98)¹⁶⁴ and several neuropsychological assessments used together to inform a binary ascertainment of delirium present or absent based on DSM-4¹ criteria. The DRS-R98 is a well-validated scale that assesses multiple dimensions of mental status change and quantifies delirium severity. The DRS-R98 has 16 domains, which are divided into two. The first part comprises 13 domains of mental status assessment that are rated according to their presence according to a three-point severity scale, or are absent. The second part comprises three items that are concerned with the diagnosis of delirium, namely 'temporal onset of symptoms', 'fluctuation of symptom severity' and 'physical disorder'. The first two of these three require information on the change from baseline. As per the instruction manual, the DRS-R98 was supplemented by short neuropsychological assessments of attention and other domains, including Digit Span,¹⁵⁷ the Observational Scale for Level of Arousal,⁴⁷ the Richmond Agitation Sedation Scale¹⁶⁵ and the DelApp computerised objective attentional assessment test.¹⁶⁶ The final DSM-4 ascertainment of delirium used a standardised process utilising the above assessment results, with final verification by consensus of the chief investigator and two independent adjudicators (ZT and SDS), who were blind to the 4AT or the CAM results. Cases judged as indeterminate on final verification were excluded from analyses. DSM-4¹ was used rather than DSM-5² because the study protocol and application were developed around the time of the launch of DSM-5; therefore, the DSM-4 was well-established,

BOX 2 Reference standard assessment for delirium

Binary ascertainment of delirium based on DSM-4¹ criteria.

Informed by the DRS-R98,¹⁶⁴ scored via interview; information from case notes and informant history; and cognitive tests, including Digit Span,¹⁵⁷ the Observational Scale for Level of Arousal,⁴⁷ the Richmond Agitation Sedation Scale,¹⁶⁵ the Vigilance A test,¹⁶⁷ the DelApp objective attentional assessment,¹⁶⁶ and simple object naming and orientation questions.

The DRS-R98 includes inspection of case notes and informant history.

Predefined positivity cut-off point.

Delirium (DSM-4 criteria) present or absent by consensus among the chief investigator and two adjudicators (SS and ZT) based on above assessments as recorded in the case record form.

Time taken: 20 minutes.

whereas DSM-5 criteria had not been tested in research around the time that the study protocols were finalised. However, DSM-4 and DSM-5 criteria are not markedly different.¹⁶⁸ The main difference is in the terminology used in criterion A, with DSM-4 using the phrase 'disturbance of consciousness', then further specifying attentional deficits, and DSM-5 using the phrase 'disturbance in attention' and introducing the term 'awareness'. How these features should be operationalised is not precisely covered in either manual. Therefore, how a test such as the 4AT performs in relation to a DSM-4 or DSM-5 reference standard will depend to some extent on how the reference standard operationalises the criteria included in DSM-4 or DSM-5.¹⁶⁹ If similar processes are used to inform criterion A for both DSM-4 and DSM-5, these criteria will score similar numbers of patients as delirium positive.¹⁶⁸

4AT

The 4AT (*Table 4*) comprises four items. Item 1 concerns an observational assessment of the level of patient alertness. Items 2 and 3 are brief cognitive tests: the AMT4,¹⁷⁰ 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 the Months Backwards test, in which the patient is asked to state the months of year in reverse order, starting with December. Only items 1–3 are administered at the bedside, and the combined typical duration is < 2 minutes. Item 4 concerns acute change in mental status, a core diagnostic feature of delirium; this information is obtained from the case notes, from the GP letter and/or from an informant, that is, from any of, or a combination of, these sources.

Confusion Assessment Method

The CAM⁴⁰ is a diagnostic algorithm drawn from the DSM-3-R delirium definition.¹⁷¹ Using the CAM, the tester rates the following four DSM-3-R features as present or absent: (1) acute change and fluctuating course, (2) inattention, (3) disorganised thinking and (4) altered level of consciousnesses. The CAM algorithm requires that features 1 and 2 are both positive; if they are, then features 3 and 4 are assessed, and if one or both of features 3 or 4 is positive, then the whole CAM is positive, indicating the presence of delirium. The tester scores the features using a combination of interviewing the patient, cognitive testing (the CAM requires that a cognitive test be performed before the features are scored), examining the case notes, and seeking informant history, if required. It should be noted that the questionnaires used to assess cognition are not specified in the CAM manual, although some suggested tests are listed. Feature 1 is assessed by the same process as item 4 of 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

TABLE 4 The 4AT

4AT item	Score
1. Alertness	
Normal	0
Mild sleepiness	1
Clearly abnormal	4
2. AMT4 (date of birth, current age, place patient is in now, current year)	
No mistakes	0
One mistake	1
Two or more mistakes/untestable	2
3. Attention (Months Backwards)	
≥ 7 correct	0
< 7 correct or refuses to start	1
Untestable	2
4. Acute change or fluctuating course	
No	0
Yes	4
Maximum score (sum of scores from each item)	12
Predefined positivity cut-off score (for secondary analysis, 0–12 as severity marker, and 1–3 as indicator of cognitive impairment)	0–3: no delirium 4–12: delirium
Time taken: 1–2 minutes	

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 disorganised 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 of the 4AT. In this study, for the pre-CAM cognitive assessment we used 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 clock-drawing (Box 3). All of these questions are used in routine clinical practice at the bedside.

Informant Questionnaire for Cognitive Decline in the Elderly

The IQCODE¹⁷² was used to assess if a patient had a pre-existing cognitive impairment. The IQCODE is a very widely-used validated questionnaire, including in delirium studies,^{173–176} that allows estimation of whether or not an individual has a pre-existing cognitive impairment. The recommended cut-off point of 3.44 was used.¹⁷⁷ It is administered to the nearest relative or carer and takes 5 minutes to complete. Consent was sought from the nearest relative or carer before the IQCODE was administered. Appropriate informants are not always available; thus, reasonable efforts were made to identify such informants within 4 weeks of patient recruitment.

Ordering of reference standard delirium assessment, 4AT and CAM

The reference standard assessment for delirium was administered to all patients by the researcher who conducted the capacity assessment and consenting process. A different researcher administered either the 4AT or the CAM. The rationale for this is that the capacity and consenting process provides information to the researcher that exceeds what the 4AT or the CAM provides. This is not a concern for the reference

BOX 3 The CAM: pre-scoring cognitive testing procedure**Pre-scoring cognitive testing procedure**

All participants underwent a brief interview (asking them to discuss their hospital stay, etc.), then the following cognitive testing procedure:

1. Days of the week backwards.
2. Counting from 20 down to 1.
3. Orientation (What is the day today?; Is it day or night just now?; What is the year?; What was your last meal?; How long have you been in hospital? What city are we in?; What is the name of the hospital?; What floor are we on?).
4. Memory registration: three words stated and then participant asked to recall these. Up to three trials until all three words recalled or three trials repeated.
5. Clock drawing: asked to draw a clock with hands at 10 past 11; if participant was unable to do the test, 30-second interval with on-specific conversation was used as a gap instead.
6. Delayed recall. Recall of the three words from question 4.

Predefined positivity cut-off point

As per the CAM manual, once the cognitive testing had been completed, and information about onset was gathered, raters used observation during the interview, and information from case notes and informants to score each feature of the CAM positive or negative, and then reached a delirium positive or negative score according to the CAM algorithm.

Time taken: \approx 10 minutes.

standard assessment, which is aimed at providing a thorough assessment to optimise diagnostic accuracy. The order of the two assessments [(4AT or CAM assessment) and reference standard assessment] was randomised in the ratio 1 : 1. Participants were also randomised in the ratio 1 : 1 to be assessed using the 4AT or the CAM. Randomisation, using a secure online system, occurred immediately after consenting. Access to the secure system for each randomisation was via a unique username and password for each researcher; the full list of randomised allocations was held within the secure online system and concealed from researchers. The randomisation sequence was created using computer-generated pseudo-random numbers, and was stratified by study site, with block allocation. The two assessments took place within a maximum of 2 hours of each other, with a target interval of 15 minutes. Researchers were blinded to each other's assessments. The result of the reference standard assessment was recorded in the case notes, and communicated to the clinical team after the assessments were complete. *Figure 2* shows the study assessment flow chart.

When possible, the IQCODE was administered following the above tests (within 4 weeks of the patient joining the study).

Outcomes data and health-care utilisation data at 12 weeks

Data on length of stay, institutionalisation and mortality were recorded up to 12 weeks. Patient resource use was derived from medical and social care records, when available, and, if necessary, via patient or carer self-report. In each case, records were retrieved or inspected in detail to extract the specified data. Standardised data collection forms were utilised, and data were then entered into the main database case record form (CRF). The optional self-report resource-use questionnaire, to be used if the required information was not available from the patient's records, included questions on patient health and social care utilisation, with a maximum recall period of 16 weeks. The self-report resource use questionnaire

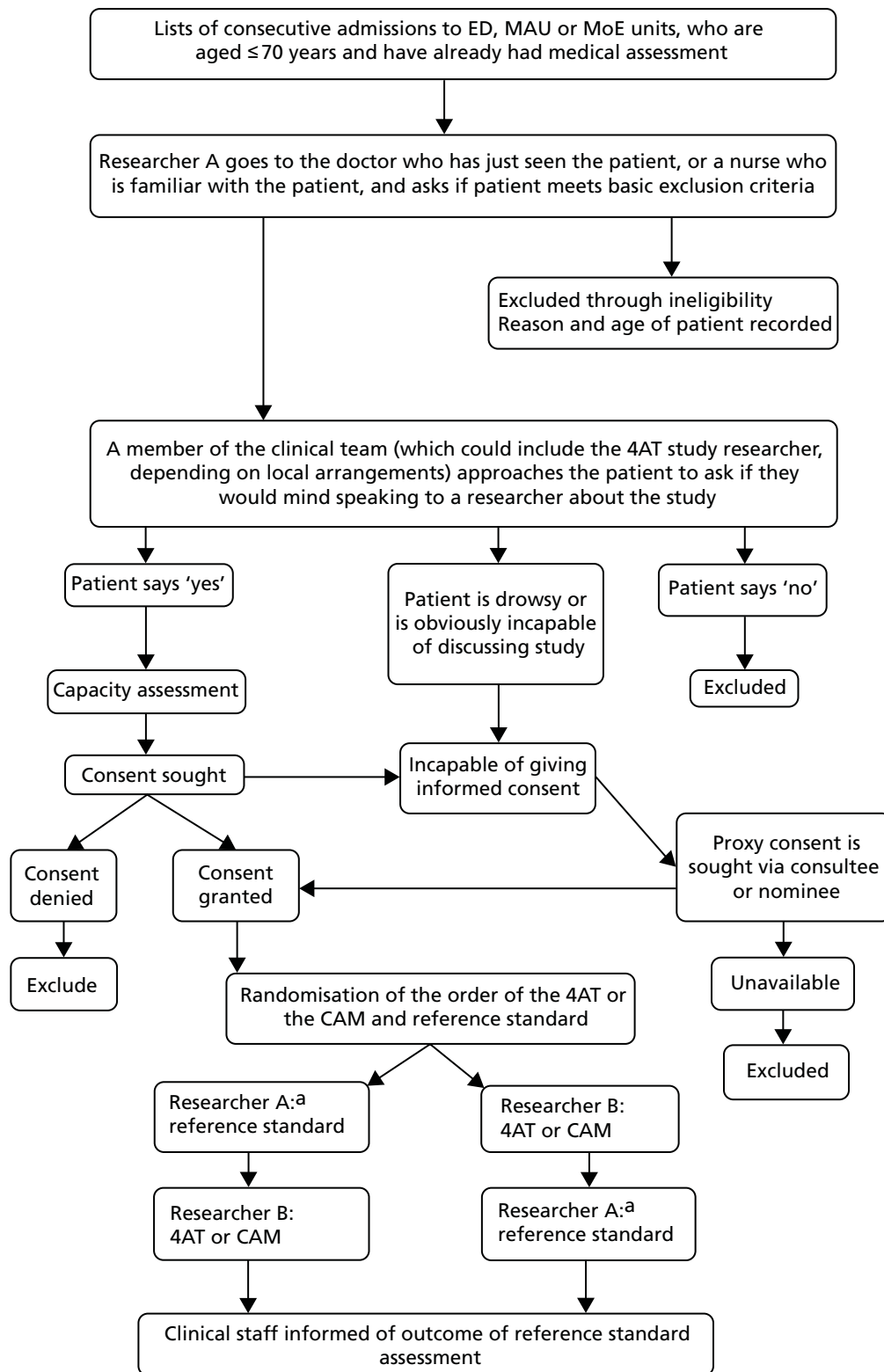


FIGURE 2 Diagnostic accuracy study: overview flow chart. a, Researcher A (i.e. the individual doing the consenting) will also always perform the reference standard assessment. Researcher A's status will not be tied to a particular individual throughout the study.

(contact authors) was developed for this study to be used by the patient or proxy respondent using guidance from the Database of Instruments for Resource Use Measurement.¹⁷⁸ If necessary, the questionnaire was administered at approximately 12 weeks by one of the researchers in the study team, either face to face, for patients who were still hospitalised, or by telephone.

Research staff and training

Researchers were nurses or trained graduate clinical research associates. All underwent a systematic training process comprising three initial modules with reading material and videos covering (1) the patient and carer experience of delirium, (2) basic knowledge of the features of delirium, clinical diagnosis and treatment approaches, and (3) specific coverage of delirium assessment, including cognitive testing, differentiating delirium from dementia, and the tools and tests used in the present study. These modules took 1 or 2 days each to complete. After these modules, researchers underwent face-to-face training with the chief investigator, and underwent role-play practice and supervised bedside practice with hospitalised patients for the reference standard assessment (guided by, and referring to, the DRS-R98 manual¹⁶⁴) and the CAM (the latter training using instructions from the CAM training manual¹⁵⁶). Researchers also underwent additional brief familiarisation sessions with the 4AT.

Analysis

The detailed statistical analysis plan was agreed prior to database lock,¹⁶⁰ blinded to the randomised allocations. All analyses were performed by statisticians (JS, CW and VA).

Methods: measures of diagnostic accuracy

Primary objective

We calculated positive predictive values (PPVs) and negative predictive values (NPVs), sensitivity and specificity (with exact binomial 95% CI) for 4AT versus delirium reference standard. We constructed a receiver operating characteristic (ROC) curve and reported area under the curve and 95% CIs to verify the appropriateness of the cut-off point of > 3 on the 4AT score (this cut-off point was used as it was specified in the original tool design).

Secondary objectives

(a) 4AT versus CAM

We estimated the difference in proportions (4AT – CAM) and its 95% CI for each of sensitivity, specificity, PPV and NPV. The statistical significance of differences was assessed using Fisher's exact test. The overall performance of 4AT and CAM was each summarised using Youden's Index (sensitivity minus false-positive rate) and the diagnostic odds ratio of sensitivity to specificity.

(b) 4AT as a screen for general cognitive impairment

The 4AT as index test with history of dementia and/or the IQCODE as reference standard.

(c) 4AT versus clinical outcomes

Descriptive statistics of clinical outcomes (length of stay, institutionalisation, and mortality) were calculated for the groups with and without 4AT scores above the cut-off point of 3. Logistic regression modelling was used to predict mortality and institutionalisation; Kaplan–Meier curves and Cox proportional hazards models were used to predict hospital length of stay. Logistic regression and Cox models also adjusted for age, gender and the presence of dementia.

Individual items of the 4AT: individual items were investigated as a predictor of delirium diagnosis using the same methods as for the primary objective.

(d) Delirium severity

The Spearman's rank-order correlation and its 95% CI were calculated between the total 4AT and DRS-R98 scores.

Subgroup analyses

Predefined subgroup analyses of measures of diagnostic accuracy for the primary objective and secondary objective (a) were performed to determine the impact of (1) time from presentation to recruitment before or after 4 hours (ED) or 24 hours (medical admissions), and (2) time between index test and reference standard less or more than 30 minutes.

Sensitivity analyses

Predefined sensitivity analyses of the primary objective and secondary objective (a) were carried out when the reference standard was indeterminate. In turn, delirium was classified as present and then as absent. A post hoc sensitivity analysis was also carried out using the reference standard DRS-R98 classification recorded at the time of the original bedside by the researchers, rather than using the chief investigator consensus adjudicated assessment. A further post hoc sensitivity analysis assumed that any patients with a missing result for the index test (4AT or CAM) had delirium.

Missing data

Individuals with missing data for the reference diagnostic test, CAM or 4AT were removed from formal statistical analysis. If items of the CAM or 4AT could not be assessed, an overall delirium diagnosis was derived, when possible, based on recorded items. There was no other imputation of missing delirium diagnoses. Individuals with missing data on an outcome variables were omitted from the analysis of that outcome.

Results**Descriptive statistics**

A total of 4928 patients were eligible, of whom 843 (17.2%) were randomised across the three sites and two withdrew, leaving 841 with data for analysis (*Figure 3*).

The initial DRS-R98-informed reference standard identified 96 (11.7%) patients as delirium present, and 723 (88.3%) as delirium absent [22 were missing because they withdrew from the study ($n = 18$) or had been discharged home ($n = 3$), or at patient request ($n = 1$)]. Final delirium ascertainment by the consensus group reclassified seven patients from delirium present to delirium absent, and 17 from delirium absent to delirium present. Thirty-two patients were classified as indeterminate and, therefore, were excluded from analyses (and two additional records were excluded because of an incomplete assessment that was interrupted for clinical reasons, and because of an error in which researchers completed the wrong part of the assessment resulting in an overall invalid assessment). Therefore, analyses were carried out for 785 patients (delirium present, $n = 95$, 12.1%; delirium absent, $n = 690$, 87.9%).

Descriptive data are presented in *Table 5*. Participants were tested in Edinburgh ($n = 372$: 4AT, $n = 187$; CAM, $n = 185$; could not assess CAM score, $n = 3$), Sheffield ($n = 125$: 4AT, $n = 64$; CAM, $n = 61$; missing 4AT, $n = 2$) and Bradford ($n = 279$: 4AT, $n = 141$; CAM, $n = 138$; could not assess, CAM $n = 2$; missing 4AT, $n = 1$). Only a minority of IQCODE questionnaires were completed (22%), mainly due to lack of availability of informants at the time of the assessments.

Primary objective**Diagnostic test accuracy of 4AT**

Overall delirium prevalence, as determined by the reference standard, was 12.1% (95 out of 785). *Table 6* provides a cross-tabulation of the index test results. Reference standard delirium prevalence in the 392

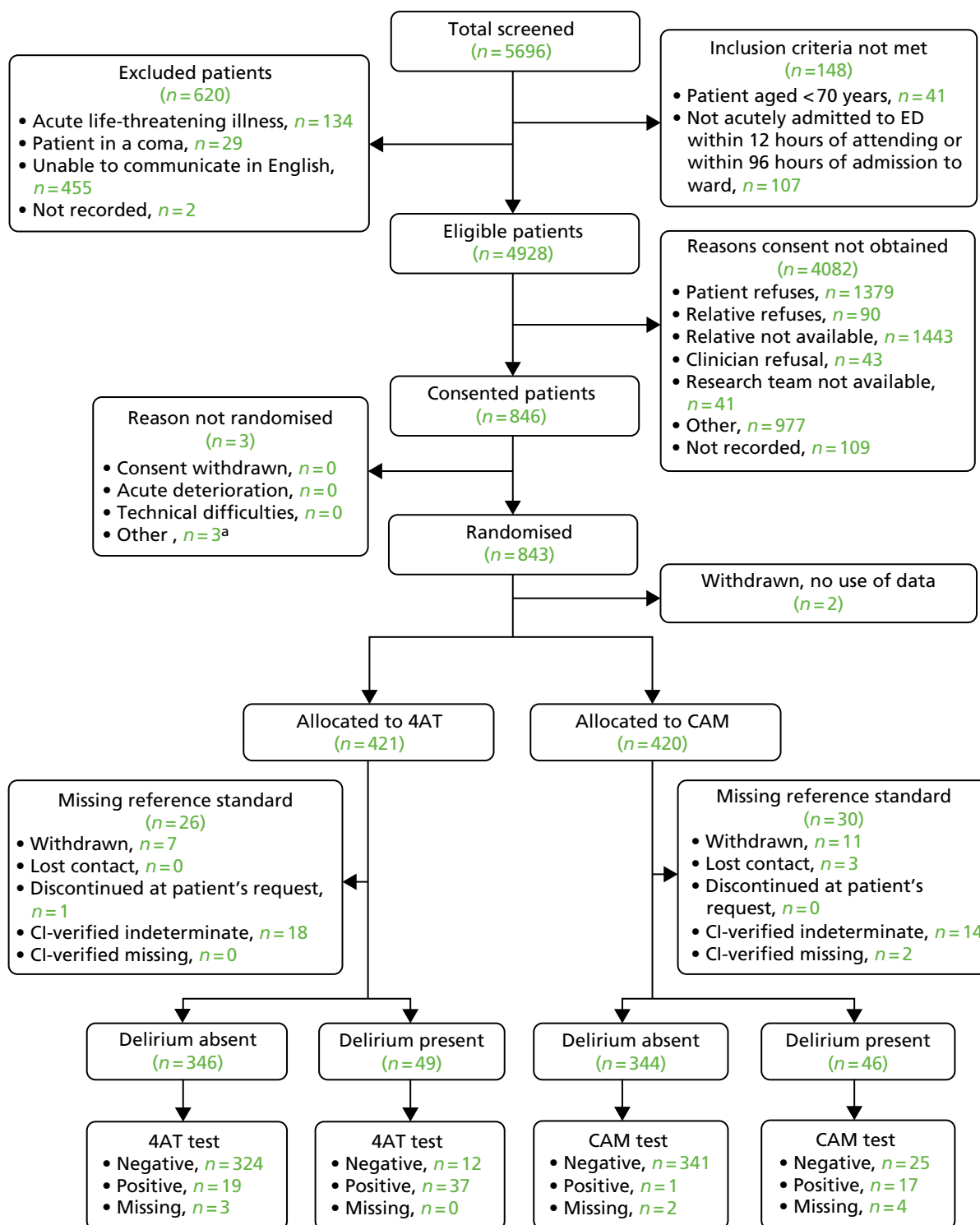


FIGURE 3 The STARD diagram of flow of participants through the study (total across all three). a, Two had moved ward, one had an issue with consultee. CI, chief investigator.

TABLE 5 Baseline characteristics stratified by delirium status

Characteristic	Total (N = 785)	Delirium status		p-value
		Present (N = 95)	Absent (N = 690)	
Age (years)				
Mean (SD)	81.4 (6.4)	83.5 (6.9)	81.1 (6.3)	0.0007
Median (Q1–Q3)	81.0 (77.0–86.0)	84.0 (78.0–89.0)	81.0 (77.0–86.0)	
Gender, n (%)				
Male	349 (44.5)	34 (35.8)	315 (45.7)	0.0697
Female	436 (55.5)	61 (64.2)	375 (54.3)	
Ethnicity, ^a n (%)				
White	773 (98.5)	91 (95.8)	682 (98.8)	0.0463
Mixed or multiple ethnic groups	2 (0.3)	1 (1.1)	1 (0.1)	
Asian, Asian Scottish or Asian British	5 (0.6)	1 (1.1)	4 (0.6)	
African	1 (0.1)	0 (0.0)	1 (0.1)	
Caribbean or black	4 (0.5)	2 (2.1)	2 (0.3)	
Dementia diagnosis, ^b n (%)				
Missing	1 (< 1)	0 (0)	1 (< 1)	< 0.0001
Yes	71 (9.1)	25 (26.3)	46 (6.7)	
No	713 (90.9)	70 (73.7)	643 (93.3)	
IQCODE score of ≥ 3.44 , ^b n (%)				
Missing	613 (78)	56 (7)	557 (71)	< 0.0001
Yes	70 (40.7)	29 (74.4)	41 (30.8)	
No	102 (59.3)	10 (25.6)	92 (69.2)	
Location of first assessment, n (%)				
ED	53 (6.8)	10 (10.5)	43 (6.2)	0.2624
Acute unit	665 (84.7)	76 (80.0)	589 (85.4)	
Hospital ward	67 (8.5)	9 (9.5)	58 (8.4)	

a p-value from Fisher's exact test of white vs. other ethnicities.

b Missing category not included in chi-squared test.

Numbers are n (%) or mean (SD). p-value from chi-squared test (categorical variables) or t-test (continuous).

patients who had a valid 4AT assessment was 12.5% ($n = 49$). The main diagnostic test accuracy tests are shown in *Table 7*. The area under the ROC curve across all cut-off points on the 4AT was 0.90 (95% CI 0.84 to 0.96) (see *Appendix 11*).

Secondary objectives

Comparison between 4AT and CAM

Among the 384 patients who had a valid CAM, the prevalence of delirium was 10.9% ($n = 42$), as assessed by the reference standard. Delirium prevalence using the 4AT only as a diagnostic test was 14.3% ($n = 56$) and using the CAM only as a diagnostic test was 4.8% ($n = 18$).

TABLE 6 Cross-tabulation of index test results (4AT and CAM) by results of reference standard

Test	Delirium status, <i>n</i> (%)		Total
	Present	Absent	
4AT			
4AT score of > 3	37 (75.5)	19 (5.5)	56 (14.3)
4AT score of ≤ 3	12 (24.5)	324 (96.5)	336 (85.7)
Missing 4AT	0	3	3
CAM			
CAM: delirium	17 (37.0)	1 (0.3)	18 (4.8)
CAM: no delirium	25 (54.3)	341 (99.4)	366 (94.1)
Could not assess CAM	4 (8.7)	1 (0.3)	5 (1.3)
Missing CAM	0	1	1

TABLE 7 Diagnostic test accuracy of the 4AT for diagnosis of delirium (defined by reference standard assessment)

Cognitive test scores	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Area under ROC curve ^a
4AT (score of > 3)	75.51% (61.13% to 86.66%)	94.46% (91.48% to 96.63%)	66.07% (52.19% to 78.19%)	96.43% (93.84% to 98.14%)	0.895

^a Refers to overall area under the curve across all cut-off points on the 4AT. Numbers are estimates (95% CI).

The CAM had a sensitivity of 40% (95% CI 27% to 57%) and a specificity of 100% (95% CI 98% to 100%) in the subset of participants in whom it was assessable. Youden's Index of the CAM was 0.40 and of the 4AT was 0.70. The odds ratio of sensitivity to specificity for the CAM was 232 (95% CI 30 to 1812); for the 4AT it was 53 (95% CI 24 to 117) (*Appendix 12, Table 19*).

4AT items as a predictor of general cognitive impairment

The AMT4 and attention items of the 4AT showed high specificity but low sensitivity, and high negative predictive value but low positive predictive value, for a diagnosis of dementia (*Tables 8 and 9*). All patients were included in this analysis, regardless of delirium status.

Diagnostic test accuracy of individual items of the 4AT for diagnosis of delirium

A score of 4 on item 1 (abnormal alertness, scored as either 0 if the patient has normal alertness or 4 if the patient has abnormal alertness) had a sensitivity of 31% (95% CI 18% to 45%) and a specificity of 99% (95% CI 98% to 100%).

Item 2 is the AMT4, in which patients are asked to state their date of birth, their age, the current year, and the place that they are currently in. Two or more errors (including untestable status) on this item, giving a score of 2, had a sensitivity of 41% (95% CI 27% to 56%) and a specificity of 96% (95% CI 94% to 98%) compared with one or no errors. One or more errors (including untestable status), giving a score of 1 or 2, had a sensitivity of 63% (95% CI 48% to 77%) and a specificity of 83% (95% CI 78% to 87%) compared with no errors.

TABLE 8 Components of the 4AT cognitive items stratified by dementia status, defined as a documented history of dementia and/or a score of ≥ 3.44 on the IQCODE

Cognitive test components of the 4AT	Dementia status, <i>n</i> (%)		Total
	Present	Absent	
AMT4			
Missing	0 (0)	3 (< 1)	3 (< 1)
No mistakes	15 (31.9)	287 (83.2)	302 (77.0)
One mistake	10 (21.3)	47 (13.6)	57 (14.5)
Two mistakes	22 (46.8)	11 (3.2)	33 (8.4)
Attention			
Missing	0 (0)	3 (< 1)	3 (< 1)
Achieves ≥ 7 months correctly	12 (25.5)	273 (79.1)	285 (72.7)
Starts but scores < 7 months/refuses to start	24 (51.1)	65 (18.8)	89 (22.7)
Untestable	11 (23.4)	7 (2.0)	18 (4.6)

See Table 4 for details of 4AT scoring.

TABLE 9 Diagnostic test accuracy of the 4AT cognitive components for diagnosis of dementia, defined as a documented history of dementia and/or a score of ≥ 3.44 on the IQCODE

Scores	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Youden's Index
AMT4 (1 or 2 vs. 0)	68.09% (52.88% to 80.91%)	83.19% (78.82% to 86.98%)	35.56% (25.74% to 46.35%)	95.03% (91.94% to 97.19%)	0.51
AMT4 (2 vs. 1 or 0)	46.81% (32.11% to 61.92%)	96.81% (94.37% to 98.40%)	66.67% (48.17% to 82.04%)	93.04% (89.89% to 95.44%)	0.44
Attention (1 or 2 vs. 0)	74.47% (59.65% to 86.06%)	79.13% (74.46% to 83.30%)	32.71% (23.95% to 42.45%)	95.79% (92.76% to 97.81%)	0.54
Attention (2 vs. 1 or 0)	23.40% (12.30% to 38.03%)	97.97% (95.86% to 99.18%)	61.11% (35.75% to 82.70%)	90.37% (86.92% to 93.17%)	0.21

Youden's Index is equal to sensitivity + specificity – 1: a value of zero indicates no value, and a value of 1 indicates a perfect test. Numbers are estimates (95% CI). See Table 4 for details of 4AT scoring.

A score of 1 or 2 on item 3 (Months Backwards test), indicating that the patient was able to start reciting the months of the year backwards but was not able to reach 7 months correctly in sequence, or that they were able to convert but refused to do the test or were 'untestable', gave a sensitivity of 71% (95% CI 57% to 83%) and a specificity of 79% (95% CI 74% to 83%). A score of 2, indicating that the patient was 'untestable', had identical performance to that of item 1. This is likely to be because a score of 2 on item 3 is given when the patient is 'untestable', and, in most cases, this score is given when the patient is unable to respond because of altered arousal.

Item 4 (acute change or fluctuating course) showed a sensitivity of 76% (95% CI 61% to 87%) and a specificity of 96% (95% CI 93% to 98%). These results are almost identical to the overall performance of the 4AT. None of other individual items of the 4AT showed good diagnostic test accuracy alone. These results indicate that, in this study, virtually all positive cases on the 4AT had a positive item 4 (acute change or fluctuating course). The full results of the diagnostic test accuracy for the individual items are show in Appendix 13.

Delirium severity

The 4AT score correlated moderately (Spearman's correlation coefficient 0.57, 95% CI 0.50 to 0.63) with the DRS-R98 total score as a measure of delirium severity (see *Appendix 14, Figure 9*).

Outcomes

Mortality was higher in patients with positive 4AT scores: 16.1% of those with a 4AT score of > 3 had died by the 12-week follow-up, compared with 9.2% of those with a 4AT score of < 3. The odds ratio was higher, at 2.00 (95% CI 0.85 to 4.70). The median stay was longer for patients with a 4AT score of > 3 (median 5 days, IQR 2.0–14.0 days) than for those with a 4AT score of ≤ 3 (median 2 days, IQR 1.0–6.0 days; $p = 0.0009$), with a hazard ratio of time to discharge of 0.64 (95% CI 0.46 to 0.88). Patients with a 4AT score of > 3 had a significantly greater length of stay and increased mortality than those with a 4AT score of ≤ 3 (*Table 10*; see *Appendix 15, Figure 10*), adjusted for age, gender and dementia status. Numbers newly admitted to institutions during the 12-week period of analysis were too small to allow analysis.

Subgroup analyses

Time from presentation to recruitment before or after 4 hours in emergency department or 24 hours in medical admissions

There was no statistically significant difference in the diagnostic test accuracy of the 4AT between those recruited early and those recruited later after initial presentation (Fisher's exact test p -values: sensitivity, $p = 0.19$; specificity, $p = 0.75$; PPV, $p = 0.47$; NPV, $p = 0.24$).

Time between index test and reference standard of less than or more than 30 minutes

The proportion of participants assessed as delirium positive with a gap between index test and reference standard of < 30 minutes was higher than that of those assessed with a gap of > 30 minutes, but there was no statistically significant difference in the performance of either test between early and late assessment (Fisher's exact test p -values: sensitivity, $p = 0.16$; specificity, $p = 0.24$; PPV, $p = 1.00$; NPV, $p = 0.56$).

Post hoc sensitivity analyses

Missing index test

If delirium was scored as present where the index test result was missing, this did not substantially alter the diagnostic test accuracy of the 4AT or the CAM. The sensitivity of the 4AT was 76% (95% CI 61% to 87%), and the specificity was 94% (95% CI 91% to 96%). The sensitivity of the CAM was 46% (95% CI 31% to 62%), and the specificity was 99% (95% CI 97% to 100%).

TABLE 10 The 4AT as a predictor of clinical outcomes

Clinical outcome	4AT score				Total	p -value	Effect (95% CI) ^a
	≤ 3		> 3				
	<i>N</i>	<i>n</i> (%) or median (Q1–3)	<i>N</i>	<i>n</i> (%) or median (Q1–3)			
Length of stay (days)	335	2.0 (1.0–6.0)	55	5.0 (2.0–14.0)	3.0 (1.0–7.0)	0.0009	0.64 (0.46 to 0.88)
Mortality	336	31 (9.2%)	56	9 (16.1%)	40 (10.2%)	0.0261	2.00 (0.85 to 4.70)
Institutionalisation	206	2 (1.0%)	34	1 (2.9%)	3 (1.3%)		

Q, quartile.

^a Adjusted for age, gender and dementia status.

p -value from assessing association between clinical outcomes and total 4AT score (Spearman's correlation/Mann–Whitney U-test). Cox proportional hazards model for length of stay (calculated as time from presentation to discharge, censored at 12-week follow-up or death) to estimate hazard ratio and 95% CI. Logistic regression model for mortality to estimate odds ratio and 95% CI. Hazard ratios and odds ratios calculated for 4AT score of > 3 vs. 4AT score of ≤ 3. An odds ratio of > 1 indicates an increase in the odds of death. A hazard ratio of < 1 is a negative outcome, indicating an increased time to discharge. No analyses were performed for institutionalisation because of the number of missing data.

Using the researcher reference standard as the Delirium Rating Scale-Revised-98 classification, not the chief investigator adjudicated classification

If the DRS-R98 classification, as performed by the researchers carrying out the reference delirium assessments, was used as the reference standard, the sensitivity of the 4AT increased to 83% (95% CI 70% to 93%), with little change in the specificity, which was 94% (95% CI 91% to 96%). The sensitivity (40%, 95% CI 25% to 56%) and specificity (99%, 95% CI 98% to 100%) of the CAM did not change materially.

Discussion

The main objective of this study was to assess the diagnostic accuracy of the 4AT against a reference standard assessment. The main findings were that the 4AT showed reasonable sensitivity (76%) and high specificity (95%). The area under the curve was high, at 0.90. The CAM showed lower sensitivity than the 4AT, at 40%. The specificity of the CAM was close to 100%. The cognitive test items of the 4AT did show some relationship with general cognitive impairment as ascertained in the reference standard assessment by a documented previous diagnosis of dementia and/or a positive IQCODE score. However, the pattern of the results was that these items had a high specificity but a moderate to good sensitivity, suggesting that an overall 4AT score of 1–3 (resulting from combinations of positive scores in the AMT4 and Attention items) should raise concerns about the possibility of general cognitive impairment, but does not exclude such impairment. These findings are broadly consistent with those reported in a recent study.⁶⁰ The bedside assessment items in the 4AT (items 1–3) individually showed varying relationships with reference standard diagnosis, with the majority showing relatively low sensitivity and relatively high specificity. Scores on the 4AT were moderately correlated with the DRS-R98 severity score. Although the 4AT was not designed to assess severity, this analysis was included to determine if useful clinical information about delirium severity could be derived from 4AT scores. These results indicate that higher scores, which are generated in the presence of altered arousal and/or worse cognitive impairment, may, to some extent, indicate worse severity. The final secondary objective was to examine the relationship between 4AT scores and outcomes. Consistent with previous studies linking the presence of delirium with poor outcomes, positive 4AT scores were significantly associated with increased length of stay and mortality. The number of patients recorded as being institutionalised during the 12-week study period was small and did not permit formal analysis.

At the time that the present study commenced, there were two published diagnostic accuracy studies of the 4AT.^{55,56} The 4AT was already in use in multiple centres, and initial clinical service evaluation reports provided informal feedback that it was considered useful in practice. Since the present study commenced, further external studies of the diagnostic accuracy of the 4AT have been published.^{57–59,61,158} The findings of the present study are broadly consistent with those of existing published studies, albeit with the present findings showing somewhat lower sensitivity and higher specificity than these published studies. The clinical use of the 4AT has continued to increase, with some centres reporting more than 10,000 uses of the tool (see *Chapter 3*). The website (www.the4AT.com) has had increasing traffic and the 4AT has been translated into multiple languages. The 4AT is also included in several national guidelines and position statements nationally and internationally.

Diagnostic accuracy studies of delirium tools vary in multiple ways, and these differences can potentially affect the findings. Differences include the population studied, the stage of the admission at which the assessment was carried out, the inclusion and exclusion criteria (especially whether or not patients with reduced arousal are included, with many studies excluding patients who are unable to communicate verbally¹⁷⁹), and the reference standard used. The last is likely to be an underestimated source of difference among studies. Neufeld *et al.*¹⁸⁰ found surprising variability in delirium reference standard assessments used in diagnostic accuracy studies of shorter delirium assessment tools. Some studies simply stated that the delirium reference standard was performed according to a standardised set of criteria, such as DSM-4, without providing any other information about how this was achieved; other studies provided much more

detail about how the reference standard assessment was performed. Notably, many reference standard assessments in published studies did not use cognitive testing as part of the assessment process.¹⁸⁰ In the present study, the reference standard used a standardised process that incorporated the best-validated detailed delirium assessment instrument, the DRS-R98, supplemented by several cognitive tests. The researcher performing the assessment gave an initial judgement on the presence or absence of delirium according to DSM-4 following both the consenting process and the reference standard assessment (because the researcher carrying out the reference standard assessment was always the one who had performed the consenting). However, the final delirium ascertainment in all cases was performed by the chief investigator in combination with two other members of the study team based on the recorded findings of the DRS-R98 items and the cognitive tests and any notes from the assessors. This was to help ensure consistency of approach. The sensitivity of the 4AT was higher (83%) against the researcher assessment of delirium present or absent than the final delirium ascertainment (76%), with little difference in the specificities. It is possible that the researcher's impression of delirium present or absent created a different sensitivity with respect to the 4AT because the objective assessments recorded in the reference standard did not fully capture other elements of the researcher's interaction with the patient, including the consenting process. Indeed, the researcher's assessment might reflect clinical practice more closely than making a judgement based on the remote inspection of cognitive test data.

However, to some extent all delirium reference standard assessments involve arbitrary judgements about thresholds for positivity on given features. Studies have shown that the prevalence of delirium in a given study varies according to how reference standard assessments are performed¹⁶⁸ and, notably, how the official diagnostic criteria themselves have changed over time¹⁸¹ and, when compared, show different prevalences of delirium in the same populations.^{168,182–184} Therefore, given the sources of variability in diagnostic accuracy studies, it is reasonable to conclude that the present findings are broadly consistent with those of other studies of the 4AT and, thus, that its use in clinical practice continues to be supported by the evidence base. The caveat with all such brief tools, however, is that they do not provide a diagnosis, and clinical judgement by a suitably qualified practitioner is always required. As judged against the reference standard used in the present study, the 4AT showed a sensitivity of 76%. Against this standard, some of the 4AT assessments resulted in false negatives. The consequences of false negatives with respect to delirium could include a failure to investigate and treat the acute causes of delirium. Another consequence is that the lack of a diagnosis means that inaccurate information is provided to patients and families. The implications of these findings are that staff involved in delirium screening using the 4AT should be informed that the 4AT is a rapid assessment tool that may not always detect reference standard delirium, and that clinical judgement is required alongside its use. The specificity of the 4AT was high in the present study (as judged against the reference standard). This means that the rate of false positives was low. A false positive delirium diagnosis can lead to unnecessary investigations and, sometimes, treatments (e.g. the presence of delirium may alter the threshold at which treatment for possible infection is initiated). Additionally, a false-positive diagnosis would lead to inaccurate information being given to patients and carers. Again, the implication of this is that staff need to use screening or assessment tools with the awareness that these do not provide a definitive diagnosis, and that judgement by a clinician with appropriate expertise is needed to provide this diagnosis.

The 4AT includes some general cognitive testing items, the AMT4 and the Months Backwards test. Individually, these short tests have some scientific literature supporting their use as brief cognitive tests, albeit with mixed sensitivities and specificities across studies.^{59,96,170,185–188} The present findings provide some support, albeit non-definitive, for the validity of these tests as indicators of general cognitive impairment, including dementia. However, as with the 4AT as a whole, in clinical practice further assessment is required before firm conclusions can be drawn. A further caveat is that in acute populations, even in the absence of delirium, there is often a degree of reversible cognitive dysfunction,¹⁸⁹ which limits the interpretability of cognitive tests in the acute setting in relation to chronic cognitive impairment.

In this study, the CAM showed low sensitivity and very high specificity. At 40%, the sensitivity was lower than that in most published studies, although some studies have also shown low sensitivities.^{146,149–151,154} There are many possible reasons for the low sensitivity found in the present study. All research staff were trained in

using the CAM, but none was medically qualified. Previous studies have sometimes found that professional background affects the sensitivity of the CAM.¹⁵⁴ Professional background might have a tendency to affect the extent of overall exposure to delirium and also the familiarity with the process of making a formal diagnosis in clinical practice, with a relative minority of non-medically qualified professionals having experience of this. The low sensitivity did not appear to be related to practice in a particular site, because the sensitivity was low in all three sites [although the overall number of cases of delirium diagnosed using the reference standard in one site (Sheffield) was small and thus it is difficult to draw a conclusion from these numbers alone].

We found that virtually all of the positive 4AT scores that were aligned with the positive reference standard diagnosis were for item 4 (acute onset or fluctuating course). This may be partly an artefact of the research process, because patients with a score of 4 on item 1 and/or a combined score of 4 from items 2 and 3 would be more likely to lack the capacity to provide consent. Therefore, these patients would have had consent given on their behalf by a legal proxy, and that proxy would have probably have provided information informing item 4. Patients with a positive 4AT score arising from items 1 and/or 2 and 3 would probably not have been recruited into this study without the involvement of a legal proxy. This situation differs from clinical practice, in that informant history is not present in a substantial minority of patients, especially in the very early stages of the assessment period.⁵⁹ The 4AT was designed to allow for a positive score based on bedside assessment alone, but, in the present study, the consenting process meant that this situation was unlikely to arise.

The present study found a moderate correlation between the total 4AT score and the DRS-R98. The 4AT scoring system was not designed to be linear; rather, it was designed to have a threshold that can be crossed by three different means: a positive score on the binary level of alertness item (item 1), or a combination of worse scores or untestability on item 2 (AMT4) plus untestability on item 3, or a positive score on the binary acute change or fluctuating course item 4. Nevertheless, it is reasonable to assume that worse scores on items 1–3 do reflect, to some extent, a more severe syndrome. Indeed, some scales measuring delirium severity, such as the Memorial Delirium Assessment Scale¹⁹⁰ and the Confusion Assessment Method – Severity (CAM-S) scale,¹⁹¹ assess the extent of the level of arousal change and the degree of cognitive impairment. The present findings do not suggest that the 4AT can replace more formal assessments of delirium severity, but clinicians may find it useful to use information beyond the binary interpretation of ‘delirium likely to be present or absent’ that the threshold score of 3 or 4 in the 4AT provides.

Delirium is known to be associated with worse outcomes, including length of stay, new institutionalisation and mortality.⁸ In the present study, 4AT positive scores were associated with increased length of stay, and mortality at 12 weeks (although the latter was after adjustment for confounders). These findings provide some validation for the 4AT as an indicator of delirium.

This study had a number of limitations that should be acknowledged. Recruitment proved challenging, with only 17% of those eligible consented. This probably meant that the recruited sample did not fully reflect the target population. The sample was mostly white, with very limited representation from other ethnic groups. We recruited under the target of 900, and numbers were further affected by the exclusion of some patients who had indeterminate reference standard results, although the performance of the 4AT and the CAM was still estimated with acceptable precision. The prevalence of delirium in patients aged ≥ 70 years at the early stages of hospital admission is likely to range from 10% to 20%.^{3,192,193} In the present study, according to the final reference standard ascertainment, the prevalence of delirium was 12%. This prevalence is likely to be slightly lower than that in the target population; however, the difference is small, and arguably does not substantially affect clinical generalisability. The study was performed in three different sites. Although this provides several advantages, it also means that there is the possibility of increased variability in how patients were recruited and how the index assessments and reference standards were performed. In one site (Sheffield), the rates of delirium were low, probably reflecting the challenges associated with the target population that were predominantly in the ED. It should also be acknowledged that, given the fluctuating nature of delirium, the gap between assessments of potentially 2 hours means that assessments could have different findings. We used the CAM as a comparator, but shorter alternatives, such as the 3D-CAM¹⁷⁹

(which takes a median of 3 minutes to complete), have recently been developed and, in some settings, may be used instead of the CAM. We also acknowledge that it is possible that researcher bias influenced how the different index assessments (4AT or CAM) were scored. The researchers were aware of the origin of the 4AT and this may have biased their responses. However, the CAM was administered according to standard recommendations, with multiple meetings reviewing any difficulties in test administration. The study was overseen by a Trial Steering Committee, which scrutinised the study protocol and the conduct of the study. The dementia ascertainment was carried out through a combination of documenting existing formal dementia diagnosis (which is imperfect because around one-third of dementia is not diagnosed) and the use of the IQCODE. Unfortunately, the number of IQCODE questionnaires completed was smaller than expected.

Conclusions

This diagnostic accuracy study had the primary purpose of determining the performance of the 4AT against a reference standard assessment. The results are broadly consistent with those of the already published studies of the performance of the 4AT, although the sensitivity was somewhat lower and the specificity was somewhat higher than in previous studies. No tool can provide perfect performance, especially in the case of delirium, a condition with multiple domains of abnormality that vary in severity and that fluctuate. Tools such as the 4AT can provide a useful means of providing a consistent approach to screening, especially in busy environments where speed and practicality are important. However, scores from such tools should not be regarded as diagnostic, and both individual practitioners and systems of care need to allow for this. No individual tool can be considered diagnostic. In all cases, whether the tool is designed for rapid assessment or is more detailed, it is necessary that the clinician judge the evidence against the diagnostic criteria for delirium and make a decision about the diagnosis. The tools allow for standardisation of collection of key information used to inform the clinician diagnosis.

In this study, the CAM had a substantially lower sensitivity and a somewhat higher specificity than the 4AT. The present findings are consistent with those of some previous studies suggesting that the performance of the CAM may be affected by the experience and background of the rater. Screening for cognitive impairment in general, which in the acute setting mostly comprises delirium and dementia, has traditionally been carried out using cognitive tests alone, such as the AMT10 or the MMSE. The 4AT was designed to function as a single instrument that had the primary function of detecting delirium, but also incorporated some general cognitive screening. The results of this study provide some support for the validity of the general cognitive screening items within the 4AT as specific rather than sensitive measures. The total score of the 4AT showed some relationship with more formal severity scoring, meaning that clinicians might consider those with higher scores as showing more severe syndrome. In conclusion, the results support the use of the 4AT as a delirium screening instrument that detects delirium with reasonable accuracy, although, as with every brief tool, scores need to be verified by clinical judgement by suitably trained practitioners, perhaps in combination with additional assessments.

Chapter 6 Health economics

Introduction

Delirium is associated with increased mortality, reduced quality of life and increased health-care costs.^{4,8,62,194–199}

As a prerequisite to estimating the cost-effectiveness of a delirium screening tool such as the 4AT, it is necessary to first characterise survival, cost and quality of life in patients with and without a diagnosis of delirium in a current UK population. The objectives of this economic evaluation within the 4AT study were to estimate the observed 12-week cumulative health-care-associated costs in these two groups of patients (delirium or no delirium) in the NHS in the three study sites. The costs were calculated by assigning unit costs to all NHS inpatient and outpatient activity, community care and additional services by all involved health-care professions.^{200,201} The analysis was framed with a view to informing any subsequent health economic modelling. Originally it was planned that costs would be calculated for groups identified as 4AT-delirium false positives, true positives, true negatives and false negatives. However, this was revised because in the main diagnostic accuracy study protocol the reference standard diagnostic process result was recorded in the case notes (thereby exerting an influence over practice) rather than the result of the 4AT or the CAM. For the purposes of this study, the reference standard was assumed to be 100% correct. The CAM was selected as a comparison in the main diagnostic test accuracy study because, at the time of the grant submission, it was the tool most commonly advocated for delirium screening in the UK NHS, and was probably the most used tool at that time (albeit used in a minority of cases because of low use of any tool for delirium).

When reading this chapter, it should be noted that the first section is a cost analysis using data collected within the study, and the second section is a cost-effectiveness analysis generated through a health economic model.

Methods

Resource use, costs and data collection

All significant resource consumption, giving special scrutiny to that which was thought to maybe differ in those diagnosed with delirium, was captured within the study, taking a NHS and Personal Social Services perspective. The main driver of the costs associated with patient care was expected to be the in-hospital time in the different wards. The time horizon of the base-case analysis for inclusion of costs was the 12-week period from randomisation.

The measurement of health-care resource consumption used only data collected as part of the trial. Patients' use of hospital services was obtained from the Health Economics Questionnaire as entered into a section of the CRF, which had been completed via telephone call with the patient/carer and/or through scrutiny of the clinical notes 12 weeks after randomisation. The initial length of stay (length of stay for the initial hospital stay within the 12 weeks captured) was generated by using the day of randomisation and the discharge day directly from the clinical record.

Research nurses conducted a telephone interview, if this was thought of as necessary because health-care records were incomplete, with each of the patients or their legal proxy 12 weeks after randomisation. Thus, health-care activity and resource use were captured by research nurses based on a combination of scrutiny of the clinical notes and the telephone interview.

Methods for analysing and handling missing data

Observed means (95% CIs) of costs are presented for all patients. An analysis of the costs occurring during the 12 weeks is presented. Preparation of the data, assigning costs and all further calculations were performed in Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA).

To align with the population used for the main statistical analysis, patients with missing reference standard diagnostic information were omitted. Any other activity presumed to be missing was assumed equivalent to zero cost.

Unit cost assignment

All resource use was valued in monetary terms using appropriate standard UK unit costs assigned at the time of analysis. NHS Scotland Information Services Division (ISD) unit costs from 2015 to 2016 were used for inpatient/day-case, outpatient, A&E and other elective or non-elective hospital-based services in the 12 weeks after randomisation.²⁰⁰ Unit costs for community health services were taken from *Unit Costs of Health and Social Care 2016*.²⁰¹

Cost calculation

Unit costs were assigned to inpatient stays specific to the admitting specialty. When details of the admitting specialty were missing, patients were manually sorted into the corresponding ward according to their medical diagnosis. In a first step, the inpatient, day-case, outpatient or rehabilitation costs were calculated so that all hospital-related costs could be summed. Subsequently, the costs occurring from the use of various community services (e.g. district nurse, GP, health visitors, psychological services, social workers, home carers, chiropractors, Meals on Wheels) were calculated and added as a separate cost considered in addition to the secondary care costs in order to generate the total cost per patient. Costs were calculated separately for:

- Scotland – national tariff costs (inpatient) were based on the ward days and whether the case was elective or emergency. As far as possible, all unit costs were taken from ISD²⁰⁰ and Personal Social Services Research Unit (PSSRU).²⁰¹
- England – in the absence of costs per hospital day for the different wards, the average day cost for inpatient stay, as well as day-cases, was used. For the outpatient costs, the *NHS Reference Costs 2015–2016* were used.²⁰²

Sensitivity analysis

In a second cost analysis, undertaken as a sensitivity analysis, the costs based solely on the initial length of stay were calculated based on the *Unit Costs of Health and Social Care 2016*²⁰¹ for the initial stay, assuming a constant daily unit cost. Because the ward of the initial stay was unknown, the English and Scottish costs of the initial stay were calculated using the *Unit Costs of Health and Social Care 2016* multiplied by the length of stay.

All health-care activity was costed irrespective of whether or not it was delirium related.

Scottish hospital costs

The Scottish costs taken from the CRFs, including every hospital stay and additional activity that was incurred during the 12 weeks after randomisation, were split into inpatient, day-case and outpatient costs (*Table 11*). The Scottish inpatient costs are *costs per day* in the different wards taken from the ISD *Scottish Health Service Costs* from 2015 to 2016.²⁰⁰ In the details/assumptions column of *Table 11* are the assumptions made for assigning the patients to different wards. This was necessary because the data set (from the CRFs) did not provide information on the exact nature of the wards, and because many of the wards are known to be general. In the case where two different wards were grouped into one (because there were unclear descriptions of where the patients had been), additional assumptions are described. For example, patients are often described as 'renal', which has been split into urology and nephrology according to NHS costing standard definitions. In these cases, the average cost of both ward types was calculated for inpatient per

TABLE 11 Scottish hospital cost estimates

Hospital costs	Inpatient cost per day (£)	Day case (£)	Outpatient (£)	Details/assumptions
A&E	1594		126	Summary of ED, A&E, emergency department, etc.
Cardiology	718	1471	140	Summary of cardiac problems, pain in chest, etc.
Dermatology	631	907	144	Summary of dermatology, derma, skin, skin disease, etc.
Ear, nose and throat	1343	1171	110	Summary of all diagnoses in a related category
Gastroenterology	538	642	218	Summary of gastro, gastroenterology, pain in stomach, belly, etc.
General medicine	442	744	225	In addition to general medicine diagnoses, patients categorised as 'medical unit' are included
General practice	361	320	62	Additionally including diabetes
General surgery	773	866	143	Summary of all unspecified surgeries and general surgeries
Geriatric assessment	333	907	176	Summary of geriatrics, medicine of the elderly, elderly, geriatric medicine, etc.
Geriatric long stay	267			Geriatric long stays (not including shorter ones): weekly cost £1868 (> £267 per day)
Gynaecology	1319	933	156	Summary of all gynaecology-related diagnoses
Haematology	955	742	324	Summary of all diagnosis related to the blood, referred to as blood issue, etc.
ICU	2019			Summary of intensive, care, intense care, ICU, etc.
Neurology	1117	1031	220	Summary of all neurological issues combined with stroke
Oncology	944	944	439	Summary of clinical and medical oncology (860/1028) (241/636), onco, oncology, etc.
Ophthalmology	1539	1291	143	Mostly stated as such
Oral surgery	1677	1057	117	Including some dental issues
Orthopaedics	902	1429	114	Summary of podiatry, ortho, vertebral fractures, obstetrics, obstetric specialist and orthopaedics
Plastic surgery	1533	1049	157	Including activity most closely matching this classification
Rehabilitation medicine	500	907	260	Including activity most closely matching this classification
Renal	716	610	167	Summary of urology and nephrology (778/654)/(635/585)/(145/189) and renal
Respiratory medicine	500	751	191	Summary of breathing problems, COPD, inability to breathe properly, etc.
Rheumatology	712	724	211	Mostly stated as such
Thoracic	1378		82	Additionally including thoracic surgeries
Vascular	582	1130	147	Summary of vascular medicine and surgery

COPD, chronic obstructive pulmonary disease.
Source of costs is *Scottish Health Service Costs 2015–16*.²⁰⁰

day, day cases and outpatient activities. In the frequent case of a missing allocation to a specific ward, the patients were allocated according to their medical diagnosis. This is an approximation that reflects not necessarily the geographical ward and associated costs, but the nature of the care that the patient is receiving according to their diagnosis. This care is often delivered in a general medical ward or a geriatric ward. The CRFs were completed by researchers on the study based on the follow-up telephone call and through the extraction of data from the electronic medical case notes. Given this combination, some details of derivations from the terms used are documented in the 'details/assumptions' column.

The costs of the initial hospital stay (unknown ward) were calculated separately using the same calculation for Scotland and England. A fixed cost per day was set²⁰¹ and multiplied by the length of stay.

English hospital costs

The English hospital costs (Table 12) were handled in a similar manner to the Scottish costs, split as they were into inpatient, day-case and outpatient costs. However, one of the big differences is that the English inpatient costs were defined as costs per stay rather than per day, leading to very different values being displayed in the tables. The calculations incorporated from the initial stay, and including all stays from the CRFs, are based on the *Unit Costs of Health and Social Care 2016*.²⁰¹

Costs of tests in hospital

The cost of the different tests and imaging, which were included in the CRFs, are calculated separately using Scottish and English costs (see Appendix 16). The Scottish costs used the *ISD Scottish Health Service Costs 2015–16*,²⁰⁰ and the English costs used the *NHS Reference Costs 2015–16*.²⁰²

Community health service costs

To calculate the full 12-week costs, the community NHS service costs were included. All community service costs were priced (see Appendix 17) according to the *Unit Costs of Health and Social Care 2016*.²⁰¹ When assumptions had to be made to estimate the cost per hour, these are stated in the column as 'additional'. When feasible, the hourly costs for specific professions were assumed to be the same regardless of whether the patient was seen in a clinic or at home. If there were different values stated in the *Unit Costs of Health and Social Care 2016*, these values were taken accordingly. As the CRFs were often filled in precisely (stating 5, 10, 15 minutes), for the hourly calculation the price of an 'hour of direct contact/per hour of face-to-face contact' was chosen rather than an 'hour worked'.

Costs of additional activities

To decide whether or not to cost all additional activities, the types and frequencies of additionally mentioned activities, such as 'daily full package of care', 'housework help' and 'impact team', were captured (see Appendix 18). The costs for these services were decided not to be formally calculated as such costs were very likely to be so small that they were insignificant. In addition, for some of those listed services, the free-text description recorded in the CRFs was insufficiently precise to be able to make proper assumptions about the service being referred to.

Results

Analysis was undertaken considering two different costs: (1) including all costs identified during the 12-week period as listed in the CRFs and (2) based on the costs of the index inpatient stay alone (sensitivity analysis).

Cost analysis using original researcher delirium ascertainment

Cumulative costs are presented for the average per-patient costs for all patients with the original categorisation into delirium/no delirium as by the researchers. Because all patients were assumed to be treated according to the reference standard (because this was the information that had to be included in the case notes based on the protocol), it was not possible to assess different costs for true positives, true negatives, false positives and false negatives. This statement is not necessarily true for all patients, as some of them switched from the

TABLE 12 English hospital cost estimates

Hospital costs	Inpatient cost per stay (£)	Day case (£)	Outpatient (£)	Source	Details/assumptions
A&E	2926	719	146.68	**	Summary of ED, A&E, emergency department, etc.
Cardiology	2926	719	155.235	**	Summary of cardiac problems, pain in chest, etc.
Dermatology	2926	719	101.63	**	Summary of dermatology, derma, skin, etc.
Ear, nose and throat	2926	719	96.87	**	Summary of all diagnoses falling in a related category
Gastroenterology	2926	719	136.57	**	Summary of gastro, gastroenterology, pain in stomach, belly, etc.
General medicine	2926	719	167.05	**	In addition to general medicine diagnoses, patients categorised as 'medical unit' are included
General practice	2926	719	158.53	**	Additionally including diabetes
General surgery	2926	719	130.60	**	Summary of all unspecified surgeries and general surgeries
Geriatric assessment	2926	719	220.72	**	Summary of geriatrics, medicine of the elderly, elderly, geriatric medicine, etc.
Geriatric long stay	2926			**	Geriatric long stays (not including shorter ones)
Gynaecology	2926	719	133.01	**	Summary of all gynaecology-related diagnosis
Haematology	2926	719	160.58	**	Summary of all diagnosis related to the blood, referred to as blood issues, etc.
ICU	2926	719	363.36	**	Summary of intensive, care, intense care, ICU, etc.
Neurology	2926	719	175.60	**	Summary of all neurological issues combined with stroke
Oncology	2926	719	151.12	**	Summary of clinical and medical oncology, onco, oncology, CA, etc.
Ophthalmology	2926	719	90.64	**	Mostly stated as such
Oral surgery	2926	719	111.47	**	Including some dental issues
Orthopaedics	2926	719	117.01	**	Summary of podiatry, ortho, vertebral fractures, obstetrics, obstetric specialist and orthopaedics
Plastic surgery	2926	719	139.61	**	Summary of all diagnoses falling in one of these a related categories
Rehabilitation medicine	2926	719	125.20	**	Summary of all diagnoses falling in one of these a related categories
Renal	2926	719	127.99	**	Summary of urology and nephrology and renal
Respiratory medicine	2926	719	154.77	**	Summary of breathing problems, COPD, inability to breath properly, etc.
Rheumatology	2926	719	142.74	**	Mostly stated as such
Thoracic	2926	719	212.94	**	Additionally including thoracic surgeries
Vascular	2926	719	153.10	**	Summary of vascular medicine and surgery

COPD, chronic obstructive pulmonary disease.

Source of costs is *Unit Costs of Health and Social Care 2016*²⁰¹ and reference costs.²⁰³

Inpatient costs: NHS non-elective inpatient stays.

Day-case costs: day-case HRG data, weighted average of all stays.

Outpatient costs: *NHS Reference Costs 2015 to 2016*²⁰² (national average).

'delirium' category to 'no delirium' and vice versa after the final central reference standard ascertainment process, by which point about 30 patients had been labelled as 'indeterminate'. Therefore, the final costs are shown in two tables: first, the average cost of the patients as categorised by the researchers who carried out the initial assessment, and, second, the cost of the final delirium ascertainment, which was determined by analysis of the CRFs by the chief investigator and two colleagues (see *Chapter 5*).

Starting with the cumulative 12-week costs, as categorised by the local researchers, it can be seen that the cost of the initial inpatient stay (sensitivity analysis) for patients with delirium is more than double that for those without (*Table 13*). Taking into account the costs during the 12 weeks following randomisation, it can be seen that, using ISD (Scottish) costs, the costs of people with delirium are also more than one-third higher than the costs of those without delirium.⁴ The same trend, although with lower magnitude, can be observed if the calculation is based on PSSRU costs (overall UK costs).²⁰¹

Cost analysis using final reference standard ascertainment

The costs, calculated after corrections, of the final reference standard ascertainment are slightly different, although they show the same trend as the costs based on the reference standard diagnosis as categorised by site (*Table 14*). The new indeterminate cost group (those patients for whom the diagnosis was considered uncertain based on the reference standard assessment) shows costs that are between those of the delirium and no delirium groups. This probably reflects the fact that patients in this group were a mix of those with delirium and those without delirium. The values from the initial stay (sensitivity analysis) again show a greater difference between patients with delirium and those without.

TABLE 13 Costs (£) at 12 weeks for delirium vs. no delirium from the original researcher assessments

Site	Delirium status, cost (95% CI)	
	Present	Absent
Scottish costs initial stay	2810 (2734 to 2886)	1277 (1267 to 1287)
Scottish costs at 12 weeks (CRF)	7559 (7362 to 7755)	4215 (4175 to 4254)
English costs initial stay	2810 (2734 to 2886)	1277 (1267 to 1287)
English costs at 12 weeks (CRF)	5216 (5107 to 5326)	4320 (4295 to 4346)

All costs are the mean cost per patient. The allocation to delirium present or delirium absent is based on the reference standard as categorised by the original researcher assessments (95% CI).

TABLE 14 Costs (£) at 12 weeks for delirium vs. no delirium by the final reference standard ascertainment

Site	Delirium status, cost (95% CI)		
	Present	Absent	Indeterminate
Scottish costs initial stay	2934 (2862 to 3006)	1239 (1229 to 1249)	1276 (1210 to 1341)
Scottish costs at 12 weeks (CRF)	6889 (6723 to 7054)	4230 (4188 to 4272)	5177 (4890 to 5465)
English costs initial stay	2934 (2862 to 3006)	1239 (1229 to 1249)	1276 (1210 to 1341)
English costs at 12 weeks (CRF)	4854 (4754 to 4953)	4264 (4239 to 4290)	4350 (4169 to 4531)

All costs are per patient. The allocation to delirium present, delirium absent or intermediate is based on the final reference standard ascertainment.

Health economic model: cost-effectiveness analysis

The aim of the 4AT cost-effectiveness analysis is to compare the cost-effectiveness of the 4AT with that of the CAM. To calculate cost-effectiveness, it was necessary to build on the within-trial analysis, which could include only the costs according to the reference standard diagnosis. To be able to compare the 4AT and the CAM in terms of cost-effectiveness, the estimated costs of true-positive, true-negative, false-positive and false-negative cases have to be known, as does the estimated quality of life associated with these categories.

It should be noted that the initial plans to undertake a meta-analysis of diagnostic test accuracy parameters were abandoned on completion of the literature survey in the field. It became apparent that quantitative estimation was not going to be feasible given the lack of previously published relevant studies. The approach outlined here relied instead on the formal elicitation of expert opinion.

To calculate the cost-effectiveness of the 4AT and the CAM, a health economic model was developed.

Model structure

In the basic structure, the model was a decision tree parameterised with the reference standard delirium and non-delirium probabilities, divided into branches dictated by the sensitivity and the specificity of the 4AT and the CAM (Figure 4).

In a next step, the model was parameterised with the costs and quality-adjusted life-years (QALYs) in order to calculate the incremental cost-effectiveness ratio (ICER). The utilisation and interpretation of the ICER outcomes is described hereafter. The ICER was calculated by dividing the difference in mean cost between the comparators by the difference in mean QALYs between those two groups:

$$\text{ICER} = \frac{C_i - C_c}{E_i - E_c} = \Delta C / \Delta E, \quad (1)$$

in which C_i and E_i are the expected cost and effectiveness of the new intervention or the new diagnostic test (4AT), and C_c and E_c are the expected cost and effectiveness of the comparator strategy, which is usually the treatment or diagnostic test in place, in this case the CAM. Therefore, ΔC shows the incremental costs and ΔE represents the incremental effectiveness of the 4AT and the CAM.

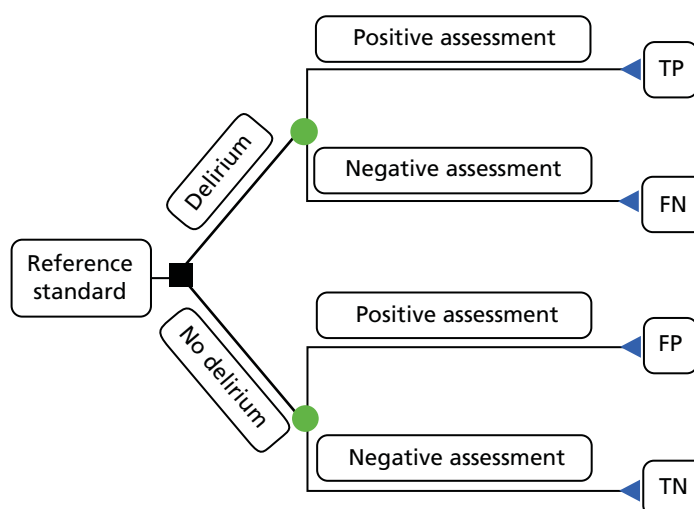


FIGURE 4 Decision tree. FN, false negative; FP, false positive; TN, true negative; TP, true positive.

If a new intervention is more effective and less costly, or if a new intervention is less effective and more costly, then the decision about whether or not to introduce the new intervention is straightforward, that is, the decision is clearly in favour of introducing it in the first scenario and against it in the second scenario. In these cases, the ICER calculation is playing a subordinate role. The ICER is important if we expect a trade-off between additional effect for additional cost and less effect for less cost. If a new intervention is more costly and more effective than the standard intervention, then the ICER is describing how much the intervention costs for an additional unit of benefit, in this study for a QALY. If it is cheap to gain an additional unit of benefit, it is still likely that the intervention will be introduced. In all cases in which the intervention is less costly and less effective, the value of the ICER represents how much money would be saved by renouncing a QALY. In this case it is the opposite way around, saying that if a small loss in benefits saves a considerable amount of resources it may be worth it. In a practical setting, this decision can be made based on the decision-maker's willingness to pay, which varies between countries and interventions. In the UK, NICE adopts a willingness-to-pay threshold value of £20,000 per QALY. That means if a new intervention has an ICER value of < £20,000 per additional QALY or is worth > £20,000 saved per QALY lost, then it is likely to be considered a cost-effective use of NHS resources. Values outside these ranges are conventionally not considered cost-effective.

Model parameters

The costs of true-positive and true-negative patients, as well as the values for the sensitivity and specificity of the different tests, were taken from the within-trial analysis. The data for the costs can be found in *Tables 11–14*. All other values were gathered through a new survey of experts conducted for this study. This step was necessary, as a systematic review of the literature in order to identify parameters did not provide sufficient information. These values included mortality estimates, quality-of-life estimates and costs associated with delirium, particularly with respect to false-negative and false-positive diagnoses. Nevertheless, the literature search provided information that was used to construct an expert interview and the subsequent survey.

Expert survey

A survey (see *Appendix 19*) was developed through consultation with the clinical advisors and chief investigator of the 4AT study. The survey elicited parameter estimates for mortality, quality of life and costs associated with delirium, particularly with respect to false-negative and false-positive diagnoses. The draft survey was then scrutinised for internal validity by members of the study team and rewritten. After this, an online survey using the Bristol Online Survey tool was set up and launched through two mechanisms: (1) a mailing list of recognised clinical experts and (2) via social media targeting health professionals. Of the 53 expert answers, 46 were included in the final analysis. The remaining seven were excluded as the responses were mostly incomplete.

Values informing the parameters

The values from the expert survey as well as from the within-trial analysis inform the parameters that subsequently flow into the model. The values and the sources of the different parameters can be found in *Appendix 20*.

Reporting

The base-case results are reported using the Scottish cost estimates and the 'best estimates' from the expert elicitation. Alternative cost estimates were used in a sensitivity analysis, as were high and low estimates of parameters obtained from expert elicitation.

Methods for the sensitivity analysis

Three sets of sensitivity analyses were conducted to test the robustness of the results to alternative assumptions and to understand the uncertainty present in the estimates. First, a scenario analysis (an alternative to the base-case analysis with different assumptions or sets of parameter values) was conducted using English costs estimates instead of Scottish estimates. In addition, for both Scottish and English estimates, other parameter estimates were taken at the 'highest estimates' or 'lowest estimates' (from the expert elicitation), respectively.

Second, a one-way sensitivity analysis was conducted in which the values of individual model parameters were varied over a range of $\pm 25\%$ of the base-case values. The differences in costs, and the differences in QALYs, between alternatives in each one-way sensitivity analysis are reported using tornado plots. Finally, a probabilistic sensitivity analysis was conducted to investigate uncertainty in the cost-effectiveness estimates due to sampling uncertainty of the parameter estimates. The probabilistic sensitivity analysis includes a gamma distribution for costs and beta distributions for probability and utility values and 1000 iterations. The results of the probabilistic sensitivity analysis are reported using a scatterplot on the cost-effectiveness plane and a cost-effectiveness acceptability curve.

Results from the health economic model

All of the results have the common problem seen with the economic evaluation of diagnostic tests, namely that the incremental costs and QALYs are very small and cluster close to zero (see *Appendix 21*). This makes the results of the cost-effectiveness analysis difficult to interpret, with wildly fluctuating ICERs. In general, the CAM produced slightly higher health-care costs than the 4AT, with only very small differences in QALYs.

The results of the base-case analysis show that the difference in costs between the 4AT and the CAM is £90.35, indicating that the 4AT resulted in lower total health-care costs over 12 weeks. The difference in QALYs was -0.00053 , indicating slightly worse health outcomes using the 4AT than using the CAM over 12 weeks. The ICER of £170,533 means that, for each QALY lost, £170,533 would be gained in resources. Using conventional thresholds, the 4AT would be considered cost-effective relative to the CAM.

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

Scenario analysis using the lowest and highest estimates from the expert elicitation exercise produced different results. At the lowest estimates, the ICER was £24,289, which is in a range that may be considered cost-effective, depending on other factors. At the highest estimates, the 4AT dominates the CAM, being more effective and less expensive.

One-way sensitivity analysis

The results of the one-way sensitivity analysis indicate that in relation to differences between the 4AT and the CAM in total cost the model was most sensitive to the costs of false-negative classified cases, the specificity of both the 4AT and the CAM and the costs of true-negative classified cases. The model was less sensitive to the sensitivity of the 4AT and the CAM, the true proportions of positive and negative cases and other cost estimates.

In relation to differences between the 4AT and the CAM in QALY outcomes, the model was most sensitive to 4AT and CAM specificity and the QALY estimates for each of the four diagnostic classifications. The model was less sensitive to the sensitivity of the 4AT and the CAM and the true proportions of positive and negative cases (see *Appendices 20* and *21*).

Probabilistic sensitivity analysis

The results of the probabilistic sensitivity analysis indicate that there is considerable uncertainty in the estimated cost-effectiveness of the 4AT compared with the CAM. In *Figure 11* (see *Appendix 22*), many points are spread across all four quadrants of the plane, indicating that there is the non-negligible probability that each alternative may be more or less costly and/or more or less effective. The cost-effectiveness acceptability curve (see *Appendix 23, Figure 12*) displays the probability that the 4AT would be considered a cost-effective alternative to the CAM at various willingness-to-pay thresholds for additional QALYs. The results indicate that the probability of cost-effectiveness is not greatly influenced by the threshold. This is because increasing the threshold makes scenarios in which the 4AT is more expensive but more effective more acceptable; however, it makes scenarios in which the 4AT is less effective but less

expensive less acceptable. Scenarios in which one test dominates (more effective and less expensive) the other are unaffected by the threshold value.

Discussion

Our economic analysis had the main purpose of showing the differences in estimated costs between patients with and patients without delirium. We also used a modelling approach to estimate the cost-effectiveness of the 4AT, within a methodological framework of diagnostic test evaluation.

Discussion of the cost analysis

Our main finding from the cost analysis was that delirium is associated with higher costs, taking into account the 12-week costs as well as those of the initial hospital stay. This is true if it is calculated using English costs as well as if it is calculated using Scottish costs. The latter showed a bigger effect; this could be because the average cost per stay was taken into account in the English costs, whereas the Scottish calculation was more detailed, being based on the single days that were estimated to be spent in the different specialties or receiving care for specific conditions.

Previous studies have also shown that a diagnosis of delirium confers a higher health-care cost than is observed in patients without delirium, for example as shown by Leslie *et al.*²¹ in 2008, who looked at the 1-year financial costs of care related to delirium. The costs of delirium arise from multiple possible sources, including increased risk of falls, other medical complications such as pressure sores, increased staff time, new institutionalisation and increased length of stay.⁴ There may also be substantial indirect costs, for example the effects on relatives and staff in addition to those on patients.¹³ Additionally, delirium is associated with a higher risk of long-term cognitive impairment^{10,11,66} and other psychiatric comorbidities.^{204,205} The prevention of delirium has been shown to be cost-effective.^{4,206} However, the treatment of delirium is under-researched, so there remains a lack of unequivocal trial evidence that it is effective. Likewise, it is unknown if the prompt detection of delirium, leading to earlier treatment, is cost saving.

Because all patients were treated according to the reference standard diagnosis, it was not possible to split the cost analysis further than separating costs for people with delirium and costs for those without. Therefore, it was also not possible to define the actual costs for 4AT and CAM as well as split those into true positives, true negatives, false positives and false negatives. Another limitation of this study was that the local researcher reference standard diagnosis did not fully agree with the final reference standard diagnosis, but the original reference standard was the one recorded in the clinical case notes. A further issue to consider is that the reference standard process probably had a tendency to produce an earlier and more accurate diagnosis than is normal in standard care. In fact, owing to the underdetection of delirium observed in many studies,⁷ the study process possibly led to increased rates of delirium diagnosis in the recruited population.

Discussion of the cost-effectiveness analysis

To calculate the cost-effectiveness of the 4AT and the CAM, and to address some of the limitations from the cost analysis, a health economics model was conducted. From the cost analysis, it could have been hypothesised that the cost for false positives is likely to be lower than the cost for false negatives and that the latter is considerably higher than the cost for true positives. This assumption was confirmed by taking into account expert opinion in our model, with results showing the cost for a false-positive case to be £4653.00, whereas the cost for a false negative case is £8955.70. The additional costs resulting from a missed diagnosis of delirium (cost false negative – cost true positive) were estimated at £2066.70 in the base-case analysis.

In terms of QALYs, there is very little difference between the 4AT and the CAM. The ICER of > £170,000 is stating that, for each QALY lost, £170,000 could be saved, which could be spent elsewhere. Equivalently,

it would not be cost-effective to switch from the 4AT to the CAM because the cost per additional QALY gained would be > £170,000, which is clearly above the NICE value of £20,000 per QALY gained.²⁰⁷

This study had several limitations from an economic point of view, starting with relatively limited numbers that meant that there was insufficient power to analyse the economic outcomes. Additionally, there were no data available within the trial about quality of life and, therefore, most parameters rely on expert opinion. These expert parameters can, on the other hand, be seen as a strength of the study, as they included estimates of a diverse sample of experts, including specialists from geriatrics, critical care, general medicine, general surgery, psychiatry, cardiology, A&E and palliative care.

This cost-effectiveness analysis demonstrates the potential of the 4AT as a screening test for delirium detection. It also highlights the need for a better understanding of the clinical pathway of delirium, as a considerable part of this analysis had to rely on expert opinion. Improved detection of delirium may have the potential to produce large cost savings within the health-care system, but substantial research will be required to realise this potential.

As is common in the economic evaluation of diagnostic testing strategies, the clustering of incremental costs and incremental QALYs either side of zero leads to an unstable ICER that fluctuates wildly with small changes in the underlying model assumptions. In addition, there is clear potential for screening tools, such as the 4AT, to be cost-effective; large incremental changes resulting in a clearly cost-effective strategy may rely on a more consistently demonstrated impact on the clinical pathway. Linkage of a 4AT result to a clearly defined treatment protocol that contains a clearly demonstrated effect on longer-term clinical and cost outcomes may be a necessary prerequisite to very high-value care.

Chapter 7 General discussion

This project had two phases. The first phase aimed to evaluate the acceptability and usefulness of the 4AT delirium assessment tool in the context of current UK practice, and the second phase was concerned with a diagnostic accuracy study and an evaluation of UK health-care costs associated with delirium care, including the estimated costs of making versus missing the diagnosis. With respect to the 4AT, the main results were that respondents to the survey mostly considered it to be usable and effective in its existing form, and the diagnostic accuracy study found that it had reasonable sensitivity and high specificity. These findings support the clinical use of the 4AT in settings where rapid assessment is required by staff who may not have specialist training.

The objective of survey A was to provide a recent and comprehensive understanding of the practice, knowledge and attitudes regarding delirium assessment among doctors, nurses, physiotherapists and occupational therapists in the UK. With > 2000 respondents, the survey is, to our knowledge, the largest in the world on delirium to date. There were several interesting findings. The overwhelming majority gave the clear opinion that it is important to make a formal diagnosis of delirium, including explicitly distinguishing it from dementia. Moreover, nearly all respondents stated that they thought that delirium treatment was effective in improving care. On this basis, it might be expected that tools and systems of care would be in place to support both detection and treatment. However, respondents indicated that although most of their units had some form of delirium guidelines, these were mostly not followed, with much delirium, therefore, underdetected and undertreated. Another important finding from the survey is that practitioners and colleagues in their organisations use a wide variety of terms to describe delirium, and mostly do not use the term 'delirium'. This finding has implications for education and training. Further support for the need for education training is that many respondents lacked confidence in their ability to detect delirium. Of note, doctors varied considerably in their opinions about whether or not acute-onset drowsiness, leading to patients not being able to produce speech (but not being comatose), was compatible with a diagnosis of delirium. This is significant in the light of expert opinion and DSM-5,² which hold that these patients would, in general, fulfil diagnostic criteria for delirium.²⁰⁸ Survey A also provided interesting results with respect to the perceived roles of different staff. In particular, the majority of nurses did not believe that they had the responsibility to generate a formal diagnosis of delirium, rather having the responsibility for screening for delirium. This finding has implications for the design of systems in which nurses might do much of the assessment for delirium, and thus in which a diagnosis made by a nurse could result in more prompt initial treatment. Future studies could build on the questions used in survey A in the assessment of knowledge of and attitudes towards delirium assessment.

The second of the two surveys was smaller, with 100 participants. This survey was valuable in that it asked practitioners with experience in using the 4AT to comment on its usability and other aspects. Generally speaking, opinions on the tool were favourable, with no consistent major issues identified. Some of the opinions expressed related to broader challenges in delirium assessment, such as the frequent lack of availability of an informant to provide a history regarding any change in mental status. However, the survey did suggest that the provision of training in the 4AT would be valuable. The implication of this is that members of the study team will produce additional training documentation and training video which will be added to the main 4AT website (www.the4at.com).

Alongside survey B, we gathered information on use of the 4AT in clinical settings and its presence in policy documents and web-based resources for clinicians. Judging by the available evidence, the use of the 4AT in clinical practice has grown rapidly since its release in 2011. It is now present in multiple national and international policy documents. We are also aware, through e-mail contacts and publications, that it is the standard tool used for delirium assessment in multiple organisations across the UK. Use of the 4AT is now a mandatory part of the best practice tariff for the care of acute fracture patients in the UK (this will influence the care of 50,000 patients per year). With respect to use in acute medical settings, two case studies from London and Sunderland provided examples; as well as demonstrating multiple uses of the 4AT in routine

clinical work, feedback from these sites has largely been positive with respect to the practicality and clinical value of the tool. Nevertheless, aligned with some of the findings from survey B, some training issues were evident with respect to delirium in general but also to the 4AT, suggesting that the availability of training specific to the 4AT (as well as broader training efforts) would be beneficial. Additionally, the 4AT has been used as part of broader quality improvement work on delirium detection and care across nine hospitals in Scotland led by Healthcare Improvement Scotland (Scottish Government). This has directly led to an estimated many thousands of uses of the 4AT, not just in these nine sites but in many others across Scotland, with residual ongoing clinical use.

The qualitative studies yielded valuable insights into the barriers to implementation of delirium assessment. The observational studies highlighted the complexity and pace of working in acute care settings, as well as the range of levels of experience of staff. These findings, although not new, must be borne in mind prominently in any initiative to improve delirium detection. Thus, proposed solutions involving substantial time, or certain special conditions such as having sufficient staff specifically trained in the use of particular tools such as the CAM, may not ultimately be effective. One of the key novel findings was the issue of prioritisation; that is, staff, having limited time, prioritise various elements of the multiple numbers of assessments (nutrition, skin assessments, etc.) that they are expected to do according to their perceived importance. For some staff, the actions following a positive delirium assessment are not clear and, moreover, the results of such assessments may not feature in interdisciplinary discussions. Delirium assessments may, therefore, sometimes be de-prioritised and, ultimately, not be conducted. Related to this issue is the apparent variability in knowledge of delirium. The findings from the present study, as well as from several previous surveys,³⁵ suggest that the fundamental lack of knowledge of delirium is a critical factor in preventing the effective implementation of delirium assessment. This knowledge is important not only in informing the correct use of assessment tools and providing excellent management of delirium, but also in encouraging engagement with delirium assessment. The implications are that any initiative to introduce delirium screening needs to be accompanied by training in delirium. Overall, qualitative studies urge a whole-systems approach, in which not only are staff trained in the basics of delirium and informed about the use of assessment tools, but the organisation and its culture also support delirium assessment and treatment, and there is a clear action plan following delirium diagnosis.

In the main diagnostic accuracy study, the 4AT was found to have a sensitivity of 76% and a specificity of 95%. Taken in the context of the other existing diagnostic accuracy studies, which, on the whole, have shown to have higher sensitivity and lower specificity, the 4AT can be considered a useful tool for delirium assessment in acute care. These conclusions are supported by the other part of this project, notably survey B, and by the widespread adoption of the 4AT by multiple external organisations and health-care professionals. The much lower sensitivity of the CAM was surprising, given that in most previous studies the sensitivity was higher; however, the present findings are not unique, with some other studies also showing low sensitivity. Our study findings suggest that the 4AT is more sensitive than the CAM; this is of interest because the 4AT and CAM were conducted in the same settings and patient cohorts. Interestingly, in recognition of the fact that the CAM takes too long for routine clinical use, since the present study was commenced alternative and shorter forms of the CAM have been developed with the explicit aim of shortening it. These include the 3D-CAM¹⁷⁹ and the Brief CAM.¹⁹³ Additionally, some other short tools have been tested, for example a two-question set (involving months of the year backwards and asking patients what day of the week it is) followed by a longer delirium assessment if the answer to one or both of these is wrong.²⁰⁹ There is also ongoing work with other short tests, such as the nursing delirium screen (NuDesc).²¹⁰ The choice of tool in a given context will depend on the time available, the desired performance characteristics, the training available, and so on.

We have been able to corroborate previous work that suggests that delirium adds substantially to health-care costs. An accurate estimate of these excess costs in a UK context will be a valuable resource for future research and service planning. The formal assessment of cost-effectiveness in the context of this study was difficult for several reasons. Because patient management took place on the basis of a reference standard diagnosis, we were not able to measure costs and other outcomes as they would be if care were based on the 4AT score. Our modelling approach relied on a comparison of diagnostic properties; a fundamental requirement of this approach is information on the costs and outcomes associated with

false-negative and false-positive diagnoses. As data on these were not available from our study, and owing to the sparsity of the published literature, we were forced to rely on expert opinion, with all of its inherent flaws.

Despite the limitations, the economic analysis was able to demonstrate the potential of the 4AT as a cost-effective screening tool. To determine how it should best be used to maximise cost-effectiveness, further research is required. This should specifically focus on resolving some of the uncertainties around costs, mortality and quality of life experienced by patients who are managed after screening with the 4AT. Given the challenges of performing a formal Phase III randomised controlled trial in this context, it might be that observational studies of practice and outcomes in real-world settings can do much to provide the basis of a better informed economic model.

Research implications

1. The fundamental lack of knowledge of delirium across the health-care system and across disciplines constitutes a critical barrier to better delirium assessment and care. Research on the presence of delirium in undergraduate and postgraduate training of health professionals, as well as its presence in mandatory continuing professional development programmes, should be studied. Some studies have begun to address this,³⁶ but a more complete picture would allow for proper planning to ensure that this hitherto neglected part of health-care education is addressed. Related to this, the present studies indicate multiple types of learning needed among professionals. There have been important advances in delirium education research in the last few years.^{37,38} However, given the complexity of the educational challenge, working towards a range of validated education approaches for professionals in different disciplines would be valuable.
2. There is consensus that delirium is an important and common problem, and that delirium assessment and care, as well as efforts to reduce the risk of delirium, are essential in providing excellent care. Yet the results of survey A, as well as of other studies and audits, demonstrate that health-care organisations show a variable approach to delirium care. Although education of individual professionals is critical in improving delirium care, it is also essential that organisations have a coherent approach to delirium. Research on what policies on delirium care exist among health-care organisations in the UK and beyond would inform planning and policy-making.
3. The present study on the diagnostic accuracy of the 4AT provides encouraging results, and several other studies on the 4AT, as detailed in *Chapter 1*, also provide support. Additional studies in different settings, such as in palliative care, care homes and surgical populations, would be of value.
4. The findings from health economics analysis are in line with those of previous studies showing that delirium is associated with a substantial increase in costs. Further work detailing the costs of delirium, in particular the costs of delirium based on real-world analysis rather than on expert opinion (as was used in the present study), would help inform policy-makers about the value of providing effective systems for delirium care.
5. The 4AT is an initial assessment tool and does not provide a diagnosis. The diagnosis might be based largely on the 4AT findings in conjunction with clinical judgement, or the clinician might seek additional information before coming to a conclusion. The processes by which the use of a tool such as the 4AT leads to a diagnosis needs to be better studied, in particular whether additional information and formal assessment against DSM-5 criteria are used, or if the diagnosis largely relies on the 4AT.

Clinical implications

The studies in this project suggest that the 4AT is useful and practical for detecting delirium in acute hospital care. The 4AT is intended to be used as part of the assessment at first encounter with a high-risk patient, and also where delirium is suspected; it is not intended to be used in daily monitoring for new-onset delirium. Therefore, although the 4AT has a place in the range of validated tools for delirium, practitioners

might consider using alternative tools for day-to-day monitoring. The results of the surveys and qualitative studies strongly suggest that there are significant training needs for clinical staff, which should be considered alongside addressing introduction of the 4AT and other tools. Moreover, the findings suggest a lack of clarity about the roles different members of the MDT have in detecting delirium. These findings suggest that making consistent delirium detection a reality requires not only tools and skilled and knowledgeable staff, but also institutional support and the development of systems of care for delirium that can be applied following a diagnosis. There is a clear need for a more consistent approach to the treatment of delirium once it has been identified. One such approach is the TIME (Think, Investigate, Manage, Engage and Explore) bundle produced by Healthcare Improvement Scotland,²¹¹ which covers the first 2 hours after diagnosis. Another more comprehensive approach derived from expert consensus is the Comprehensive Delirium Management Pathway, produced by the Scottish Delirium Association (www.scottishdeliriumassociation.com).

Acknowledgements

This study was supported by the National Institute for Health Research Health Technology Assessment programme. Alasdair MJ MacLulich and Susan D Shenkin are members of the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-council Lifelong Health and Wellbeing Initiative (grant number MR/K026992/1).

The authors would like to thank all of the patients and carers who give up their time to take part in the study. We would also like to thank all of the NHS staff who supported the study in the three sites. The authors acknowledge the help of the Trial Steering Committee. The committee was chaired by Professor Miles Witham, and the other members were Professor Phyo Myint, Dr Julia Boyd, Dr Linda Wolff, Ms Barbara Lamb, Dr Daniel Davis, Ms Ann Stewart, Ms Jo-Anne Robertson, Ms Fiona Redmond, Professor Christopher Weir and Professor Alasdair MacLulich. We thank all of the administrative staff in the three study sites for their support. We thank Dr Valentia Assi for contributing to the statistical analyses. We acknowledge the help of the Edinburgh Clinical Trials Unit, including Mr Allan Walker (database programmer) and Dr Lorraine Smith. We thank Ms Miranda Odam, Ms Mia Paderanga and Ms Louise Ross from the Emergency Medicine Research Group of Edinburgh (EMERGE). We thank the research nurses and other research staff in the clinical research facilities who took part in the recruitment and testing.

The protocol can be obtained from the corresponding author.

Contributions of authors

Alasdair MJ MacLulich (Professor of Geriatric Medicine) was chief investigator and co-site lead for Edinburgh. He conceived and planned the study and acted as study lead. He led writing of the report and all aspects of study design, conduct and analysis, and designed and led the training of research staff.

Susan D Shenkin (Senior Lecturer in Geriatric Medicine) was co-site lead for Edinburgh. She provided expertise in geriatric medicine, and contributed to protocol design and statistical analysis and supervision of staff, and drafted sections of the report.

Steve Goodacre (Professor of Emergency Medicine) was site lead for Sheffield. He provided expertise in emergency medicine, and contributed to protocol design, staff training and supervision, and writing of the report.

Mary Godfrey (Reader in Health and Social Care) provided expertise in qualitative research, contributed to the design, conduct and analysis of the qualitative studies, staff training and supervision, and drafted sections of the report.

Janet Hanley (Reader in Nursing Studies) provided expertise in qualitative research, contributed to the design, conduct and analysis of the qualitative studies, staff training and supervision, and drafted sections of the report.

Antaine Stiobhairt (Research Associate in Psychology) contributed to the design and analysis of the qualitative studies, recruited and interviewed patients for these studies, and drafted sections of the report.

Elizabeth Lavender (Qualitative Researcher) contributed to the design and analysis of the qualitative studies, recruited and interviewed patients for these studies, and drafted sections of the report.

Julia Boyd (Project Manager) provided expertise in study management, contributed to protocol design and staff training and supervision, and was project manager.

Jacqueline Stephen (Statistician) was a trial statistician. She contributed to the development of the statistical analysis plan, undertook the statistical analyses and drafted sections of the report.

Christopher Weir (Statistician) was the trial statistician. He co-designed the protocol and led the design of the statistical analysis plan, oversaw the statistical analyses, and contributed to the writing of the report.

Allan MacRaid (Research Nurse) contributed to study design, participated in patient recruitment and testing, and participated in staff training.

Jill Steven (Research Nurse) contributed to study design, participated in patient recruitment and testing, and participated in staff training.

Polly Black (Research Nurse) contributed to study design, participated in patient recruitment and testing, and participated in staff training.

Katharina Diernberger (Health Economist) provided expertise in health economics, co-designed the health economics analysis, carried out the analysis of the health economics data, and drafted sections of the report.

Peter Hall (Health Economist) provided expertise in health economics, co-designed the health economics analysis, supervised the analysis of the health economics data, and drafted sections of the report.

Zoë Tiegies (Research Psychologist) provided expertise in psychology, and contributed to design of the reference standard assessment and to data analysis.

Christopher Fox (Professor of Old Age Psychiatry) provided expertise in dementia, and contributed to protocol design and writing of the report.

Atul Anand (Clinical Research Fellow in Geriatric Medicine) contributed to protocol design and writing of the report.

John Young (Professor of Elderly Care Medicine) was Bradford site lead. He contributed to protocol design and writing of the report.

Najma Siddiqi (Consultant Psychiatrist) provided expertise in psychiatry and contributed to protocol design and writing of the report.

Alasdair Gray (Professor of Emergency Medicine) provided expertise in emergency medicine, was co-site lead for Edinburgh, and contributed to staff training and supervision and writing of the report.

Publication

Shenkin SD, Fox C, Godfrey M, Siddiqi N, Goodacre S, Young J, *et al*. Delirium detection in older acute medical inpatients: a multicentre prospective comparative diagnostic test accuracy study of the 4AT and the confusion assessment method. *BMC Med* 2019;**17**:138. <https://doi.org/10.1186/s12916-019-1367-9>

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review and appropriate agreements being in place. Exclusive use will be retained until the publication of major outputs.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. 4th edn. Washington, DC: American Psychiatric Association; 2000.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th edn. 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 Ageing* 2006;**35**:350–64. <https://doi.org/10.1093/ageing/af1005>
4. National Institute for Health and Care Excellence (NICE). *Delirium: Diagnosis, Prevention And Management*. London: NICE; 2010. URL: www.nice.org.uk/CG103 (accessed 1 December 2017).
5. Goldberg SE, Whittamore KH, Harwood RH, Bradshaw LE, Gladman JR, Jones RG, Medical Crises in Older People Study Group. The prevalence of mental health problems among older adults admitted as an emergency to a general hospital. *Age Ageing* 2012;**41**:80–6. <https://doi.org/10.1093/ageing/afr106>
6. Timmons S, Manning E, Barrett A, Brady NM, Browne V, O'Shea E, *et al*. Dementia in older people admitted to hospital: a regional multi-hospital observational study of prevalence, associations and case recognition. *Age Ageing* 2015;**44**:993–9. <https://doi.org/10.1093/ageing/afv131>
7. Collins N, Blanchard MR, Tookman A, Sampson EL. Detection of delirium in the acute hospital. *Age Ageing* 2010;**39**:131–5. <https://doi.org/10.1093/ageing/afp201>
8. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet* 2014;**383**:911–22. [https://doi.org/10.1016/S0140-6736\(13\)60688-1](https://doi.org/10.1016/S0140-6736(13)60688-1)
9. Davis DH, Muniz Terrera G, Keage H, Rahkonen T, Oinas M, Matthews FE, *et al*. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. *Brain* 2012;**135**:2809–16. <https://doi.org/10.1093/brain/aws190>
10. Davis DH, Muniz-Terrera G, Keage HA, Stephan BC, Fleming J, Ince PG, *et al*. Association of delirium with cognitive decline in late life: a neuropathologic study of 3 population-based cohort studies. *JAMA Psychiatry* 2017;**74**:244–51. <https://doi.org/10.1001/jamapsychiatry.2016.3423>
11. MacLulich AM, Beaglehole A, Hall RJ, Meagher DJ. Delirium and long-term cognitive impairment. *Int Rev Psychiatry* 2009;**21**:30–42. <https://doi.org/10.1080/09540260802675031>
12. Witlox J, Eurelings LS, de Jonghe JF, Kalisvaart KJ, Eikelenboom P, van Gool WA. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. *JAMA* 2010;**304**:443–51. <https://doi.org/10.1001/jama.2010.1013>
13. Partridge JS, Martin FC, Harari D, Dhesi JK. The delirium experience: what is the effect on patients, relatives and staff and what can be done to modify this? *Int J Geriatr Psychiatry* 2013;**28**:804–12. <https://doi.org/10.1002/gps.3900>
14. Babine RL, Hyrkäs KE, Bachand DA, Chapman JL, Fuller VJ, Honess CA, Wierman HR. Falls in a tertiary care hospital-association with delirium: a replication study. *Psychosomatics* 2016;**57**:273–82. <https://doi.org/10.1016/j.psych.2016.01.003>
15. Jackson TA, Wilson D, Richardson S, Lord JM. Predicting outcome in older hospital patients with delirium: a systematic literature review. *Int J Geriatr Psychiatry* 2016;**31**:392–9. <https://doi.org/10.1002/gps.4344>
16. Royal College of Psychiatrists. *Who Cares Wins*. London: Royal College of Psychiatrists; 2005.

17. Jackson TA, MacLulich AM, Gladman JR, Lord JM, Sheehan B. Undiagnosed long-term cognitive impairment in acutely hospitalised older medical patients with delirium: a prospective cohort study. *Age Ageing* 2016;**45**:493–9. <https://doi.org/10.1093/ageing/afw064>
18. Jackson TA, Gladman JR, Harwood RH, MacLulich AM, Sampson EL, Sheehan B, Davis DH. Challenges and opportunities in understanding dementia and delirium in the acute hospital. *PLOS Med* 2017;**14**:e1002247. <https://doi.org/10.1371/journal.pmed.1002247>
19. Fong TG, Jones RN, Shi P, Marcantonio ER, Yap L, Rudolph JL, et al. Delirium accelerates cognitive decline in Alzheimer disease. *Neurology* 2009;**72**:1570–5. <https://doi.org/10.1212/WNL.0b013e3181a4129a>
20. Rockwood K, Cosway S, Carver D, Jarrett P, Stadnyk K, Fisk J. The risk of dementia and death after delirium. *Age Ageing* 1999;**28**:551–6. <https://doi.org/10.1093/ageing/28.6.551>
21. Leslie DL, Marcantonio ER, Zhang Y, Leo-Summers L, Inouye SK. One-year health care costs associated with delirium in the elderly population. *Arch Intern Med* 2008;**168**:27–32. <https://doi.org/10.1001/archinternmed.2007.4>
22. Kakuma R, du Fort GG, Arsenault L, Perrault A, Platt RW, Monette J, et al. Delirium in older emergency department patients discharged home: effect on survival. *J Am Geriatr Soc* 2003;**51**:443–50. <https://doi.org/10.1046/j.1532-5415.2003.51151.x>
23. Bush SH, Tierney S, Lawlor PG. Clinical assessment and management of delirium in the palliative care setting. *Drugs* 2017;**77**:1623–43. <https://doi.org/10.1007/s40265-017-0804-3>
24. Fuller V. Delirium recall – an integrative review. *J Clin Nurs* 2016;**25**:1515–27. <https://doi.org/10.1111/jocn.13155>
25. Morandi A, Lucchi E, Turco R, Morghen S, Guerini F, Santi R, et al. Delirium superimposed on dementia: a quantitative and qualitative evaluation of patient experience. *J Psychosom Res* 2015;**79**:281–7. <https://doi.org/10.1016/j.jpsychores.2015.07.010>
26. Aligeti S, Baig MR, Barrera FF. Terminal delirium misdiagnosed as major psychiatric disorder: palliative care in a psychiatric inpatient unit. *Palliat Support Care* 2016;**14**:307–10. <https://doi.org/10.1017/S147895151500098X>
27. Adunsky A, Levy R, Mizrahi E, Arad M. Exposure to opioid analgesia in cognitively impaired and delirious elderly hip fracture patients. *Arch Gerontol Geriatr* 2002;**35**:245–51. [https://doi.org/10.1016/S0167-4943\(02\)00044-4](https://doi.org/10.1016/S0167-4943(02)00044-4)
28. Morandi A, Lucchi E, Turco R, Morghen S, Guerini F, Santi R, et al. Delirium superimposed on dementia: a quantitative and qualitative evaluation of informal caregivers and health care staff experience. *J Psychosom Res* 2015;**79**:272–80. <https://doi.org/10.1016/j.jpsychores.2015.06.012>
29. Day J, Higgins I. Adult family member experiences during an older loved one's delirium: a narrative literature review. *J Clin Nurs* 2015;**24**:1447–56. <https://doi.org/10.1111/jocn.12771>
30. Department of Health and Social Care. *Using the Commissioning for Quality and Innovation (CQUIN) Payment Framework: Guidance on New National Goals for 2012–13*. London: Department of Health and Social Care; 2012.
31. Elie M, Rousseau F, Cole M, Primeau F, McCusker J, Bellavance F. Prevalence and detection of delirium in elderly emergency department patients. *CMAJ* 2000;**163**:977–81.
32. Han JH, Zimmerman EE, Cutler N, Schnelle J, Morandi A, Dittus RS, et al. Delirium in older emergency department patients: recognition, risk factors, and psychomotor subtypes. *Acad Emerg Med* 2009;**16**:193–200. <https://doi.org/10.1111/j.1553-2712.2008.00339.x>

33. Stelmokas J, Gabel N, Flaherty JM, Rayson K, Tran K, Anderson JR, Bieliauskas LA. Delirium detection and impact of comorbid health conditions in a post-acute rehabilitation hospital setting. *PLOS ONE* 2016;**11**:e0166754. <https://doi.org/10.1371/journal.pone.0166754>
34. Davis D, MacLulich A. Understanding barriers to delirium care: a multicentre survey of knowledge and attitudes amongst UK junior doctors. *Age Ageing* 2009;**38**:559–63. <https://doi.org/10.1093/ageing/afp099>
35. Jenkin RP, Al-Attar A, Richardson S, Myint PK, MacLulich AM, Davis DH. Increasing delirium skills at the front door: results from a repeated survey on delirium knowledge and attitudes. *Age Ageing* 2016;**45**:517–22. <https://doi.org/10.1093/ageing/afw066>
36. Fisher JM, Gordon AL, MacLulich AM, Tullo E, Davis DH, Blundell A, et al. Towards an understanding of why undergraduate teaching about delirium does not guarantee gold-standard practice – results from a UK national survey. *Age Ageing* 2015;**44**:166–70. <https://doi.org/10.1093/ageing/afu154>
37. Teodorczuk A, Mukaetova-Ladinska E, Corbett S, Welfare M. Deconstructing dementia and delirium hospital practice: using cultural historical activity theory to inform education approaches. *Adv Health Sci Educ Theory Pract* 2015;**20**:745–64. <https://doi.org/10.1007/s10459-014-9562-0>
38. Teodorczuk A, Mukaetova-Ladinska E, Corbett S, Welfare M. Reconceptualizing models of delirium education: findings of a grounded theory study. *Int Psychogeriatr* 2013;**25**:645–55. <https://doi.org/10.1017/S1041610212002074>
39. World Health Organisation. *International Classification of Disease, 11th Revision (ICD-11)*. 2018. URL: www.who.int/classifications/icd/en
40. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegel AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med* 1990;**113**:941–8. <https://doi.org/10.7326/0003-4819-113-12-941>
41. Bradley EH, Schlesinger M, Webster TR, Baker D, Inouye SK. Translating research into clinical practice: making change happen. *J Am Geriatr Soc* 2004;**52**:1875–82. <https://doi.org/10.1111/j.1532-5415.2004.52510.x>
42. Chester JG, Beth Harrington M, Rudolph JL, VA Delirium Working Group. Serial administration of a modified Richmond Agitation and Sedation Scale for delirium screening. *J Hosp Med* 2012;**7**:450–3. <https://doi.org/10.1002/jhm.1003>
43. Yates C, Stanley N, Cerejeira JM, Jay R, Mukaetova-Ladinska EB. Screening instruments for delirium in older people with an acute medical illness. *Age Ageing* 2009;**38**:235–7. <https://doi.org/10.1093/ageing/afn285>
44. 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**:31–5. <https://doi.org/10.1093/ageing/26.1.31>
45. Eeles EM, White S, Bayer T. Inability to complete cognitive assessment is a marker of adverse risk. *Age Ageing* 2009;**38**:633–4. <https://doi.org/10.1093/ageing/afp115>
46. Tieges Z, Evans JJ, Neufeld KJ, MacLulich AM. The neuropsychology of delirium: advancing the science of delirium assessment. *Int J Geriatr Psychiatry* 2018;**33**:1501–11. <https://doi.org/10.1002/gps.4711>
47. Tieges Z, McGrath A, Hall RJ, MacLulich AM. Abnormal level of arousal as a predictor of delirium and inattention: an exploratory study. *Am J Geriatr Psychiatry* 2013;**21**:1244–53. <https://doi.org/10.1016/j.jagp.2013.05.003>
48. Hall RJ, Meagher DJ, MacLulich AM. Delirium detection and monitoring outside the ICU. *Best Pract Res Clin Anaesthesiol* 2012;**26**:367–83. <https://doi.org/10.1016/j.bpa.2012.07.002>

49. De J, Wand AP. Delirium screening: a systematic review of delirium screening tools in hospitalized patients. *Gerontologist* 2015;**55**:1079–99. <https://doi.org/10.1093/geront/gnv100>
50. 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**:255–9. <https://doi.org/10.1016/j.jpsychores.2008.05.024>
51. Tamune H, Yasugi D. How can we identify patients with delirium in the emergency department? A review of available screening and diagnostic tools. *Am J Emerg Med* 2017;**35**:1332–4. <https://doi.org/10.1016/j.ajem.2017.05.026>
52. Jackson TA, Naqvi SH, Sheehan B. Screening for dementia in general hospital inpatients: a systematic review and meta-analysis of available instruments. *Age Ageing* 2013;**42**:689–95. <https://doi.org/10.1093/ageing/aft145>
53. Ringdal GI, Ringdal K, Juliebø V, Wyller TB, Hjermsstad MJ, Loge JH. Using the Mini-Mental State Examination to screen for delirium in elderly patients with hip fracture. *Dement Geriatr Cogn Disord* 2011;**32**:394–400. <https://doi.org/10.1159/000335743>
54. Tiegies Z, Brown LJ, MacLulich AM. Objective assessment of attention in delirium: a narrative review. *Int J Geriatr Psychiatry* 2014;**29**:1185–97. <https://doi.org/10.1002/gps.4131>
55. Lees R, Corbet S, Johnston C, Moffitt E, Shaw G, Quinn TJ. Test accuracy of short screening tests for diagnosis of delirium or cognitive impairment in an acute stroke unit setting. *Stroke* 2013;**44**:3078–83. <https://doi.org/10.1161/STROKEAHA.113.001724>
56. Bellelli G, Morandi A, Davis DH, Mazzola P, Turco R, Gentile S, *et al*. Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age Ageing* 2014;**43**:496–502. <https://doi.org/10.1093/ageing/afu021>
57. 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. <https://doi.org/10.2147/NDT.S97228>
58. De J, Wand AP, Smerdely PI, Hunt GE. Validating the 4A's test in screening for delirium in a culturally diverse geriatric inpatient population. *Int J Geriatr Psychiatry* 2017;**32**:1322–9. <https://doi.org/10.1002/gps.4615>
59. Hendry K, Quinn TJ, Evans J, Scortichini V, Miller H, Burns J, *et al*. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. *Age Ageing* 2016;**45**:832–7. <https://doi.org/10.1093/ageing/afw130>
60. O'Sullivan D, Brady N, Manning E, O'Shea E, O'Grady S, O'Regan N, Timmons S. Validation of the 6-Item Cognitive Impairment Test and the 4AT test for combined delirium and dementia screening in older emergency department attendees. *Age Ageing* 2018;**47**:61–8. <https://doi.org/10.1093/ageing/afx149>
61. Baird L, Spiller JA. A quality improvement approach to cognitive assessment on hospice admission: could we use the 4AT or Short CAM? *BMJ Open Qual* 2017;**6**:e000153. <https://doi.org/10.1136/bmjopen-2017-000153>
62. Weinrebe W, Johannsdottir E, Karaman M, Füsgen I. What does delirium cost? An economic evaluation of hyperactive delirium. *Z Gerontol Geriatr* 2016;**49**:52–8. <https://doi.org/10.1007/s00391-015-0871-6>
63. Akunne A, Davis S, Westby M, Young J. The cost-effectiveness of multi-component interventions to prevent delirium in older people undergoing surgical repair of hip fracture. *Eur J Orthop Surg Traumatol* 2014;**24**:187–95. <https://doi.org/10.1007/s00590-013-1170-9>

64. Akunne A, Murthy L, Young J. Cost-effectiveness of multi-component interventions to prevent delirium in older people admitted to medical wards. *Age Ageing* 2012;**41**:285–91. <https://doi.org/10.1093/ageing/afr147>
65. Cole MG, Bailey R, Bonnycastle M, McCusker J, Fung S, Ciampi A, *et al.* Partial and no recovery from delirium in older hospitalized adults: frequency and baseline risk factors. *J Am Geriatr Soc* 2015;**63**:2340–8. <https://doi.org/10.1111/jgs.13791>
66. Pandharipande PP, Girard TD, Ely EW. Long-term cognitive impairment after critical illness. *N Engl J Med* 2014;**370**:185–6. <https://doi.org/10.1056/NEJMc1313886>
67. Numan T, van den Boogaard M, Kamper AM, Rood PJT, Peelen LM, Slooter AJC, Dutch Delirium Detection Study Group. Recognition of delirium in postoperative elderly patients: a multicenter study. *J Am Geriatr Soc* 2017;**65**:1932–8. <https://doi.org/10.1111/jgs.14933>
68. El Hussein M, Hirst S, Salyers V. Factors that contribute to underrecognition of delirium by registered nurses in acute care settings: a scoping review of the literature to explain this phenomenon. *J Clin Nurs* 2015;**24**:906–15. <https://doi.org/10.1111/jocn.12693>
69. Bellelli G, Nobili A, Annoni G, Morandi A, Djade CD, Meagher DJ, *et al.* Under-detection of delirium and impact of neurocognitive deficits on in-hospital mortality among acute geriatric and medical wards. *Eur J Intern Med* 2015;**26**:696–704. <https://doi.org/10.1016/j.ejim.2015.08.006>
70. van Velthuisen EL, Zwakhalen SM, Mulder WJ, Verhey FR, Kempen GI. Detection and management of hyperactive and hypoactive delirium in older patients during hospitalization: a retrospective cohort study evaluating daily practice. *Int J Geriatr Psychiatry* 2018;**33**:1521–9. <https://doi.org/10.1002/gps.4690>
71. Mistarz R, Elliott S, Whitfield A, Ernest D. Bedside nurse-patient interactions do not reliably detect delirium: an observational study. *Aust Crit Care* 2011;**24**:126–32. <https://doi.org/10.1016/j.aucc.2011.01.002>
72. Arendts G, Love J, Nagree Y, Bruce D, Hare M, Dey I. Rates of delirium diagnosis do not improve with emergency risk screening: results of the Emergency Department Delirium Initiative Trial. *J Am Geriatr Soc* 2017;**65**:1810–15. <https://doi.org/10.1111/jgs.14904>
73. Bellelli G, Morandi A, Zanetti E, Bozzini M, Lucchi E, Terrasi M, *et al.* Recognition and management of delirium among doctors, nurses, physiotherapists, and psychologists: an Italian survey. *Int Psychogeriatr* 2014;**26**:2093–102. <https://doi.org/10.1017/S1041610214001653>
74. Cadogan FL, Riekerk B, Vreeswijk R, Rommes JH, Toornvliet AC, Honing ML, Spronk PE. Current awareness of delirium in the intensive care unit: a postal survey in the Netherlands. *Neth J Med* 2009;**67**:296–300.
75. Devlin JW, Bhat S, Roberts RJ, Skrobik Y. Current perceptions and practices surrounding the recognition and treatment of delirium in the intensive care unit: a survey of 250 critical care pharmacists from eight states. *Ann Pharmacother* 2011;**45**:1217–29. <https://doi.org/10.1345/aph.1Q332>
76. Eastwood GM, Peck L, Bellomo R, Baldwin I, Reade MC. A questionnaire survey of critical care nurses' attitudes to delirium assessment before and after introduction of the CAM-ICU. *Aust Crit Care* 2012;**25**:162–9. <https://doi.org/10.1016/j.aucc.2012.01.005>
77. Elliott SR. ICU delirium: a survey into nursing and medical staff knowledge of current practices and perceived barriers towards ICU delirium in the intensive care unit. *Intensive Crit Care Nurs* 2014;**30**:333–8. <https://doi.org/10.1016/j.iccn.2014.06.004>

78. Ely EW, Stephens RK, Jackson JC, Thomason JW, Truman B, Gordon S, *et al.* Current opinions regarding the importance, diagnosis, and management of delirium in the intensive care unit: a survey of 912 healthcare professionals. *Crit Care Med* 2004;**32**:106–12. <https://doi.org/10.1097/01.CCM.0000098033.94737.84>
79. Forsgren LM, Eriksson M. Delirium – awareness, observation and interventions in intensive care units: a national survey of Swedish ICU head nurses. *Intensive Crit Care Nurs* 2010;**26**:296–303. <https://doi.org/10.1016/j.iccn.2010.07.003>
80. Hamdan-Mansour AM, Farhan NA, Othman EH, Yacoub MI. Knowledge and nursing practice of critical care nurses caring for patients with delirium in intensive care units in Jordan. *J Contin Educ Nurs* 2010;**41**:571–6. <https://doi.org/10.3928/00220124-20100802-01>
81. Jildenstål PK, Rawal N, Hallén JL, Berggren L, Jakobsson JG. Perioperative management in order to minimise postoperative delirium and postoperative cognitive dysfunction: results from a Swedish web-based survey. *Ann Med Surg* 2014;**3**:100–7. <https://doi.org/10.1016/j.amsu.2014.07.001>
82. Luetz A, Balzer F, Radtke FM, Jones C, Citerio G, Walder B, *et al.* Delirium, sedation and analgesia in the intensive care unit: a multinational, two-part survey among intensivists. *PLOS ONE* 2014;**9**:e110935. <https://doi.org/10.1371/journal.pone.0110935>
83. Mac Sweeney R, Barber V, Page V, Ely EW, Perkins GD, Young JD, *et al.* A national survey of the management of delirium in UK intensive care units. *QJM* 2010;**103**:243–51. <https://doi.org/10.1093/qjmed/hcp194>
84. Morandi A, Davis D, Taylor JK, Bellelli G, Olofsson B, Kreisel S, *et al.* Consensus and variations in opinions on delirium care: a survey of European delirium specialists. *Int Psychogeriatr* 2013;**25**:2067–75. <https://doi.org/10.1017/S1041610213001415>
85. Özşaban A, Acaroglu R. Delirium assessment in intensive care units: practices and perceptions of Turkish nurses. *Nurs Crit Care* 2016;**21**:271–8. <https://doi.org/10.1111/nicc.12127>
86. Patel RP, Gambrell M, Speroff T, Scott TA, Pun BT, Okahashi J, *et al.* Delirium and sedation in the intensive care unit: survey of behaviors and attitudes of 1384 healthcare professionals. *Crit Care Med* 2009;**37**:825–32. <https://doi.org/10.1097/CCM.0b013e31819b8608>
87. Shehabi Y, Botha JA, Boyle MS, Ernest D, Freebairn RC, Jenkins IR, *et al.* Sedation and delirium in the intensive care unit: an Australian and New Zealand perspective. *Anaesth Intensive Care* 2008;**36**:570–8. <https://doi.org/10.1177/0310057X0803600423>
88. Van Eijk MM, Kesecioglu J, Slooter AJ. Intensive care delirium monitoring and standardised treatment: a complete survey of Dutch intensive care units. *Intensive Crit Care Nurs* 2008;**24**:218–21. <https://doi.org/10.1016/j.iccn.2008.04.005>
89. Sri-on J, Tirrell GP, Vanichkulbodee A, Niruntarai S, Liu SW. The prevalence, risk factors and short-term outcomes of delirium in Thai elderly emergency department patients. *Emerg Med J* 2016;**33**:17–22. <https://doi.org/10.1136/emered-2014-204379>
90. Shipway DJ, Partridge JS, Foxton CR, Modarai B, Gossage JA, Challacombe BJ, *et al.* Do surgical trainees believe they are adequately trained to manage the ageing population? A UK survey of knowledge and beliefs in surgical trainees. *J Surg Educ* 2015;**72**:641–7. <https://doi.org/10.1016/j.jsurg.2015.01.019>
91. Sinvani L, Kozikowski A, Pekmezaris R, Akerman M, Wolf-Klein G. Delirium: a survey of healthcare professionals' knowledge, beliefs, and practices. *J Am Geriatr Soc* 2016;**64**:e297–e303. <https://doi.org/10.1111/jgs.14544>
92. Rowley-Conwy G. Critical care nurses' knowledge and practice of delirium assessment. *Br J Nurs* 2017;**26**:412–17. <https://doi.org/10.12968/bjon.2017.26.7.412>

93. Selim AA, Ely EW. Delirium the under-recognised syndrome: survey of healthcare professionals' awareness and practice in the intensive care units. *J Clin Nurs* 2017;**26**:813–24. <https://doi.org/10.1111/jocn.13517>
94. Trogrlić Z, Ista E, Ponssen HH, Schoonderbeek JF, Schreiner F, Verbrugge SJ, *et al.* Attitudes, knowledge and practices concerning delirium: a survey among intensive care unit professionals. *Nurs Crit Care* 2017;**22**:133–40. <https://doi.org/10.1111/nicc.12239>
95. Richardson SJ, Davis DHJ, Bellelli G, Hasemann W, Meagher D, Kreisel SH, *et al.* Detecting delirium superimposed on dementia: diagnostic accuracy of a simple combined arousal and attention testing procedure. *Int Psychogeriatr* 2017;**29**:1585–93. <https://doi.org/10.1017/S1041610217000916>
96. Meagher J, Leonard M, Donoghue L, O'Regan N, Timmons S, Exton C, *et al.* Months backward test: a review of its use in clinical studies. *World J Psychiatry* 2015;**5**:305–14. <https://doi.org/10.5498/wjp.v5.i3.305>
97. Morandi A, Di Santo SG, Cherubini A, Mossello E, Meagher D, Mazzone A, *et al.* Clinical features associated with delirium motor subtypes in older inpatients: results of a multicenter study. *Am J Geriatr Psychiatry* 2017;**25**:1064–71. <https://doi.org/10.1016/j.jagp.2017.05.003>
98. Bellelli G, Morandi A, Di Santo SG, Mazzone A, Cherubini A, Mossello E, *et al.* 'Delirium Day': a nationwide point prevalence study of delirium in older hospitalized patients using an easy standardized diagnostic tool. *BMC Med* 2016;**14**:106. <https://doi.org/10.1186/s12916-016-0649-8>
99. Bo M, Bonetto M, Bottignole G, Porrino P, Coppo E, Tibaldi M, *et al.* Length of stay in the emergency department and occurrence of delirium in older medical patients. *J Am Geriatr Soc* 2016;**64**:1114–19. <https://doi.org/10.1111/jgs.14103>
100. Scottish Government. *Scottish Standards of Care for Hip Fracture Patients 2016*. Edinburgh: Scottish Government; 2016. URL: www.shfa.scot.nhs.uk/_docs/20161109_SSC_for_Hip_Fracture_Patients.pdf (accessed 5 December 2017).
101. Healthcare Improvement Scotland. *Delirium Toolkit*. Edinburgh: Healthcare Improvement Scotland. URL: www.healthcareimprovementscotland.org/our_work/person-centred_care/opac_improvement_programme/delirium_toolkit.aspx (accessed 5 December 2017).
102. Care Inspectorate and NHS Scotland. *Managing Falls and Fractures in Care Homes for Older People – Good Practice Resource. Revised Edition*. 2016. URL: www.careinspectorate.com/images/documents/2737/2016/Falls-and-fractures-new-resource-low-res.pdf (accessed 5 December 2017).
103. Barrett E. *Dementia Revealed: What Primary Care Needs to Know. A Primer for General Practice*. Dementia Partnerships; 2014. URL: <http://dementiapartnerships.com/resource/dementia-the-view-from-primary-care/> (accessed 5 December 2017).
104. Royal College of Emergency Medicine. *Assessing for Cognitive Impairment in Older People: Clinical Audit 2014–15*. London: Royal College of Emergency Medicine; 2015. URL: www.rcem.ac.uk/docs/Previous%20Audits/CEM8463-RCEM%20Older%20People%202014-15%20National%20Audit%20Report.pdf (accessed 5 December 2017).
105. Public Health England. *Falls and Fracture Consensus Statement: Resource Pack*. London: Public Health England; 2017. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/628732/Falls_and_fracture_consensus_statement_resource_pack.pdf (accessed 5 December 2017).
106. London Major Trauma System. *London Major Trauma System: Management of Elderly Major Trauma Patients - Second Edition*. London: London Major Trauma System; 2017. URL: www.c4ts.qmul.ac.uk/downloads/london-major-trauma-system-elderly-trauma-guidancessecond-editiondecember-2018.pdf (accessed 29 May 2019).

107. US Department of Veterans Affairs. *Geriatrics and Extended Care: Delirium Information for VA Health Care Professionals*. Washington, DC: US Department of Veterans Affairs. URL: www.va.gov/geriatrics/Guide/LongTermCare/Delirium_professionals.asp# (accessed 5 December 2017).
108. American Nurses Association. *Delirium: Prevent, Identify, Treat*. Silver Spring, MD: American Nurses Association. URL: www.nursingworld.org/practice-policy/work-environment/health-safety/delirium/ (accessed 5 December 2017).
109. Government of Ireland. *Early Identification and Initial Management of Delirium in the Emergency Department/Acute Medical Assessment Unit*. Dublin: Government of Ireland; 2016. URL: www.hse.ie/eng/services/publications/Clinical-Strategy-and-Programmes/Early-Identification-and-Initial-Management-of-Delirium-in-the-Emergency-Department.pdf (accessed 5 December 2017).
110. Australian Commission on Safety and Quality in Health Care. *Delirium Clinical Care Standard*. Sydney, NSW: Australian Commission on Safety and Quality in Health Care; 2016. URL: www.hcasa.asn.au/documents/254-delirium-clinical-care-standard/file (accessed 5 December 2017).
111. Health Navigator New Zealand. *Delirium*. URL: www.healthnavigator.org.nz/health-a-z/d/delirium/?tab=13127 (accessed 5 December 2017).
112. Google LLC. Google Analytics. 2018. URL: <https://analytics.google.com/analytics/web> (accessed 1 March 2019).
113. Healthcare Improvement Scotland. *Identifying Critical Success Factors for Improved Outcomes for People With Dementia and Their Carers in Acute Care. A Focus on NHS Grampian*. Edinburgh: Healthcare Improvement Scotland. URL: <https://ihub.scot/news-events/identifying-critical-success-factors-for-improved-outcomes-for-people-with-dementia-and-their-carers-in-acute-care-a-focus-on-nhs-grampian/> (accessed 1 March 2019).
114. Healthcare Improvement Scotland. *Improving Older People's Acute Care: Impact Report*. Edinburgh: Healthcare Improvement Scotland; 2015. URL: <http://healthcareimprovementscotland.org/his/idoc.ashx?docid=98dacacf-39da-4f15-beff-d6a472ff1e1c&version=-1> (accessed 5 December 2017).
115. Royal College of Physicians. *National Hip Fracture Database (NHFD) Annual Report 2016*. London: Royal College of Physicians; 2016. URL: www.nhfd.co.uk/2016report (accessed 1 March 2019).
116. Ninewells Hospital and Medical School. *Think Delirium: Recognition of Delirium in an Acute Medical Unit*. Dundee: Ninewells Hospital and Medical School. URL: www.acutemedicine.org.uk/wp-content/uploads/2013/10/aqi%2066%20-%20recognition%20of%20delirium%20in%20an%20amu.pdf (accessed 1 March 2019).
117. Beveridge L. *Improving Identification and Management of Delirium in the Orthopaedic Trauma Unit*. URL: <https://learn.nes.nhs.scot/1147/quality-improvement-zone/learning-programmes/scottish-quality-and-safety-sqs-fellowship-programme/posters-fellowship/fellowship-cohort-6/improving-identification-and-management-of-delirium-in-the-orthopaedic-trauma-unit> (accessed 5 December 2017).
118. Derbyshire Healthcare NHS Foundation Trust. *Annual Report and Accounts 2015/16*. Derby: Derbyshire Healthcare NHS Foundation Trust. URL: http://dchs.nhs.uk/annual_reports (accessed 1 March 2019).
119. Royal College of Psychiatrists. URL: <http://uclpstoneuprod.blob.core.windows.net/cmsassets/140401-%20Who's%20the%20Daddy%20King's%20College%20Hospital%20project%20handout.pdf>
120. Bélanger L, Ducharme F. Patients' and nurses' experiences of delirium: a review of qualitative studies. *Nurs Crit Care* 2011;**16**:303–15. <https://doi.org/10.1111/j.1478-5153.2011.00454.x>

121. Hosie A, Agar M, Lobb E, Davidson PM, Phillips J. Palliative care nurses' recognition and assessment of patients with delirium symptoms: a qualitative study using critical incident technique. *Int J Nurs Stud* 2014;**51**:1353–65. <https://doi.org/10.1016/j.ijnurstu.2014.02.005>
122. Godfrey M, Smith J, Green J, Cheater F, Inouye SK, Young JB. Developing and implementing an integrated delirium prevention system of care: a theory driven, participatory research study. *BMC Health Serv Res* 2013;**13**:341. <https://doi.org/10.1186/1472-6963-13-341>
123. Dahlke S, Phinney A. Caring for hospitalized older adults at risk for delirium: the silent, unspoken piece of nursing practice. *J Gerontol Nurs* 2008;**34**:41–7. <https://doi.org/10.3928/00989134-20080601-03>
124. Brajtman S, Higuchi K, McPherson C. Caring for patients with terminal delirium: palliative care unit and home care nurses' experiences. *Int J Palliat Nurs* 2006;**12**:150–6. <https://doi.org/10.12968/ijpn.2006.12.4.21010>
125. Andersson EM, Hallberg IR, Edberg AK. Nurses' experiences of the encounter with elderly patients in acute confusional state in orthopaedic care. *Int J Nurs Stud* 2003;**40**:437–48. [https://doi.org/10.1016/S0020-7489\(02\)00109-8](https://doi.org/10.1016/S0020-7489(02)00109-8)
126. Rogers AC, Gibson CH. Experiences of orthopedic nurses caring for patients with acute confusion. *J Orthopaedic Nurs* 2002;**6**:9–17. <https://doi.org/10.1054/joon.2001.0210>
127. Lou MF, Dai YT. Nurses' experience of caring for delirious patients. *J Nurs Res* 2002;**10**:279–90. <https://doi.org/10.1097/01.JNR.0000347609.14166.84>
128. Stenwall E, Jönhagen ME, Sandberg J, Fagerberg I. The older patient's experience of encountering professional carers and close relatives during an acute confusional state: an interview study. *Int J Nurs Stud* 2008;**45**:1577–85. <https://doi.org/10.1016/j.ijnurstu.2008.02.001>
129. Bradley EH, Webster TR, Schlesinger M, Baker D, Inouye SK. Patterns of diffusion of evidence-based clinical programmes: a case study of the Hospital Elder Life Program. *Qual Saf Health Care* 2006;**15**:334–8. <https://doi.org/10.1136/qshc.2006.018820>
130. van de Steeg L, Langelaan M, Ijkema R, Nugus P, Wagner C. Improving delirium care for hospitalized older patients. A qualitative study identifying barriers to guideline adherence. *J Eval Clin Pract* 2014;**20**:813–19. <https://doi.org/10.1111/jep.12229>
131. Spradley JP. *Participant Observation*. New York, NY: Holt, Rinehart and Winston; 1980.
132. Lofland J. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. 4th edn. Belmont, CA: Wadsworth/Thomson Learning; 2006.
133. Emerson RM, Fretz RI, Shaw LL. *Writing Ethnographic Fieldnotes*. Chicago, IL: University of Chicago Press; 1995. <https://doi.org/10.7208/chicago/9780226206851.001.0001>
134. Charmaz K. *Constructing Grounded Theory*. Thousand Oaks, CA: Sage; 2014.
135. Barron EA, Holmes J. Delirium within the emergency care setting, occurrence and detection: a systematic review. *Emerg Med J* 2013;**30**:263–8. <https://doi.org/10.1136/emered-2011-200586>
136. LaMantia MA, Messina FC, Hobgood CD, Miller DK. Screening for delirium in the emergency department: a systematic review. *Ann Emerg Med* 2014;**63**:551–60.e2. <https://doi.org/10.1016/j.annemergmed.2013.11.010>
137. Lewis LM, Miller DK, Morley JE, Nork MJ, Lasater LC. Unrecognized delirium in ED geriatric patients. *Am J Emerg Med* 1995;**13**:142–5. [https://doi.org/10.1016/0735-6757\(95\)90080-2](https://doi.org/10.1016/0735-6757(95)90080-2)
138. Ryan DJ, O'Regan NA, Caoimh RÓ, Clare J, O'Connor M, Leonard M, et al. Delirium in an adult acute hospital population: predictors, prevalence and detection. *BMJ Open* 2013;**3**:e001772. <https://doi.org/10.1136/bmjopen-2012-001772>

139. Department of Health and Social Care. *Living Well With Dementia: A National Dementia Strategy*. London: Department of Health and Social Care; 2009. URL: www.gov.uk/government/publications/living-well-with-dementia-a-national-dementia-strategy (accessed 10 July 2016).
140. NHS Wales. *1000 Lives +: Improving Dementia Care*. 2010. URL: www.1000livesplus.wales.nhs.uk/mh-dementia (accessed 1 December 2017).
141. Scottish Government. *Scotland's National Dementia Strategy 2013–2016*. URL: www.gov.scot/Topics/Health/Services/Mental-Health/Dementia (accessed 10 July 2016).
142. Royal College of Psychiatrists. *Report of the National Audit of Dementia Care in General Hospitals*. London: Royal College of Psychiatrists; 2011.
143. Young J, Hood C, Gandesha A, Souza R. *National Audit of Dementia Care in General Hospitals 2012–2013: Second Round Audit Report and Update*. London: Healthcare Quality Improvement Partnership; 2013.
144. Godfrey M, Young J, Shannon R, Skingley A, Woolley R, Arrojo F. The Person, Interactions and Environment Programme to improve care of people with dementia in hospital: a multisite study. *Health Serv Deliv Res* 2018;**6**(23).
145. Traynor V, Cordato N, Burns P, Xu Y, Britten N, Duncan K, et al. Is delirium being detected in emergency? *Australas J Ageing* 2016;**35**:54–7. <https://doi.org/10.1111/ajag.12255>
146. Shi Q, Warren L, Saposnik G, Macdermid JC. Confusion assessment method: a systematic review and meta-analysis of diagnostic accuracy. *Neuropsychiatr Dis Treat* 2013;**9**:1359–70. <https://doi.org/10.2147/NDT.S49520>
147. Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry* 2000;**15**:1021–7. [https://doi.org/10.1002/1099-1166\(200011\)15:11<1021::AID-GPS234>3.0.CO;2-6](https://doi.org/10.1002/1099-1166(200011)15:11<1021::AID-GPS234>3.0.CO;2-6)
148. Hospital Elder Life Program. *Delirium Instruments*. URL: www.hospitalelderlifeprogram.org/delirium-instruments (accessed 1 September 2017).
149. Rolfson DB, McElhane JE, Jhangri GS, Rockwood K. Validity of the confusion assessment method in detecting postoperative delirium in the elderly. *Int Psychogeriatr* 1999;**11**:431–8. <https://doi.org/10.1017/S1041610299006043>
150. Pompei P, Foreman M, Cassel CK, Alessi C, Cox D. Detecting delirium among hospitalized older patients. *Arch Intern Med* 1995;**155**:301–7. <https://doi.org/10.1001/archinte.1995.00430030095011>
151. Rockwood K, Cosway S, Stolee P, Kydd D, Carver D, Jarrett P, O'Brien B. Increasing the recognition of delirium in elderly patients. *J Am Geriatr Soc* 1994;**42**:252–6. <https://doi.org/10.1111/j.1532-5415.1994.tb01747.x>
152. Aakerlund LP, Rosenberg J. Postoperative delirium: treatment with supplementary oxygen. *Br J Anaesth* 1994;**72**:286–90. <https://doi.org/10.1093/bja/72.3.286>
153. Radtke FM, Franck M, Schneider M, Luetz A, Seeling M, Heinz A, et al. Comparison of three scores to screen for delirium in the recovery room. *Br J Anaesth* 2008;**101**:338–43. <https://doi.org/10.1093/bja/aen193>
154. Inouye SK, Foreman MD, Mion LC, Katz KH, Cooney LM. Nurses' recognition of delirium and its symptoms: comparison of nurse and researcher ratings. *Arch Intern Med* 2001;**161**:2467–73. <https://doi.org/10.1001/archinte.161.20.2467>
155. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS – towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010;**81**:932–7. <https://doi.org/10.1016/j.resuscitation.2010.04.014>

156. Inouye SK. *The Short Confusion Assessment Method (Short CAM): Training Manual and Coding Guide*. Boston, MA: Hospital Elder Life Program; 2014.
157. Simon SE, Bergmann MA, Jones RN, Murphy KM, Orav EJ, Marcantonio ER. Reliability of a structured assessment for nonclinicians to detect delirium among new admissions to postacute care. *J Am Med Dir Assoc* 2006;**7**:412–15. <https://doi.org/10.1016/j.jamda.2006.02.006>
158. Infante MT, Pardini M, Balestrino M, Finocchi C, Malfatto L, Bellelli G, *et al.* Delirium in the acute phase after stroke: comparison between methods of detection. *Neurol Sci* 2017;**38**:1101–4. <https://doi.org/10.1007/s10072-017-2832-x>
159. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;**6**:e012799. <https://doi.org/10.1136/bmjopen-2016-012799>
160. Shenkin SD, Fox C, Godfrey M, Siddiqi N, Goodacre S, Young J, *et al.* Protocol for validation of the 4AT, a rapid screening tool for delirium: a multicentre prospective diagnostic test accuracy study. *BMJ Open* 2018;**8**:e015572. <https://doi.org/10.1136/bmjopen-2016-015572>
161. 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**:34–7. <https://doi.org/10.1111/j.1365-2044.2004.03526.x>
162. Great Britain. *Mental Capacity Act 2005*. London: The Stationery Office; 2005.
163. Great Britain. *Adults with Incapacity (Scotland) Act 2000*. London: The Stationery Office; 2000.
164. Trzepacz PT, Mittal D, Torres R, Canary K, Norton J, Jimerson N. Validation of the Delirium Rating Scale-revised-98: comparison with the delirium rating scale and the cognitive test for delirium. *J Neuropsychiatry Clin Neurosci* 2001;**13**:229–42. <https://doi.org/10.1176/jnp.13.2.229>
165. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal PV, Keane KA, *et al.* The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med* 2002;**166**:1338–44. <https://doi.org/10.1164/rccm.2107138>
166. Tiegies Z, Stioibhairt A, Scott K, Suchorab K, Weir A, Parks S, *et al.* Development of a smartphone application for the objective detection of attentional deficits in delirium. *Int Psychogeriatr* 2015;**27**:1251–62. <https://doi.org/10.1017/S1041610215000186>
167. Hart RP, Levenson JL, Sessler CN, Best AM, Schwartz SM, Rutherford LE. Validation of a cognitive test for delirium in medical ICU patients. *Psychosomatics* 1996;**37**:533–46. [https://doi.org/10.1016/S0033-3182\(96\)71517-7](https://doi.org/10.1016/S0033-3182(96)71517-7)
168. Meagher DJ, Morandi A, Inouye SK, Ely W, Adamis D, MacLulich AJ, *et al.* Concordance between DSM-IV and DSM-5 criteria for delirium diagnosis in a pooled database of 768 prospectively evaluated patients using the delirium rating scale-revised-98. *BMC Med* 2014;**12**:164. <https://doi.org/10.1186/s12916-014-0164-8>
169. Davis DH, Barnes LE, Stephan BC, MacLulich AM, Meagher D, Copeland J, *et al.* The descriptive epidemiology of delirium symptoms in a large population-based cohort study: results from the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). *BMC Geriatr* 2014;**14**:87. <https://doi.org/10.1186/1471-2318-14-87>
170. Swain DG, O'Brien AG, Nightingale PG. Cognitive assessment in elderly patients admitted to hospital: the relationship between the shortened version of the Abbreviated Mental Test and the Abbreviated Mental Test and Mini-Mental State Examination. *Clin Rehabil* 2000;**14**:608–10. <https://doi.org/10.1191/0269215500cr368oa>
171. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd (revised) edn. Washington, DC: American Psychiatric Association; 1987.

172. Jorm AF. A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. *Psychol Med* 1994;**24**:145–53. <https://doi.org/10.1017/S003329170002691X>
173. Jackson TA, MacLulich AM, Gladman JR, Lord JM, Sheehan B. Diagnostic test accuracy of informant-based tools to diagnose dementia in older hospital patients with delirium: a prospective cohort study. *Age Ageing* 2016;**45**:505–11. <https://doi.org/10.1093/ageing/afw065>
174. Krogseth M, Wyller TB, Engedal K, Juliebø V. Delirium is an important predictor of incident dementia among elderly hip fracture patients. *Dement Geriatr Cogn Disord* 2011;**31**:63–70. <https://doi.org/10.1159/000322591>
175. Neerland BE, Hall RJ, Seljeflot I, Frihagen F, MacLulich AM, Raeder J, et al. Associations between delirium and preoperative cerebrospinal fluid C-reactive protein, interleukin-6, and interleukin-6 receptor in individuals with acute hip fracture. *J Am Geriatr Soc* 2016;**64**:1456–63. <https://doi.org/10.1111/jgs.14238>
176. Morandi A, Han JH, Meagher D, Vasilevskis E, Cerejeira J, Hasemann W, et al. Detecting delirium superimposed on dementia: evaluation of the diagnostic performance of the Richmond Agitation and Sedation Scale. *J Am Med Dir Assoc* 2016;**17**:828–33. <https://doi.org/10.1016/j.jamda.2016.05.010>
177. Jorm AF. The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a review. *Int Psychogeriatr* 2004;**16**:275–93. <https://doi.org/10.1017/S1041610204000390>
178. 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**:867–72. <https://doi.org/10.1111/j.1524-4733.2010.00788.x>
179. Marcantonio ER, Ngo LH, O'Connor M, Jones RN, Crane PK, Metzger ED, Inouye SK. 3D-CAM: derivation and validation of a 3-minute diagnostic interview for CAM-defined delirium: a cross-sectional diagnostic test study. *Ann Intern Med* 2014;**161**:554–61. <https://doi.org/10.7326/M14-0865>
180. Neufeld KJ, Nelliot A, Inouye SK, Ely EW, Bienvenu OJ, Lee HB, Needham DM. Delirium diagnosis methodology used in research: a survey-based study. *Am J Geriatr Psychiatry* 2014;**22**:1513–21. <https://doi.org/10.1016/j.jagp.2014.03.003>
181. Neufeld KJ. Delirium classification by the Diagnostic and Statistical Manual – a moving target. *Int Psychogeriatr* 2015;**27**:881–2. <https://doi.org/10.1017/S1041610215000502>
182. Sepulveda E, Franco JG, Trzepacz PT, Gaviria AM, Meagher DJ, Palma J, et al. Delirium diagnosis defined by cluster analysis of symptoms versus diagnosis by DSM and ICD criteria: diagnostic accuracy study. *BMC Psychiatry* 2016;**16**:167. <https://doi.org/10.1186/s12888-016-0878-6>
183. Sepulveda E, Franco JG, Trzepacz PT, Gaviria AM, Viñuelas E, Palma J, et al. Performance of the Delirium Rating Scale-Revised-98 against different delirium diagnostic criteria in a population with a high prevalence of dementia. *Psychosomatics* 2015;**56**:530–41. <https://doi.org/10.1016/j.psych.2015.03.005>
184. Laurila JV, Pitkala KH, Strandberg TE, Tilvis RS. Impact of different diagnostic criteria on prognosis of delirium: a prospective study. *Dement Geriatr Cogn Disord* 2004;**18**:240–4. <https://doi.org/10.1159/000080022>
185. Hare M, Wynaden D, McGowan S, Speed G. Assessing cognition in elderly patients presenting to the emergency department. *Int Emerg Nurs* 2008;**16**:73–9. <https://doi.org/10.1016/j.ienj.2008.01.005>
186. Schofield I, Stott DJ, Tolson D, McFadyen A, Monaghan J, Nelson D. Screening for cognitive impairment in older people attending accident and emergency using the 4-item Abbreviated Mental Test. *Eur J Emerg Med* 2010;**17**:340–2. <https://doi.org/10.1097/MEJ.0b013e32833777ab>

187. Locke T, Keat S, Tate M, Bown A, Hart A, Ghosh R. Assessing the performance of the four question abbreviated mental test in the acute geriatric setting. *Acute Med* 2013;**12**:13–17.
188. Dyer AH, Briggs R, Nabeel S, O'Neill D, Kennelly SP. The Abbreviated Mental Test 4 for cognitive screening of older adults presenting to the emergency department. *Eur J Emerg Med* 2017;**24**:417–22. <https://doi.org/10.1097/MEJ.0000000000000394>
189. Inouye SK, Zhang Y, Han L, Leo-Summers L, Jones R, Marcantonio E. Recoverable cognitive dysfunction at hospital admission in older persons during acute illness. *J Gen Intern Med* 2006;**21**:1276–81. <https://doi.org/10.1111/j.1525-1497.2006.00613.x>
190. Breitbart W, Rosenfeld B, Roth A, Smith MJ, Cohen K, Passik S. The Memorial Delirium Assessment Scale. *J Pain Symptom Manage* 1997;**13**:128–37. [https://doi.org/10.1016/S0885-3924\(96\)00316-8](https://doi.org/10.1016/S0885-3924(96)00316-8)
191. Inouye SK, Kosar CM, Tommet D, Schmitt EM, Puelle MR, Saczynski JS, *et al*. The CAM-S: development and validation of a new scoring system for delirium severity in 2 cohorts. *Ann Intern Med* 2014;**160**:526–33. <https://doi.org/10.7326/M13-1927>
192. Han JH, Wilson A, Ely EW. Delirium in the older emergency department patient: a quiet epidemic. *Emerg Med Clin North Am* 2010;**28**:611–31. <https://doi.org/10.1016/j.emc.2010.03.005>
193. Han JH, Wilson A, Vasilevskis EE, Shintani A, Schnelle JF, Dittus RS, *et al*. Diagnosing delirium in older emergency department patients: validity and reliability of the delirium triage screen and the brief confusion assessment method. *Ann Emerg Med* 2013;**62**:457–65. <https://doi.org/10.1016/j.annemergmed.2013.05.003>
194. Holmes J, House A. Psychiatric illness predicts poor outcome after surgery for hip fracture: a prospective cohort study. *Psychol Med* 2000;**30**:921–9. <https://doi.org/10.1017/S0033291799002548>
195. Thomason JW, Shintani A, Peterson JF, Pun BT, Jackson JC, Ely EW. Intensive care unit delirium is an independent predictor of longer hospital stay: a prospective analysis of 261 non-ventilated patients. *Crit Care* 2005;**9**:R375–81. <https://doi.org/10.1186/cc3729>
196. Bourdel-Marchasson I, Vincent S, Germain C, Salles N, Jenn J, Rasoamanarivo E, *et al*. Delirium symptoms and low dietary intake in older inpatients are independent predictors of institutionalization: a 1-year prospective population-based study. *J Gerontol A Biol Sci Med Sci* 2004;**59**:350–4. <https://doi.org/10.1093/gerona/59.4.M350>
197. Zywiell MG, Hurley RT, Perruccio AV, Hancock-Howard RL, Coyte PC, Rampersaud YR. Health economic implications of perioperative delirium in older patients after surgery for a fragility hip fracture. *J Bone Joint Surg Am* 2015;**97**:829–36. <https://doi.org/10.2106/JBJS.N.00724>
198. Brown CH, Laflam A, Max L, Lyman D, Neufeld KJ, Tian J, *et al*. The impact of delirium after cardiac surgical procedures on postoperative resource use. *Ann Thorac Surg* 2016;**101**:1663–9. <https://doi.org/10.1016/j.athoracsur.2015.12.074>
199. Zaubler TS, Murphy K, Rizzuto L, Santos R, Skotzko C, Giordano J, *et al*. Quality improvement and cost savings with multicomponent delirium interventions: replication of the Hospital Elder Life Program in a community hospital. *Psychosomatics* 2013;**54**:219–26. <https://doi.org/10.1016/j.psym.2013.01.010>
200. ISD. *Scottish Health Service Costs*. URL: www.isdscotland.org/Health-Topics/Finance/Costs/ (accessed 1 September 2017).
201. Curtis L, Burns A. *Unit Costs of Health and Social Care 2016*. Canterbury: PSSRU, University of Kent; 2016.

202. Department of Health and Social Care. *NHS Reference Costs 2015 to 2016*. London: Department of Health and Social Care; 2016. URL: www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016 (accessed 1 September 2017).
203. NHS Improvement. *Reference Costs*. 2017. URL: <https://improvement.nhs.uk/resources/reference-costs/> (accessed 1 March 2019).
204. Langan C, Sarode DP, Russ TC, Shenkin SD, Carson A, Maclullich AMJ. Psychiatric symptomatology after delirium: a systematic review. *Psychogeriatrics* 2017;**17**:327–35. <https://doi.org/10.1111/psyg.12240>
205. Wade D, Hardy R, Howell D, Mythen M. Identifying clinical and acute psychological risk factors for PTSD after critical care: a systematic review. *Minerva Anestesiol* 2013;**79**:944–63.
206. Leslie DL, Inouye SK. The importance of delirium: economic and societal costs. *J Am Geriatr Soc* 2011;**59**(Suppl. 2):241–3. <https://doi.org/10.1111/j.1532-5415.2011.03671.x>
207. National Institute for Health and Care Excellence (NICE). *The Guidelines Manual: 7 Assessing Cost Effectiveness*. London: NICE. URL: www.nice.org.uk/process/pmg6/chapter/assessing-cost-effectiveness (accessed 13 December 2017).
208. European Delirium A, American Delirium S. The DSM-5 criteria, level of arousal and delirium diagnosis: inclusiveness is safer. *BMC Med* 2014;**12**:141. <https://doi.org/10.1186/s12916-014-0141-2>
209. Fick DM, Inouye SK, Guess J, Ngo LH, Jones RN, Saczynski JS, Marcantonio ER. Preliminary development of an ultrabrief two-item bedside test for delirium. *J Hosp Med* 2015;**10**:645–50. <https://doi.org/10.1002/jhm.2418>
210. Hargrave A, Bastiaens J, Bourgeois JA, Neuhaus J, Josephson SA, Chinn J, *et al*. Validation of a nurse-based delirium-screening tool for hospitalized patients. *Psychosomatics* 2017;**58**:594–603. <https://doi.org/10.1016/j.psych.2017.05.005>
211. Bond P, Goudie K. Identifying and managing patients with delirium in acute care settings. *Nurs Older People* 2015;**27**:28–32. <https://doi.org/10.7748/nop.27.9.28.s19>
212. NHS England and Monitor. *National Tariff Information Workbook 2014/15*. 2013. URL: www.gov.uk/government/publications/national-tariff-information-workbook-201415 (accessed 1 March 2019).

Appendix 1 Survey A

Survey A

Delirium - National Survey of UK Healthcare Practitioners

Survey Legend

* = Mandatory question

○ = One response only

□ = Multiple responses accepted

R = Responses randomised

**Need to finish off which ones were * and R
and R(excluding bottom option)**

We greatly appreciate your interest!

Who are the intended respondents?

This survey is intended for UK-based doctors, nurses, physiotherapists and occupational therapists that come into contact with adult patients with delirium (or 'acute confusional state') as part of their routine work.

Please do not fill in this survey if you are not a UK health practitioner.

Why this survey?

There is some variation among healthcare practitioners regarding how delirium is detected and assessed. As part of our efforts to improve delirium detection and assessment, we are interested in finding out more about these variations.

What is delirium?

Broadly speaking, delirium is an acute deterioration in cognition or other mental functions caused by an acute medical problem, drug side-effects or other acute causes.

What does the survey cover?

Attitudes and knowledge around delirium detection and assessment.

Who is funding this survey?

The UK National Institute for Health Research, Health Technology Assessment programme.
Link to NIHR webpage on the study.

Which institutions are involved?

Team members are employed by the NHS (Edinburgh, Sheffield, Bradford, Leeds), the University of Edinburgh, the University of Sheffield, the University of Leeds and the University of East Anglia.

Are my responses anonymous?

The survey is completely anonymised. No personal identifiable information will be collected as part of the survey. No computer location information will be collected. There are some broad questions on your professional background, the type of setting you work in, and the country in which you work.

The results of this survey will be published and some completely anonymised responses may be quoted in these publications.

How long will the survey take?

10-15 minutes.

If you are interested in helping with the survey, please read the first question below and if appropriate, click 'I agree':

Do you agree to participate in this survey and consent to the potential use of your anonymised responses as described above? *

I agree

I do not agree

Asterisks (*) are shown beside questions that require an answer to allow the survey to proceed.

For those working in community-based palliative care, please refer to your work settings where the survey uses the term 'unit'.

Please click Next to begin.

Which clinical setting(s) do you currently work in? (primarily, but tick more than one option if your time is divided)

Emergency department

Acute assessment / medical assessment unit

Intensive care

Medicine of the elderly

Orthopaedics

Rehabilitation ward

Oncology

Stroke

Hospice / Palliative care

Internal medicine specialist ward (e.g. cardiology, respiratory, gastroenterology, nephrology, endocrinology, neurology)

Old age mental health ward

Liaison mental health

Surgical ward (not including orthopaedics)

Other (please specify)

What nation do you work in? *R

England

Northern Ireland

Scotland

Wales

Other (please specify)

How many years have you been working (approximate full-time equivalent) since obtaining your primary professional qualification?**Which field is your primary professional qualification in? *R**

Medicine

Nursing

Occupational therapy

Physiotherapy

Other (please specify)

===== Begin filter: For respondents who chose 'Medicine' in the previous question =====

Which stage of your career are you at? *R

FY1 – 2

ST1 – 2

ST3+

Consultant

Specialist doctor

Associate specialist

GP

Other (please specify)

Which of the following specialties are you trained in (i.e. to CCT/CCST level) or undergoing training in?

General Practice

Surgical specialty

Intensive Care

Palliative Medicine

Emergency Medicine

General (Internal) Medicine

Geriatric Medicine

General Adult Psychiatry

Anaesthetics

Old Age Psychiatry

Acute Medicine

Liaison Psychiatry

Other (please specify)

The following four questions concern various aspects of delirium assessment. Please answer according to your own views. These questions are not intended to test knowledge but rather to understand more about existing variations in clinical opinion.

In your opinion, in order to allow bedside assessment for delirium, does a patient need to have a sufficiently high level of consciousness to produce verbal responses?

Yes

No

Don't know

If a patient presents with recent onset drowsiness and is not producing verbal responses, but is responding intermittently to one-stage commands, in practice which of the following terms would you most likely use to describe this state?

Coma

Obtundation

Delirium

Stupor

Encephalopathy

Other (please specify)

A patient has recent-onset drowsiness and is not producing verbal responses, but is responding intermittently to one-stage commands. In your opinion, how likely is it that the patient has delirium?

Very unlikely 0 – 20%	Unlikely 21 – 40%	Neither likely nor unlikely 41 – 60%	Likely 61 – 80%	Very Likely 81 – 100%

Please answer the following questions using the 5-point scale provided.

	Never/rarely (0 – 20%)	Sometimes (21 – 40%)	About half of the time (41 – 60%)	Frequently (61 – 80%)	Almost always/ always (81 – 100%)
In patients with cognitive impairment, how often do you seek a history of mental status changes from collateral sources (e.g. family, GP, etc.)?					
When you have detected delirium, how often do you record this using the term "delirium" in the case notes?					

===== End filter: For respondents who chose 'Medicine' in earlier question =====

In your opinion, of all the cases of delirium in your unit, what percentage of patients have their delirium diagnosed and documented?

0 – 20%	21 – 40%	41 – 60%	61 – 80%	81 – 100%	(I don't know)

Please indicate the extent to which you think the following potential causes contribute to under-detection of delirium in your unit.

	Very small	Small	Moderate	Large	Very large	(Don't know)
Lack of consensus on which screening tools/criteria are used						
Detection is considered the duty of certain healthcare practitioners only						
Lack of staff knowledge of delirium						
Time constraints						
Delirium detection not part of routine assessment						
Difficulty in discriminating delirium from dementia						
Lack of staff confidence in assessment						

Please list anything else you think contributes to under-detection of delirium in your unit, and the extent to which you think they contribute.

Please indicate how often the following terms are used in your unit to describe patients with acute deterioration in cognition or other mental functions caused by an acute medical problem, drug side-effects or other acute causes.

	Never/rarely (0 – 20%)	Sometimes (21 – 40%)	About half of the time (41 – 60%)	Frequently (61 – 80%)	Almost always/ always (81 – 100%)
Septic encephalopathy					
Delirium					
Confusion					
Acute confusional state / Acute confusion					

Please list any other terms and how often they are used in your unit to describe the clinical syndrome described above.

Are there guidelines in place for any of the following in your unit?

	No	Yes
Delirium detection		
Delirium management		

In your opinion, how often are these guidelines followed?

	Never/rarely (0 – 20%)	Sometimes (21 – 40%)	About half of the time (41 – 60%)	Frequently (61 – 80%)	Almost always/ always (81 – 100%)	Don't know	N/A
Delirium detection							
Delirium management							

Add comments here.

Are there routine audits of any of the following in your unit?

	No	Yes
Delirium detection		
Delirium management		

Do you think there has been an increase in awareness of delirium among colleagues in your specialty in the last three years?

Yes

No

Don't know

N/A (if qualified less than three years)

Do you think there has been an increase in how often delirium is mentioned in the following domains, in the last three years?

	No	Yes	Don't know
Clinical Journals			
Twitter			
Facebook			
Clinically-related websites			
Professional conferences			
Training events			
General media coverage (e.g. BBC, newspapers)			
Training curriculum			

Any comments or additional domains?

What is your level of confidence in your own ability to detect delirium? *

(Can't decide)	Very low	Low	Moderate	High	Very High

Have you ever used a tool to detect delirium? (i.e. a specific assessment tool or cognitive test)

No

Yes

Please rank the following tools/criteria in order of how often you have used them to detect delirium in your clinical practice.

Drag-and-drop the options below into the appropriate order, where 1 = most often.

Please mark the items that you have never used as 'N/A' using the option on the right; this could apply to all the items listed

Confusion Assessment Method (CAM)	N/A
Delirium Observation Scale (DOS)	N/A
Confusion Assessment Method – Intensive Care Unit (CAM-ICU)	N/A
Nursing Delirium Scale (NU-DESC)	N/A
Delirium Rating Scale (DRS/DRS-R-98)	N/A
4AT	N/A
Other 1 (please give details below)	N/A
Other 2 (please give details below)	N/A

Many more delirium assessment tools exist. Please add any other tools that you have used.

When you are assessing a patient for possible delirium, how often do you use bedside cognitive testing?

Never/rarely (0 – 20%)	Sometimes (21 – 40%)	About half of the time (41 – 60%)	Frequently (61 – 80%)	Almost always/ always (81 – 100%)

When you use a cognitive test as part of assessment for possible delirium, which test(s) do you use?

Drag-and-drop the options below into the appropriate order, where 1 = most often .

Please mark the items that you have never used as 'N/A' using the option on the right; this could apply to all the items listed

Abbreviated Mental Test – 10 item	N/A
Abbreviated Mental Test – 4 item	N/A
Months of the year backwards	N/A
Days of the week backwards	N/A
Counting from 20 down to 1	N/A
Serial 7s	N/A
Digit span	N/A
Orientation to time, place, person	N/A
Vigilance test (e.g. SAVEAHEART)	N/A
Mini-Mental State Examination	N/A
Montreal Cognitive Assessment (MoCA)	N/A
Other 1 (please give details below)	N/A
Other 2 (please give details below)	N/A

Please add any other cognitive tests that you have used.

Please briefly describe your approach if/when you do not use a delirium assessment tool or cognitive test to detect delirium.

Please indicate your level of agreement with the following statements.

	Strongly disagree	Disagree	Neither disagree nor disagree	Agree	Strongly agree
Making a formal diagnosis of delirium is important to provide good delirium care					
Delirium treatment improves patient outcomes					
Distinguishing between delirium and dementia is important in providing good care					

In your opinion, which of the following tasks should each healthcare practitioner consider their duty in regards to delirium detection?

	Flagging potential cases	Screening high risk patients	Making a formal diagnosis
Doctors			
Nurses			
Physiotherapists			
Occupational Therapists			

Please use this space to comment on any of the issues raised in this survey or additional issues surrounding the detection and assessment of delirium that have not been addressed.

We may wish to contact some participants to ask them to expand on some of their answers. If you're happy to share another few minutes of your time please enter your name and email address below. (optional)

By providing contact details your results will no longer be anonymous; however this information will only be used to contact you and will not be used for any other purpose.

If you do not wish to provide contact details please press next.

Name	<input type="text"/>
Email address	<input type="text"/>

Your responses have been submitted

Thank you - your time and effort is very much appreciated!

This survey is part of a larger study which aims to develop and validate the 4AT - a new triage tool for delirium - developed at the University of Edinburgh.

We would be very grateful if you could help us reach more UK health practitioners.

Please consider:

- Forwarding the email you received to relevant medical, nursing, physiotherapy and occupational therapy staff in your department, Trust/Health board and contacts list. (e.g. Emergency department, ICU, AAU/MAU, Elderly Care, Trauma & Orthopaedics , Oncology, Stroke services, Palliative care/hospices).
- Displaying an A4 poster highlighting the survey in your staff area (simply email astiobha@staffmail.ed.ac.uk for a copy)

Clicking 'Done' will bring you to www.the4at.com/ where you will find more information about the 4AT.

Appendix 2 Survey B

Survey Legend

* = Mandatory question

○ = One response only

□ = Multiple responses accepted

R = Responses randomised

RL = Responses randomised except for bottom option

The 4AT Delirium Assessment Tool: An International Survey of its Use in Clinical Practice

We greatly appreciate your interest!

Who are the intended respondents?

This survey is intended for healthcare practitioners worldwide that use the 4AT to screen patients for delirium (or 'acute confusional state').

Please do not fill in this survey if you have never used the 4AT.

We recently carried out a general survey relating to delirium detection and assessment in the UK. Practitioners who completed that survey are also invited to complete this one. We are very grateful for the support we have received so far.

Why this survey?

As part of our efforts to improve delirium detection and assessment, we plan to carry out a large-scale validation study of the 4AT, but first we would like to find out more about how it is currently used clinically and how it may be improved.

What does the survey cover?

Current practice surrounding the use of delirium screening tools, particularly the 4AT, and your opinions of the 4AT itself.

Who is funding this survey?

The UK National Institute for Health Research, Health Technology Assessment programme. Click [here](#) to see the NIHR webpage on the study.

Which institutions are involved?

Team members are employed by the NHS (Edinburgh, Sheffield, Bradford, Leeds), the University of Edinburgh, the University of Sheffield, the University of Leeds and the University of East Anglia.

Are my responses anonymous?

The survey is completely anonymised. No personal identifiable information will be collected as part of the survey. No computer location information will be collected. There are some broad questions on your professional background and the type of setting you work in.

The results of this survey will be submitted as part of a report and published in academic journals. Some completely anonymised responses may be quoted in these publications.

How long will the survey take?

About 10 minutes.

We highly recommend having a copy of the 4AT in front of you while completing this survey. Please click [here](http://www.the4at.com) to view or download a copy from www.the4at.com

Do you agree to participate in this survey and consent to the potential use of your anonymised responses as described above?

I agree

I do not agree

Asterisks (*) are shown beside questions that require an answer to allow the survey to proceed.

Please click Next to begin.

Which clinical setting(s) do you currently work in? (primarily, but tick more than one option if your time is divided) * *RL*

Emergency department

Acute assessment / medical assessment unit

Intensive care

Medicine of the elderly

Orthopaedics

Rehabilitation ward

Oncology

Stroke

Hospice / Palliative care

Internal medicine specialist ward (e.g. cardiology, respiratory, gastroenterology, nephrology, endocrinology, neurology)

Old age mental health ward

Liaison mental health

Surgical ward (not including orthopaedics)

Other (please specify)

Which nation do you work in? * *RL*

England

Northern Ireland

Scotland

Wales

Other (please specify)

How many years have you been working (approximate full-time equivalent) since obtaining your primary professional qualification? ***Which field is your primary professional qualification in? * *R***

Medicine

Nursing

Occupational therapy

Physiotherapy

Other (please specify)

Which stage of your career are you at? * RL

FY1 – 2

ST1 – 2

ST3+

Consultant

Specialist doctor

Associate specialist

Other (please specify)

Which of the following specialties are you trained in (i.e. to CCT/CCST level) or undergoing training in? * R

Geriatric Medicine

Acute Medicine

Emergency Medicine

Anaesthetics

Intensive Care

Palliative Medicine

Old Age Psychiatry

Liaison Psychiatry

General Adult Psychiatry

General Practice

Surgical specialty

General (Internal) Medicine

Other (please specify)

What is your level of confidence in your own ability to detect delirium? *

(Can't decide)	Very low	Low	Moderate	High	Very High

Please answer the following questions using the 5-point scale provided

	Never / rarely (0 – 20%)	Sometimes (21 – 40%)	About half of the time (41 – 60%)	Frequently (61 – 80%)	Almost always / always (81 – 100%)
How often do you assess patients for delirium?					
When assessing for delirium, how often do you use a scale or other tool?					

Please rank the following tools/criteria in order of how often you have used them to detect delirium in your clinical practice.

Drag-and-drop the options below into the appropriate order, where 1 = most often .

Please mark the items that you have never used as 'N/A' using the option on the right; this could apply to all the items listed

4AT	N/A
Confusion Assessment Method (CAM)	N/A
Confusion Assessment Method – Intensive Care Unit (CAM-ICU)	N/A
Nursing Delirium Scale (NU-DESC)	N/A
Delirium Rating Scale (DRS/DRS-R-98)	N/A
Delirium Observation Scale (DOS)	N/A
Other 1 (please give details below)	N/A
Other 2 (please give details below)	N/A

Many more delirium assessment tools exist. Please add any other tools that you have used.

Approximately how long has it been since you first used the 4AT?

<1 month

1 – 6 months

7 – 12 months

>1 year

Please add any comments here.

How often do you use the 4AT in patients at risk of delirium?

Never/rarely (0 – 20%)	Sometimes (21 – 40%)	About half of the time (41 – 60%)	Frequently (61 – 80%)	Almost always/ always (81 – 100%)

When using a tool to screen for delirium, what factors influence your decision to use the 4AT instead of another tool?

When using a tool to screen for delirium, what factors influence your decision to use another tool instead of the 4AT?

Is the 4AT used as part of routine assessment by you or others in your unit?

No

Yes

Please add any comments here.

Approximately what proportion of patients in your unit undergo the 4AT?

(Don't know)	0 – 20%	21 – 40%	41 – 60%	61 – 80%	81 – 100%

Do you think that use of the 4AT as part of routine assessment is feasible in your unit?

No

Yes

Please explain why you think so.

To what extent are the following necessary in order for healthcare practitioners to use the 4AT effectively?

	(Don't know)	None / very little	Some	A moderate amount	Quite a bit	An extensive amount
Knowledge of delirium						
Training in the use of the 4AT						

Please indicate the extent to which you think the following potential barriers prevent the 4AT from being used more regularly in your unit.

	Very small	Small	Moderate	Large	Very large
Time constraints					
Existing use of / familiarity with an alternative tool					
Lack of staff confidence in using the tool					
Lack of perceived need to use a delirium screening tool					
Lack of staff knowledge of delirium					

Are there any other factors which prevent the 4AT from being used more regularly? Please state why and the extent to which they contribute.

Please mark the responses below that best reflect your typical experience in using the 4AT with the following groups.

If you have never used the 4AT on a particular group of patients choose 'N/A'. *R*

	Very easy	Easy	Neither easy nor difficult	Difficult	Very Difficult	N/A
Drowsy patients who can not produce verbal responses						
Patients with dementia who are alert and able to converse						
Patients who are agitated and distressed						

Approximately how long does it typically take you to complete all three of the bedside components of the 4AT (items 1 - 3: 'Level of alertness', 'Months backwards' and 'AMT4')?

<1 minute

1 – 2 minutes

3+ minutes

Please add any comments here

Approximately how long does it typically take you to do item 4 ('Acute change or fluctuating course')?

	Time [drop down menu with option for “<1 minute” to “10+ minutes”]
From:	_____
To:	_____

Please add any comments here

Would you suggest any changes to item 1 ('Alertness')?

Yes

No

If so please describe which changes, and why.

Would you suggest any changes to item 2 ('AMT4')?

Yes

No

If so please describe which changes, and why.

Would you suggest any changes to item 3 ('Attention')?

Yes

No

If so please describe which changes, and why.

Would you suggest any changes to item 4 ('Acute change or fluctuating course')?

Yes

No

If so please describe which changes, and why.

Would you suggest any changes to the scoring system?

Yes

No

If so please describe which changes, and why. (Is there an alternative system you could suggest that would work better?)

Would you suggest any changes to the current guidance notes?

Yes

No

If so please describe which changes, and why.

What is your general opinion of the 4AT?

Broadly speaking, how has the 4AT been received by your colleagues?

	(Don't know)	Very Poorly	Poorly	Neutral	Well	Very well
Medical						
Nursing						
Other (if applicable, specify below)						

Why do you think this is so?

Please use this space to comment on any of the issues raised in this survey or any concerns that have arisen through use of the 4AT.

We may wish to contact some participants to ask them to expand on some of their answers. If you're happy to share another few minutes of your time please enter your name and email address below. (optional)

By providing contact details your results will no longer be anonymous; however this information will only be available to members of the research team and will only be used to contact you in connection with this survey and will not be used for any other purpose.

If you do not wish to provide contact details please press next.

Name


Email
address

Your responses have been submitted

Thank you - your time and effort is very much appreciated!

We would be very grateful if you could help us reach more 4AT users.

Please consider:

- **Forwarding the email you received to 4AT users in your department, Trust/Health board and contacts list, particularly junior doctors and nurses.**
- **Displaying an A4 poster highlighting the survey in your staff area (simply email  for a copy)**

Clicking 'Done' will bring you to www.the4at.com/ where you will find more information about the 4AT.

Appendix 3 Survey B: respondent characteristics

TABLE 15 Survey B: respondent characteristics

Characteristic	Medicine (tertiles of years of experience)				Specialty			Total
	Lower	Mid	Upper	Overall	Nursing	Occupational therapy	Physiotherapy	
Frequency (% of sample)	17	12	35	64	32	3	1	100
Years qualified, median (IQR)	1.50 (1–2.5)	9.50 (8–11.50)	16 (12–20)	12 (4.75–16.26)	18 (10–30)	5 (3–7.5)	7 (NA)	12 (6–19.25)
Setting (n)								
Acute inpatient medical	10	11	28	49	22	2	1	74
ED	4	3	4	11	3	–	–	12
ICU	1	–	–	1	1	–	–	3
Rehabilitation unit	1	–	5	6	–	–	–	6
Surgical ward (excluding orthopaedics)	4	1	3	8	4	1	0	13
Psychiatry	–	1	3	5	3	1	0	9
Other ^a	1	–	3	4	2	–	–	6

^a Acute medicine ($n = 2$), community psychiatry ($n = 2$), general practice ($n = 1$), immediate care ($n = 1$).

All estimates relate to frequency of respondents except where stated otherwise. Cells for settings are not equal to the total sample size as respondents could choose multiple settings. Acute inpatient medical includes acute assessment unit/MAU, medicine of the elderly, orthopaedics, oncology, stroke, and internal medicine specialist ward (e.g. cardiology, respiratory, gastroenterology, nephrology, endocrinology, neurology).

Appendix 4 Survey B: assessment tools used and the frequency of use among respondents

TABLE 16 Survey B: assessment tools used and the frequency of use among respondents

Tool	Proportion used by, % (n)	Frequency of use, median rank (IQR)
4AT	100 (90/90)	1 (1–1)
CAM	67.78 (61/90)	2 (2–3)
DOS	33.33 (30/60)	3.5 (3–4)
CAM-ICU	32.22 (29/90)	3 (2–3)
DRS/DRS-R98	27.78 (25/90)	5 (4–5)
NU-DESC	20.00 (18/90)	6 (6–7)
Respondent-specified tools		
SQUID	(2)	
MDAS	(1)	
In-house tool	(1)	

CAM-ICU, Confusion Assessment Method for the ICU; DOS, Delirium Observation Scale; DRS, Delirium Rating Scale; MDAS, Memorial Delirium Assessment Scale; NU-DESC, Nursing Delirium Scale; SQUID, Single Question in Delirium. All estimates are shown as % (proportion) of the total responses for each tool. Rank of 1 = used most often. Ranking for delirium screening tools ranged from 1 to 8. 'Respondent-specified tools' are tools that were not listed in the survey as response options but were specified by participants in an open-text field that followed the question. Accurate percentages and ranks could not be generated for these tools.

Appendix 5 User opinions of the 4AT from survey B

Question 14

When respondents were asked which factors influence their decision to use the 4AT instead of another tool, several respondents reported that it is the recommended tool of use in their unit/trust ($n = 25$), and that it is easy ($n = 23$) and brief to use ($n = 20$). Respondents also reported the presence of the tool in paperwork ($n = 11$), familiarity ($n = 5$), the ease with which the items can be remembered ($n = 5$), the fact that it is free to use ($n = 2$), that it is used by others in their unit ($n = 2$), that it is based on useful clinical markers of delirium ($n = 2$) and because little training is required for staff ($n = 2$). A number of respondents referred directly to the strengths of the 4AT over alternative tools, stating that it 'takes factors other than patient questions into account'; is 'more clear for learners than CAM'; has the 'ability to quantify attentional deficit, ability to include the drowsy/untestable'; 'helps to differentiate Delirium and Dementia'; and tool users have 'Instant access online to jog memory of components'.

Question 15

Factors that influence the decision of respondents to use another tool instead of the 4AT include that a more detailed assessment was needed ($n = 11$), another tool had been recommended ($n = 8$), they were familiar with another tool ($n = 7$), they had limited time ($n = 3$) and the tests had already been carried out ($n = 2$). Respondents also stated that they sometimes specifically want the results of the CAM or will continue to use the same tool used to establish a baseline. Importantly, these comments largely refer to systemic and cultural factors rather than to the tool itself. Additional noteworthy comments include 'speed of use of AMT4', 'Lack of corroborative history', and '4AT does not discriminate well in patients with pre-existing dementia'. These comments suggest that these respondents have limited understanding of delirium diagnosis and its importance, indicating that appreciation of the 4AT might rely on a basic knowledge of delirium among those who use it.

Question 19

Part 1

When respondents were asked how much knowledge of delirium is necessary for health-care practitioners to use the 4AT effectively, 7.14% (6/84) said 'none/very little', 58.33% (49/84) said 'some' or 'a moderate amount' and 34.52% (29/84) said 'quite a bit' or 'an extensive amount'.

Part 2

When respondents were asked how much training in the use of the 4AT is necessary for health-care practitioners to use the tool effectively, 16.67% (14/84) stated 'none/very little', 59.52% (50/84) stated 'some' or 'a moderate amount' and 23.81% (20/84) stated 'quite a bit' or 'an extensive amount'. Although the 4AT was designed to require minimal knowledge of delirium and training, these results demonstrate that many 4AT users believe that at least moderate levels of both are necessary for it to be used effectively.

Question 18

When respondents were asked if they thought that use of the 4AT as part of routine assessment was feasible in their unit, a majority of 95.29% (81/85) said 'yes', with many referring to the ease (39.47%, 15/38) and

speed of use (28.95%, 11/38). One comment again draws attention to the fact that routine use relies on broader systemic factors:

It worked and improved detection when done as part of an improvement project. However there has been some difficulty embedding it in routine practice involving change of paperwork etc.

ST3+, MAU/MoE

Among the 4.71% (4/85) of respondents who stated 'no', only one explained why they thought this was the case and said that they 'already use a validated tool'. This highlights the need to explicitly state the advantages of the 4AT over alternative tools to encourage changes to clinical practice.

Questions 20 and 21

Respondents believed that the largest barrier preventing the 4AT from being used more regularly in their unit was a 'lack of perceived need to use a delirium screening tool', with 43.37% (36/83) considering this to have a 'large' or 'very large' effect. This was followed by 'lack of staff confidence in using the tool' (27.71%, 23/83), 'time constraints' (26.51%, 22/83), 'lack of staff knowledge of delirium' (24.10%, 20/83) and finally by 'existing use of/familiarity with an alternative tool' (14.46%, 12/83; see *Table 3*). These findings again suggest that delirium education is necessary for appreciation and use of the 4AT. Additional noteworthy barriers highlighted by respondents include:

Not being aware the tool is available and not having a short teaching session to explain delirium and how the tool can help highlight if a patient has changed from their baseline.

Nurse (3 years), Orthopaedics

Lack of confidence in what to do when you've made the diagnosis!

Consultant, MoE

Once completed no action seems to be taken.

Nurse (12 years), MoE

Location of 4AT in the admission pack, who is to fill it in is not clear in certain units.

FY1-2, MAU/Rehab/GP

The component items are easy to remember but the scoring is difficult to remember and so raters need to have the scoring sheet with them.

Lecturer, MoE/stroke

Clinical assessment for delirium predominates. This is likely because it is common in our unit and nurses and medical staff look out for it.

Consultant, MoE

Change in practice takes re-enforcement.

Nurse (32 years), Internal medicine specialist/surgical ward

Question 32

When respondents were asked how the 4AT had been received by their medical colleagues, only 4.92% (3/61) said 'very poorly' or 'poorly', whereas 49.18% (30/61) said 'neutral' and 45.90% (28/61) said 'well' or 'very well'. Response rates were very low to the question of how the 4AT had been received by nursing staff; however, all respondents who answered this item ($n = 9$) said 'neutral' ($n = 4$), 'well' ($n = 4$) or

'very well' ($n = 1$). When asked to explain why they thought this was so, one respondent stated 'concerns over validation'. The remaining noteworthy comments were we summed up by one respondent, who stated:

1) Yet another process to complete! 2) Lack of interest/awareness by non-geriatrician. 3) Is it unnecessary duplication? Ultimately if all patients > 75yrs have AMTs on admission we would then proceed to MMSE, MOCA +/or ACE III as require.

ST3+, MAU/MoE

Questions 31 and 32

When respondents were asked for their general opinion of the 4AT, 84.29% (59/70) had positive views, 14.29% (10/70) had neutral or mixed views and one 1.43% (1/70) had a negative view. Although those with positive views generally liked it because it was quick, easy and suitable for different inpatient environments, those who gave mixed opinions provided particularly valuable insights:

Easy to use. Not convinced as a reliable screen but better than AMTS.

Consultant, ED/MAU/MoE/Rehab/Internal medicine specialist/Surgical ward

Good, most probably easier for non-expert to use than the CAM. Unclear how it helps more experienced staff.

ST3+, MAU/MoE

Good quick bedside screen BUT some non-specialist staff do not fill in correctly, esp acute/fluct course box.

Consultant, MoE

I don't use it as clinical assessment is more reliable. I can see its use in AMU [acute medical unit] and non-MoE specialties where staff are less familiar with delirium – however it needs to lead into a care bundle to assist with management/prevention.

Consultant, MoE

I like the 4AT find it easy to use, and easy for others non specialist with moderate training. I would like to use the 4AT Trust wide for delirium but Nice Guidance appears to suggest CAM as best practice validated tool.

Nurse (18 years), MAU

It's fine – but if you make it any more complex, you risk it being ignored or worse, completed as a 'best guess' by clinicians who are already too busy to care.

GP, ED

It's pretty good. I think the major challenge is clinical staff (beyond the already converted) being aware of both the tool and the general idea of delirium as a diagnosable, treatable and important condition.

Consultant, MAU/MoE

Pragmatic, quick tool. Performs best if staff doing it have some knowledge of delirium. Also needs to be tied to some kind of action plan if the score is positive.

Consultant, MoE/Orthopaedics

Simple tool that can be used to assist in detection of delirium but should be used with clinical judgement as well.

Nurse (20 years), MoE

Question 33

Few respondents made use of an open-comments field in which they were asked to comment on any of the issues raised in this survey or any concerns that have arisen through use of the 4AT; however, a number of those who did made noteworthy comments:

The 4AT needs more publicity if it is to be implemented routinely. In a surgical setting, target junior doctors as senior clinicians have little time/interest in diagnosing delirium.

FY1 – 2, Surgical

I think it is important to keep an open mind and encourage diversity in the use of tools available for delirium assessment. I have serious concerns regarding how the 4AT is being promoted and enforced in Scotland.

Nursing (30 years), MoE

Staff forget that the 4AT is a screening instrument and should apply their clinical judgement with regards to diagnosis of delirium.

Consultant, Liaison MH

Appendix 6 Interviewees in qualitative study by hospital and department

TABLE 17 Interviewees in qualitative study by hospital and department

Hospital setting	4AT 'virgin' sites: Avonfield and Cranford (n)	4AT 'experienced' site: Denbury (n)
ED	6	3
MAU/EAU	4	6
Ward	9 (2 surgical and 7 older people wards)	15 (11 older people wards, 2 orthopaedic trauma, 1 cardio and 1 stroke)

Appendix 7 Interviewees in qualitative study by profession and seniority

TABLE 18 Interviewees in qualitative study by profession and seniority

Profession and seniority	4AT 'virgin' sites: Avonfield and Cranford (n)	4AT 'experienced' site: Denbury (n)
Consultant/senior doctor	4	7
Junior doctor	6	3
Senior/specialist nurse	3	4
Physiotherapist: frailty unit	1	–
Occupational therapist: ward		1
Staff nurse	4	9
Health support worker	1	–

Appendix 8 Topic guide for qualitative staff interviews: identifying delirium: knowledge, screening practices and barriers to identifying delirium in acute hospitals

Objectives

The study objectives were to:

1. Examine screening practices for delirium at specific points in the patient journey into and through hospital;
2. Explore staff knowledge of delirium and the value attached to identifying delirium including the strategies called forth to manage it at different points in the patient journey;
3. Determine organisational, environmental and system level barriers that impact use of screening tools to identify patients with possible delirium across the patient journey;
4. Consider staff use of screening tools generally and the 4AT specifically, the contexts in which it is used, speed and ease of use and value attached to use.

Introduction

Introduce study; confidentiality; timing; recording and consent.

Role

- Job title
- Work unit
- Experience
- Summary of role

Location on Care Pathway/Patient Journey

- a. Patient journey into the ED/MAU etc to the ward (focus will depend on the location of the informant)

Patients coming into the ED (or...specific to each point in the patient journey)

- Process for triage and diagnostics in respect of patients coming in via different routes (e.g. ambulance, independently mobile)
- What happens and length of time at each point in the process
- Decision-making on destination in respect of older people – does dementia/confusion enter into decision-making
- Problems/consequences – of organisational constraints and co-morbidities

Routine information gathering:

Explore what kind of information is collected and recorded routinely about older patients at each point in the journey, including what is asked of patients/relatives at each stage of the patient journey (A&E; MAU; Ward). [Collect relevant documentation]. Probe:

- What routine information is collected and recorded at each stage in the patient journey (at A&E; at MAU; and at Ward level)? How and by whom?
- What information is communicated to which staff at each stage
- How is it communicated
- Adequacy of information as perceived by staff at different points in the journey
- Problems with collecting/recording/ communicating information across the patient journey? What is routinely collected/usually or often missing? And cognitive impairment? Delirium? How and why?

Identifying cognitive impairment and delirium (at each point in the patient pathway)

Can you talk me through the process for identifying people with a cognitive impairment? And delirium? – use the following as probes

- Are older patients typically screened for cognitive impairment? (Screening tools employed and by whom)
- Are older patients typically screened for delirium? All? Some? Specific triggers? (observation of behavioural cues; use of systematic tools? (what and by whom)
- What are the contexts in which individual older patients are screened for cognitive impairment generally and for delirium specifically (all, some?)
- How and who makes the decision as to which patients should be screened for delirium
- What cues trigger assessment of delirium and how does it occur? Who does it?
- How is information collected and from whom about the patient's cognitive state prior to current episode

Knowledge of Delirium

Is 'delirium' a term that you would use? What term would you use? What does this mean to you? Probe

- Own understanding of delirium? What does this [whatever term is used] mean to you
- Understanding of delirium [whatever term is used] among team members
- Acquisition of knowledge about delirium
- Perceived value attached to delirium screening (staff member/team/hospital)
- Does it/should it affect the treatment/management plan? How and in what respects does it do so?
- Training received about delirium (What training? Covered in professional training? Professional development? Hospital mandatory training?)

Use of 4AT

Reflecting back now on your use of the 4AT in your work setting, can we go through what happened? Probe

- Context in which it was used; how decided which patient to use it with
- Ease/difficulty of use
- What contributed to ease/difficulty of use from your experience
- Did you have to think about what the responses meant/were they self-evident
- What level of knowledge is required to use it effectively
- How would you rate it in comparison with other screening tools
- What would need to happen to encourage you/your team to use it routinely
- Conversely what might inhibit you/your team in using it routinely
- How might use of the tool change practice on managing delirium
- Perceived costs versus benefits(time/resources)
- What would need to happen to change the balance of benefits and costs

Organisational, environmental and system barriers to screening for cognitive impairment and delirium

**Could you take me through a recent admission of a patient who appeared to be ‘confused’?
Probe**

- What was it that alerted you to the ‘confusion’?
- Is this how a patient with ‘confusion’ typically presents?
- What is typical?
- What happened next? Typical or not?
- Perceived dilemma/difficulty of distinguishing between dementia and acute confusion? What would you look out for in making such a distinction? What makes it feasible/practicable to assess this in your work setting?
- Value attached to identifying delirium in this context? Is the distinction between dementia and delirium useful in making a decision about what to do? Did it make a difference to the assessment/managing the person’s care in this case?
- How do you usually go about identifying delirium in [this setting]? And does it make a difference to what action follows for the patient?
- What are the barriers within this environment to screening and assessing for delirium? (organisational, environmental, knowledge)

Anything else?

Thank you for your time

Appendix 9 Comments and examples of suggested changes to the 4AT

These comments are drawn from the qualitative study interviews.

We employed the ‘think aloud’ method to elicit comments from study participants about the 4AT. Presented with a copy of the tool, each participant went through it, offering their reflections on its use in their particular setting.

Title, presentation and layout

Is that the same as the AMT? It's confusing and then there's the AMT4 referred to as well.

Junior doctor, ED

It looks pretty user friendly and that it won't take long to do which is key; otherwise people just won't do it.

Senior nurse, surgical ward

The guidance notes in this form aren't helpful. They don't explain things clearly. But if I'm busy with a patient or doing clerking I wouldn't read them anyway. Guidance notes aren't enough; it needs a teaching session to explain why it should be done and not only how to complete the form.

Senior doctor, MAU

Which patients should be targeted?

It wouldn't be feasible to use this on everyone coming into the ED. I don't like the idea of using an age cut-off either. Should it be patients at risk of delirium?

Consultant, ED

Specific items

Alertness

I'm not trying to be overly pedantic but altered level of alertness in normal waking hours, yes it could well be delirium; but altered alertness for someone coming in to A&E in the middle of the night having fallen . . . we need to be careful about overinterpretation.

Junior doctor, Care of older people ward

Just by knowing the patient, you would know if someone were more drowsy; you'd notice it naturally.

Staff nurse, Care of older people ward

Attention

Date of birth might not always be the best one. There are cultural issues. Elderly Asian men and women may not have a formal record of their date of birth. Isn't there a similar sort of question to use instead?

Senior nurse, Care of older people ward

Acute change or fluctuating course

That's not difficult when you know their baseline – these are the essential ones. Easy when you know the patient.

Junior doctor, Surgical ward

That would be really difficult if you've got someone who is confused in front of you and there is no one who knows the person with them. It might take time to find out. It's not the question as such; it's being able to find the answer. If you don't have that information to hand how would you score it. What would you do? That's the critical question for delirium.

Junior doctor, ED

We do drum it into staff that they are allowed to ring GPs and they are allowed to ring relatives and the care homes to find out what the patient is normally like, and get old notes. So there's lots of different ways you can try and find out if this is a patient's normal state or whether this is a significant change.

Consultant, ED

Scoring

How do you score patients who have clearly improved but have had an acute onset/fluctuating course over the previous 2 weeks? Would you use the 4AT to review change? Would that be right? There's no guidance?

Senior nurse, Care of older people ward

Appendix 10 The STARD flow diagrams of recruitment at each site

Edinburgh

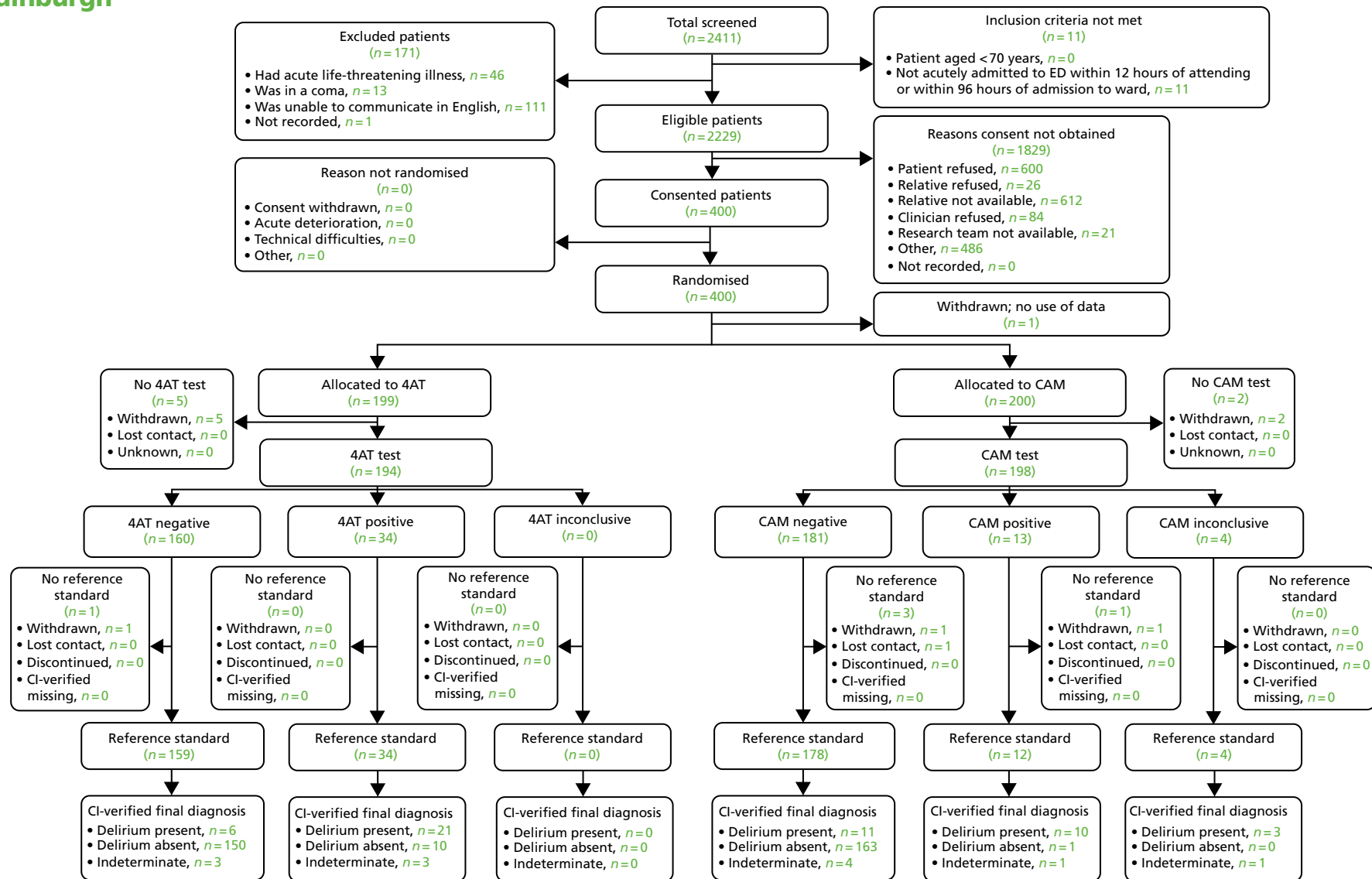


FIGURE 5 The STARD diagram of flow of participants through the Edinburgh site. CI, chief investigator.

Sheffield

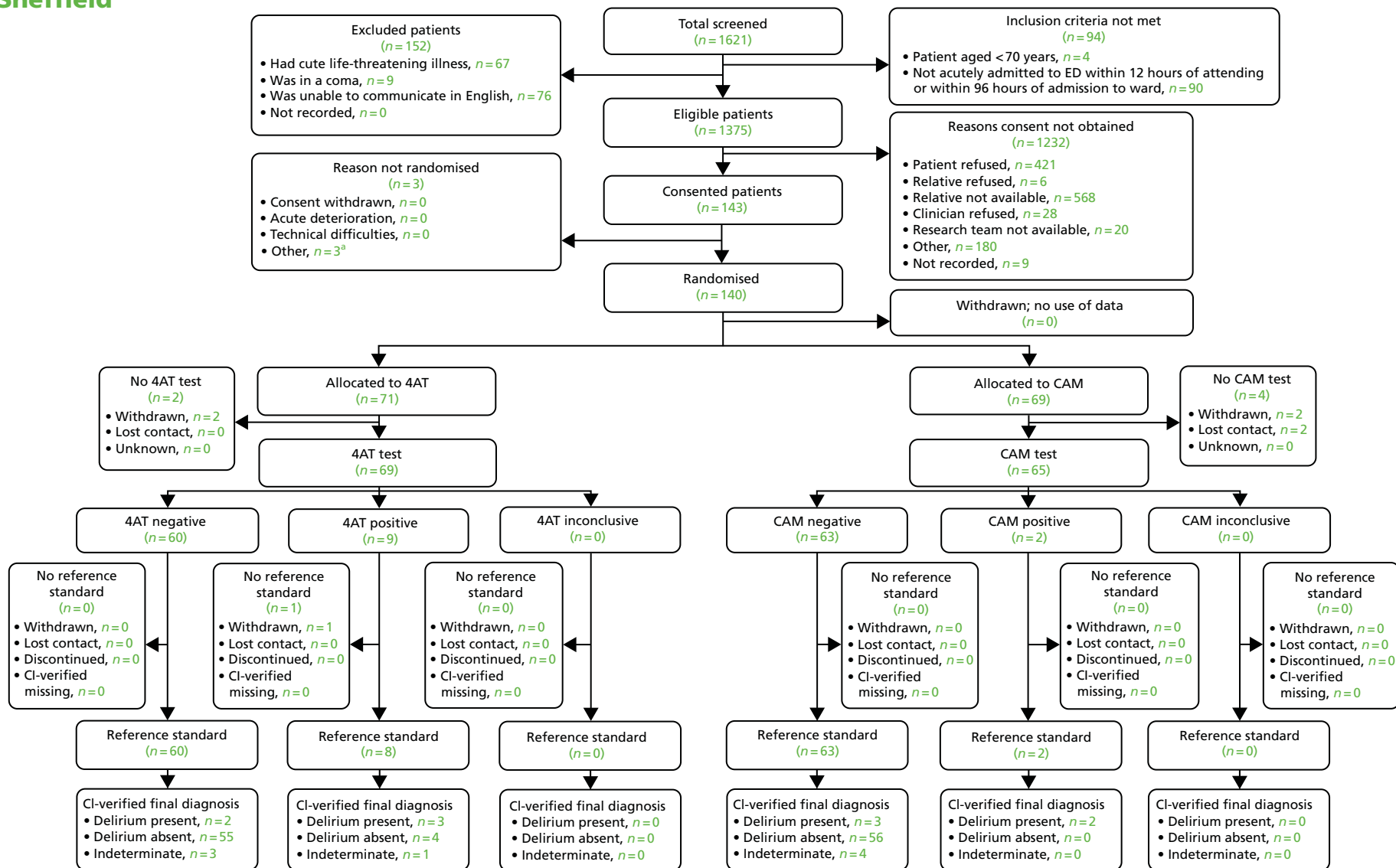


FIGURE 6 The STARD diagram of flow of participants through the Sheffield site. a, Two had moved ward, one had an issue with consultee. CI, chief investigator.

Bradford

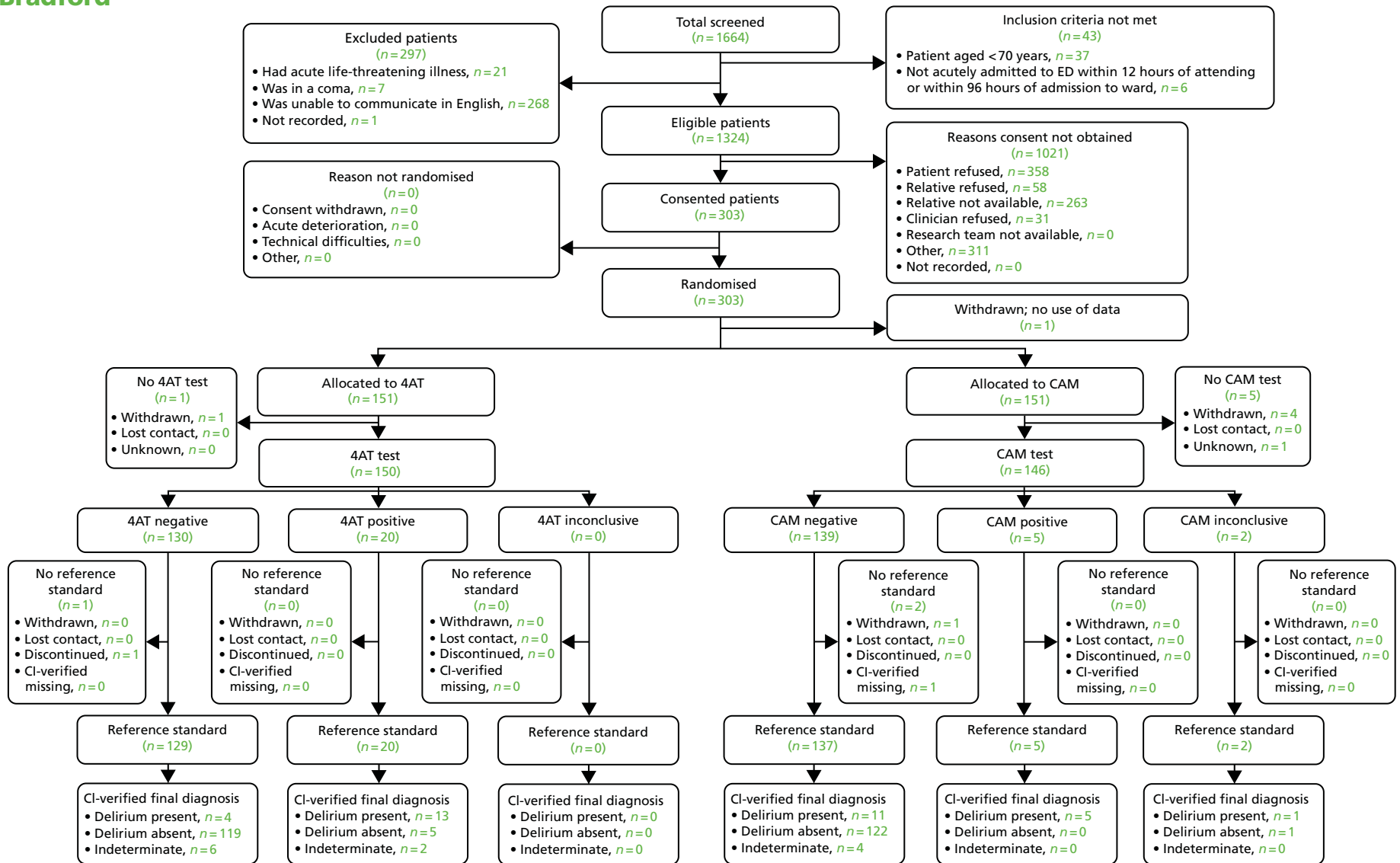


FIGURE 7 The STARD diagram of flow of participants through the Bradford site. CI, chief investigator.

Appendix 11 Receiver operating characteristic curve for 4AT diagnostic accuracy

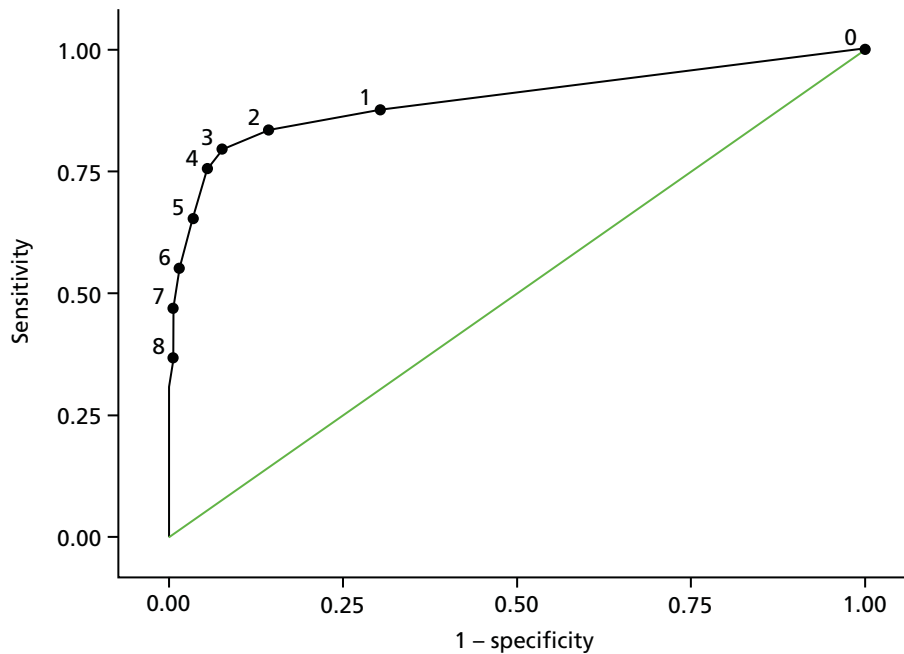


FIGURE 8 The ROC curve for 4AT diagnostic accuracy.

Appendix 12 Diagnostic test accuracy of 4AT versus CAM for diagnosis of delirium

TABLE 19 Diagnostic test accuracy of 4AT vs. CAM for diagnosis of delirium

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Diagnostic OR (95% CI)	Youden's Index
4AT (score of > 3)	75.51% (61.13% to 86.66%)	94.46% (91.48% to 96.63%)	66.07% (52.19% to 78.19%)	96.43% (93.84% to 98.14%)	52.57 (23.65 to 116.85)	0.70
CAM – delirium	40.48% (25.63% to 56.72%)	99.71% (98.38% to 99.99%)	94.44% (72.71% to 99.86%)	93.17% (90.08% to 95.53%)	231.71 (29.63 to 1811.89)	0.40
Fisher's exact test <i>p</i> -value	0.0012	<0.0001	0.0297	0.0629		
OR	4.53 (1.85 to 11.11)	0.05 (0.01 to 0.38)	0.11 (0.01 to 0.93)	1.98 (0.98 to 4.01)		
Difference in proportions	35.03% (14.66% to 53.17%)	-5.25% (-12.70% to 2.23%)	-28.37% (-53.44% to -1.72%)	3.26% (-4.14% to 10.64%)		

OR, odds ratio.

Numbers are estimates (95% CI). Difference in proportions is for 4AT – CAM.

Youden's Index is equal to sensitivity + specificity – 1: a value of zero indicates no value, and a value of 1 indicates a perfect test.

Appendix 13 Diagnostic test accuracy of individual items of 4AT for diagnosis of delirium

TABLE 20 Diagnostic test accuracy of individual items of the 4AT for diagnosis of delirium

Item	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Youden's Index
Alertness (4 vs. 0)	30.61% (18.25% to 45.42%)	99.13% (97.47% to 99.82%)	83.33% (58.58% to 96.42%)	90.91% (87.53% to 93.62%)	0.30
Acute change (4 vs. 0)	75.51% (61.13% to 86.66%)	95.61% (92.87% to 97.52%)	71.15% (56.92% to 82.87%)	96.46% (93.90% to 98.16%)	0.71
AMT4 (1 or 2 vs. 0)	63.27% (48.29% to 76.58%)	82.80% (78.38% to 86.64%)	34.44% (24.74% to 45.20%)	94.04% (90.74% to 96.43%)	0.46
AMT4 (2 vs. 1 or 0)	40.82% (27.00% to 55.79%)	96.21% (93.61% to 97.97%)	60.61% (42.14% to 77.09%)	91.92% (88.60% to 94.52%)	0.37
Attention (1 or 2 vs. 0)	71.43% (56.74% to 83.42%)	79.01% (74.31% to 83.20%)	32.71% (23.95% to 42.45%)	95.09% (91.90% to 97.29%)	0.50
Attention (2 vs. 1 or 0)	30.61% (18.25% to 45.42%)	99.13% (97.47% to 99.82%)	83.33% (58.58% to 96.42%)	90.91% (87.53% to 93.62%)	0.30

Numbers are estimates (95% CI). Youden's Index is equal to sensitivity + specificity – 1: a value of zero indicates no value and a value of 1 indicates a perfect test. Delirium is defined by the reference standard assessment.

Appendix 14 Scatterplot of Delirium Rating Scale-Revised-98 scores and 4AT scores

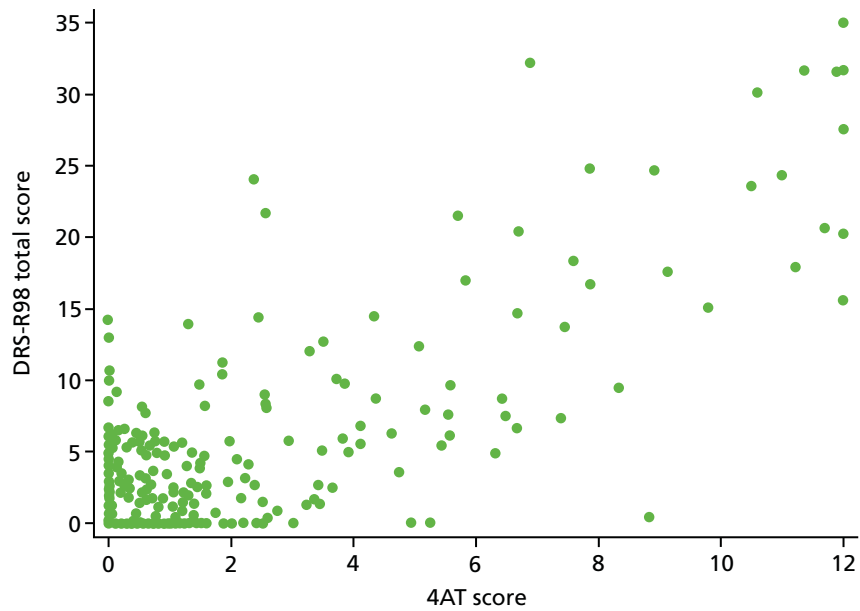
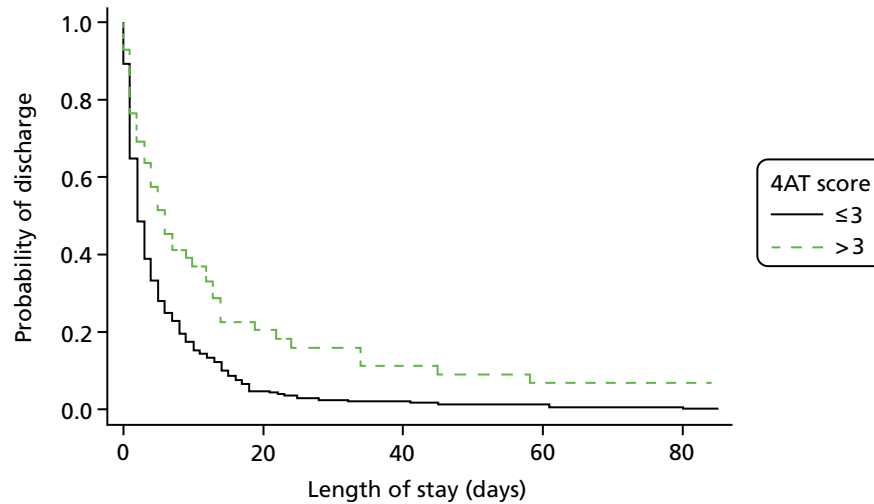


FIGURE 9 Scatterplot of DRS-R98 scores and 4AT scores.

Appendix 15 Kaplan–Meier plot for 4AT and length of stay



Number not yet discharged (Number of events)									
4AT score of ≤ 3	335	(315)	13	(7)	5	(2)	3	(1)	2
4AT score of > 3	55	(41)	9	(4)	5	(2)	2	(0)	2

FIGURE 10 Kaplan–Meier plot for 4AT and length of stay.

Appendix 16 Costs of tests: health economics analysis

TABLE 21 Costs of tests: health economics analysis

Test	Location, cost (£)	
	Scotland	England
Computed tomography	90	129
Electrocardiography	140	106
Magnetic resonance imaging	210	196
Physiotherapy	44	48
Ultrasound	56	72
Radiography	55	30 ^a

a Assumption based on National Tariff Information Workbook 2014/15 – non-mandatory costs.²¹² Source of Scottish costs is *Scottish Health Service Costs 2015–16*.²⁰⁰ Source of English costs is *NHS Reference Costs 2015 to 2016*.²⁰²

Appendix 17 Community health service costs

TABLE 22 Community health service costs

Social service costs per hour	Cost (£)	Source	Additional
Care attendant home	30	PSSRU 2016 ²⁰¹	
Care manager home	40	PSSRU 2016 ²⁰¹	
Care support worker home	30	PSSRU 2016 ²⁰¹	
Chiropodist home	32	PSSRU 2016 ²⁰¹	Community-based health-care staff qualification band 4 or 5
Community psychiatrist clinic	138	PSSRU 2016 ²⁰¹	
Community psychiatrist home	135	PSSRU 2016 ²⁰¹	Consultant
Community psychiatrist mental health home	44	PSSRU 2016 ²⁰¹	
Dietitian clinic	63	PSSRU 2016 ²⁰¹	Clinic and home assumed to be the same price
Dietitian home	63	PSSRU 2016 ²⁰¹	Clinic and home assumed to be the same price
District nurse clinic	86	PSSRU 2016 ²⁰¹	Band 5 in hospital per hour of patient contact
District nurse home	61	PSSRU 2016 ²⁰¹	Hourly costs of patient contact
GP clinic	104	PSSRU 2016 ²⁰¹	
GP home	216	PSSRU 2016 ²⁰¹	
Health visitor home	30	PSSRU 2016 ²⁰¹	Assumption, based on the cost for similar professions
Home-care worker home	24	PSSRU 2016 ²⁰¹	Per hour face to face
Meals on wheels home	3	AESOPS	Per meal: no hourly price
Occupational health home	44	PSSRU 2016 ²⁰¹	
Physiotherapist clinic	64	PSSRU 2016 ²⁰¹	Clinic and home assumed to be the same price
Physiotherapist home	64	PSSRU 2016 ²⁰¹	Clinic and home assumed to be the same price
Practice nurse clinic	86	PSSRU 2016 ²⁰¹	Band 5 in hospital per hour patient contact
Practice nurse home	61	PSSRU 2016 ²⁰¹	Consultant
Psychologist clinic	64	PSSRU 2016 ²⁰¹	
Social worker home	79	PSSRU 2016 ²⁰¹	

Appendix 18 Costs of additional activities

TABLE 23 Costs of additional activities

Additional activities mentioned	Minutes	Hours/visits
Cleaner	20,850	375.0
Phlebotomist	160	2.7
Speech and language therapist	60	1.0
Daily full package of care	2400	40.0
Hospital @ Home		37 visits
Podiatrist	300	5.0
Chest, heart and stroke liaison nurse	420	7.0
Housework help	4320	72.0
IMPACT team	450	7.5
Breathing nurse	90	1.5
Tissue viability and leg ulcer service	60	1.0
Continence service		Three visits
Diabetes hub		Two visits
Virtual ward: community support services	9870	164.5
Community gateway services: neighbourhood therapy	90	1.5
Cardiac rehabilitation nurse	150	2.5
Optician		Two visits
Dentist		One visit
GP telephone appointment	150	2.5
Cardiac rehabilitation team	75	1.3
Audiology, hearing aid	30	0.5
Step-down, frail elderly		Seven visits
Steady Steps	240	4.0
Pulmonary rehab team	1090	18.2
Intermediate care team		One visit
Stroke nurse		Eight visits
Acupuncture	180	3.0
Carer for daily medications	840	14.0
Home-care worker	720	12.0
POA Edinburgh North Care Service		One visit
Community support nurse	60	1.0
Continence services	25	0.4
Nurse practitioner	135	2.3

continued

TABLE 23 Costs of additional activities (continued)

Additional activities mentioned	Minutes	Hours/visits
Diabetes nurse specialist	90	1.5
Stoma nurse	180	3.0
Heart failure nurse	60	1.0
Community matron	920	15.3
Respiratory team	20	0.3
Community support team: falls assessment	40	0.7
COPD clinic	120	2.0

BDCT, Bradford District Care Trust; COPD, chronic obstructive pulmonary disease; POA, Psychiatry of Old Age.

Appendix 19 Health economics survey

4AT Health Economics Model Questionnaire

Page 1: Introduction

We are running a study to model the implications of routine screening for delirium in patients admitted to hospital under acute medical or geriatric specialties. A screening tool may lead to false positives (FP) – patients who are incorrectly labelled as having delirium in whom unnecessary investigation or treatment may be initiated. It may also lead to false negatives (FN) – patients who are labelled as not having delirium when in fact they do, which may lead to false reassurance and delayed or missed diagnosis. Mislabelling patients with or without delirium may have consequences for mortality, healthcare costs and patient quality of life.

In the absence of comprehensive data, we are also seeking the opinion of people experienced in delirium care to help inform some estimates in delirium detection and treatment. We will be asking for your spontaneous guess about how delirium detection and care effect duration and intensity. There is no need to consult books or other literature. Even if you are uncertain or find these questions difficult to answer, please just give your best guess.

The study is being led by Katharina Diernberger (Edinburgh Clinical Trials Unit, University of Edinburgh) and Dr Peter Hall (Edinburgh Clinical Trials Unit - University of Edinburgh, Cancer Research UK Edinburgh Centre - MRC Institute of Genetics & Molecular Medicine) in collaboration with Prof Alasdair MacLulich (Edinburgh Delirium Research Group - University of Edinburgh)

All responses are anonymised. Please send any queries to: [REDACTED] or [REDACTED]

Page 2: Mortality rates (1)

Please consider the average 12-week mortality rates in all patients aged between 80 and 85 who are admitted to acute medical or geriatric hospital wards. Please consider absolute numbers not relative risks. **Please assume an estimated 12-week mortality of 14 out of 100 for patients suffering from (all cases/detected or not) delirium and 10 out of 100 for patients without.**

	Best guess	lowest guess	highest guess
Out of 100 people who are incorrectly labelled with a diagnosis of delirium (FP), how many will die within 12 weeks as a direct consequence of that incorrect label?	<input type="text"/>	<input type="text"/>	<input type="text"/>
Out of 100 delirium sufferers in whom delirium is missed (FN), how many will die within 12 weeks as a direct consequence of that missed or delayed diagnosis?	<input type="text"/>	<input type="text"/>	<input type="text"/>

Do you have any additional comments?

Page 3: Mortality rates (2)

Please be aware that in this question the values which should be assumed are different to the last question. Please consider the average 12-week mortality rates in all patients aged between 80 and 85 who are admitted to acute medical or geriatric hospital wards. Please consider absolute numbers not relative risks. **Please assume an estimated 12-week mortality of 18 out of 100 for patients suffering from (all cases/detected or not) delirium and 8 out of 100 for patients without.**

	Best guess	lowest guess	highest guess
Out of 100 people who are incorrectly labelled with a diagnosis of delirium (FP), how many will die within 12 weeks as a direct consequence of that incorrect label?	<input type="text"/>	<input type="text"/>	<input type="text"/>
Out of 100 delirium sufferers in whom delirium is missed (FN), how many will die within 12 weeks as a direct consequence of that missed or delayed diagnosis?	<input type="text"/>	<input type="text"/>	<input type="text"/>

Do you have any additional comments?

Page 4: Costs

Please consider a cohort of patients aged between 80 and 85 who are admitted to acute medical or geriatric hospital wards.

	Best guess	lowest guess	highest guess
Assume the average hospital cost of a person suffering from delirium and being treated appropriately (TP) is 100%. What would you estimate as the total cost in % if the delirium is missed?	<input type="text"/>	<input type="text"/>	<input type="text"/>
Now assume the average hospital cost of a person who is not suffering from delirium (TN) is 100%. What would you estimate as the total cost in % if the person is incorrectly investigated or treated for delirium?	<input type="text"/>	<input type="text"/>	<input type="text"/>

Do you have any additional comments?

Page 5: Routine Care

Now assume all patients are treated according to a reference standard (perfectly accurate) diagnosis, which is assumed to promptly identify 100% of all patients suffering from delirium.

	Best guess	lowest guess	highest guess
In routine standard care, out of 100 patients with a reference standard diagnosis of delirium, how many would be missed completely in your type of ward? Please state the number out of 100.	<input type="text"/>	<input type="text"/>	<input type="text"/>
In routine care, for patients who are getting a diagnosis but being delayed, how long on average would you estimate for the diagnosis to be delayed? Please write your answer in days and/or hours.	<input type="text"/>	<input type="text"/>	<input type="text"/>

Do you have any additional comments?

Please choose the main type of healthcare setting you are working in.

<input type="radio"/> A&E	<input type="radio"/> Cardiology	<input type="radio"/> Critical Care
<input type="radio"/> Diagnostic imaging	<input type="radio"/> ENT	<input type="radio"/> Gastroenterology
<input type="radio"/> General surgery	<input type="radio"/> Geriatrics	<input type="radio"/> Gynaecology
<input type="radio"/> Haematology	<input type="radio"/> Nephrology	<input type="radio"/> Neurology
<input type="radio"/> Oncology	<input type="radio"/> Orthopaedics	<input type="radio"/> Rheumatology
<input type="radio"/> Urology	<input type="radio"/> other	<input type="radio"/> Other

If you selected Other, please specify:

Page 6: Quality of Life (from a health economic perspective)

The last part of the survey concerns the impact of delirium on quality of life. Please answer spontaneously based on your experience where on the provided scale patients' quality of life can be placed. A Quality of Life (QoL) score can represent various defined health states usually between 0 and 1 where 0 indicates death and 1 full physical and mental health. Average quality of life is observed to decrease with age whereby the average QoL of 80 to 85 year olds is estimated to be 0.7. In order to make the placement easier **a scale is provided (link below)** showing different health states/medical conditions with the referring utility weights taken from the literature.

Please be aware that the scale is a mix of acute and chronic conditions. It is not meant to be comprehensive but rather to give some orientation.

Please open the following link in a new tab (right click - open link in a new tab) as it might be helpful for answering the remaining questions:

https://static.onlinesurveys.ac.uk/med/account/93/survey/28075/question/orientation_scalefor_qol1.pdf

In the following questions, please consider 3 time periods:

Period 1: During an episode of delirium with or without an accurate diagnosis

Period 2: Hospital stay following resolution of delirium (or removal of incorrect label)

Period 3: Following hospital discharge (assuming resolution of delirium)

In order to answer the “% of care”, please consider the overall care staff time, number-, difficulty-, duration- of intervention, care intensity.

For all people who are not suffering from delirium and are labelled/treated correctly (TN)

the % of care is assumed to be 100% during the hospital stay. The quality of life is assumed to be 0.7 during the hospital stay and within the following 12 weeks.

People who are suffering from delirium and are labelled/treated correctly (TP) - Be aware that "Best guess duration" and "% of care" just can be filled in for Period_1 and Period_2. - The quality of life estimates (QoL) are supposed to be filled in for Period_1 and Period_3 only.

	Best guess duration(days)	% of care	Best guess QoL (0-1)	lowest QoL estimate	highest QoL estimate
Period_1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Period_2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Period_3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

People who are suffering from delirium with a missed diagnosis (FN) - Be aware that "Best guess duration" and "% of care" just can be filled in for Period_1 and Period_2. - The quality of life estimates (QoL) are supposed to be filled in for Period_1 and Period_3 only.

	Best guess duration(days)	% of care	Best guess QoL (0-1)	lowest QoL estimate	highest QoL estimate
Period_1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Period_2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Period_3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

People with an incorrect diagnosis of delirium (FP) - Be aware that "Best guess duration" and "% of care" just can be filled in for Period_1 and Period_2. - The quality of life estimates (QoL) are supposed to be filled in for Period_2 (as patients are not actually suffering from delirium, but any other condition) and Period_3 only.

	Best guess duration(days)	% of care	Best guess QoL (0-1)	lowest QoL estimate	highest QoL estimate
Period_1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Period_2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Period_3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Page 7: Thank you for completing our questionnaire!

Your answers have been successfully submitted.

Thank you for taking time to fill out the survey today, we really appreciate your time and effort.

Appendix 20 Model parameters

TABLE 24 Model parameters

Parameter	Value	Source
Parameters used to derive decision tree node parameters		
<i>Proportion</i>		
With delirium by reference standard (%)	0.1341	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Without delirium by reference standard (%)	0.8659	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
<i>4AT test</i>		
Sensitivity (%)	0.7551	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Specificity (%)	0.9456	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
<i>CAM test</i>		
Sensitivity (%)	0.4048	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Specificity (%)	0.9971	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Decision-tree node parameters		
<i>4AT</i>		
Probability of being TP	0.10	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Probability of being FP	0.05	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Probability of being TN	0.82	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Probability of being FN	0.03	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
<i>CAM</i>		
Probability of being TP	0.05	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Probability of being FP	0.00	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Probability of being TN	0.86	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
Probability of being FN	0.08	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
<i>12-week cost</i>		
TP	6889.00	Within-trial analysis – based on values from <i>Chapter 5</i> ^a
FP	4653.00	Based on 12-week Scottish costs and expert elicitation
TN	4230.00	Within-trial analysis – 12-week cost as above (Scotland)
FN	8955.70	Based on 12-week Scottish costs and expert elicitation
<i>QALY</i>		
TP	0.09	Derived from multiple responses in the expert elicitation
FP	0.13	Derived from multiple responses in the expert elicitation
TN	0.15	Derived from multiple responses in the expert elicitation
FN	0.09	Derived from multiple responses in the expert elicitation

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

^a Sensitivity and specificity of the tests and proportion of patients with and patients without delirium based on the reference standard.

Appendix 21 Results from the health economics model

TABLE 25 Differences in costs and effectiveness and ICERs

	Cost (£)		QALY		Difference in		
	4AT	CAM	4AT	CAM	Costs (£)	QALYs	ICER (£)
<i>Base case</i>							
Scottish costs – 12 weeks – best estimates (costs and QALYs)	4680	4770	0.14050	0.14103	-90.35	-0.00053	170,553
<i>Scenario analysis</i>							
Lowest cost and QALY estimates	4614	4659	0.13374	0.13557	-44.49	-0.00183	24,289
Highest cost and QALY estimates	4765	4882	0.14585	0.14459	-117.35	0.00127	-92,744
English costs – 12 weeks – best estimates (costs and QALYs)	4416	4478	0.14050	0.14103	-61.52	-0.00053	116,133
Lowest cost and QALY estimates	4364	4399	0.13374	0.13557	-34.94	-0.00183	19,070
Highest cost and QALY estimates	4488	4557	0.14585	0.14459	-69.10	0.00127	-54,609
First stay (English/Scottish) – best estimates (costs and QALYs)	1506	1554	0.14050	0.14103	-47.96	-0.00053	90,522
Lowest cost and QALY estimates	1481	1507	0.13374	0.13557	-25.92	-0.00183	14,147
Highest cost and QALY estimates	1537	1602	0.14585	0.14459	-64.47	0.00127	-50,952

Appendix 22 Cost-effectiveness plane

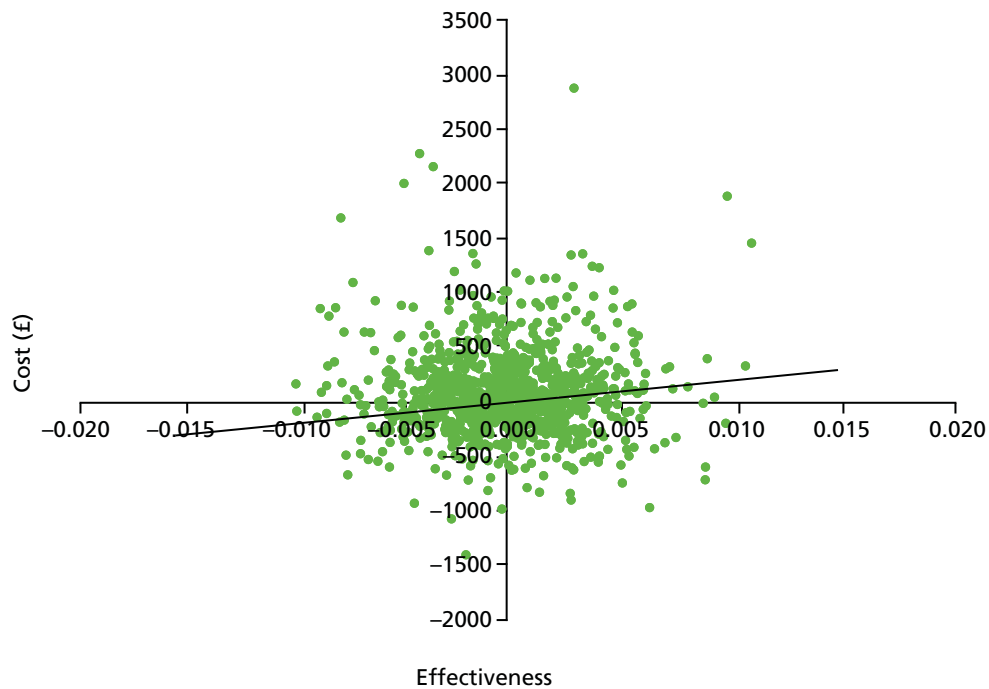


FIGURE 11 Cost-effectiveness plane.

Appendix 23 Cost-effectiveness-acceptability curve

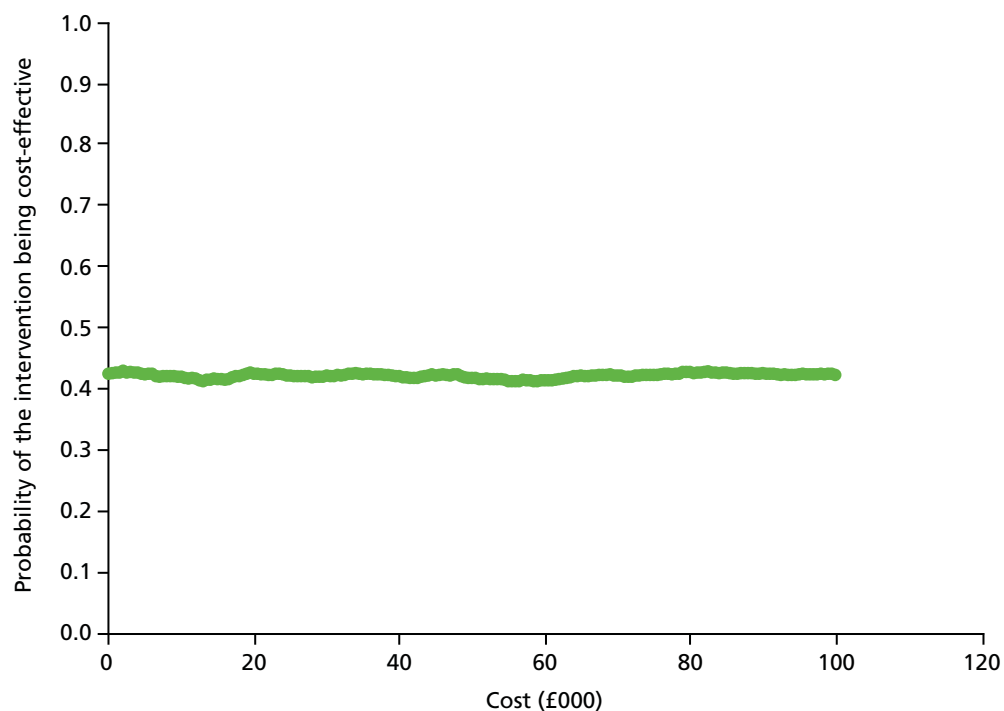


FIGURE 12 Cost-effectiveness-acceptability curve.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library