



This is a repository copy of *Comparison of case note review methods for evaluating quality and safety in health care*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/10704/>

Article:

Hutchinson, A., Coster, J.E., Cooper, K.L. et al. (8 more authors) (2010) Comparison of case note review methods for evaluating quality and safety in health care. *Health Technology Assessment*, 14 (10). pp. 1-170. ISSN 1366-5278

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

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.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

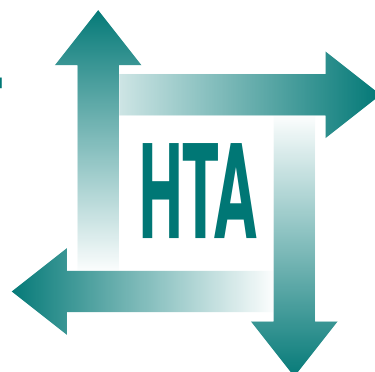
Comparison of case note review methods for evaluating quality and safety in health care

A Hutchinson, JE Coster, KL Cooper,
A McIntosh, SJ Walters, PA Bath,
M Pearson, TA Young, K Rantell,
MJ Campbell and J Ratcliffe



February 2010
DOI: 10.3310/hta14100

Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk





How to obtain copies of this and other HTA programme reports

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (www.hta.ac.uk). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our despatch agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per issue and for the rest of the world £3 per issue.

How to order:

- fax (with **credit card details**)
- post (with **credit card details** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you to either print out your order or download a blank order form.

Contact details are as follows:

Synergie UK (HTA Department)
Digital House, The Loddon Centre
Wade Road
Basingstoke
Hants RG24 8QW

Email: orders@hta.ac.uk

Tel: 0845 812 4000 – ask for ‘HTA Payment Services’
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put ‘HTA Order’ on the fax header

Payment methods

Paying by cheque

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

Paying by credit card

You can order using your credit card by phone, fax or post.

Subscriptions

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

How do I get a copy of HTA on DVD?

Please use the form on the HTA website (www.hta.ac.uk/htacd/index.shtml). *HTA on DVD* is currently free of charge worldwide.

The website also provides information about the HTA programme and lists the membership of the various committees.

Comparison of case note review methods for evaluating quality and safety in health care

A Hutchinson,^{1*} JE Coster,¹ KL Cooper,¹
A McIntosh,¹ SJ Walters,² PA Bath,³
M Pearson,⁴ TA Young,⁵ K Rantell,²
MJ Campbell² and J Ratcliffe⁵

¹Section of Public Health, ScHARR, University of Sheffield, UK

²Section of Health Services Research, ScHARR, University of Sheffield, UK

³Department of Information Studies, University of Sheffield, UK

⁴Clinical Effectiveness and Evaluation Unit, Royal College of Physicians,
London, UK

⁵Section of Health Economics and Decision Sciences, ScHARR, University of
Sheffield, UK

*Corresponding author

Declared competing interests of authors: none

M Pearson is now at the Department of Medicine, University of Liverpool.

Published February 2010

10.3310/hta14100

This report should be referenced as follows:

Hutchinson A, Coster JE, Cooper KL, McIntosh A, Walters SJ, Bath PA, *et al.* Comparison of case note review methods for evaluating quality and safety in health care. *Health Technol Assess* 2010; **14**(10).

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

NIHR Health Technology Assessment programme

The Health Technology Assessment (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 research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series 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 referees 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.

The research reported in this issue of the journal was commissioned by the National Coordinating Centre for Research Methodology (NCCRM), and was formally transferred to the HTA programme in April 2007 under the newly established NIHR Methodology Panel. The HTA programme project number is 06/91/02. The contractual start date was in June 2004. The draft report began editorial review in March 2009 and was accepted for publication in May 2009. The commissioning brief was devised by the NCCRM who specified the research question and study design. 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 referees 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.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley CBE
Series Editors: Dr Martin Ashton-Key, Dr Aileen Clarke, Professor Chris Hyde,
Dr Tom Marshall, Dr John Powell, Dr Rob Riemsma and Professor Ken Stein

ISSN 1366-5278

© 2010 Queen's Printer and Controller of HMSO

This journal may be freely reproduced for the purposes of private research and study and 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: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk), on behalf of NETSCC, HTA.

Printed on acid-free paper in the UK by the Charlesworth Group.

MR



Abstract

Comparison of case note review methods for evaluating quality and safety in health care

A Hutchinson,^{1*} JE Coster,¹ KL Cooper,¹ A McIntosh,¹ SJ Walters,² PA Bath,³ M Pearson,⁴ TA Young,⁵ K Rantell,² MJ Campbell² and J Ratcliffe⁵

¹Section of Public Health, ScHARR, University of Sheffield, UK

²Section of Health Services Research, ScHARR, University of Sheffield, UK

³Department of Information Studies, University of Sheffield, UK

⁴Clinical Effectiveness and Evaluation Unit, Royal College of Physicians, London, UK

⁵Section of Health Economics and Decision Sciences, ScHARR, University of Sheffield, UK

*Corresponding author

Objectives: To determine which of two methods of case note review – holistic (implicit) and criterion-based (explicit) – provides the most useful and reliable information for quality and safety of care, and the level of agreement within and between groups of health-care professionals when they use the two methods to review the same record. To explore the process–outcome relationship between holistic and criterion-based quality-of-care measures and hospital-level outcome indicators.

Data sources: Case notes of patients at randomly selected hospitals in England.

Review methods: In the first part of the study, retrospective multiple reviews of 684 case notes were undertaken at nine acute hospitals using both holistic and criterion-based review methods. Quality-of-care measures included evidence-based review criteria and a quality-of-care rating scale. Textual commentary on the quality of care was provided as a component of holistic review. Review teams comprised combinations of: doctors ($n=16$), specialist nurses ($n=10$) and clinically trained audit staff ($n=3$) and non-clinical audit staff ($n=9$). In the second part of the study, process (quality and safety) of care data were collected from the case notes of 1565 people with either chronic obstructive pulmonary disease (COPD) or heart failure in 20 hospitals. Doctors collected criterion-based data from case notes and used implicit review methods to derive textual comments on the quality of care provided and score the care overall. Data were analysed for intrarater consistency, inter-rater reliability between pairs of staff using intraclass correlation coefficients (ICCs) and completeness of criterion data capture, and comparisons were made within and between staff

groups and between review methods. To explore the process–outcome relationship, a range of publicly available health-care indicator data were used as proxy outcomes in a multilevel analysis.

Results: Overall, 1473 holistic and 1389 criterion-based reviews were undertaken in the first part of the study. When same staff-type reviewer pairs/groups reviewed the same record, holistic scale score inter-rater reliability was moderate within each of the three staff groups [intraclass correlation coefficient (ICC) 0.46–0.52], and inter-rater reliability for criterion-based scores was moderate to good (ICC 0.61–0.88). When different staff-type pairs/groups reviewed the same record, agreement between the reviewer pairs/groups was weak to moderate for overall care (ICC 0.24–0.43). Comparison of holistic review score and criterion-based score of case notes reviewed by doctors and by non-clinical audit staff showed a reasonable level of agreement (p -values for difference 0.406 and 0.223, respectively), although results from all three staff types showed no overall level of agreement (p -value for difference 0.057). Detailed qualitative analysis of the textual data indicated that the three staff types tended to provide different forms of commentary on quality of care, although there was some overlap between some groups. In the process–outcome study there generally were high criterion-based scores for all hospitals, whereas there was more interhospital variation between the holistic review overall scale scores. Textual commentary on the quality of care verified the holistic scale scores. Differences among hospitals with regard to the relationship between mortality and quality of care were not statistically significant.

Conclusions: Using the holistic approach, the three groups of staff appeared to interpret the recorded care differently when they each reviewed the same record. When the same clinical record was reviewed by doctors and non-clinical audit staff, there was no significant difference between the assessments of quality of care generated by the two groups. All three staff groups performed reasonably well when using criterion-based review, although the quality and type of information provided by doctors was of greater

value. Therefore, when measuring quality of care from case notes, consideration needs to be given to the method of review, the type of staff undertaking the review, and the methods of analysis available to the review team. Review can be enhanced using a combination of both criterion-based and structured holistic methods with textual commentary, and variation in quality of care can best be identified from a combination of holistic scale scores and textual data review.



Contents

Contents	v	Appendix 5 Reviewer training scenarios to assist in recognising variation in care quality using holistic review	101
List of abbreviations	vii	Appendix 6 Record review for safety and quality study	103
Executive summary	ix	Appendix 7 COPD – correlations between holistic mean overall scale scores and outcome variables	119
1 Introduction	1	Appendix 8 Heart failure – correlations between holistic mean overall scale scores and outcome variables	121
2 Assessing quality of care from hospital case notes: comparison of reliability and utility of holistic (implicit) and criterion-based (explicit) methods	3	Appendix 9 COPD – correlations between holistic mean phase scale scores and outcome variables	123
Background	3	Appendix 10 Heart failure – correlations between holistic mean overall scale scores and outcome variables	125
Methods	5	Appendix 11 COPD – correlations between holistic mean criterion scores and outcome variables	127
Results	16	Appendix 12 Heart failure – correlations between mean criterion scores and outcome variables	129
Discussion	34	Appendix 13 Comparison of holistic and criterion-based review methods using structured clinical records in stroke care ...	131
3 What is the relationship between information on quality of care from case notes and hospital-level outcomes of care?	41	Appendix 14 The place of trigger tool methodology in case note review for quality and safety	141
Background	41	Health Technology Assessment reports published to date	145
Methods	42	Health Technology Assessment programme	167
Results	47		
Discussion	61		
4 Overall conclusions of the research	65		
Implications for reviewing quality of care .	65		
Relationships between quality and outcome of care	66		
5 Future research agenda	67		
Acknowledgements	69		
References	71		
Appendix 1 COPD review criteria	75		
Appendix 2 Heart failure review criteria .	83		
Appendix 3 Validity of review criterion questionnaire (COPD)	91		
Appendix 4 A holistic review data collection page	99		





List of abbreviations

ADE	adverse drug event	MI	myocardial infarction
CI	confidence interval	NICE	National Institute for Health and Clinical Excellence
COPD	chronic obstructive pulmonary disease	NPSA	National Patient Safety Agency
HCC	Healthcare Commission	RCP	Royal College of Physicians
HES	Hospital Episode Statistics	RCP CEEu	Royal College of Physicians Clinical Effectiveness and Evaluation Unit
HRG	Healthcare Resource Group	SD	standard deviation
HSMR	Hospital Standardised Mortality Ratio	SMR	standardised mortality ratio
ICC	intraclass correlation coefficient	SpR	specialist registrar
IHI	Institute for Healthcare Improvement	TIA	transient ischaemic attack

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.



Executive summary

Purpose

The purpose of the first part of the study was twofold. First, to determine which of two methods of case note review provide the most useful and reliable information for reviewing quality and safety of care, and for what purpose. Second, to determine the level of agreement within and between groups of health-care professionals (doctors, nurses and other clinically trained staff, and non-clinical audit staff) when they use the two methods to review the same record.

The results were also expected to influence the methods of data capture for the second part of the study, which explored the process–outcome relationship between holistic and criterion-based quality-of-care measures (process measures) and hospital-level outcome indicators, grouped by mortality level.

Methods

In the first part of the study, retrospective multiple reviews of 684 case notes were undertaken using both holistic (implicit) and criterion-based (explicit) review methods. Quality-of-care measures included evidence-based review criteria and a quality-of-care rating scale. Textual commentary on the quality of care was provided as a component of holistic review. Data collection was conducted in nine randomly selected acute hospitals in England, by hospital staff trained in case note review. These local review teams comprised combinations of three staff types: doctors ($n = 16$), specialist nurses ($n = 10$) and clinically trained audit staff ($n = 3$) ($n = 13$ in total), and non-clinical audit staff ($n = 9$).

During the second part of the study, process (quality and safety) of care data were collected from the case notes of 1565 people with either chronic obstructive pulmonary disease (COPD) or heart failure in 20 randomly selected hospitals in England. Doctors collected criterion-based data from case notes and used implicit review methods to derive textual comments on the quality of care provided and score the care overall.

Analysis methods

Intra-rater consistency, inter-rater reliability between pairs of staff using intraclass correlation coefficients (ICCs), completeness of criterion data capture, within- and between-staff group comparison, and between-review-method comparison. To explore the process–outcome relationship, a range of publicly available health-care indicator data were used as proxy outcomes in a multilevel analysis.

Results

A total of 1473 holistic reviews and 1389 criterion-based reviews were undertaken in the first part of the study.

When same staff-type reviewer pairs/groups reviewed the same record, holistic scale score inter-rater reliability was moderate within each of the three staff groups (ICC 0.46–0.52), and inter-rater reliability for criterion-based scores was moderate to good (ICC 0.61–0.88). When different staff-type pairs/groups reviewed the same record, agreement between the reviewer pairs/groups was weak to moderate for overall care (ICC 0.24–0.43).

Comparison of holistic review score and criterion-based score of case notes reviewed by doctors and by non-clinical audit staff showed a reasonable level of agreement between the two methods (p -values for difference 0.406 and 0.223, respectively), although results from all three staff types showed no overall level of agreement (p -value for difference 0.057).

Detailed qualitative analysis of the textual data provided by reviewers indicated that the three staff types tended to provide different forms of commentary on quality of care, although there was some overlap between non-clinical audit staff and the nursing group and between the nursing group and the doctors. Thus the non-clinical audit staff mainly reported facts from the case notes. Nurses and clinical audit staff provided commentaries that were mainly about process of care, together with

some implicit judgements about the quality of care provided. Information from the doctors tended to be more focused on technical aspects of care, making rather more explicit judgements on quality of care.

In the process–outcome study there generally were high criterion-based scores for all of the hospitals, while there was rather more inter-hospital variation between the holistic review overall scale scores. Rich textual commentary on the quality of care verified the holistic scale scores. While there were trends towards hospitals that had lower mortality also having higher quality-of-care scores, none of these differences was statistically significant. There was only limited correlation between the outcome indicators and the criterion-based or holistic scale scores for either condition across the 20 hospitals.

Conclusions

Using a holistic approach to review case notes, groups of the same staff type can achieve reasonable repeatability within their professional groups when asked to rate quality of care on a scale. But there is little agreement between the three staff types when using holistic review methods to rate quality of care for the same clinical record, possibly because the different staff types are exploring different aspects of quality of care, as the qualitative analysis suggests.

All three staff groups have reasonable to high levels of consistency when using criterion-based review and, because there tend to be low levels of missing values in the data collected by all three staff types, there is little to choose between the staff groups in terms of reviewer effectiveness.

When the same clinical record was reviewed by the doctors, and by the non-clinical audit staff, using first holistic and then criterion-based methods, there is no significant difference between the

assessments of quality of care generated by the two methods. This suggests that although the two methods are exploring quality of care differently, they can allow similar levels of quality ratings to be made. When measuring quality of care from case notes, therefore, consideration needs to be given to three important factors: the method of review, the type of staff to undertake the review, and the methods of analysis available to the review team.

It is likely that review of quality of care can be enhanced by using a combination of both criterion-based (explicit) methods and structured holistic (implicit) methods, which will identify both evidence-based elements of care and the nuances of care that are almost always a component of care in long-term conditions. Free textual commentary on the quality of care provided is a valuable asset in judging care, but it is complex to analyse and is likely to remain as a research tool in this field of health-care evaluation.

Variation in quality of care can be identified from a combination of holistic scale scores and textual data review to provide a rich means of understanding the outcome of care on an individual patient basis.

Although there are some correlations between quality-of-care scores and hospital-level outcome data, there is no clear relationship between the process of care and hospital-level outcomes for the two indicator conditions in this study. This probably reflects the complexity of the process–outcome relationship at the group level. Available hospital-level outcome indicator data are probably insufficiently sensitive to reflect the quality of care recorded in patient case notes. Furthermore, high-quality care may be given even when the patient's outcome is poor, and vice versa. These findings may be pointing to process measures as being more useful than outcome measures when reviewing the care of people who have chronic disease or multiple conditions.

Chapter I

Introduction

The study had two main aims, which were agreed in response to a call for proposals from the National Coordinating Centre for Research Methodology (now part of the NIHR Health Technology Assessment programme). These aims were:

1. To compare the validity and reliability of two methods of case note review of quality and safety of care. That is, to explore which of two methods of case note review – holistic (implicit) review or criterion-based (explicit) review – is the most effective method of reviewing quality of care, and under what circumstances and by which type of staff. Methodological questions include a comparison of reliability of holistic (implicit) and criterion-based (explicit) methods.
2. To investigate whether there appears to be a link between the quality of medical care, as recorded in case notes, and the outcome of hospital care, for two chronic medical conditions. Methodological questions include an exploration of how implicit and explicit

case note review might be used to explore the relationship between process of care and risk-adjusted outcomes.

Since the aims are linked but address two different aspects of case note review – the methodology of case note review and process–outcome relationship – this report is presented in two main parts (Chapters 2 and 3), and the overall summary and future research agenda presented as a whole (Chapters 4 and 5). *Figure 1* shows where the linkages lie between the two studies, particularly through the choice of review methods, type of reviewers and methods of selection of hospitals.

Additionally, there are two small studies that were also commissioned as part of the programme of work. One uses the review methods to explore their value in the context of structured record-keeping in stroke care. The second study explores the literature relating to the use of trigger tools when reviewing paper-based case notes for quality and safety. These studies are presented in Appendices 13 and 14, respectively.

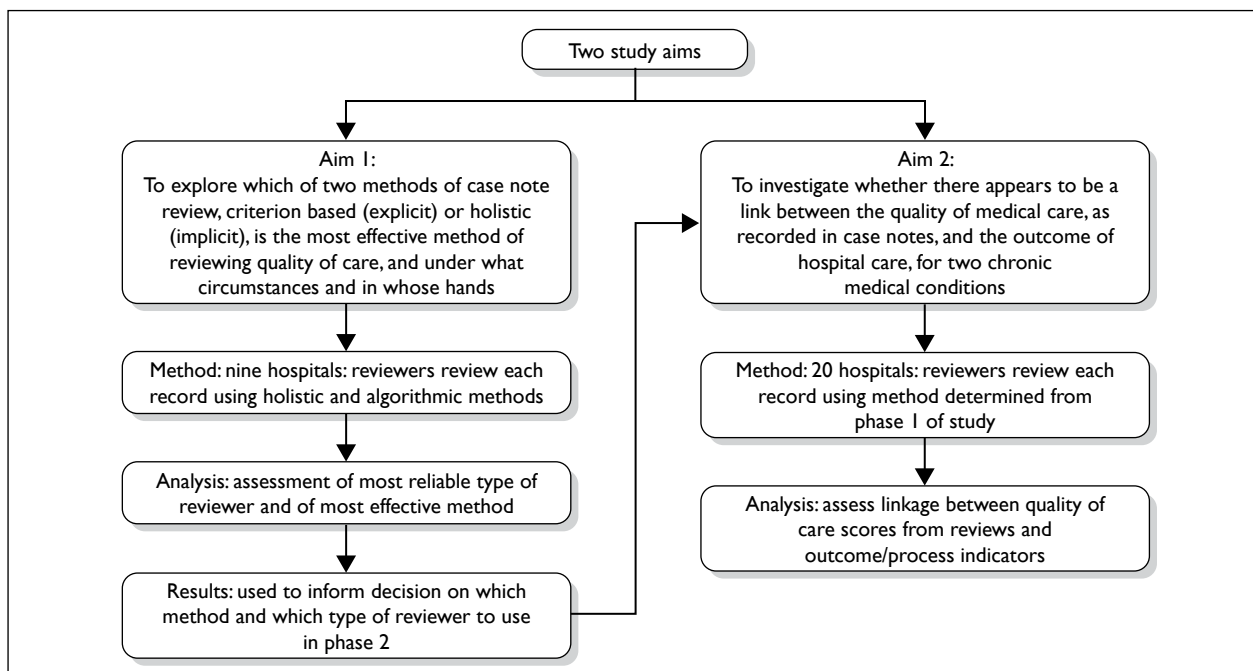


FIGURE 1 Aims of the two studies and links between them.

Chapter 2

Assessing quality of care from hospital case notes: comparison of reliability and utility of holistic (implicit) and criterion-based (explicit) methods

Background

Review of the quality of care as described in written case notes has become a standard means of assessing variation from quality standards and for identifying adverse incidents, either concerning individuals or groups of patients.

Quality of care is currently assessed from clinical records by collecting data using two principal approaches: holistic review (sometimes called implicit review) and criterion-based (explicit) review. Both of these approaches have recognised strengths and weaknesses, whether they are being used for performance monitoring and assessment or in a research setting.

Although attempts to systematise the review of quality of care began nearly a century ago with the work of Codman in 1912,¹ much of the development of case note review methodology began in North America in the 1970s with the work of the Peer Review Organisations, which used implicit review methods (sometimes called 'holistic methods') to determine variations in the standards of care provided by hospitals.² Subsequently, variants of these holistic (implicit) methods to review the quality of care of hospital patients were used as the basis for determining adverse event and medical error rates in three large epidemiological studies in New York state,³ Australia,⁴ and in Colorado and Utah.⁵ Holistic review was widely used, subsequently, in clinical audit in the UK.

Clinical staff in the UK are accustomed to looking through a set of patient records in order to form an opinion on the quality of care delivered. This holistic approach uses professional judgement and has the advantage that it requires no prior assumptions about the individual case, can be applied to any condition, can extend to examining any aspect of care, and, at least in experienced

hands, may be relatively quick to perform. However, the standards against which quality is judged holistically are implicit, being to a considerable extent dependent on the reviewer's personal knowledge and perspective, and thus are subjective. As a result the use of implicit professional judgements as the basis for reviewing quality and identifying variations from good practice has been increasingly criticised.

Research has identified a range of assumptions about what is being measured by holistic (implicit) review, and problems have been identified with the reliability and the validity of the approach. Weingart *et al.*⁶ conducted a retrospective record review of 1025 case notes to compare explicit and implicit review methods when examining quality of care. Their study found that implicit reviews by physicians tend to take a global approach, including an assessment of the severity of the case, but are less likely than nurses to take into account any process issues that may lead to reduced quality of hospital care. This finding was supported by the results of a study by Gibbs and colleagues,⁷ who compared quality of care for patients selected using higher- and lower-than-expected mortality rates. The authors highlighted the insensitivity of implicit methods when used for detecting hospital-level differences and reported that implicit chart reviews are not successful at discovering differences in quality of care.

Ashton *et al.*⁸ found that not only can implicit review be highly idiosyncratic and reviewer dependent, but also it can result in lower levels of inter-rater reliability than explicit methods at patient level. Moreover, reanalysis of data from the Utah and Colorado Medical Practice Study has contributed to concerns that holistic record review may have low reliability, with the finding that different implicit review strategies produced different estimates of the total number

of adverse events and negligent adverse events.⁹ Despite attempts to reduce levels of subjectivity in holistic review (for example by such means as providing extensive training for physician reviewers), a number of other concerns remain about the value of review methods that are based principally on professional judgement. Inter-rater reliability between reviewers has been identified as being particularly problematic, with Hofer and colleagues¹⁰ finding levels of between 0.25 (low) and 0.45 (modest) in a study of a range of diseases and service settings. It may also be that the choice of methods of assessing reliability may have some effect on the results of studies, since the kappa statistic is influenced by the prevalence of events.^{11,12} Additionally, the individual consistency of reviewers has been questioned,¹³ and the individual reviewer's bias towards harshness or leniency has been considered as problematic in comparing results between reviewers.¹⁴ Fischhoff's¹⁵ initial work on hindsight bias has recently been reiterated as a confounding factor in implicit review.¹⁶ For these reasons, criterion-based review, using predefined criteria, has been proposed as a more reliable means of assessing quality from clinical records.^{17,18}

Criterion-based (explicit) methods of review are an acknowledged alternative method to holistic review and have been widely used in the UK and in the USA. Standardised methods for developing explicit evidence-based review criteria were proposed by an Agency for Health Care Policy and Research working party in 1995¹⁸ and were further developed by Hadorn and colleagues.¹⁹ Criterion-based review allow comparison of care against explicit standards (such as those derived from national clinical guidelines). It requires the definition of unambiguous questions to construct variables that can be retrieved from the case records and, although only predefined questions can be addressed, the variables have good reproducibility.

Derivations of these methods, using locally based standards of care as a template for identifying variations from care standards, were used in a large UK study of general practice during the 1980s.²⁰ Subsequently, a number of structured methods for developing review criteria for explicit quality review of case notes have been developed in the UK, including methods for developing criteria directly from evidence-based clinical guidelines.²¹ These methods all seek to determine the rate of conformance with the criteria within a single patient's care and are aggregated across a group

of patients, recorded as a percentage. Patient preferences and clinical choices, based on the severity and the anticipated outcome of the case, are allowed for in estimating conformance and are not considered to be 'violations' of a standard of care. Criteria can thus be developed for any condition where there are externally agreed explicit standards of care.

This approach is becoming part of UK health policy. All clinical practice guidelines now being published by the National Institute for Health and Clinical Excellence (NICE) are now being accompanied by evidence-based review criteria to support review of clinical quality (see, for example, criteria published with the NICE guideline on the management of chronic heart failure²²).

Clinical audit in UK hospitals has adopted these objective, criterion-based approaches,²³⁻²⁵ using explicit standards that are not profession dependent and have shown, for example, substantial variations in organisation and clinical care between hospitals.²³

Nevertheless, criterion-based review has been criticised as an insensitive method that may not identify unexpected factors that might influence outcomes of care,^{26,27} so that implicit review may have still retain some advantages. In some North American studies mixed methods have been adopted,^{6,9} where nurses used criterion-based review to identify a subset of problematic cases for subsequent holistic review by doctors, although this two-stage approach carries a risk of hindsight bias, such that those cases identified as problematic by nurses might be reviewed more harshly by the physicians.^{14,17} Rubenstein and colleagues²⁸ proposed a structured form of implicit review in which a series of clear questions were asked of the reviewers, as distinct from seeking specific data items as in explicit review, and Pearson *et al.*²⁹ used this method to review the quality of nursing care.

Thus the decision on methods is not necessarily settled by the adoption of the criterion-based approach, which may fail to identify the nuances of health-care variation. Mohammed *et al.*²⁷ reviewed the quality of care of 50 patients with stroke from each of four hospitals reported as having the 'best' mortality outcomes for stroke in the West Midlands area of the UK, and four reported as having the 'worst' mortality outcomes. After adjusting for case mix using W-scores, the researchers identified a number of issues that affected outcome. Some influences were predictable, such as the

organisation of care. Some were unexpected, such as the influence on outcomes of 'do not resuscitate' orders. The authors suggested that these unexpected influences would have been identified only by expert reviewing using holistic methods.

Decisions on which of the two review methods to use, and under which circumstances, are also clouded by the results from studies that have used mixed methods. Thus Weingart and colleagues⁶ have suggested that nurses and doctors may use different types of information on which to make judgements of care quality (and thus may come to different judgements about an individual case). On the other hand, Keeler *et al.*³⁰ used explicit and implicit methods and sickness (risk)-adjusted outcomes to review quality in different types of hospitals. They reported similar quality-of-care ratings for the specific hospitals when using the two methods. Any differences in quality were thought to be the result of differences in the characteristics of the hospitals rather than the result of using different methods of record review.

Overall, then, there is real lack of clarity about choice of method for case note review – which method, in whose hands and for what purpose. Building on the international evidence, this study was designed to explore these issues further.

Study aim and research questions

The first study aim agreed with the research commissioners was: to compare the validity and reliability of two methods of case note review of quality and safety of care.

Research questions were:

- Do holistic (implicit) and criterion-based (explicit) methods of case note review identify the same variations in quality within the same record?
- Do holistic (implicit) and criterion-based (explicit) methods of case note review identify the same variations in quality across groups of records for the same clinical condition?
- To what extent do holistic and criterion-based methods of case note review provide similar results when used by reviewers from similar professional groups?
- To what extent do holistic and criterion-based methods of case note review provide similar results when used by reviewers from different professional groups?
- Which method of case note review and which staff type would be most appropriate for phase two of the study (on the relationship between recorded process of care and outcomes of care)?

Methods

Choice of conditions, review methods, settings and staff

The overall research approach was to investigate holistic and criterion-based case note review methods across hospitals, with a range of risk-adjusted levels of 28-day mortality, using two chronic illnesses as tracer conditions. Quality of care was assessed using each of the two review methods by reviewers from three professional groups. Each case note was reviewed using both methods and by between one and four reviewers.

Choice of clinical condition for review

The initial research brief for this study encouraged research teams to consider using a limited range of tracer conditions, mainly concerned with the care for people with chronic conditions. Three conditions were initially proposed for the study – care for people with chronic obstructive pulmonary disease (COPD), heart failure and stroke. Following discussions with the research commissioners, COPD and heart failure were the two conditions chosen for the study.

Chronic obstructive airways disease

About 10% of admissions through hospital UK Emergency Medicine Departments are for people with COPD, which has a high mortality rate at 3 months after index admission. A NICE guideline with review criteria was produced at the commencement of the study.³¹ An extensive set of review criteria were available from the national Royal College of Physicians (RCP) COPD audit, including a limited number that were considered predictive of survival.

Chronic heart failure

People with heart failure often have repeated episodes of hospital readmission. To support our choice of heart failure as one of the two study conditions, we took into account the availability of an evidence-based guideline,²³ together with a limited set of audit review criteria that had recently become available from NICE and was produced by the RCP Clinical Effectiveness and Evaluation Unit (RCP CEEU). The guideline and review criteria also provided a basis for developing, within the study, an externally referenced set of review criteria

for safety and quality assessment for heart failure management. There were no national audit data available.

Cases for review

There are relatively few admissions per hospital per year of new cases of heart failure or COPD and much of the diagnostic work-up is undertaken in the primary care or outpatient setting. We therefore chose to study cases of admission for an exacerbation of either of these two tracer conditions and excluded admissions for diagnostic work-up.

These working definitions for data collection were:

- *Exacerbation of COPD* An exacerbation is a sustained worsening of the patient's symptoms from their usual stable state, which is beyond normal day-to-day variations, and is acute in onset. Commonly reported symptoms are worsening breathlessness, cough, increased sputum production and change in sputum colour.³¹
- *Exacerbation of heart failure* An exacerbation of heart failure is a sustained worsening of the patient's symptoms from their usual stable state, which is beyond normal day-to-day variations, and is acute in onset. Commonly reported symptoms are worsening breathlessness, tiredness and swelling of the feet and/or ankles.²³

Choosing the number of case notes for review

In choosing the number of case notes for review we were unable to use prior hypotheses to assist in determining how many case notes would be required for the reliability studies. We considered using van Belle's method³² of calculating the number of events (e.g. identified opportunities for error as being event = 20 times the number of parameters) from which to assist with this calculation, but subsequently found that it was practically impossible to model the range of opportunities for error presenting in these complex care pathways. In addition, it was decided that the study was more likely to find variations in care rather than identifiable adverse events, and that in health care there could be very large opportunities for error. We therefore took a pragmatic decision to select approximately 50 case notes per condition per hospital. This number also fitted with the custom and practice of the RCP CEEu, in which about 60 case notes per hospital form the basis for review in national clinical audits, and this number

of case notes had previously provided sufficient data for studies of inter-rater reliability.²⁴

For this first phase of the study we therefore sought to obtain sets of 50 case notes from consecutive admissions for each condition in each of eight hospitals. That is, 800 case notes in total.

Selection and recruitment of hospitals and staff

A four-stage process was used to recruit eight study hospitals in England. First, Hospital Episode Statistics³³ on 28-day mortality data for COPD and heart failure were accessed through the East Midlands Public Health Observatory. Hospitals were excluded from the selection process if they reported less than 200 inpatient cases per year for either condition, effectively excluding smaller or specialist acute hospitals. There were 136 hospitals in the final data set.

Second, 28-day mortality data for the two study conditions for each hospital was combined, using simple averaging, to create an average 28-day mortality ratio for each hospital. Third, these were then ranked from the lowest mortality to the highest, and the data was split into four quartiles, each of 34 hospitals. And finally, from this ranking, hospitals were randomly selected in each of the lowest- and the highest-mortality quartiles.

Combinations of review methods and proposed numbers/types of staff

In our initial research proposal we sought to create review teams in each hospital, comprising two types of personnel: clinical staff (for example nurses and staff working in clinical audit departments) and doctors in senior stages of their specialist training [medical specialist registrars (SpRs)]. This choice of types of staff was made in order to test some of the assumptions in the literature⁶ that medically qualified reviewers undertake holistic case note reviews differently from other personnel. Different combinations of reviewer type would review the same records to test inter-rater reliability within types of reviewer and between types of reviewer. Each case note would be reviewed twice by each reviewer: first using holistic methods and then criterion-based methods. We chose this sequence to reduce the bias on the holistic review results that might occur from a reviewer having previously examined the case notes to undertake criterion review.

Subsequent to these initial decisions on numbers of case notes and staff, a more limited set of reviews

was agreed with the research commissioning panel, because the costs of undertaking a full set of two reviews of each of 50 case notes across the eight hospitals using four reviewers per condition proved too great for the available study budget. This second proposal still retained the ability to make comparisons between types of reviewer and review methods, albeit with a smaller number of reviewers in total. *Table 1* indicates the type of reviewers and frequency of review that were proposed. The number next to the code for clinical audit/nursing staff (CA) and physicians (P) indicates the number of reviewers sought for each condition in each hospital. Each reviewer was expected to evaluate 50 sets of case notes.

For each of the two conditions the framework in *Table 1* was used to calculate the number of reviews and reviewers for both the holistic and the criterion-based review methods, so that the total proposed number of case note reviews for the eight hospitals was 3600, using 400 case notes overall for each of the two conditions (800 in all).

Recruitment of hospitals and staff

Recruitment of study participants was a complex and time-consuming process, as participation required the agreement the COPD and heart failure clinical teams at each hospital and was also dependent on the availability of hospital staff to review records for the study. A total of eight hospitals were required for participation in this first part of the study. Because we expected difficulties in recruiting hospitals, mainly a lack of availability of staff within hospitals to review records, a total of 20 randomly selected hospitals (10 in the lower-

mortality quartile and 10 in the higher-mortality quartile) were contacted and invited to participate in the study when only eight were actually required to participate in this phase.

From the initial 20 hospitals contacted, the study recruited five hospitals in each quartile, including one reserve hospital per mortality quartile in an attempt to ensure that a minimum number of eight hospitals were available for the analysis. One reserve hospital subsequently dropped out, leaving a total of nine hospitals in the study – four in the lower-mortality quartile and five in the higher-mortality quartile (*Figure 2*).

Hospitals were invited to participate through contact with one of the specialists in COPD and one in heart failure management. In each selected hospital, consultants specialising in each of the two conditions were approached jointly by the RCP CEEu and the University of Sheffield, with a request that they act as sponsors for the study. Their role was to recruit review staff, within the hospital, who would undertake the data collection for the study. Recruitment of the selected hospitals was completed when two specialists in a hospital agreed to act as sponsor and there were enough staff to undertake the reviews (*Figure 3*).

The proposed number of reviewers required for each condition at each hospital varied from one to four (see *Table 1*), and this request sometimes proved difficult to meet in some hospitals. Among the reasons affecting the recruitment of reviewers included whether the hospital had a dedicated audit department and the change-over time of

TABLE 1 Proposed numbers of reviewers and types of staff for each review method (each staff member to undertake 50 reviews)

	Upper-mortality hospitals				Total
	Hospital 1	Hospital 2	Hospital 3	Hospital 4	
COPD	CA(1)	CA(1)	CA(2) P(1)	CA(2) P(2)	
Heart failure	CA(2) P(1)	CA(1) P(2)	CA(2)	CA(1)	
Total reviews (from 400 case notes)	200	200	250	250	900
	Lower-mortality hospitals				Total
	Hospital 5	Hospital 6	Hospital 7	Hospital 8	
COPD	CA(1)	CA(1)	CA(2)	CA (2)	
Heart failure	CA(2) P (1)	CA(1) P(2)	CA(2) P(1)	CA(1) P(2)	
Total reviews (from 400 case notes)	200	200	250	250	900
Overall total of reviews for each review method					1800

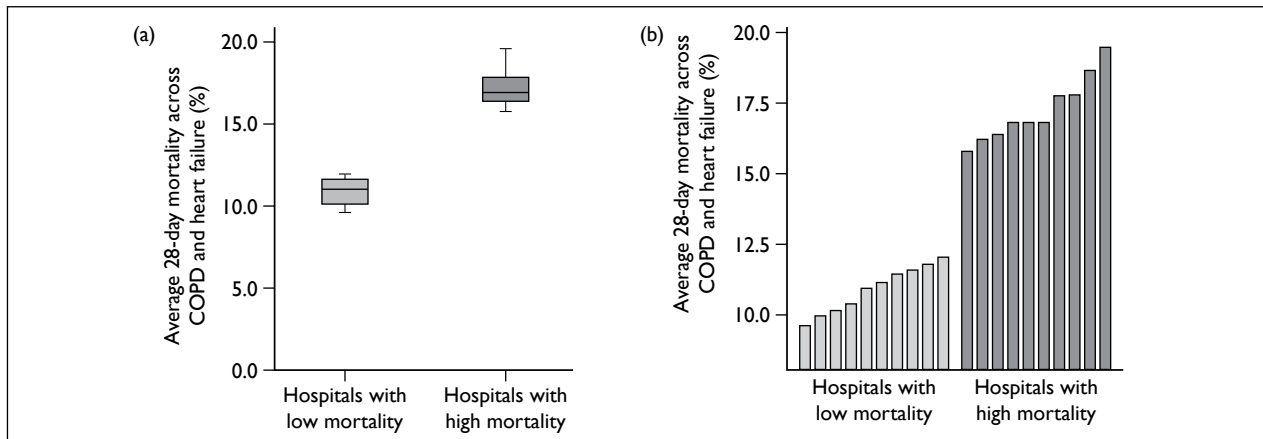


FIGURE 2 Distribution of higher and lower study mortality hospitals. (a) Box plot of COPD/heart failure mortality, showing differences between higher- and lower-mortality groups. (b) Bar chart showing individual hospital-level mortality rates.

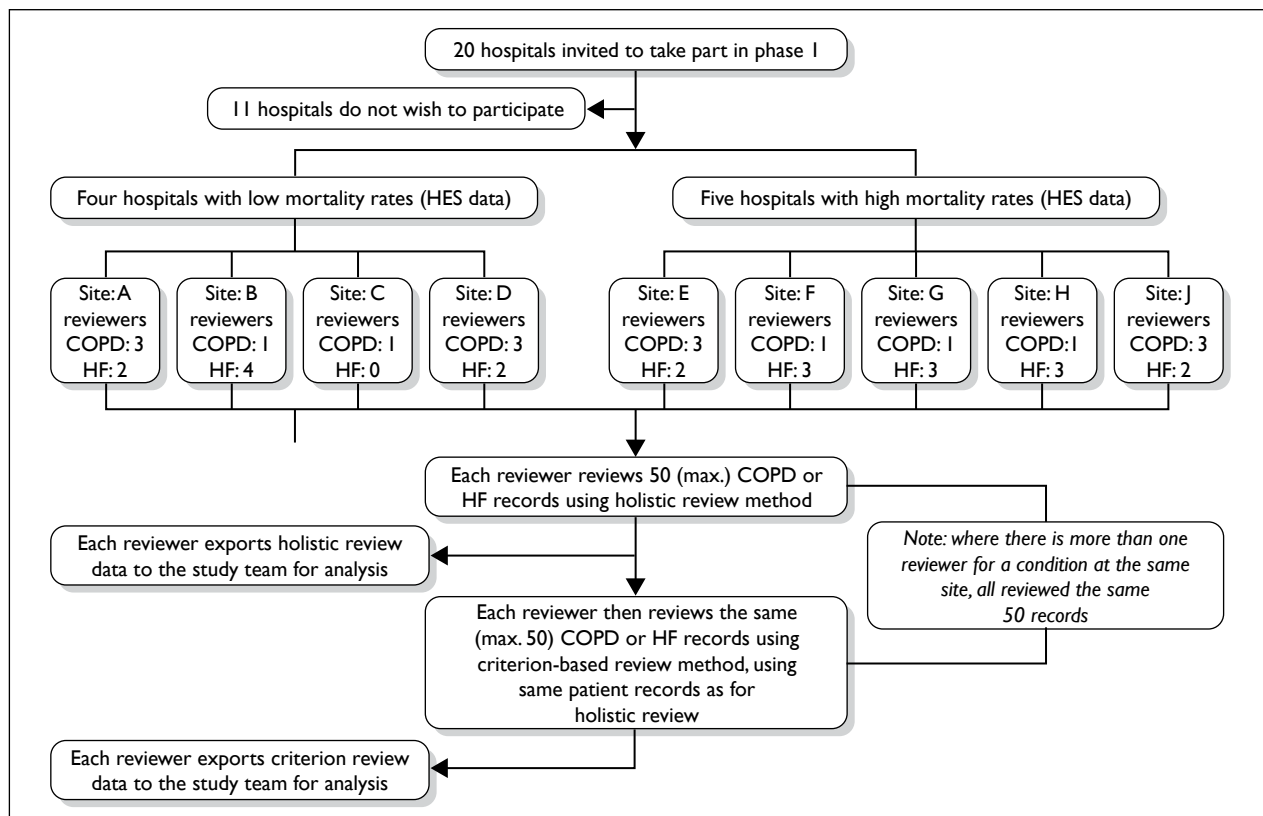


FIGURE 3 Recruitment to the study. HF, heart failure.

SpR training posts. At the end of the recruitment period, three types of hospital staff (reviewers) were engaged in the study – doctors in specialist training, other staff with a clinical background (of whom many were nurses specialising in the care of one of the two tracer conditions) and non-clinical audit staff. Across the nine participating hospitals, the reviewers comprised 16 doctors, 10 specialist nurses (together with one clinically trained audit

person, one pharmacist and one physiotherapist) and nine non-clinical audit staff (i.e. 38 reviewers in total).

Data capture methods

Holistic review

The concept of structured implicit review²⁸ has been found to be valuable in North American

implicit review studies as a means of reducing the variability previously found in inter-rater reliability studies.^{10,11} Structured implicit review attempts to place a framework on data collection by providing headings that can be used in the ‘holistic story’. However, US-based authors working for the RAND Corporation²⁸ chose to use what might be termed a mid-point between criterion-based and textual holistic review, using structured questions that were not as specific as review criteria, but which organised responses to the questions in such a way that might be considered to ask closed questions of the data. See, for example, *Box 1*.

In this study, the concept of structured holistic review was developed to provide reviewers undertaking holistic reviews with a limited structure, but one that was at the same time not so directive as the structured implicit review framework developed by the RAND teams. In doing so, this allowed for different levels of health-care quality to be identified – from excellent care, to care not provided, to the identification of adverse incidents.

Data was captured under three phases of care and for care overall:

- care during the investigation/assessment phase
- care during the initial management phase
- care during the pre-discharge phase
- quality of care overall.

Using this structured holistic framework, reviewers were asked to provide two forms of assessment of quality and safety of care. First, reviewers provided a written assessment of the quality and safety of care of each patient, using information from the case notes (paper and/or electronic records) of the most recent episode of inpatient care for an exacerbation of the illness.

BOX 1 Examples of ‘closed’ structured questions

Was the length of stay appropriate, given the patient’s status at discharge and postdischarge plans?

Definitely yes	___ 1
Probably yes	___ 2
Probably no	___ 3
Definitely no	___ 4

Adapted from Rubenstein *et al.* (1991).²⁸

For the phases of care, reviewers were guided by two prompts:

- Please comment on the care received by the patient during this phase.
- From the records, was there anything in particular worth noting?

Second, reviewers were asked to rate the care received by the patient for each of three phases of care – admission/investigations, initial management and pre-discharge care. Each phase was rated on a six-point scale (1 = unsatisfactory, 6 = very best care), and a definition was provided for each of the points on the scale:

1. Care fell short of current best practice in one or more significant areas resulting in the potential for, or actual, adverse impact on the patient.
2. Care fell short of current best practice in more than one significant area, but is not considered to have the potential for adverse impact on the patient.
3. Care fell short of current best practice in only one significant area, but is not considered to have the potential for adverse impact on the patient.
4. This was satisfactory care, falling short of current best practice in only more than two minor areas.
5. This was good care, which fell short of current best practice in only one or two minor areas.
6. This was excellent care and met current best practice.
7. The format of the questions is set out in *Box 2*.

Next, in assessing the quality of care overall, reviewers were asked to comment on the care received by the patient overall.

An overall quality-of-care rating was requested for each patient review on a 10-point scale (1 = unsatisfactory, 10 = very best care, using only the two anchor points on the scale) to provide for a global rating of care quality. This was given a wider, more fine grained scale so that reviewers could assimilate their perceptions of care for all of the phases of care to give an ‘in the round’ assessment (*Box 3*).

Assessing the quality of recording in the case notes

Evaluation of the quality of care through case note review is critically dependent on the quality of recording in the case notes, together with that

BOX 2 Investigations/examination (for example)

We are interested in comments about the quality of care the patient received and whether it was in accordance with current best practice (for example your professional standards). You may also wish to comment from your own professional viewpoint. If there is any other information that you think is important or relevant that you wish to comment on then please do so.

Please comment on the care received by the patient during this phase.

From the records, was there anything in particular worth noting?

Please rate the care received by the patient during this phase.

Please tick only one box:

Unsatisfactory Very best care

BOX 3 Overall assessment

Please comment on the care received by the patient overall.

Please rate the care received by the patient overall.

Please tick only one box:

Unsatisfactory overall Very best care overall

in associated data sources, such as computerised pathology and radiology results. It might be hypothesised that a poor record could prevent a high-quality retrospective critical review of care. Alternatively, there might be a relationship between poor case notes and poor quality of care. Factors enhancing the use of the record for case note review include the extent to which information is recorded and placed in the case notes, the detail or otherwise of the entry and the legibility of the entry.

It was anticipated that most of the information relevant to the study would be recorded on paper-based case notes but that systems would vary from hospital to hospital, for example in the extent to which the principal case notes provided a holistic record of care or whether medical notes and nursing notes might be held separately.

Reviewers were therefore asked to assess the quality of each record at the end of the holistic review, using a six-point rating scale (1 = inadequate, 6 = excellent):

1. The patient record contains gaps in three or more significant areas.
2. The patient record contains gaps in two significant areas.

3. The patient record contains gaps in one significant area.
4. The patient records are satisfactory and only contain gaps in three or more minor areas.
5. The patient records are good and only contains gaps in one or two minor areas.
6. The patient records are excellent.

Reviewers were asked to complete their assessment in the form shown in *Box 4*.

Review criteria development for COPD and heart failure

Criterion-based review does not seek judgements of care – it requires the reviewer to only identify and record specific items of care. The purpose of review criteria when used in clinical audit is to gather data on which to make a judgement about the quality of care provided by an institution. However, for the purpose of this study, although the quality of care provided by the hospital was useful information, the prime objective was to investigate the extent to which data collection of a case note review method was reliable and in which type of staff’s hands it was most reliable.

This objective meant that the number of review criteria used for each of the two conditions could be limited to a smaller number rather than, for

BOX 4 *Quality of recording*

We are interested in your view about the quality of the patient records in enabling good quality care to be provided.

Please tick only one box:

Inadequate Excellent

example, the full set used by national clinical audit projects (for instance the RCP COPD audit comprised about 75 clinical criteria in total).³⁴

The review criteria were developed using established methods for developing explicit evidence-based review criteria from clinical guidelines.^{19,20,22} That is, for each of the two conditions, the first draft of the criteria were developed from the evidence base in the relevant national clinical guideline^{23,31} and subsequently validated using expert opinion.

COPD review criteria

Information to form the first draft set of criteria came from the national clinical guideline for the management of COPD,³¹ the limited associated set of review criteria from the guideline and, third, from the national RCP clinical audit for COPD.³⁴ From the guideline recommendations and the available review criteria the project team identified a subset of criteria that might be useful in the study.

Refinement of the set was undertaken in three stages. First, the criteria were reviewed to determine whether the required data were likely to be available from case note review. This excluded a number of review criteria used in the national RCP audit,³⁴ which were concerned with organisational effectiveness. Thirty-eight criteria remained.

Second, a questionnaire was sent to a selected group of respiratory physicians to seek their views on the value of the criteria for measuring quality of care. Seventeen senior physicians and specialist nurses ranked the criteria as:

- essential
- desirable
- non-essential.

Eleven criteria were removed as a result of this process.

Third, the structure and wording of each criterion in the data set was reviewed to ensure that it was clear, logical and could be captured from case

notes. At the end of this process there were 37 criteria for COPD care (see Appendix 1).

Heart failure review criteria

A similar approach was taken to the production of heart failure review criteria. A draft set of criteria was developed from information in the national clinical guideline for the management of heart failure²³ and from the limited associated set of criteria for the guideline. Discussion within the project team identified a subset of criteria that might be of value in the study.

Refinement of the set was undertaken in three stages. First, the criteria were reviewed to determine whether the required data were likely to be available from case note review. There were 34 criteria.

Second, a questionnaire was sent to a selected group of cardiovascular physicians and specialist cardiovascular nurses to seek their views on the value of the criteria for measuring quality of care. Ten replies were received. One criterion was removed as a result of this process.

Third, the structure and wording of each criterion in the data set was reviewed to ensure that it was clear, logical and could be captured from clinical records. At the end of this process there were 33 criteria for heart failure care (see Appendix 2).

An example of the external review questionnaire used for COPD can be found at Appendix 3. A similar style of questionnaire was used for review of heart failure.

Developing data capture tools

In order to facilitate the work of the reviewers and data transfer to the study team, data capture was developed through an electronic format based on Microsoft ACCESS[®]. Holistic data capture forms were developed from the format outlined in Holistic review (above), using separate screens for key data, case history data, phases of care and overall care (see Appendix 4, *Figure 33*, for an example). The database was constructed so that information could

be transferred to the study team either by e-mail or by CD, first removing all identifiable data to preserve anonymity for patients and staff. The hospital staff retained access to the full data set to provide for local analysis and audit should they so wish. Criterion-based review data collection fields were created in the same way as those for holistic data.

Because there was considerable variety of local systems and versions of Microsoft ACCESS® in the study hospitals, copies of Microsoft ACCESS® were purchased and made available to the reviewers where required. Provision was also made for staff to collect data on a paper form where data-processing facilities were difficult to access. For these records, data entry was undertaken by the research team from anonymised paper records.

Data were collected from consecutive admissions over a period of 6 months before the review process started in each hospital, a time period which varied slightly, but was approximately between January and July 2005.

Reviewer training and case note selection support

The study sought to provide all reviewers with standardised training in case note review, the emphasis here being to train in the data capture methods. Each reviewer was provided with copies of clinical guidelines for COPD and heart failure care as a means of ensuring that all reviewers had an explicit database of the standards of care expected for the two conditions.^{23,31} Other than providing the guidelines, in this part of the study there was no intention to try to influence each reviewer's own implicit standards for quality of care – that is, each reviewer would have their own, internal, standards for the care that they were reviewing.

During a day-long training session, reviewers were provided with an introduction to the two review methods (particularly as most reviewers were not familiar with the holistic method), together with review software training. Quality-of-care variation was discussed using four theoretical scenarios from stroke care that contained aspects of good and poor care (see Appendix 5). Stroke care was chosen for training to avoid biasing the reviewers in their view of quality of care for the two study conditions.

The challenges of finding information in paper-based records and dealing with missing data were also considered along with a discussion about the means of obtaining case notes from the hospital

records departments. Particular attention was paid to identification of case notes of admissions for exacerbation of known COPD or heart failure (rather than new cases or admissions for a main condition that was not related to the study), and to selecting case notes from the most recent admission.

During the data collection period a telephone helpline was made available throughout office hours if reviewers had any queries or required advice about the data collection. The study team also contacted each reviewer regularly throughout the study period to track progress with the reviews, and liaised with relevant hospital medical record departments if reviewers had problems obtaining records.

Analysis methods

Overall approach

The quantitative analysis was designed to investigate the extent of reliability between individual reviewers and groups of reviewers of the same, and different, professional backgrounds, using measures of internal (intra-rater) consistency, and between-reviewer (inter-rater) reliability for holistic quality-of-care scale scores and criterion-based scores. Correlation and regression analyses were undertaken.

Detailed qualitative analysis of the textual data provided on the phases of care and the overall care was undertaken to explore the relationship between the holistic scale scores for each case and the narrative assessment. This analysis was also used to explore any differences between the results from the different professional groups undertaking the reviews.

Holistic scale score analysis

To assess intra-rater consistency (that is, whether reviewers were internally consistent in their ratings of care) for each individual review, the mean scale score rating was calculated across the three phases of care (admission/investigations, initial management and pre-discharge). The Pearson correlation coefficient was calculated between the mean rating of the three phases (each on a six-point scale) and the overall rating (on a 10-point scale) within each review. The purpose of this analysis was to examine the consistency of the reviewer's scoring across the phases of care and in the final overall care to discover, for example, whether some reviewers might provide quite low

scores for one or more phases of care and then a rather higher score for overall care.

Intraclass correlation coefficients (ICCs) were used as the principal measure of agreement.³⁵ However, although the kappa statistic is susceptible to prevalence (in this case, of 'opportunity for error' rates per set of records),^{10,11} kappa scores were also computed as measures of agreement for overall scores (see *Table 6b* below), as this measure of agreement is more commonly used in the literature.

To assess inter-rater reliability between ratings of the same record by different reviewers, raw ratings were converted to ranks to adjust for variation in the range of scores used by different reviewers and ICCs were calculated on these ranks.

Measuring reliability between reviewer pairs

The ICC gives the correlation between any two measurements or ratings for the same subject or patient, using randomly chosen methods or reviewers. ICCs are based on continuous data, unlike kappa statistics, which require the data to be categorical. ICCs were used to assess the reliability between reviews of the same patient records carried out by pairs or groups of reviewers (e.g. two nurses or two doctors) at the same hospital and were calculated first between the holistic quality-of-care ratings allocated by the two reviewers, and, second, between the criterion-based scores.

When undertaking the holistic (implicit) review, each reviewer rated the overall quality of care received by each patient against a 10-point scale. It is possible that different reviewers may have interpreted the rating scale differently (e.g. one reviewer may tend to give higher or lower ratings than another). Therefore, each reviewer's ratings were converted to a rank. For example, if a reviewer reviewed 50 records then the ratings were ranked from 1 to 50. (In the event of tied ratings, the average rank was used.) The reliability between these ranked ratings for each pair of reviewers was then assessed using ICCs.

For the criterion-based review, care was assessed against a set of condition-specific criteria for either COPD or heart failure care. ICCs were used to assess the inter-rater reliability between the overall number of criteria (as unranked criteria scores) noted by each reviewer as having been met.

Average reliabilities per staff type

To provide an overview of the average reliability for each staff type (e.g. doctors versus doctors, nurses versus nurses), a pooled or overall mean ICC was calculated across all the reviewer pairs in each staff group. Because some reviewer pairs had reviewed more records than others, each ICC was weighted when calculating the overall mean ICC, with the weight being proportional to the inverse of the variance of the ICC estimate.

Sites with more than two reviewers of different types

For sites where there were two reviewers of one staff type plus one of another type (e.g. two doctors and one nurse), we wished to avoid counting the same nurse twice in the comparison with the doctors. Therefore, the mean of the two doctors' scores for each record was calculated (and the mean holistic scores converted to a rank). An ICC was calculated between the mean score from the two doctors and the score from the nurse. This approach was used whenever there were odd numbers of a reviewer type in this analysis. The ICC was then combined with the doctor–nurse ICCs from other sites to calculate an overall mean ICC for doctors versus nurses, weighting by inverse variance as described above. At site B there were three, rather than two, doctors reviewing the same records. Therefore, a single ICC was calculated among all three doctors' scores at this site.

For the purpose of the analysis, care was rated on a three-point scale: (1) care fell short of current best practice (unsatisfactory); (2) satisfactory; and (3) good or excellent care. We considered whether it might be possible to reduce down the scale score data to a binary 'poor' or 'good' score to enable direct comparisons on a 2×2 table between the two review methods, but this approach would reduce the spread of judgements even further from the six- or 10-point scales and would not accommodate the range of judgements offered by the reviewers.

Criterion-based review

Data were scored in two ways, first to assess the completeness of the data and to assess the effectiveness of each reviewer type at completing the data collection form, and, second, to calculate a quality-of-care score for each review.

An 'effectiveness of reviewer' score for each record review comprised one point allocated for each data field completed by the reviewer (irrespective of whether the criterion was recorded as being met or

not being met), and one point subtracted for every data field left blank by the reviewer. These scores were converted to a percentage.

Quality-of-care scores for each record comprised the percentage of the criteria identified by the reviewer as having been met. ICCs were used to estimate inter-rater reliability for overall scores by pairs or triplets of staff reviewing the same records. Because some phases of care generated only a small number of criteria, ICCs were not computed for phases of care.

Intraclass correlation coefficient estimates from the different combinations of reviewers were pooled using a weighting that was inversely proportional to the variance of the estimate.³⁶

Comparison of holistic scale scores and criterion-based review

Inter-rater reliability results for each of the two review methods were compared. Additionally, an estimate of the within-staff-type consistency across the two review methods was calculated using *p*-values for differences between the overall holistic quality-of-care ratings and the percentage of criteria recorded as being met.

Comparison of quality scores with hospitals grouped by mortality level

In the original call for proposals for this study it was suggested that the quality-of-care scores might be risk-adjusted by severity of illness of each case. However, we agree with Daley and colleagues³⁷ that risk adjustment remains a controversial and difficult subject. Pitches *et al.*³⁸ undertook a systematic review of 36 studies, which included 51 'process versus risk-adjusted mortality relationships', exploring the extent to which variations in risk-adjusted mortality rates were associated with differences in quality of care. They found a positive correlation in only 51% of the relationships, with no correlation in 31% and an unexpected correlation in a further 18%, in what was a very heterogeneous set of studies.

A range of approaches to risk adjustment were considered in the initial phases of the study. Because of the complexity of data capture and the level of workload that could reasonably be asked of the reviewers, individual scoring of the risk for individual patients proved impossible. The chosen approach was therefore to compare aggregate quality-of-care results between the hospitals in the low- and high-mortality-rate groups.

Analysis of holistic textual data

The comments made by the reviewers in their holistic reviews were in two free text areas. The first asked them to comment on the care received during a particular phase of care or for care overall. The second asked if there was anything in particular worth noting from the records about the care (see *Box 2*, above).

Reviewers provided textual data when commenting on each of the four phases of care and the overall assessment of care. These data were analysed in two ways in order to address the question of whether, as suggested by Weingart and colleagues,⁶ different staff types were concerned with different elements of care when making their holistic assessments. Data from the past medical history heading was excluded from the analysis because there were few comments, and all were about the case notes rather than about care.

Content analysis

The primary approach was a content analysis drawing on grounded theory.³⁹ The textual responses provided insights into the different ways that different individuals and different professional groups interpreted the task, as well as their interpretation of care provided. By analysing textual responses we were able to investigate similarities and differences between individuals about their interpretation of the same record, construct pictures of how professional groups interpreted the task and viewed care provision, and give an indication of the concepts that they used.

Categorising and coding types of comment made by reviewers

Following familiarisation with the textual responses, it became clear that different types of comments were given, reflecting, in large part, different reviewer types. A categorisation was developed that identified these different responses, irrespective of professional background. These categories could be thought of as hierarchical, to a degree, if set in the context of what might constitute an ideal review. Thus, at the lower end of the hierarchy were no comments, or limited comments, about the record rather than the care, ranging through different types of comment about care to the higher end of the hierarchy where the most discerning reviews picked up more complex issues. At the upper end there was a clear cluster of issues commented on overall, displaying a fairly sophisticated degree of reviewing. While the concepts emerged from the data, the labels attached to these categories were

developed by the research team. The concept terms were not used by the reviewers.

An initial set of codes was developed by five analysts in group discussion. Output from three reviewers was then reviewed by two pairs of analysts (each separately), one pair examining COPD comments and the other examining heart failure comments. A fifth analyst examined all of the comments. Each pair of analysts discussed their experience of using the comments and compared their results. The results of the initial analysis and commentary on the utility of the coding framework were then discussed by the group, moderated by the fifth analyst and refinements were made to the coding frame.

These categories were then used to code all responses that made up the responses for each review. Since some of the responses were made up of a number of separate comments, the code given was the 'highest-level' category used in each of the comments. For each phase of care, and for care overall, up to four codes were allocated by the analyst. The analysis reported here refers to the highest level of code allocated by a reviewer for overall care for each set of case notes and refers only to the data collection item 'Please comment on the care the patient received'. Thirteen coding

categories were developed and these were also grouped into three broader categories (*Box 5*).

This categorisation was subsequently used to help identify the types of reviewing undertaken by different professional groups, which, in turn, assisted the decision on which group or groups of professionals best matched our requirements in the review process. Together with this categorisation of the type of reviewing being undertaken, the textual analysis was also used to identify specific issues raised by reviewers and to see if they varied by professional group and by individual for the same record. Careful examination was undertaken of the particular words, phrases and style used in each comment, although not to the level of a discourse analysis. The constant comparisons allowed us to generate categories (themes) to identify different approaches to reviewing, different content of reviews, and contrasts between individuals and professional groups.

Resource analysis

In the initial proposal to the funding body we set out a proposal for a cost-benefit analysis of the two review methods. Because of resource constraints this did not form part of the final agreement. However, we decided to undertake a limited resource analysis in case choices on reviewer type

BOX 5 Coding categories

Code	Highest-level comment used in the each review	Broad category description
1	Blank	Codes 1–5: little or no comment about care and little or no judgement
2	No comment or other words to indicate nothing to say	
3	Description of what's in the record	
4	Judgement of record (not care they received)	
5	Description of what happened to patient (not care they received)	
6	Description of care delivered	Codes 6–8: limited comment about care and implied judgement
7	Description of omission of care	
8	Implied judgement of care (not records or patient pathway)	
9	Explicit judgement of care (not records or patient pathway)	Codes 9–13: sophisticated comments about care with explicit judgements and views
10	Questioning/query of care delivered	
11	Explanation/justification of care delivered	
12	Alternative/justification of care that should have been delivered	
13	Concerns	

might essentially be made on cost (where there were limited differences in review results between one or more types of reviewers or review methods).

The resource impact of each reviewer type was explored, based on self-reported data on the time taken to undertake each review and on annual staff cost data taken from Unit Costs of Health and Social Care 2005 (clinical and doctor reviewers) and mid-point administrative and clerical staff costs from Whitley Council pay rates (for non-clinical audit reviewers).⁴⁰ The mid-point on the scale was used as the cost for each staff type and only one cost for each staff type was used in the analysis.

Descriptive statistics for the time taken to undertake each review and the cost per review for each staff type were produced in SPSS. A mean time per review and a mean cost per review for each staff type was calculated. We also included the minimum and maximum range to look more closely at the spread of the data.

Research ethics review

A research ethics review of the study was sought from the Trent Multi-centre Research Ethics Committee on 21 July 2004, prior to the start of data collection. Because, in both phases, data were to be collected by staff working in each hospital, and the data were anonymised before transmission to the research team, the Committee considered this to be equivalent to a national audit programme. The Trent Multi-centre Research Ethics Committee response was therefore that the study did not require an ethics opinion from the Committee.

Research governance

The potential need for research governance review existed in both parts of the study. However, because the data collection was being undertaken by hospital staff, the results were available to the individual hospital and the research team were not undertaking data collection on the hospital premises, the project was seen by research governance departments as akin to the national clinical audit programmes from which learning is derived as a result of the use of anonymous collated data. No study hospital required that a full research governance review should be undertaken, although initial discussions were held with a number of research governance teams and the offer to

undertake the governance review process was made to all hospitals.

Results

Across the nine hospitals, 38 reviewers undertook 1473 holistic reviews and 1389 criterion-based reviews (a total of 684 clinical records were reviewed). The numbers of case notes reviewed by each individual ranged from nine to 50 (*Table 2*). Variation in the numbers of reviews achieved was related to job rotations, local workload pressures and difficulties in obtaining clinical records.

Quality of case note recording

The mean quality of case notes rating for COPD and heart failure were 4.3 (SD 1.2) and 4.7 (SD 0.9), respectively, on a scale of 1–6, indicating a reasonable overall quality of recording in the paper-based notes.

Analysis of holistic review scale scores

Completion rates for scale scores

Data returned by reviewers was checked for completion rates. *Tables 3 and 4* show completion rates in excess of 90% for all phases of care, except for the overall phase assessment completion rate for COPD reviews by non-clinical audit staff.

Intra-rater consistency in holistic reviews

Across all three staff types there were statistically significant correlations ($r > 0.71$, $p < 0.001$) between the mean scale score ratings that reviewers assigned to the individual phases of care and their rating of the overall quality of care, indicating a fair to good level of intra-rater consistency in rating the quality of care using holistic review scale scores (*Table 5*). Reviewers appeared to be relatively consistent in the way that they scored quality of care for the phases of care in a case and then gave an overall assessment score for the episode of care.

Inter-rater reliability for holistic review

Holistic review reliability between scale score ratings of the same record by pairs of reviewers was fair within all three staff types, although it varied from one reviewer pair to another and for some pairs was very poor (*Table 6a*). The overall weighted mean ICC was fair across all three types of reviewers, with no significant differences between staff types. *Table 6b* displays the same analysis using kappa statistics. The same trend occurs as in *Table*

TABLE 2 Summary of number of case note reviews and review staff per hospital

Site	COPD			Heart failure		
	Review staff types	Holistic reviews	Criterion-based reviews	Review staff types	Holistic reviews	Criterion-based reviews
A	Non-clinical audit	49	30	Doctor	11	11
	Non-clinical audit	49	44	Non-clinical audit	12	12
	Doctor	48	33			
B	Non-clinical audit	50	50	Non-clinical audit	49	49
				Doctor	49	47
				Doctor	49	46
				Doctor	49	46
C	Nurse/other clinical	49	19			
D	Nurse/other clinical	49	48	Nurse/other clinical	21	21
	Nurse/other clinical	50	50	Nurse/other clinical	21	21
	Doctor	34	34			
E	Non-clinical audit	42	41	Doctor	14	14
	Non-clinical audit	43	43	Doctor	14	14
	Doctor	41	37			
F	Nurse/other clinical	46	46	Non-clinical audit	9	10
				Doctor	22	14
				Doctor	48	47
G	Nurse/other clinical	35	35			
	Non-clinical audit	38	36			
	Doctor	50	50			
	Doctor	50	50			
H	Nurse/other clinical	49	50	Nurse/other clinical	50	50
				Nurse/other clinical	50	50
				Doctor	49	50
J	Nurse/other clinical	30	29	Nurse/other clinical	30	29
	Nurse/other clinical	49	29	Doctor	25	24
	Doctor	50	50			
Total	20 review staff	901	834	18 review staff	572	555

6a, that is, the doctor reviewers have a higher level of agreement than the other staff types, although the results for the nurse/clinical group and the non-clinical audit staff are somewhat lower than for the ICC analysis. This is likely to be due to differences in the two methods of analysis.

Comparisons between professional groups

Where reviewers from different staff types used holistic scale score methods to review the same record, inter-rater reliability was assessed within and between staff groups for all phases of care and

overall care (Table 7). For the phase of care findings within staff groups there was generally modest to fair agreement within pairs, particularly among the doctors, although even in this group the range was large (see, for example, the initial management results), and, as we have stated in the paragraph above (relating to Table 6a), the level of agreement varied from one reviewer pair to another and for some pairs was very poor. However, where staff from different groups reviewed the same record, agreement between the different professional groups on their assessment of the quality of care was poor to non-existent.

TABLE 3 Completion rates for COPD holistic reviews

	Admission and investigations phase (%)	Initial management phase (%)	Pre-discharge phase (%)	Overall phase (%)	Total (%)
Doctors (n = 273 reviews)					
Number of completed rating scales	269 (98.5)	265 (97)	267 (97.8)	272 (99.6)	1073 (97.4)
Missing data	4 (1.5)	8 (3)	6 (2.2)	1 (0.4)	19 (2.6)
Non-clinical audit (n = 271 reviews)					
Number of completed rating scales	260 (96)	261 (96.3)	263 (97)	227 (83.8)	1011 (93.3)
Missing data	11 (4)	10 (3.7)	8 (3)	44 (16.2)	73 (6.7)
Clinical (n = 357 reviews)					
Number of completed rating scales	341 (95.5)	332 (93)	326 (91.3)	353 (98.9)	1352 (94.7)
Missing data	16 (4.5)	25 (7)	31 (8.7)	4 (1.1)	76 (5.3)

TABLE 4 Completion rates for heart failure holistic reviews

	Admission and Investigations phase (%)	Initial management phase (%)	Pre-discharge phase (%)	Overall phase (%)	Total (%)
Doctors (n = 330 reviews)					
Number of completed rating scales	320 (97)	323 (98)	300 (91)	322 (98)	1265 (96)
Missing data	10 (3)	7 (2)	30 (9)	8 (2)	55 (4)
Non-clinical audit (n = 70 reviews)					
Number of completed rating scales	69 (99)	70 (100)	68 (97)	70 (100)	277 (99)
Missing data	1 (1)	0 (0)	2 (3)	0 (0)	3 (1)
Clinical (n = 180 reviews)					
Number of completed rating scales	170 (99)	172 (100)	170 (99)	171 (99)	708 (99)
Missing data	2 (1)	0 (0)	2 (1)	1 (1)	4 (1)

TABLE 5 Intra-rater consistency between holistic scale score ratings for phases of care and for overall care

Review staff type (number of review staff)	Number of reviews	Mean overall rating of quality of care (SD)	Mean rating of phase quality of care ^a	Pearson correlation between mean rating across three phases of care and overall rating (p-value)
Doctors (16)	593	7.8 (1.8)	4.7 (0.8)	0.77 (<0.001)
Nurses/other clinical (14)	529	7.0 (2.0)	4.4 (1.0)	0.81 (<0.001)
Non-clinical audit (9)	296	7.9 (1.3)	4.6 (0.8)	0.71 (<0.001)

a Based on the mean score across three phases of care. Overall quality of care was rated on a 1 (unsatisfactory)–10 (very best care) scale. Quality of care in each of the three phases (admission/investigations, initial management and pre-discharge) was rated on a 1 (unsatisfactory)–6 (very best care) scale.

The overall 'quality-of-care' score for both holistic and criterion-based methods used across the 684 patient records was similarly rated by the three staff types (between 70% and 79%, where 100% is very best care). Analysis of variance between the holistic overall scale ratings of the three staff types show that the nurse/other clinical group scores were significantly lower than the doctor ($p < 0.001$) and non-clinical audit groups ($p < 0.001$). The comparison of the last two groups showed no significant differences ($p = 0.352$).

Analysis of review criterion-based scores

Criterion-based reviewer effectiveness

Effectiveness scores relate to the ability of the reviewer to find and access the data in the case record. For each criterion one point was allocated for each data field completed by the reviewer (irrespective of whether the criterion was recorded as being met or recorded as not being met) and one point subtracted for every data field left blank by the reviewer. Effectiveness in capturing criterion-based data was high and similar across all three staff types (Table 8), with mean scores all being around 95% (that is, an average of approximately 1.5 data items missing for each review).

Inter-rater reliability for criterion-based review

Inter-rater reliability between criterion-based scores (that is, the percentage of criteria recorded as being met) for the same record by different reviewers ranged from moderate to good within all staff types, although with the doctors showing a significantly higher level of reliability (Table 9).

Comparison of holistic and criterion-based methods

Table 10 shows the results of a comparison between holistic review and criterion-based review methods, using 'quality-of-care' scores. Reviewers rated the overall quality of care on a 10-point scale from 1 (unsatisfactory) to 10 (very best care). This was converted to a percentage for comparison with criterion-based review data. Criterion-based quality-of-care scores are shown as percentages out of 32 criteria (where patient is a current or ex-smoker) or out of 31 criteria (where patient is a non-smoker).

Mean overall quality-of-care scores were similar for both holistic and criterion-based methods, and also for all three staff types (scores of between 70% and 79%, where 100% is excellent care).

Paired individual data was used for the comparison: that is, the score for each criterion review of a case note minus the overall score for the holistic review of the case note.

There were 1109 paired sets of case note reviews in total (some reviewers only undertook one type of review on some case notes), so the paired review numbers are smaller than the possible total of 1384 reviews. For the purposes of the analysis there are 1109 differences (criterion review score minus overall holistic review score). The confidence intervals (CIs) and p -values are adjusted for clustering by reviewer (in this case, 38 reviewers).

Estimation of the level of quality-of-care score agreement between the two methods for an individual record, using p -value for difference, shows that there was no significant difference

TABLE 6a Inter-rater reliability (ICC) between holistic overall ratings of the same record by paired reviewers of the same staff type

Reviewer pairs	Condition	Site ^a	Number of paired reviews	ICC between ranked scores (95% CI)	Weighted mean ICC ^b (95% CI)
Doctor vs doctor	Heart failure	B ^c	49	0.67 (0.54 to 0.79)	0.52 (0.41 to 0.62)
	COPD	G	48	0.33 (0.05 to 0.56)	
	Heart failure	F	18	-0.03 (-0.48 to 0.43)	
	Heart failure	E ^d	12	-0.44 (-0.80 to 0.15)	
Nurse/clinical vs nurse/clinical	Heart failure	D	21	0.74 (0.47 to 0.89)	0.46 (0.34 to 0.59)
	COPD	D	49	0.37 (0.10 to 0.58)	
	COPD	J	26	0.27 (-0.12 to 0.59)	
	Heart failure	H	48	0.22 (-0.07 to 0.47)	
Non-clinical audit staff vs non-clinical audit staff	COPD	A	48	0.47 (0.22 to 0.66)	0.47 (0.22 to 0.66)

a Only sites with more than one reviewer of the same staff type are included in this table.
 b Mean ICC per staff type, weighted by inverse variances to account for differing numbers of paired reviews.
 c A single ICC was calculated for the three doctors at site B.
 d The doctors at site E were non-specialist doctors.

TABLE 6b Kappa agreement statistics for holistic overall ratings of the same record by paired reviewers of the same staff type

Reviewer pairs	Condition	Site ^a	Number of paired reviews	Kappa (95% CI)	Weighted mean kappa ^b (95% CI)
Doctor vs doctor	Heart failure	B ^c	49	0.60 (0.48 to 0.72)	0.51 (0.40 to 0.61)
	COPD	G	48	0.25 (-0.01 to 0.51)	
	Heart failure	F	18	0.35 (-0.14 to 0.83)	
	Heart failure	E ^d	12	0.00 (-0.57 to 0.57)	
Nurse/clinical vs nurse/clinical	Heart failure	D	21	0.34 (0.02 to 0.67)	0.22 (0.08 to 0.36)
	COPD	D	49	0.26 (0.03 to 0.48)	
	COPD	J	26	0.16 (-0.17 to 0.48)	
	Heart failure	H	48	0.10 (-0.20 to 0.40)	
Non-clinical audit staff vs non-clinical audit staff	COPD	A	48	0.30 (-0.01 to 0.61)	0.30 (-0.01 to 0.61)

a Only sites with more than one reviewer of the same staff type are included in this table.
 b Mean kappa per staff type, weighted by inverse variances to account for differing numbers of paired reviews.
 c A single kappa was calculated for the three doctors at site B.
 d The doctors at site E were non-specialist doctors.
 Overall care was rated on a three-point scale: (1) care felt short of current best practice (unsatisfactory); (2) satisfactory; and (3) good or excellent care.

between the holistic and criterion-based assessments when used by the doctors (p -value for difference 0.406) and by the non-clinical audit staff (p -value for difference 0.223).

However, there was a difference (that is lack of agreement) between the two scores rated by

the nurse/other clinical group of reviewers. It is possible that this was because of the differences between criterion-based methods and holistic methods of review. The review criteria tended to be clinical measurement based, certainly in the admission and initial management phases, whereas the qualitative data from holistic reviews (see

TABLE 7 Within-staff-type ICC results and between-staff-type group comparisons of inter-rater reliability of holistic scale scores for phases of care and overall care

Reviewer pairs	Number of reviewer pairs (or triplets)	Number of case notes	Weighted mean ICC ^a between ranked scores				
			Admission/ investigations and examinations	Initial management	Pre-discharge	Overall	
Within-staff-type ICC results							
Doctor vs doctor	4	127	Weighted mean	0.58	0.70	0.46	0.52
			95% CI	0.48 to 0.68	0.63 to 0.78	0.34 to 0.59	0.41 to 0.62
			Range	-0.41 to 0.72	-0.31 to 0.81	-0.01 to 0.55	-0.44 to 0.67
Nurse/clinical vs nurse/clinical	4	144	Weighted mean	0.50	0.22	0.43	0.46
			95% CI	0.38 to 0.62	0.07 to 0.37	0.30 to 0.55	0.34 to 0.59
			Range	0.24 to 0.76	-0.12 to 0.41	-0.04 to 0.77	0.22 to 0.74
Non-clinical audit staff vs non-clinical audit staff	2	87	Weighted mean	0.35	0.10	0.39	0.47
			95% CI	0.16 to 0.54	-0.10 to 0.30	0.21 to 0.57	0.22 to 0.66
			Range	0.31 to 0.38	-0.11 to 0.27	0.32 to 0.45	0.47 to 0.47
Between-staff-type comparisons							
Doctor vs nurse/clinical	5	179	Weighted mean	0.23	0.25	0.29	0.43
			95% CI	0.09 to 0.37	0.12 to 0.39	0.16 to 0.43	0.31 to 0.54
			Range	0.03 to 0.38	0.02 to 0.41	-0.21 to 0.63	-0.06 to 0.67
Doctor vs non-clinical audit staff	6	188	Weighted mean	-0.01	0.03	0.25	0.24
			95% CI	-0.15 to 0.12	-0.11 to 0.16	0.12 to 0.38	0.12 to 0.37
			Range	-0.15 to 0.67	-0.53 to 0.45	-0.16 to 0.71	-0.39 to 0.54
Nurse/clinical vs non-clinical audit staff	1	34	Weighted mean	-0.12	0.19	0.47	0.43
			95% CI	-0.44 to 0.23	-0.15 to 0.49	0.17 to 0.70	0.11 to 0.67
			Range	-0.12 to -0.12	0.19 to 0.19	0.47 to 0.47	0.43 to 0.43
a Weighted mean ICC: estimates from the different combinations of reviewers were pooled using a weighting that was inversely proportional to the variance of the estimate.							

TABLE 8 Criterion-based reviewer effectiveness scores

Review staff type (number of review staff)	Number of reviews ^a	Mean score %, SD (95% CI) ^b	Range
Doctor (16)	477	94.9, 4.8 (93.2 to 96.5)	74.2–100.0
Nurse/other clinical (14)	443	95.2, 4.1 (93.5 to 97.0)	67.7–100.0
Non-clinical audit (9)	289	94.7, 5.0 (93.2 to 96.5)	61.3–100.0
Total (39)	1209	95.0, 4.6 (94.0 to 95.9)	61.3–100.0

a Analysis excludes patients who died.
b 95% CIs are adjusted for clustering by reviewer.

TABLE 9 Inter-rater reliability between criterion-based scores (proportion of criteria stated as being met) for the same record by different reviewers

Reviewer pairs	Condition	Site ^a	Number of paired reviews	ICC between scores (95% CI)	Weighted mean ICC ^b (95% CI)
Doctor vs doctor	Heart failure	F	14	0.96 (0.87 to 0.99)	0.88 (0.83 to 0.93)
	COPD	G	50	0.65 (0.46 to 0.79)	
	Heart failure	B	46	0.65 (0.50 to 0.77)	
	Heart failure	E ^c	12	0.64 (0.13 to 0.88)	
Nurse/clinical vs nurse/clinical	COPD	J	25	0.86 (0.71 to 0.94)	0.74 (0.66 to 0.82)
	COPD	D	48	0.70 (0.52 to 0.82)	
	Heart failure	D	21	0.69 (0.38 to 0.86)	
	Heart failure	H	50	0.27 (0.00 to 0.51)	
Non-clinical audit staff vs non-clinical audit staff	COPD	E	40	0.69 (0.49 to 0.82)	0.61 (0.47 to 0.76)
	COPD	A	29	0.33 (–0.04 to 0.61)	

a Only sites with more than one reviewer are included in reliability analysis, therefore some sites do not appear on this table.
b Mean ICC per staff type, weighted by inverse variances to account for differing numbers of paired reviews. A single ICC was calculated for the three doctors at site B, and this was combined with the other doctor pairs in the weighted mean ICC.
c Non-specialist doctors.

below) suggest that the nursing-trained reviewers were quite strongly influenced by the quality and effectiveness of care pathways. Holistic review results may therefore be demonstrating a view that is more nursing focused than do the selected review criteria.

Bland–Altman plots of the difference in score between the review methods, against the average of the two scores, can also be used to examine the size of the differences and also their distribution around zero. The plot also allows for a visual check to determine whether the differences are (or are not) related to the size of the measurement.

For the purpose of this study, the average reviewer score across both methods acts as the best estimate

of the true value. The mean difference, in review method scores, is an estimate of the average bias of one method relative to another. The standard deviation (SD) of the differences, or the 95% limits of agreement, can be used to see how well methods are likely to agree for an individual. For a systematic distribution we expect the range ($\text{mean} \pm 2 \text{SD}_{\text{DIFFERENCE}}$) to include about 95% of the observations. This range of values defines the 95% limits of agreement.

Figures 4–6 show that the reviewers tended to rate the majority of records with a mean combined holistic and criterion score of between 40 and 100, with very few records having lower scores reflecting poorer care. There is some evidence of a systematic pattern, in all three plots, which suggests that at

TABLE 10 Mean ratings/scores of overall quality of care – paired comparison of two review methods

Staff type	Number of holistic and criterion-based reviews ^a (and review staff)	Criterion-based review mean score as a percentage of total criteria ^b (95% CI)	Holistic mean rating of overall quality of care ^c (95% CI)	Mean difference (95% CI)	p-value for difference
Doctor	462 (16)	78.7 (77.1 to 80.4)	76.8 (72.2 to 81.4)	-1.9 (-6.7 to 2.9)	0.406
Nurse/other clinical	428 (14)	77.5 (75.0 to 80.1)	71.2 (66.4 to 76.0)	-6.3 (-10.5 to -2.2)	0.005
Non-clinical audit	219 (8)	75.4 (71.1 to 79.7)	78.5 (74.7 to 82.3)	3.1 (-2.4 to 8.5)	0.223
All staff	1109 (38)	77.6 (76.2 to 79.0)	75.0 (72.3 to 77.6)	-2.6 (-5.4 to 0.1)	0.057

a Only paired reviews are included in the analysis – that is, holistic and criterion review undertaken by the same reviewer on the same record.

b Scores are shown as percentages out of 32 criteria (where patient is a current or ex-smoker) or out of 31 criteria (where patient is a non-smoker).

c Reviewers rated the overall quality of care on a 10-point scale from 1 (unsatisfactory) to 10 (very best care). This was converted to a percentage for comparison with criterion-based review data.

lower average scores (up to 60) the holistic-based score tends to be less than the criterion-based score for the same reviewer/patient (i.e. negative differences). At higher average scores, above 60, then there is evidence of the opposite pattern, i.e. positive differences, which implies that the holistic-based scores are larger than the criterion-based review scores for the same patient. These patterns, at both the higher- and lower-level scores, may be reflecting the methodological differences of measuring quality of care. While the criterion-based scores are rigid – the item is either present or absent – holistic scoring allows the reviewer to make a judgement, which might be ‘harsher’ than criterion-based scoring at lower-quality levels and more ‘favourable’ at higher-quality levels, as many more factors may be taken into account in the judgement in holistic reviewing. Evidence from the following section on the analysis of textual data may also support this hypothesis.

Both *Table 10* and *Figures 4–6* suggest that there was more variation in the holistic review results than in the criterion-based review. The reason for this may be due to the differing nature of the two review methods, as criterion-based review is by its nature very structured.

Thematic analysis of holistic textual data

Textual comments on the quality of care were sought from reviewers as part of the holistic review

process and a textual analysis was undertaken where any type of response was given to either of the following questions:

- Please comment on the care received by the patient during this phase (the first box of the data collection form).
- From the records, was there anything in particular worth noting? (The second box of the data collection form.)

There was variation in the type and amount of comments given by reviewers. Some reviewers gave no response, others gave one or two words, while a list type response was given by some and extensive narratives were provided by other reviewers. The data are presented by staff type, followed by an overall summary analysis.

Non-clinical audit staff

Several non-clinical audit staff reviewers made no comment in the first box in most instances. In a few instances, those who made no comment did, however, offer a comment in the more general comments box (the second box). Among this reviewer group, the comments in the second box were sometimes about documentation rather than about the care delivered.

A relatively common approach by non-clinical audit staff was to present a list, of things that had been done or requested, or, in some instances, things that had not been done. In some instances the lists

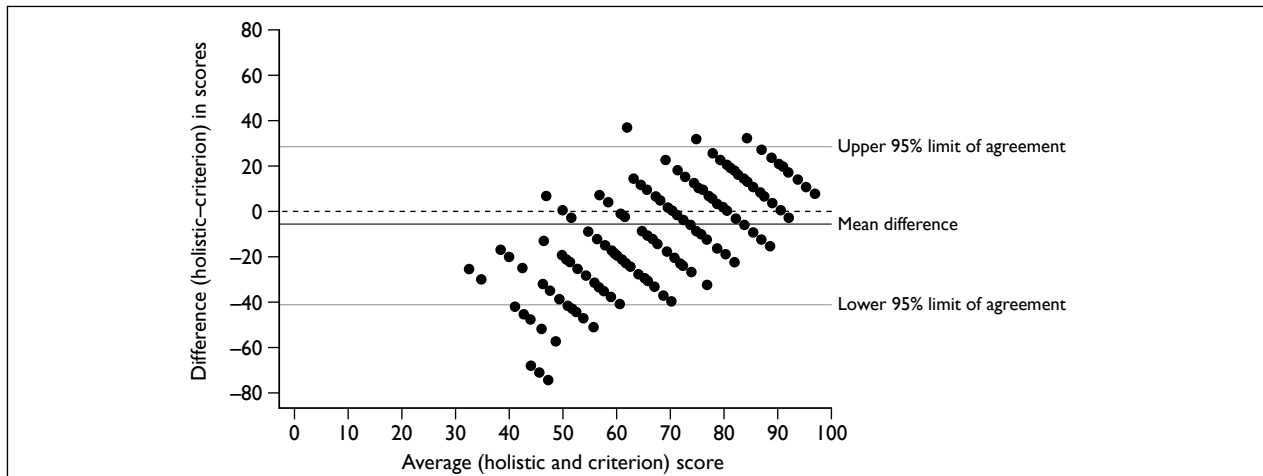


FIGURE 4 Plot of difference in holistic compared with criterion quality-of-care score versus average score – staff type: clinical.

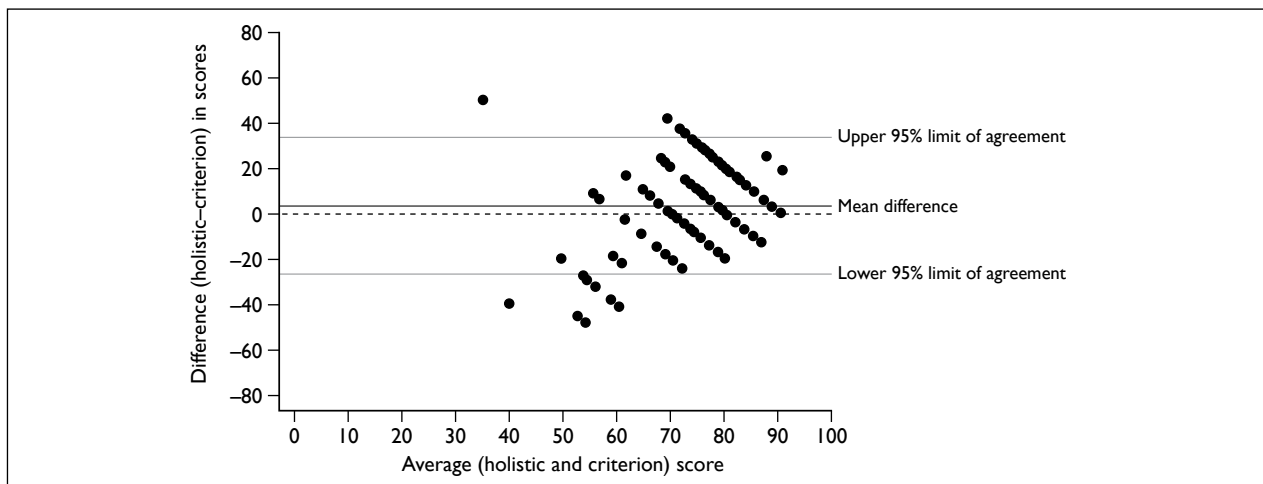


FIGURE 5 Plot of difference in holistic compared with criterion quality-of-care score versus average score – staff type: non-clinical audit.

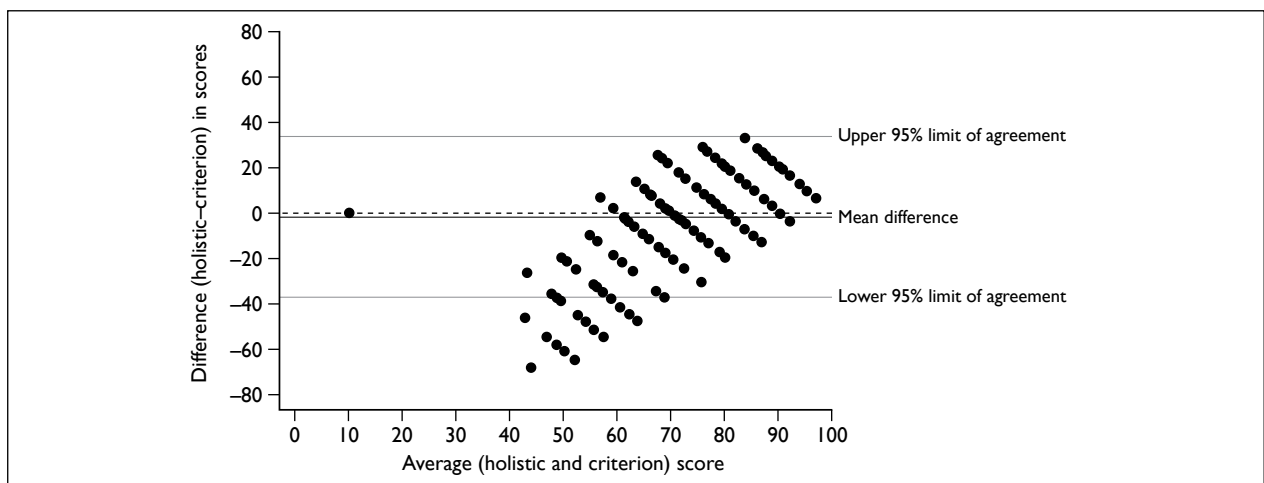


FIGURE 6 Plot of difference in holistic compared with criterion quality-of-care score versus average score – staff type: doctor.

were quite exhaustive; however, it was not possible to say whether the reviewer had included them because they had decided they were relevant to judgements about the care the patient received or simply because they looked important. It is difficult therefore to reach firm conclusions about how much selection had been exercised.

This attempt to reach conclusions was made more difficult by the fact that in some instances the comments were clearly lifted verbatim from the notes or, in other instances, a paraphrase was given, which might, perhaps, include some interpretation of what had taken place. A list of items by itself without any information about why they are important or how they related to other aspects of their care means that there is not a sufficient narrative to allow a picture to be formed of what the reviewer thinks about the issue, let alone what their view is of the quality of care was delivered.

Another approach that was observed in the comments from this group was the attention to very specific issues, such as timing of medication issues, lack of follow-up in terms of test results, details of transfers, information given to patient and family, timing of interventions/other aspects of care, including delay length if it occurred. This was sometimes the main approach taken. In other reviews it was combined with the list approach. Two reviewers did, however, pick up issues of adverse incidents or queried the care delivered in a couple of instances, but this was very much the exception rather than the rule.

On the whole, among this staff group, there was relatively little judgement directly expressed by the non-clinical reviewers about the care perceived. There was a limited amount of explicit and implied judgement, but from only relatively few of the reviewers. It was unclear in some instances how much selection had been made about what to include in the comments provided.

Nursing and other clinical staff

As with all staff groups, there was variation in the types of review comments made by nurses, although many reviewers in this group included an element of listing of what had, and had not, been done. In some instances there was only a list of actions and omissions. However, for most cases there was also other information and in many instances implied or explicit judgement for care in general or for specific aspects of care. For example, many reviews gave information about

what was done and commented on whether this was appropriate.

As with the non-clinical audit reviewers, some of the nurse reviewers mentioned documentation, but it was quite clear that they more often were concerned with care issues rather than documentation. Even the concentration of some nurse reviewers on care plans (which might be expected given the importance of this in nursing care) appeared to be more concerned with the content of the plan, rather than whether it was legible, which again tends to suggest a focus on care rather than documentation. In terms of comments about the plan, in most instances there was a judgement about the quality of the plan. Most of the comments about the plan were explicit judgements about it, and, in many instances, the views about the plan were the most explicit comments in terms of judgements made. In aspects of care other than the plan, the judgements were more often implied.

In terms of the areas of care that comments were made on, as might be expected many were about areas that might be considered the responsibility of nursing staff. For example, comments on review including observations, timing of medication administration, involvement of other nursing staff (such as specialist nurses), discharge planning, social circumstances, patient education and nutrition. However, a couple of nurse reviewers took a wider view and made comments on appropriateness of medication, investigations and teams involved in care. In some instances they also queried care provided. Several reviewers also picked up issues of concern, including potential and actual adverse events.

Overall, the nurse reviewers commented much more on the care delivered rather than the documentation, in contrast with the non-clinical audit reviewers. They also utilised the list approach, and, in some instances, it was combined with either implicit or explicit views about care delivered. While many focused on areas more traditionally thought of as nursing realms, one or two did give views (implicit, and, in one case, usually explicit) about medical as well as nursing care. It was easier to get an overall impression of the care delivered, what was delivered and views about the quality of care than with the non-clinical audit reviewers. However, the pictures were still patchy in most instances, and relied in many instances on the reader going with the implications, rather than being given an explicit view.

Doctors

As with the other two staff groups, there was considerable variation in the reviewing style and comments given by different medical staff who undertook reviews. The variation in comments provided ranged from no opinion or very brief opinions of care delivered to several lines that allowed a reasonable picture of the episode of care to be gleaned.

As with other staff groups, some element of listing what was done was evident in the comments from the medical staff. However, where this was done, the items listed were usually much fewer in number. Several reviewers gave limited comments but explicit judgements in almost all cases, so that their comments, as well as listing items, also contain some explicit judgements. If aspects of care had been omitted or delayed or the implication was that care could have been better, medical reviewers often gave further details that almost justified their implied criticism.

One of the most striking differences between this staff group and the other two groups was that almost all of the medical reviewers who gave fuller answers routinely gave explicit opinions, views or judgements about the care delivered. Medical reviewers picked up issues where the care given was queried and alternatives were suggested, and, in some instances, adverse incidents or practice said to be unsafe were picked up. These were more frequent in reviews by medical staff than in nursing/clinical groups and non-clinical staff. It was noted, however, that in instances where care was perhaps not as good as it should be, the reviewers were sometimes less willing to give an explicit judgement, certainly not without adding additional comments (and much more so than when care was deemed satisfactory). In instances where poor care was commented on, the comment often included a statement about it not being clear or mentioned if certain actions/treatments were done. This may, or may not, be a reflection of professional reluctance to criticise.

Overall, the medical reviewers gave explicit views about the care provided, often supplemented with comments that allowed a better picture of the episode of care to be gleaned. Their focus was on medical care rather than on nursing or patient-centred issues, which might be expected, and it might be considered that the comments were on the domains/items that could have a greater impact on patient outcomes.

All reviewers

There was variation in the type and amounts of comments given by reviewers. Some reviewers gave no response, others gave one or two words, while a list type response was given by some and extensive narratives were given by other reviewers. This variation might reflect different levels of understanding of care received, willingness to offer views about the care received, or both.

On the whole, it was most difficult to get a view of care from the non-clinical audit staff. From the nursing/clinical group, more information about care was given, and generally more judgements were given, although they were often implicit and in the area of nursing care rather than care overall. The medical reviewers, while they tended to focus on medical aspects of care, usually gave an explicit view about the care given and picked up issues that were likely to have an impact on patient outcomes.

Except for the information about explanation and justification of actions, proposed alternatives and views about likely impact (which was almost exclusively provided by the medical reviewers), the additional detail from lists provided by the other types of reviewers did little to help build a picture of the episode of care. In each group there was a variation in reviewing quality. The best in each group picked up issues of adverse incidents, unsafe care and gave a good account of the care episode, including their judgement about how good it was. In general the reviews provided by doctor reviewers gave a better representation of the care provided than did those of the nursing/clinical group which, in turn, was a better representation than that of the reviews from the non-clinical audit staff.

Analysis of the type and level of comment used by staff groups

Comments in all of the responses for overall care were coded using the framework provided in *Box 5* (above). The codes were then grouped into three main bands to provide an overall assessment of the similarity of types and levels of coding between the three professional groups. These groupings were:

- codes 1–5: little or no comment about care and little or no judgement
- codes 6–8: limited comment about care and implied judgement
- codes 9–13: explicit judgement of care or more sophisticated comments about care with explicit judgements and views.

The data are presented in tabular form in *Table 11* (COPD) and *Table 12* (heart failure) and the analysis takes a null hypothesis. That is, there is no association between the level of coding given and staff type (i.e. the rows and columns are independent).

For the COPD reviews, across all of the phases of care and overall care, for both conditions, there were statistically significant differences between the types of comments made by the three types of staff, with the medical staff using the highest-level codes across all phases of care and overall care (*Table 11*). There is therefore some evidence to suggest that doctors are more likely to use explicit codes than the other staff groups.

For heart failure reviews, while there are still statistically significant differences between the three groups, the differences between the groups are less strong, with similarities the doctors and nurses for phases of care and little difference between the groups for overall care (*Table 12*).

The high percentage of high-level codes in the overall section for non-clinical audit staff may be due to the fact that there were only three non-clinical audit reviewers for heart failure, and 49 out of 70 of these reviews were done by one reviewer. Unusually for non-clinical audit reviewers, this reviewer, in the overall phase, usually put a comment such as 'good all round care' or 'good nursing and medical care,' hence it has been coded 9 – explicit judgement. However, this judgement by the reviewer was not reflected in the individual phases of care. It also appears that the nurses gave more high-level comments for heart failure than those for COPD, perhaps because there were a greater proportion of specialist nurses in the heart failure group.

Figures 7 and 8 provide a graphical representation of the data.

Resource implications

Tables 13 and 14 summarise the data on resource use on holistic review and criterion-based review, respectively. For COPD holistic review (*Table 13*), doctors and nurses both took a similar amount of time to review each record (approximately 18 minutes per review). However, the non-clinical audit staff took much longer to review each COPD record using holistic methods, with a mean review time of 34 minutes. The mean total cost per review data indicates that nurses had the lowest cost

per COPD holistic review at £6.73, non-clinical staff had a cost per review of £8.53, and doctors incurred the highest cost per review at £10.16, despite taking the least time to undertake the reviews. Doctors incur higher costs per review than the other staff types because of the higher overall staff costs compared with the other staff groups in the study.

There was less variation between the different staff groups for the time taken to review heart failure records holistically (mean range 24.06–29.48 minutes). Doctors were again found to be the most expensive staff group, with non-clinical audit staff incurring the least cost per review.

Cost per review for heart failure holistic reviews were slightly lower for each staff group than for the COPD holistic reviews.

For the COPD criterion-based reviews (*Table 14*) the non-clinical staff took the most time to complete each review. However, as with the holistic review data, the doctors incurred the highest cost per review. This pattern is repeated in the heart failure criterion data.

Quality of care–hospital mortality group relationship

Of the nine study hospitals, five were grouped together in the lower-mortality group and four were in the higher-mortality group, based on a calculation from Hospital Episode Statistics data (*Figure 2*, above). This analysis explores the relationship between these two groups of hospitals, which are ranked and grouped by mortality rates, and the quality-of-care data expressed as the group mean holistic scale scores for overall care and the group mean percentage criterion score. Data are presented in tabular and graphical form.

Table 15 shows the mean difference, for each condition, between the holistic overall quality-of-care rating (based on a 10-point scale) for hospitals classified as belonging to a higher-mortality group or a lower-mortality group. Mean difference is the mean score for higher-mortality hospitals minus the mean score for lower-mortality hospitals. A negative mean difference in the table indicates that the lower-mortality group mean score is higher than the mean score for the higher-mortality group.

The result trends are different for COPD and for heart failure. For COPD, quality-of-care scores

TABLE 11 Types of comment made by different staff types for COPD ('highest-level' comment per review)

	Admission			Initial management			Pre-discharge			Overall		
	Doctor	Clinical (nurse or other)	Non- clinical audit	Doctor	Clinical (nurse or other)	Non- clinical audit	Doctor	Clinical (nurse or other)	Non- clinical audit	Doctor	Clinical (nurse or other)	Non- clinical audit
Coding bands (%)												
Codes 1–5 (little or no comment about care and little or no judgement)	5	14	41	1	12	38	4	13	47	4	16	47
Codes 6–8 (limited comment about care and implied judgement)	15	32	56	15	36	48	27	53	48	4	15	7
Codes 9–13 (explicit judgement of care ⁹ or more sophisticated comments about care with explicit judgements and views)	81	54	4	84	52	13	68	34	6	93	69	46
Number of reviews (and reviewers)												
	258 (6)	357 (8)	271 (7)	273 (6)	356 (8)	271 (7)	273 (6)	355 (8)	271 (7)	273 (6)	352 (8)	267 (7)
Chi-squared test												
	338.0 ($p < 0.001$)			306.2 ($p < 0.001$)			313.1 ($p < 0.001$)			190.3 ($p < 0.001$)		

TABLE 12 Heart failure: types of comment made by different staff types ('highest-level' comment per review)

	Admission			Initial management			Pre-discharge			Overall		
	Doctor	Clinical (nurse or other)	Non- clinical audit	Doctor	Clinical (nurse or other)	Non- clinical audit	Doctor	Clinical (nurse or other)	Non- clinical audit	Doctor	Clinical (nurse or other)	Non- clinical audit
Coding bands (%)												
Codes 1–5 (little or no comment about care and little or no judgement)	10	5	6	12	4	1	19	3	17	10	3	4
Codes 6–8 (limited comment about care and implied judgement)	6	12	77	6	12	72	21	51	64	2	8	0
Codes 9–13 (explicit judgement of care ⁹ or more sophisticated comments about care with explicit judgements and views)	84	83	17	82	84	26	60	42	19	89	89	96
Number of reviews (and reviewers)												
	329 (10)	172 (5)	70 (3)	329 (10)	172 (5)	69 (3)	324 (10)	172 (5)	70 (3)	326 (10)	172 (5)	70 (3)
Chi-squared test												
	217.9 ($p < 0.001$)			200.2 ($p < 0.001$)			81.0 ($p < 0.001$)			25.1 ($p < 0.001$)		

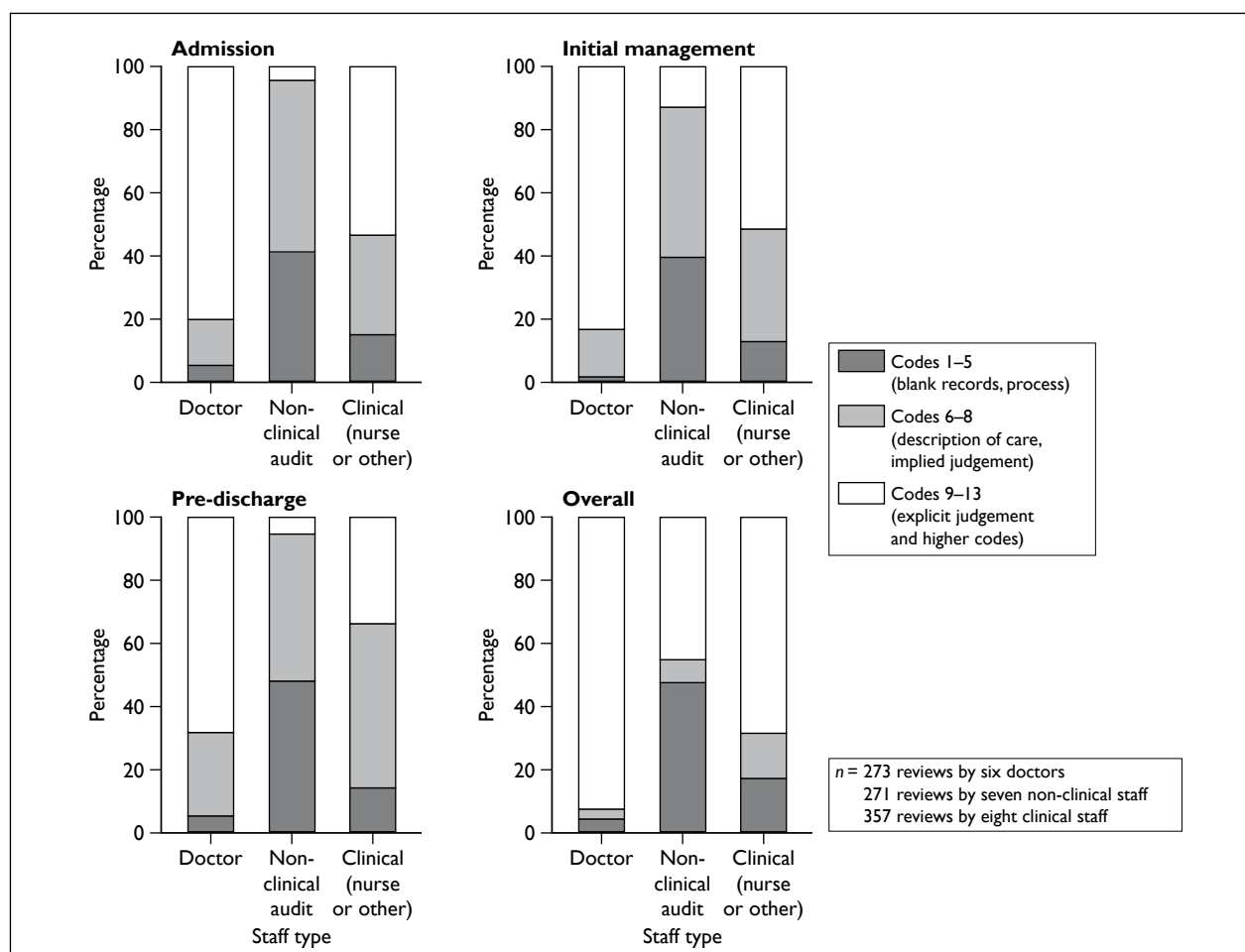


FIGURE 7 COPD coding according to types of comment.

tend to be higher in the lower-mortality group of hospitals than in the higher-mortality group, and there is a significant difference in the quality scores given by the doctors. The significant score for all three groups of staff ($p = 0.033$) is likely to be driven by the significant score for the doctors groups ($p = 0.012$). These findings might be thought of as being in an expected direction, although research shows that process–outcome relationships are by no means straightforward.³⁸

For heart failure, there are smaller mean differences in a negative direction, indicating that quality-of-care scores are higher in the higher-mortality hospital group. However, the differences are quite small and none of the differences are significant.

It is difficult to interpret the findings in Table 15 overall – there is certainly no overall trend to higher quality scores in lower-mortality hospitals. It may be that the contrasts are an example of the unexpected findings found by other researchers

(e.g. Gibbs *et al.*⁷), where higher quality-of-care scores have been found among patients with poorer outcomes.

The holistic overall care review data in Table 15 are also presented as box plots in Figures 9 and 10 to demonstrate the distribution of scores from individual reviewers (identified in the tails of the distributions by anonymous review numbers), together with medians and interquartile ranges, for the two hospital mortality groups.

Table 16 shows the mean difference for the mean of the total percentage scores from the criterion-based reviews for each condition for the higher- and lower-hospital mortality groups. These are not dissimilar findings from Table 15, with many of the criterion scores tending to be higher in the lower-mortality group of hospitals. In this analysis there is also only one significant difference found, but, this time, it is in the physician reviewers' score for heart failure – a higher quality score for the lower-mortality hospital group.

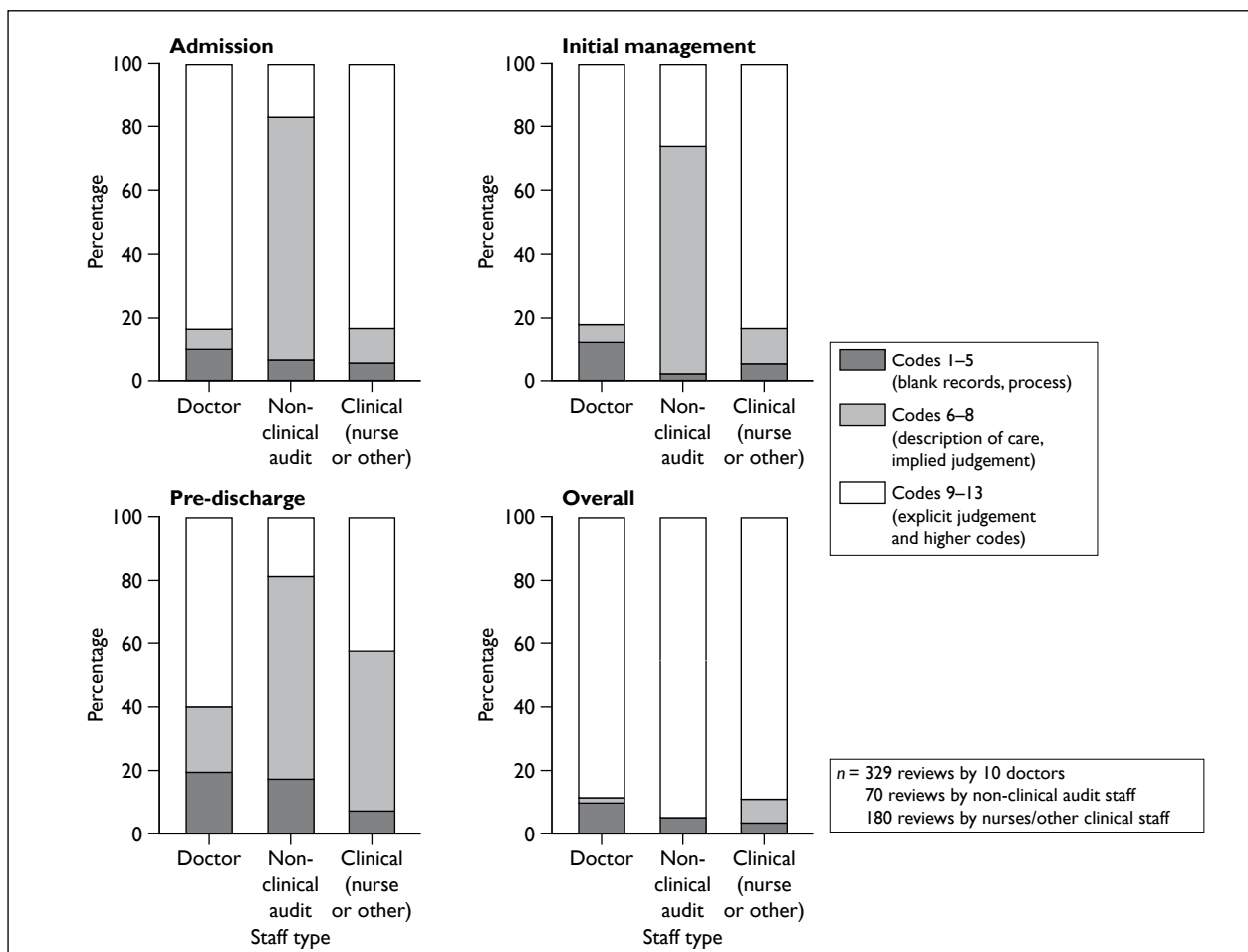


FIGURE 8 Heart failure coding according to types of comment.

Overall, there does tend to be a higher quality score for criterion review in the lower-mortality hospitals. However, note that when contrasting the results of *Tables 15 and 16*, the reviewers were using implicit judgements to score care in *Table 15*, but there were explicit standards set through the review criteria in *Table 16*. Judgement-based holistic scale scoring tends to show larger 'tails' at both ends of scoring than criterion scoring, which is unlikely to be influenced by outcome (*Figures 9 and 10*). For 'poor' care at the individual case level, reviewers using holistic methods can be very critical, as some of our textual data show, and the reverse is the case for good care.

The criterion-based review data that are presented in *Table 16* are also shown in *Figures 11 and 12*.

Feedback to study hospitals and teams

Following the completion of the reviews, each reviewer received a feedback report containing their own data. Individual hospitals or reviewers

were not identified in the report. The report contained summary statistics and frequencies for the algorithmic criterion and the holistic quality-of-care scores for all phases of care and the quality-of-records element of the review. The holistic quality-of-care scores were compared with those from other reviewers reviewing the same records. A mean quality-of-care score was calculated for each hospital and this was also provided in the report. Appendix 6 contains an example report that has been anonymised.

Each reviewer received an electronic copy of the report, as it was envisaged that the report would provide a useful basis for local audit presentations. Where requested, individual assistance was provided to reviewers wishing to present their data at local audit meetings. For example, advice was given on the meaning of the data or specific graphs/tables were produced in POWERPOINT.

From the feedback we received from the reviewers about the reports, most found the reports to be useful and interesting.

TABLE 13 Resource used during the holistic review process, by staff type

Staff type	Number of valid reviews (missing data)	Mean review time (minutes) (SD)	SD time per review	Time per review (min-max)	Mean total cost per review (£)	SD cost per review (£)	Cost per review (min-max) (£)
COPD							
Doctor	266 (7)	18.5	9.9	5–52	10.16	5.46	2.75–28.60
Nurse/other clinical	347 (10)	18.6	9.3	2–60	6.73	3.62	0.67–27.00
Non-clinical	268 (3)	34.12	16.7	5–105	8.53	4.17	1.25–26.25
All reviews	881 (20)						
Heart failure							
Doctors	309 (21)	24.1	10	5–50	12.96	5.84	2.75–24.75
Nurses/ other clinical	177 (3)	29	21	10–180	10.54	7.80	3.33–60.00
Non-clinical	63 (7)	27.1	9.5	10–50	6.78	2.37	2.50–12.50
All reviews	549 (27)						

TABLE 14 Resource used during the criterion review process, by staff type

Staff type	Number of valid reviews (missing data)	Mean review time (minutes) (SD)	SD time per review	Time per review (min-max)	Mean total cost per review (£)	SD cost per review (£)	Cost per review (min-max) (£)
COPD							
Doctors	254 (0)	19.5	7.2	6–50	9.16	5.21	2.50–32.50
Nurses/ other clinical	335 (1)	15.4	5.8	3–45	5.57	2.16	1.00–15.00
Non-clinical	243 (1)	30	26.2	10–280	6.00	2.52	1.33–18.75
All reviews	832 (2)						
Heart failure							
Doctors	311 (2)	20.8	9.9	5–50	10.44	4.89	2.58–21.67
Nurses/ other clinical	177 (2)	20.6	8.1	10–60	6.27	2.79	2.83–17.00
Non-clinical	71 (0)	27.5	12	10–50	6.88	2.99	2.50–12.50
All reviews	559 (4)						

TABLE 15 Relationship between the holistic overall quality-of-care rating (10-point scale) and mortality group

Staff type	Higher mortality hospitals (n=4)	Lower mortality hospitals (n=5)	Mean difference ^a	95% CI of the difference ^b	p-value
	Number of reviews; mean score (SD)	Number of reviews; mean score (SD)			
COPD					
Doctors	190; 7.19 (2.01)	82; 8.55 (0.788)	1.35	0.45 to 2.26	0.012 ^b
Non-clinical	80; 7.46 (1.40)	147; 7.92 (1.44)	0.46	-0.17 to 1.08	0.118
Clinical	206; 6.48 (2.24)	147; 7.23 (1.74)	0.76	-0.79 to 2.30	0.286
All staff	476; 6.93 (2.06)	376; 7.79 (1.54)	0.86	0.08 to 1.64	0.033
Heart failure					
Doctors	166; 8.11 (1.94)	156; 7.69 (1.42)	-0.42	-1.52 to 0.69	0.415
Non-clinical	9; 8.22 (0.83)	61; 8.15 (0.51)	-0.07	-1.09 to 0.94	0.781
Clinical	137; 7.53 (1.56)	42; 7.09 (2.02)	-0.43	-1.47 to 0.60	0.332
All staff	312; 7.86 (1.78)	259; 7.70 (1.42)	-0.15	-0.83 to 0.52	0.637
a Mean difference is the mean score for the lower-mortality hospitals group minus the mean score for the higher-mortality hospitals group.					
b p-Values and 95% CIs adjusted for potential clustering by reviewer.					

TABLE 16 Relationship between the mean criterion score (scaled to 100) and mortality group

Staff type	Higher mortality hospitals (n=4)	Lower mortality hospitals (n=5)	Mean difference ^a	95% CI of the difference ^b	p-value ^b
	n; mean (SD)	n; mean (SD)			
COPD					
Doctors	187; 77.33 (10.47)	67; 77.99 (7.85)	0.67	-5.63 to 6.97	0.796
Non-clinical	120; 74.26 (12.90)	124; 74.78 (9.31)	0.52	-9.09 to 10.13	0.895
Clinical	189; 73.63 (11.98)	147; 76.02 (9.73)	2.39	-5.54 to 10.33	0.499
All staff	496; 75.18 (11.77)	338; 75.96 (9.28)	0.78	-3.27 to 4.83	0.691
Heart failure					
Doctors	163; 71.41 (14.65)	150; 77.90 (9.23)	6.49	1.52 to 11.46	0.016
Non-clinical	10; 79.00 (7.74)	61; 74.96 (8.72)	-4.04	-20.39 to 12.31	0.399
Clinical	135; 80.0 (5.62)	42; 80.65 (3.86)	0.65	-2.98 to 4.27	0.664
All staff	308; 75.41 (12.12)	253; 77.64 (8.62)	2.23	-2.55 to 7.00	0.340
a Mean difference is the mean score for the lower-mortality hospitals group minus the mean score for the higher-mortality hospitals group.					
b p-Values and 95% CIs adjusted for potential clustering by reviewer.					

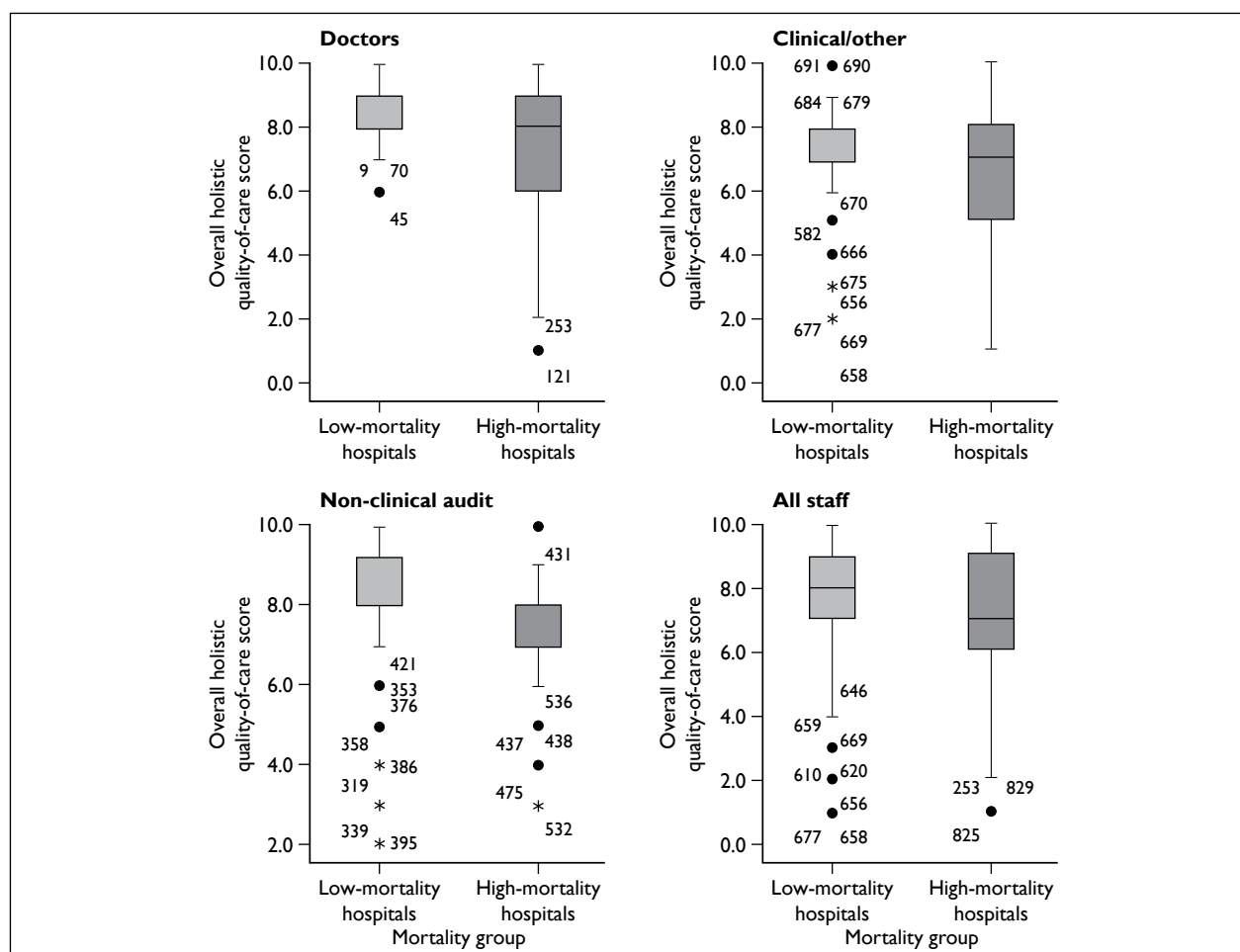


FIGURE 9 COPD – holistic overall quality-of-care box plots to show the difference between the overall quality-of-care ratings for hospitals classified as in a high-mortality group or in a low-mortality group (note: tails of the plots show the results from individual reviewers).

Discussion

Retrospective assessment of the quality and safety of care primarily depends on review of information from the clinical record, and the literature suggests that both holistic and criterion-based review methods make valuable contributions, but both also have methodological limitations.

This is the first study in the UK that has compared the two methods of review and it has additionally contrasted the results of three different professional groups. Few international studies have contrasted the review results of different professional groups using the same methods, although Weingart *et al.*⁶ compared the results of explicit (criterion-based) review undertaken by nurses with implicit review of the same record undertaken by physicians. They found that when examining medical records ‘nurse and physician reviewers often came to substantially

different conclusions’.⁶ Key results from our study include:

- Reviewers are reasonably internally consistent.
- There is some evidence of moderate to good within-group reliability for holistic review.
- All three professional groups were good at criterion review.
- There is evidence of agreement between the results of holistic and criterion review in the hands of physician reviewers.
- There is some difference in review focus between nurses and physicians.

Our most important research questions have been, first, to determine the level of agreement between health-care professionals, from different backgrounds, when they review the same record. This agreement, or reliability, relates to the repeatability of the results from the review –

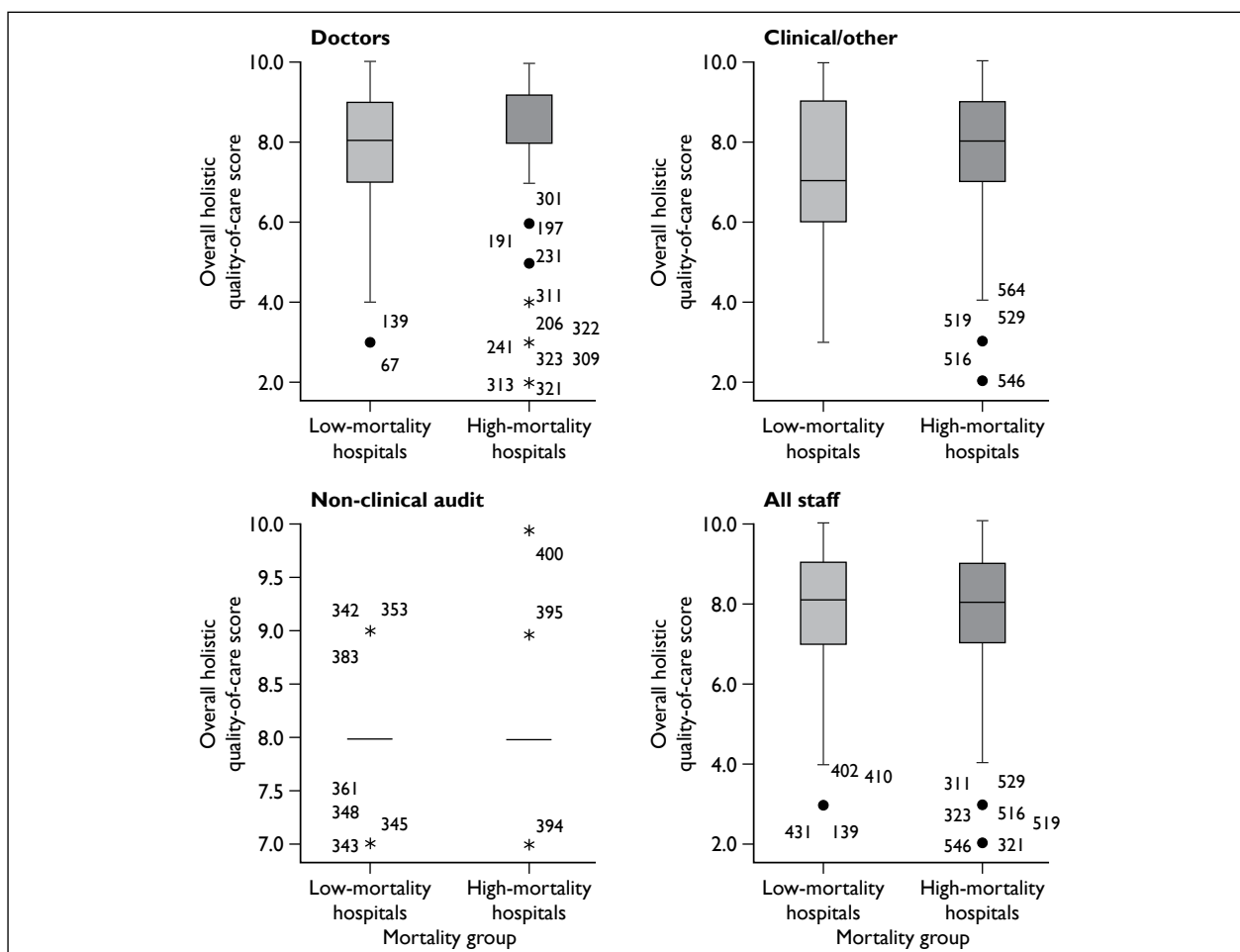


FIGURE 10 Heart failure – holistic overall quality-of-care box plots to show the difference between the overall quality-of-care ratings for hospitals classified in a high-mortality group or in a low-mortality group (note: tails of the plots show the results from individual reviewers).

whether a different reviewer would come to the same conclusion about the quality of care from the same data source, using the same method. This is clearly a practical question, as well as a research question. Second, we have used review of the same record to explore the relationship between holistic and criterion-based methods.

Reviewers undertaking holistic review, using scale scores, were relatively consistent in the scores for each case note allocated to care quality across the individual phases of care and overall for the entire episode of care. All three professional groups had moderate within-group inter-rater reliability, ranging from 0.46 (95% CI 0.34 to 0.59) to 0.52 (95% CI 0.41 to 0.62). The physician reviewers had the highest values. These results were replicated for the physician reviewers in the kappa analysis (Table 6b), although the kappa scores for the other two professional groups were considerably lower.

The ICC reliability values were rather higher than the average found in a systematic review by Lilford *et al.*¹² in which, for implicit structured review, studies of case note review concerned with causality and process of care had kappa values of below 0.4 [causality: kappa 0.39 (SD 0.19); process: kappa 0.35 (SD 0.19)].

The inter-rater reliability results of the study are also somewhat similar to those of Hofer and colleagues,¹⁰ who also used ICCs and found a reliability of 0.46 for a structured holistic review of diabetes and heart failure records by physician reviewers (although only 0.26 for COPD records). In comparison, a recent holistic assessment of patients dying in UK hospitals achieved a kappa score of 0.39 on the key indicator of quality of medical care.⁴¹ Nevertheless, our study found that there was still a considerable range of reliability scores between individual reviewer pairs, even

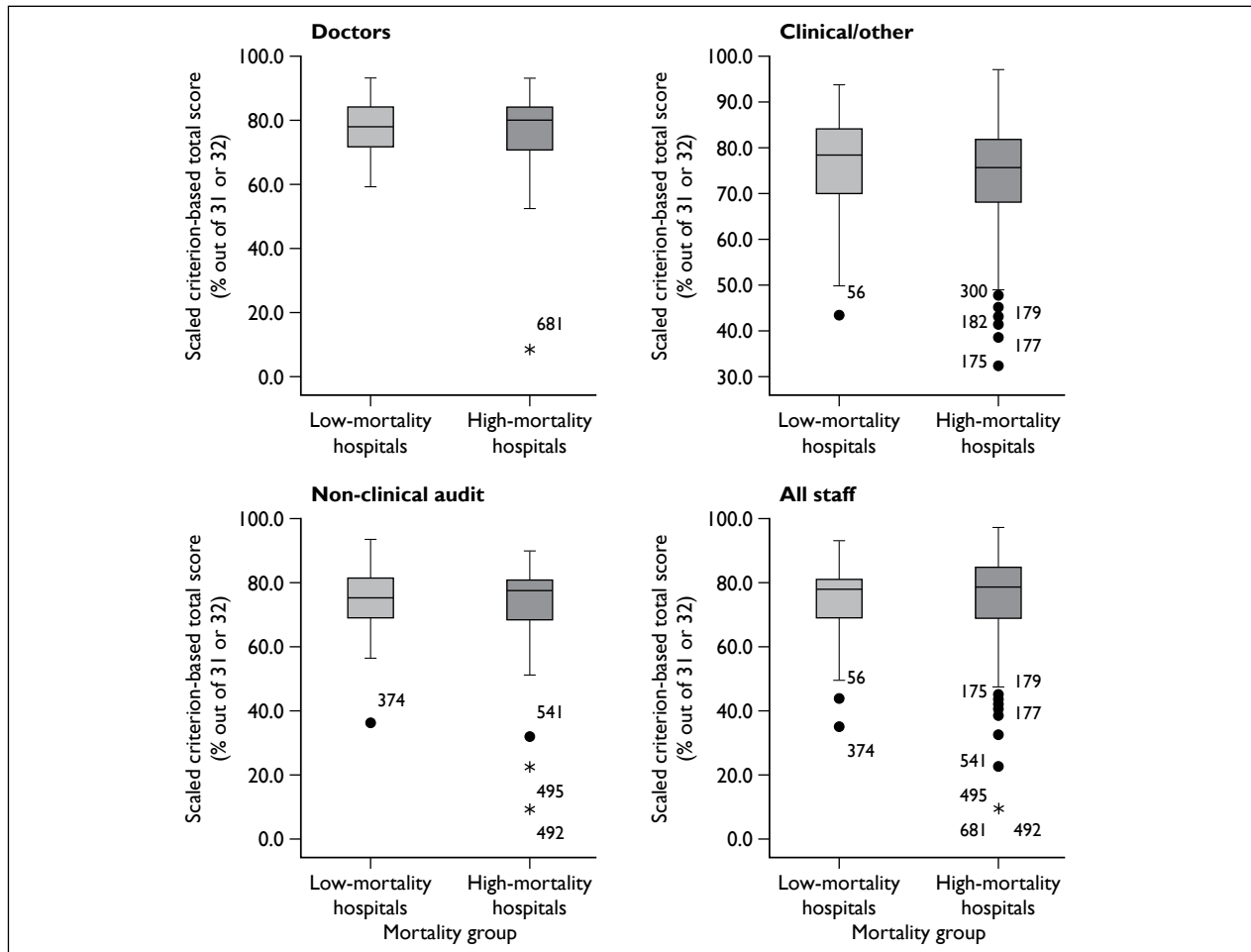


FIGURE 11 COPD criterion-based total score box plots to show the difference between the criterion-based total score for hospitals in the higher-mortality group or lower-mortality group.

between the doctor reviewers. This variation might be a reflection of either training or experience, or perhaps it represents other aspects of the causes of holistic review variation identified in the literature, such as bias or harshness in reviewing.¹²⁻¹⁶ It could also reflect our decision not to train on explicit standards of care.

Most of our reviewers had not undertaken a formal holistic review before. Training and support was provided, but it was deliberately designed not to influence the effect of implicit judgement and the training may therefore not have been sufficient to reduce the element of variation due to inexperience with the review methods. The large quality and safety review programmes in the USA (for example in the Colorado and Utah study⁵ and the study by Daley and colleagues³⁷) use senior physician reviewers who have been screened to identify and exclude those who may have particular biases. This approach might have had a positive

impact on the inter-rater reliability results, but was not an option open to this study because of the cost and availability of senior staff.

Completeness of the clinical record is a significant feature in the success of review methods, as a review can only consider what is written in the record and reviewers have to depend on the abstracted details of the case. Not recording an event does not mean that the event did not occur. It may be, for instance, that a practitioner did not record the event because he/she considered it to be too trivial. However, non-recording or very limited recording is a definite constraint on the effectiveness of records review as a means of assessing quality of care among groups of patients. On the other hand, direct observation of the quality of care, suggested as an alternative to records review,⁹ would be very expensive if it were to be used as a standard procedure. Nevertheless, reviewers judged the quality of recording in the case notes as reasonable (over 4 on a scale of 1-6),

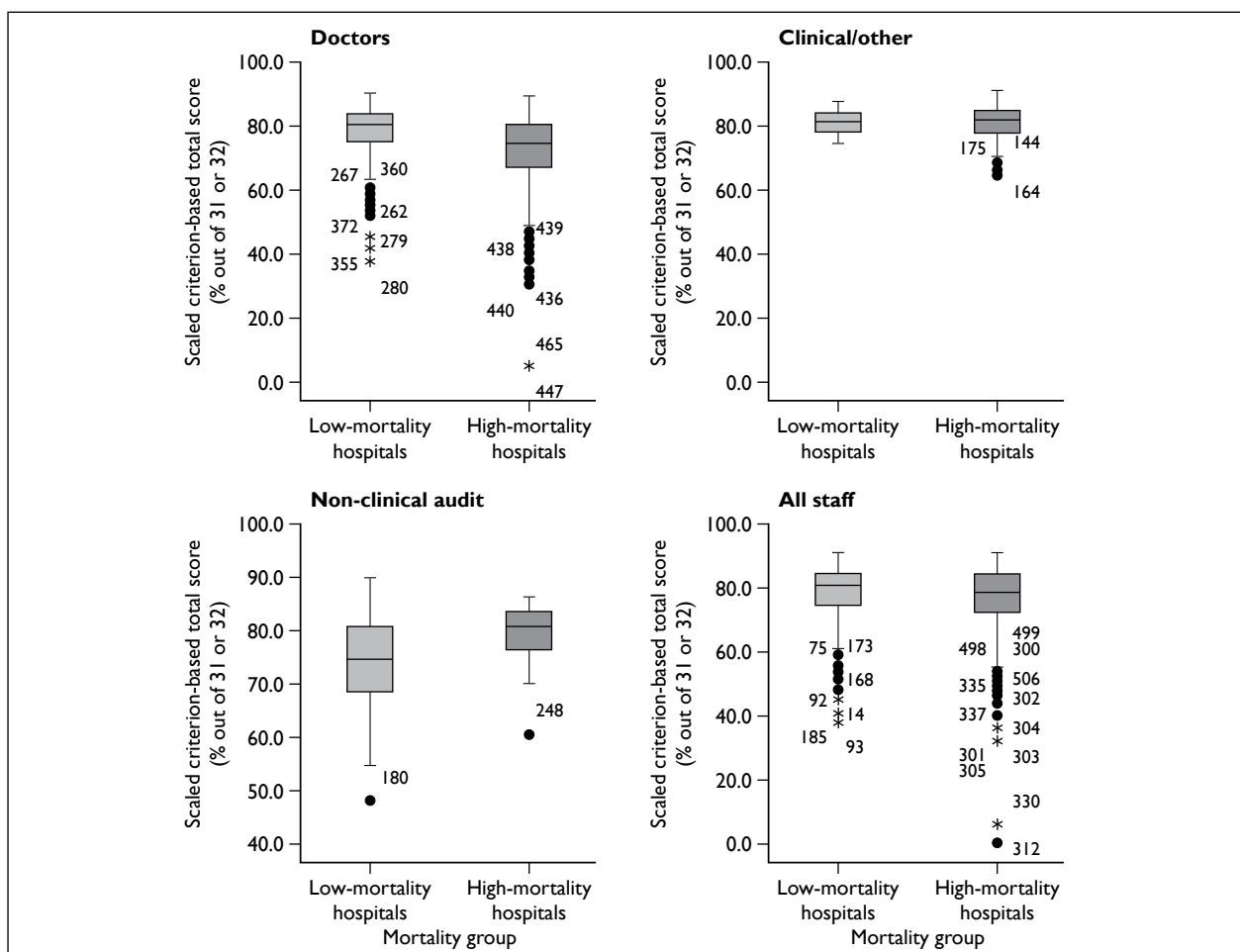


FIGURE 12 Heart failure criterion-based total score box plots to show the difference between the criterion-based total score for hospitals in the higher-mortality group or lower-mortality group.

although some notable poor exceptions were commented on in the textual reviews.

The levels of inter-rater reliability for criterion-based review ranged from moderate (0.61 for non-clinical audit staff) to quite high (clinical staff 0.74, doctors 0.88), and are similar to those found in the large UK national clinical audit programmes of stroke care^{24,25} and continence care.⁴² All three staff types performed equally well at capturing criterion-based data from records, despite differences in their backgrounds, again confirming the findings of the UK stroke care audit.^{24,25} It is unsurprising that criterion-based review has higher levels of inter-rater reliability than holistic review, as the criteria are predetermined, directly evidence based, have been subject to peer review and are explicit rather than implicit. Under these circumstances, an ICC of less than 0.75 might be deemed unsatisfactory and any large study using explicit criteria might best train reviewers with this target in mind.

Criterion-based review demonstrated that the reviewers from all three groups could identify relevant data where it existed (the effectiveness of reviewer scores were around a mean of 95%), and that, in general, sufficient case note data were available from which to assess quality of care using the review criteria. The quality-of-care scores for both criterion-based review and from holistic review using overall scale scores were similarly rated by the three staff types at between 70% and 79% (Table 10), where 100% represents excellent care.

We found that in the hands of medically trained and of non-clinical audit reviewers, although not for the nurse and other clinical group, there was a significant level of agreement between the results of holistic and criterion-based review, suggesting that the two methods are measuring related elements of quality. This may have important implications for the choice of review method when evaluating the quality of care from case notes. Criterion-based

review methods may give sufficient information on which to judge the overall quality of care, performing as a screening tool for large clinical audit studies, provided that the appropriate review criteria are chosen. On the other hand, a structured form of holistic review, such as described here, can, in the right hands, also give a reliable picture of the quality of care and pick up the nuances of quality variation that criterion-based review is unable to provide.

What additional contribution does holistic textual data make to the assessment of quality of care?

Although there was some overlap between the results from the three professional groups, across the spectrum from descriptive statements to judgements, some distinctive patterns also emerged (Tables 11 and 12, and Box 5). It is not surprising that, in general, the reviewers without any clinical training tended to provide mainly descriptions of care process and only relatively rarely offered judgements on quality. The nursing group offered some judgements, especially about process or care issues, and in general were more concerned about the nursing process and the care plan than about the interventions offered to the patient. The doctors, on the other hand, were in general more concerned with assessing the quality and safety of the therapies, and strategies for dealing with the acute illness rather than about the overall care plan.

Nurses tend to be close to the hospital medical care process because of frequent contacts with the patient, so it might have been expected that the results from the nursing/clinical staff reviewers might have been close to those of the medical reviewers. But the limited agreement between the doctors and the nurses may also reflect different internal professional standards for assessing quality and safety of care when reviewing a record. Weingart and colleagues⁶ conjectured that nurses and doctors reviewed in different ways – that nurses sought data on the routines of care, whereas doctors looked for a wider picture – and that, in general, neither group considered both dimensions. Analysis of the textual data tends to support the notion that the doctors and nurses in this study commented on different aspects of care when assessing quality. Judgements by the nurses tended to be implicit rather than explicit and they tended towards reviews of nursing process of care.

Doctors were the group who tended to mention outcomes or impact on future care, more so than the other groups, on the whole using explicit statements. Some were willing to justify why they thought care was good or unsatisfactory. Our results suggest that the hypothesis posited by Weingart *et al.*⁶ may have some validity – in reviewing quality of care, nurses tend to concern themselves with care processes and pathways, while doctors tend to be concerned with diagnosis and interventions. Each professional type may therefore identify nuances of care that the other does not.

Mean review times for both methods were quite similar for the doctor and nurse reviewers and the costs for the doctor group were consequently higher for each review. The cost differences between the nurse reviewers and the doctor reviewers were rather less for holistic review than for criterion-based review. So the decision on which reviewer type to use for a review process will principally depend on which type of information is required from the review, though with an eye to the cost differentials. Because the cost of analysis of the textual data is high and requires specialist skills, and because there is some evidence from the relationship between criterion review and holistic scale score data, it is likely that any large-scale study using holistic review would use scale scores to judge quality of care, rather than a full textual data analysis. On the other hand, for smaller scale, more detailed studies, analysis of textual commentary on quality of care provides a very rich data set on which to judge quality and safety.

Examination of the more than 100 reviews of cases judged to be unsatisfactory shows very few defined adverse events, the commonest of these being decisions to give the wrong drug.⁴³ Other, fewer, more serious events include missed or erroneous diagnoses. Much more commonly in the reviews there appear unsatisfactory aspects of care that may run as a thread throughout the hospital admission. What might be regarded individually as ‘minor shortcomings’ on their own become fused together to create an ‘event’. These shortcomings do not always translate across all of the phases of care. Sometimes, for example when care is handed over from one team to another, missed diagnoses or inappropriate or suboptimal treatment (e.g. not following evidence-based practice) is seen to be recovered by the receiving team and overall care is judged satisfactory. From our study we have been able to show that adverse events are at least as likely to be non-discrete constellations of suboptimal components of care that, taken together, put the

patient at risk. Notwithstanding the additional costs, narrative descriptions of care through holistic analysis can considerably enhance understanding of health-care quality and may be of particular value in locally based clinical audit.

What is the relationship between mortality (outcomes) and quality-of-care scores for the study hospitals?

In a systematic review of the relationship between quality and outcomes of care (risk-adjusted mortality) in hospitals or hospital units, Pitches and colleagues³⁸ found an uncertain relationship emerging from a heterogeneous group of 36 studies, including 51 quality-and-outcome relationships. About one-half of the studies demonstrated a link between better quality of care and the lower-risk hospitals, about one-quarter of studies showed a negative association and the remainder were equivocal.

We also found some mixed results (*Tables 15 and 16 and Figures 9 and 10*). Across all staff types, for holistic review of cases of COPD, there was a trend for a positive relationship between higher overall quality-of-care scores and a lower mortality ranking, with one significant difference scored by the physician reviewers. For heart failure quality of care there was a trend towards higher scores in higher-mortality hospitals, although the mean score differences were small. For the criterion-based review there was a general trend towards higher quality scores for lower-mortality hospitals. For both conditions, it was the results from only the nurse/clinical reviewers for COPD, and from only the physician reviewers for heart failure that showed a positive association with lower mortality.

It is likely that these associations are influenced by review method and staff type, as well as by the method of risk adjustment and the actual quality of care provided and recorded. It could be argued that the differences between hospitals that might be found using criterion-based review would be (and were) relatively limited because of the very structured nature of the criteria. So review criteria might be less useful in assessing differences between hospital units than the more broadly based holistic method. In contrast, triangulation of the holistic review intra-rater consistency, inter-rater reliability and qualitative analysis results suggests that the doctor reviewers, on the whole, produce information that does allow judgements about quality of care, and that for both COPD and heart

failure there is a significant positive relationship between their overall holistic scale scores and mortality ranking of the hospitals, suggesting that for these two conditions better quality is found in hospitals with lower mortality.

Hofer and colleagues¹⁰ suggested that inter-rater reliability results for chronic diseases, such as heart failure and COPD, were, to some extent, influenced by the evidence base, proposing that in their study the heart failure reviews had higher inter-rater reliabilities than COPD reviews because the evidence base for heart failure management was stronger than that for COPD. There is room for debate on this hypothesis, but, in any event, we found that the quality–mortality relationship was apparent more consistently across reviewers for COPD than for heart failure.

Study limitations

In a complex study such as this there are bound to be methodological limitations. There were only two tracer conditions used in the study, whereas comparable studies in the USA have used five or more conditions.¹⁰ Additionally, because of the nature of the research questions, only 38 reviewers in nine hospitals were involved and they came from a range of backgrounds, although with some similarity to the study by Weingart and colleagues.⁶ Results from this study, nevertheless, show enough similarities with those from Hofer *et al.*¹⁰ and Weingart *et al.*⁶ to suggest that they are meaningful.

Assessment of the quality and safety of care using, among other methods, a six- or 10-point scale, remains unusual in the literature. Evaluation of the sensitivity of the scales was not possible prior to the main data collection, and, although the intra-rater reliability results suggested that there was a fair degree of internal consistency when these scales were used, further research on the use of similar scales would be of value.

There is a potential for bias in the study method in that the design, and especially the constraints of the ethics and research governance requirements, meant that reviewers evaluated the case notes in their own hospitals (rather than, for example, reviewing the case notes in another hospital, where they would not have the possibility of reviewing cases in which they may have provided care). This is an acknowledged potential bias, although the range of holistic scores and the types of commentary that were recorded suggest that

many of the reviewers were quite robust when determining the score for a case.

The reliability estimates reported in this study are likely to be larger than the true population reliability values because of the sampling method used in the design. The variability of the results of the holistic reviews, including up to 15% of reviews that scored 1 or 2, means that the ICC will tend to be higher than if the review results had been more homogeneous. Furthermore, the ICC 'combines three features of the data (patient variability, reviewer variability and measurement variability) from which it is calculated ... it does make comparisons of ICCs between different studies difficult to interpret'.⁴⁴ Nevertheless, a recent systematic review of the inter-rater reliability reported in case note review studies¹¹ was able to make kappa comparisons and this study had higher

ICC results than the review kappa for both causality and process of care [causality: kappa 0.39 (SD 0.19); process: kappa 0.35 (SD 0.19)].

We have undertaken a relative simple statistical analysis to compare the reliability of the holistic and criterion-based reviews by using kappa and ICC statistics. These statistical methods do not allow for other facets of measurement, such as the disease or reviewer characteristics. The lack of balance in the design precludes a classic ANOVA-based generalisability analysis to assess reliability. However, a multilevel modelling approach could be used. We have not been able to undertake this, but we acknowledge that such a model might produce smaller standard errors and give narrower CI estimates for the reliability coefficients. However, we believe that this model is unlikely to change the conclusions of the study in a substantive way.

Chapter 3

What is the relationship between information on quality of care from case notes and hospital-level outcomes of care?

Background

The purpose of this second part of the study was to investigate how much of the variance in risk-adjusted outcome for two important clinical conditions could be attributed to differences in quality of clinical care in acute hospitals, as assessed through review of case notes.

Using case notes as the basis for assessing health-care quality is known to be problematic for a number of reasons, such as reliability,^{6,10,12,45} bias¹⁵ and consistency,⁸ even although case notes are still almost universally used as a primary data source for this purpose. This continuing usage is partly due to the fact that the alternatives to using case note review as the basis for quality assessment – such as direct observation, use of video or actor patients – may be even more time-consuming and expensive, and they have their own methodological challenges. Approaches to reducing uncertainty in the measurement of quality of care have been concerned with improving reliability in implicit review, such as using structured review methods,²⁸ and providing evidence-based review criteria for explicit review.^{19,21}

In addition to measuring the process of care, the putative relationship between process and outcome is also an important consideration for health-care systems and is the subject of scrutiny by public bodies.⁴⁶ Case note review has been used in a considerable number of studies to provide process of care data in order to explore the relationship with outcome of care, a relationship that appears complex. This was recently demonstrated by Pitches *et al.*,³⁸ who undertook a systematic review of 36 studies, which included 51 ‘process-versus-risk-adjusted outcome relationships’, exploring the extent to which variations in risk-adjusted mortality rates were associated with differences in quality of care. The authors found a positive correlation in only 51% of the relationships, with no correlation in 31% and an unexpected correlation in a further

18%, in what was a very heterogeneous set of studies.

In a study of 87,000 surgical operations in eight subspecialties in 44 hospitals with a range of risk-adjusted mortality, Gibbs *et al.*⁷ used structured implicit review of case notes to populate the primary outcome measure of a five-point scale that assessed overall quality of care. Overall, the authors found that quality-of-care ratings were not significantly different between hospitals with higher than, and lower than, expected mortality.

National quality-of-care audits in the UK, such as the National Stroke Audit,⁴⁷ have used explicit review criteria to assess quality of care. Using a 60-item set of criteria derived from the UK stroke audit in a longitudinal case note review study of stroke care in New Zealand, McNaughton and colleagues⁴⁸ found only weak relationships between process and outcome variables across four hospitals, and the hospital with the best process scores had the worst case mix-adjusted outcomes. A study of 20 UK maternity units found substantial change in practise over 8 years, but few associations between proxy outcomes and other explanatory variables.⁴⁹

Earlier in this phase of the study, we considered the benefits and limitations of using case notes as the basis for reviewing quality of care and examined the reliability and value of two review methods – holistic (implicit) review and criterion-based (explicit) review – in the hands of different types of health professional. We showed that the two methods provide different types of quality assessment with reasonable to good levels of reliability.

In this next stage of the study, hospital-based process of care is assessed using mixed case note review methods, with implicit review structured by phases of care, including both scale scores and structured textual data, together with explicit,

criterion-based scores; outcome is judged using a set of direct and indirect measures derived from national data sets.

Study aim and research questions

Aim

To investigate how much of the variance in outcome for two important clinical conditions (adjusted for measurable differences in risk) can be attributed to known differences in quality of clinical care in acute hospitals.

Research questions

- What is the relationship between the quality of care for individual conditions in hospitals and overall quality of care and quality markers across hospital institutions?
- What is the relationship between case mix or risk-adjusted outcome and quality of care as measured by case record review?
- Is high-quality care associated with high-quality outcomes in risk-adjusted cases?
- Is there a correlation in clinical quality between the management of different conditions in the same hospital, as measured by case note review?

Methods

Choice of settings, review methods and staff

Selection of study hospitals

In our original proposal we indicated that a total of 24 acute hospitals in England would contribute data to this process–outcome study, including the data from with the case note reviews in the initial eight (nine) hospitals in the reliability study. During the analysis of the first part of the study, it became clear that the method of data collection being selected for the outcomes study – using a single physician reviewer per condition per hospital and a compound method of reviewing – meant that the data from the reliability study could not contribute fully to the objectives of this outcome study. Thus, a total of 24 hospitals were separately recruited for this quality and outcomes study.

Selection of the hospitals used the same processes in the phase one reliability study. First, Hospital Episode Statistics³³ on 28-day mortality data for COPD and heart failure were accessed through the East Midlands Public Health Observatory. Hospitals were excluded from the selection process

if they reported less than 200 inpatient cases per year for either condition, effectively excluding smaller or specialist acute hospitals. There were 136 hospitals in the final data set.

Second, 28-day mortality data for the two study conditions for each hospital were combined, using simple averaging, to create an average 28-day mortality ratio for each hospital. Third, these were then ranked from the lowest mortality to the highest, and the data was split into four quartiles. And, finally, from this ranking, hospitals were randomly selected in each of the lowest- and the highest-mortality quartiles.

The 20 hospitals selected for the reliability study were excluded from the selection process. Thirty hospitals were then randomly selected from the remaining 116 in the data set – 15 from the lower-mortality quartile and 15 from the higher-mortality quartile. Six additional hospitals over the proposed 24 were selected to take account of the likely drop-out rate during the recruitment and fieldwork phases of the study.

At the commencement of data collection there were 25 hospitals included in the study. However, only 20 hospitals in total, 10 each in the upper- and lower-mortality groups, were able to collect data on both COPD and heart failure. Thus only 20 hospitals were included in the overall analysis.

Figures 13 and 14 demonstrate the mortality differences between the two groups of hospital included in the final analysis, each mortality group (upper and lower) containing 10 hospitals.

Selection of review conditions

This quality and outcomes study used the same two conditions as in the phase one study – namely, admissions for acute exacerbation of COPD and of heart failure – in each study hospital.

The working definitions for data collection are shown in *Box 6*.

Number of case notes for review and of reviews

Analysis of the data from the first part of the study suggested that 40 case notes would be an appropriate number to review at each hospital, for each condition, in a study that was to seek associations between recorded quality-of-care and hospital-level outcomes. Furthermore, review of 40 case notes has been shown in the UK Stroke and COPD national audits^{34,47} to be practical and

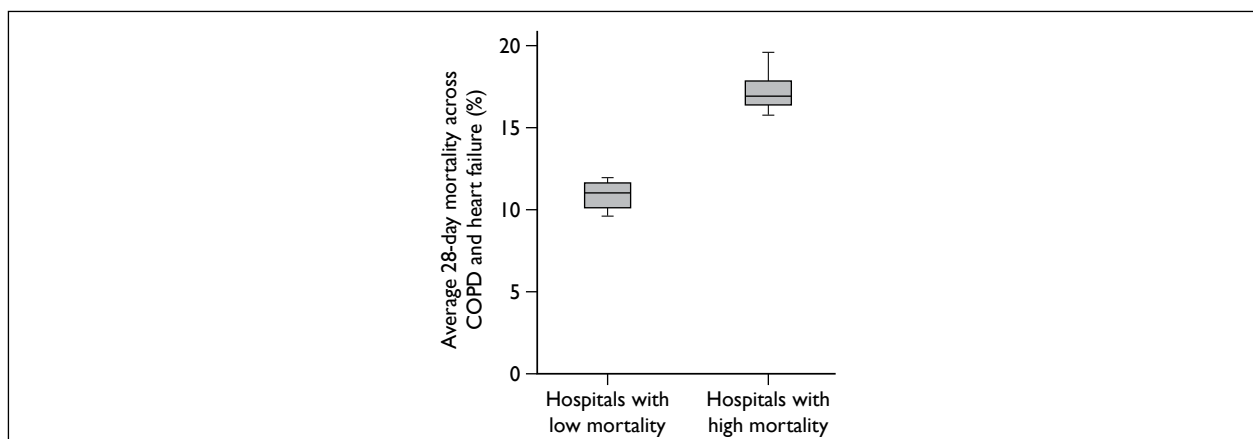


FIGURE 13 Box plot of COPD/heart failure mortality, showing differences between the higher- and lower-mortality groups.

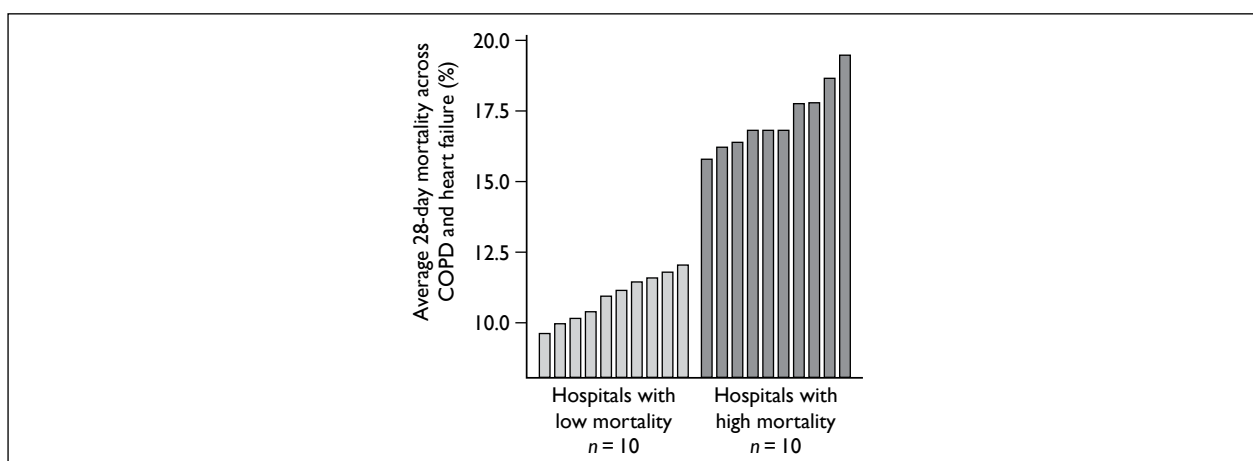


FIGURE 14 Bar chart showing individual hospital-level mortality rate differences.

BOX 6 Working definitions for data collection

Definition of an exacerbation of COPD

An exacerbation is a sustained worsening of the patient's symptoms from their usual stable state, which is beyond normal day-to-day variations, and is acute in onset. Commonly reported symptoms are worsening breathlessness, cough, increased sputum production and change in sputum colour.³¹

Definition of an exacerbation of heart failure

An exacerbation of heart failure is a sustained worsening of the patient's symptoms from their usual stable state, which is beyond normal day-to-day variations, and is acute in onset. Commonly reported symptoms are worsening breathlessness, tiredness and swelling of the feet and/or ankles.²²

to yield useful data. In contrast with the earlier analysis, it was not intended to undertake intra-rater and inter-rater reliability analyses, so it was not necessary for each set of records to be reviewed more than once.

In our proposal for phase two of the study we said that in order to test the reliability of the review process, 10 sets of case notes per condition per hospital would be double reviewed (rather than double reviewing all case notes). However, this

target of 10 double reviews proved not possible, partly because of the costs involved in undertaking the double reviews and partly because of inability to recruit second reviewers of the same staff type, beyond the 50 reviewers already required.

Choice and recruitment of reviewers

In our initial research proposal we said that only one review method would be used at this stage of the project, the choice being made following the analysis of the data in the reliability study. Analysis of the reliability study showed that while there were similarities between the criterion-based reviewing skills of the doctor and nurse, there were differences in the type of holistic data captured by the two professional groups and which formed the basis of the judgements about quality-of-care scores. Medically trained reviewers tended to judge the quality of interventions, whereas nurses tended to review from a care pathway perspective. As the study reported here sought associations between process of care and outcomes of care, it was judged that medically trained reviewers in higher-specialty training were likely to provide the most appropriate data.

Following the random selection of candidate hospitals, in each hospital a senior physician working in cardiology (heart failure care), and a senior physician working in COPD care were both asked to recruit a physician reviewer in mid-stage training in their relevant specialty.

Selection and refinement of review methods

One of the subsidiary purposes of the earlier reliability study was to determine which of two review methods was most appropriate for the quality-of-care/outcomes study – criterion-based review or holistic review. We had found that the two methods were complementary, although different,

and, therefore, that both methods might have a value in the review of care for the outcomes study. Review of care using evidence-based criterion methods would provide information on the extent to which care, overall, fitted with external standards. Structured holistic review with quality-of-care scores, together with short explanatory comments where reviewers thought it appropriate, would provide information about the nuances of care that could not be identified through the use of preconstructed review criteria. In this second part of the study it was not necessary to separate out the reviewing stages (holistic review followed by criterion-based review), as the results of the two methods were not being compared one with another, as in the earlier study. It was therefore possible to merge the holistic and criterion-based reviews into one process, so that the case notes were only reviewed once and a single data collection form could be used.

Using a form of structured implicit review, the reported quality of care provided to each patient was scored for each of three phases of care (admission/investigations, initial management and pre-discharge care) and for care overall.

In the earlier part of the study of holistic review, a six-point Likert scale was used by reviewers to rate the quality of care in each of the three phases of care, together with a 10-point Likert scale that was used to rate the overall quality of care. Although the 10-point scale was more finely grained, comparison in the phase one study between the six-point phase scores and the 10-point scale added to the complexity of the analysis without providing obvious benefits to the structured review process. A six-point scale was therefore used for all phases and overall care review in this study, with anchors of 1 = unsatisfactory and 6 = very best care. The descriptors for the six points are shown in *Box 7*.

BOX 7 Descriptors

1. Care fell short of current best practice in one or more significant areas, resulting in the potential for, or actual, adverse impact on the patient
2. Care fell short of current best practice in more than one significant area but is not considered to have the potential for adverse impact on the patient
3. Care fell short of current best practice in only one significant area but is not considered to have the potential for adverse impact on the patient
4. This was satisfactory care, only falling short of current best practice in more than two minor areas
5. This was good care, which only fell short of current best practice in one or two minor areas
6. This was excellent care and met current best practice

Two prompting questions had earlier been used to seek textual comments on the quality of care, but reviewers tended only to respond to one or other of the questions. Thus, in this study, only one question was asked of reviewers when they provided textual comment on the quality of care. *Box 8* shows the format of the questions used for phases of care and care overall.

The quality of care provided was also measured through the presence of a condition-specific set of review criteria for each of the two study conditions. A number of changes to the earlier data collection methods were required to make the criteria and data capture tools fit for purpose for the outcomes study. During the first phase study it became apparent that a limited number of the review criteria had far higher levels of missing data than other criteria. We presumed that this was because the data was routinely unavailable, or more difficult to find, or that the criteria were more difficult to understand. Whatever the reason, these criteria added nothing to the review process and were removed. Additionally, because of the potential impact of poor quality recording on the ability to assess quality of care, reviewers were asked to rank the quality of the information held in the clinical records on a Likert scale from 1 (poor) to 6 (excellent).

Reviewers were unaware of their hospital mortality ranking – whether they were in the higher- or lower-mortality quartile – and they evaluated the records within their own hospital, as would happen with a local clinical audit. No patient-identifiable data was returned to the study team or used in the analysis.

Reviewer training

Reviewers were trained to use a combination of two forms of case note review – criterion-based (explicit) review and holistic (implicit) review – and

also to provide written critical commentary on quality of care received by the patient, including on adverse events. The majority of reviewers were trained at a seminar in which the review methods were presented, using a set of anonymised case notes as the basis for small group and larger directed discussions. Topics discussed included techniques to find relevant information and group discussion to identify good and less satisfactory performance from the case notes.

Some reviewers were unable to attend the main training sessions and they received one-to-one or small group training from study research staff.

Selection of outcome data

An a priori choice of outcome measures relevant to the study was made by the multiprofessional study team. Measures were selected from a range of sources, relating to:

- clinical practice measures, such as 28-day mortality rates for COPD and heart failure (from Hospital Episode Statistics³³) and patients with myocardial infarction (MI) receiving thrombolysis within an hour⁴⁶
- hospital-level proxy clinical outcome data, such as Hospital Standardised Mortality Ratio (HSMR)⁵⁰
- proxy measures of patient safety and safety climate, for example from the NHS Staff Survey 2006⁵¹ and the National Patient Safety Agency (NPSA),⁵² used in an evaluation of incident reporting levels to the Agency (these data items were selected from a larger set as a result of previous empirical research)⁵³
- external review data, such the ability of the hospital to meet national targets for quality, collected by the Healthcare Commission (HCC) for England.⁴⁶

BOX 8 Investigations/examination (for example)

We are interested in comments about the quality of care the patient received and whether it was in accordance with current best practice (e.g. your professional standards). You may also wish to comment from your own professional viewpoint. If there is any other information that you think is important or relevant that you wish to comment on then please do so.

Please comment on the care received by the patient during this phase.

Please rate the care received by the patient during this phase.

Please tick only one box:

Unsatisfactory Very best care

BOX 9 Direct and proxy outcome hospital-level variables included in the analysis

Variable

Percentage of COPD or heart failure patients who die in hospital within 28 days
HSMR (3-year mortality) (Dr Foster)
HSMR (1-year mortality) (Dr Foster)
Numbers of incidents reported by the hospital to the NPSA per 100 bed-days
SMR for deaths in low-mortality health care-related groups (HRGs)
COPD or heart failure finished consultant NHS episodes [Hospital Episode Statistics (HES)]
COPD or heart failure bed-days (HES)
COPD or heart failure mean length of stay (HES)
COPD or heart failure mean age (HES)
HCC for England star rating (0 worst to 3 best)
Use of resources (HCC)
Patient's experience (HCC)
Quality of services (HCC)
Percentage of patients with acute MI receiving thrombolysis (HCC)
Extent to which hospital meets existing national targets (HCC)
Extent to which hospital meets new national targets (HCC)
NHS staff survey Q25a: Have seen errors in the past month (% yes)
NHS staff survey Q27b: Encouraged to report errors (mean)
NHS staff survey Q27e: The hospital takes action to ensure incident does not happen again (mean)
NHS staff survey Q24a: I know how to report (% yes)
NHS staff survey Q24b: I know the system for reporting (% yes)
NHS staff survey Q24b: I know the system for reporting (% no)
NHS staff survey Q24b: I know the system for reporting (% don't know)
NHS staff survey Q22e: Care of patient/service user is top priority (mean)
NHS staff survey Q22f: Happy with standard of care provided (mean)

The final set of outcome variables is shown in *Box 9*.

Analytical approach

Quantitative quality-of-care data from holistic scale scores and criterion scores were examined using summary statistics.

The outcomes data set for each hospital was constructed from the data available for each hospital on the data sets identified in *Box 9* (above). For the ranked mortality analyses, data

were combined for the 10 hospitals in each of the higher- or lower-mortality groups.

Correlation of the relationship between the quality-of-care (process) data and outcome data was undertaken using Pearson's correlation coefficient for continuous data, together with linear regression analysis. Multiple regression analysis was used where appropriate. Spearman's correlation coefficient was used for ordered categorical data (such as quality-of-service ratings). Levels of strength of correlation are shown in *Box 10*.

BOX 10 Levels of strength of correlation

$ r \geq 0.8$	Very strong relationship
$0.6 \leq r < 0.8$	Strong relationship
$0.4 \leq r < 0.6$	Moderate relationship
$0.2 \leq r < 0.4$	Weak relationship
$ r < 0.2$	Very weak relationship

All statistical tests were two-sided and a significance level of $p \leq 0.05$ was used for all analysis except multiple regression analysis, where a level of $p \leq 0.1$ was used.

Qualitative data from the textual commentary was examined by four reviewers, three of whom analysed data from 17 reviews each (of 40) to ensure some overlap, while a fourth reviewer reviewed a further sample of five from the total. Consistency among reviewers was checked after all had assessed output from two reviews, followed by discussion of the results. Further consistency check was undertaken between all four reviewers prior to completion of the analysis.

The results of each qualitative assessment of a reviewer's output, in which analysts identified cases being described as excellent, good, satisfactory, unsatisfactory or very unsatisfactory, were then checked against the overall scale scores to judge

whether a reviewer's quantitative scores were reflected in their qualitative description of the quality of care provided for a case.

Results

Within the 20 study hospitals, reviews of case notes were undertaken of 873 people with COPD and 692 people with heart failure (1565 in total). The numbers of reviews undertaken by each reviewer for COPD ranged from 8 to 40 (median 40, mean 35) and for heart failure from 10 to 49 (median 40, mean 35).

Reviewers were asked to assess the quality of recording in the case notes because of the potential impact of poor-quality recording on the ability to assess quality of care. On a scale of 1–6 the quality of case notes was scored satisfactory or better in 85% of the reviews.

Relationship between outcome variables and higher- and lower-mortality groups of hospitals

There were 10 hospitals in each mortality group. *Table 17* presents the mean scores for continuous outcome variables available from national data sets (excluding NHS staff survey outcomes) for the higher- and lower-mortality groups of hospitals. There was a significantly higher percentage of both

TABLE 17 Mean scores for continuous outcome variable by hospital mortality group

Variable	Higher mortality (SD)	Lower mortality (SD)	t-test (p-value)
HSMR from Dr Foster (3-year mortality)	107.1 (12.2)	99.2 (8.6)	1.67 (0.112)
HSMR from Dr Foster (1-year mortality)	106.3 (12.6)	95.6 (8.9)	2.20 (0.041)
Incidents to NPSA per 100 bed-days	0.74 (0.79)	0.70 (0.72)	0.13 (0.896)
SMR for deaths in low mortality HRGs	92.7 (41.2)	113.9 (38.0)	-1.20 (0.247)
COPD finished consultant episodes	1164 (1012)	1427 (815)	-0.63 (0.540)
COPD bed-days	7170 (5117)	7268 (3744)	-0.05 (0.962)
COPD mean length of stay	9.4 (1.8)	8.7 (2.5)	0.68 (0.507)
COPD mean age	69.2 (1.3)	69.1 (2.6)	0.12 (0.905)
Heart failure finished consultant episodes	636 (525)	765 (525)	-0.53 (0.601)
Heart failure bed-days	5298 (3425)	5680 (4025)	-0.22 (0.828)
Heart failure mean length of stay	12.8 (1.0)	12.5 (1.1)	0.59 (0.561)
Heart failure mean age	78.1 (1.5)	75.9 (2.3)	2.48 (0.024)
Patient's experience	68.2 (3.7)	67.2 (4.5)	0.58 (0.567)
Percentage of patients with acute MI receiving thrombolysis	56.3 (18.9)	56.6 (21.3)	-0.04 (0.967)

Correlations at the <0.05 significance level highlighted.

COPD and heart failure patients who died within 28 days in the higher-mortality group of hospitals than in the lower-mortality group of hospitals (as would be expected from the hospital selection criteria), and the 1-year mortality data available from the ‘Dr Foster’ analysis⁵⁰ was significantly higher for the higher-mortality group of hospitals.

Additionally, the mean age for heart failure patients was significantly higher for the higher-mortality group of hospitals than the lower-mortality group of hospitals. The differences in mortality rates between the two groups of hospitals were as expected from the selection criteria for those hospitals, so that the results serve to confirm the expected differences (see also *Figures 13 and 14*). The 1-year mortality data from the Dr Foster source also support the expected differences in mortality, and the higher mean age of people with heart failure in the higher-mortality group of hospitals may be a partial explanation for those higher mortality rates. No other significant associations were found.

Relationship between mortality groups and quality of care

Table 18 shows that most reviewers classify care in the good range [about 41% for COPD, and 46% for heart failure (score 5)], with around 19–21% in the ‘fell-short-of-best-practice’ range (score 1–3) and the lowest scores (1, unsatisfactory) under less than 5% for both conditions across the 20 hospitals.

The relationship between the holistic data (including mean scale scores for overall care, all of the phase of care scores) and the mean total criterion score, correlated with the hospital groups ranked by mortality, was analysed using the two independent samples *t*-test (*Tables 19 and 20*). Mean difference is the mean score for each case note review in the higher-mortality group of hospitals minus the mean score for the lower-mortality

group of hospitals. A negative mean difference in the table indicates that the lower-mortality group mean score is higher than the mean score for the higher-mortality group.

For both conditions the findings are therefore consistent. Across all phases of care and overall care using holistic review, and total criterion score, there was a trend towards higher mean scores for the lower-mortality hospital groups. However, when account was taken of clustering, there were no statistically significant differences in scores between the two groups of hospitals.

The similarities between the quality-of-care scores in the upper and lower hospitals are further demonstrated when the relationship between mean overall and mean phase holistic scale scores and ranked mortality is explored. *Tables 21–24* show the relationships between holistic scores (ranging from 1 = unsatisfactory to 6 = very best care) and ranked mortality. No significant differences in scores were found between the upper and lower mortality groups of hospitals.

The relationship between the mean holistic overall quality-of-care scores for COPD and heart failure in each hospital was also explored using a Bland–Altman plot (*Figure 15*) to investigate whether the two scores tended to be in the same direction in each hospital. That is, if the hospital was in the lower-mortality group (which, on average, had somewhat higher group mean scores for both COPD and heart failure) was this pattern replicated within each hospital? No clear pattern emerged between the two mean scores for each hospital when grouped for overall mortality, so that it is not possible to see a trend towards higher scores for both conditions in each of the lower-mortality hospitals, or for the lower scores for each condition in each of the hospitals in the higher-mortality group.

TABLE 18 Distribution of overall holistic quality-of-care scores – total number (and percentage) of reviews in each category

	Scale scores						Total number of reviews
	Fell short of good practice			Satisfactory	Good or better		
	1 (unsatisfactory)	2	3	4	5 (good)	6 (very best care)	
COPD (no. reviews, %)	38 (4.4)	56 (6.4)	87 (10.0)	166 (19.1)	362 (41.3)	164 (18.8)	873 (100)
Heart failure (no. reviews, %)	26 (3.8)	44 (6.3)	60 (8.7)	154 (22.2)	318 (45.9)	91 (13.1)	692 (100)

TABLE 19 Relationship between the COPD mean holistic scale scores, total criterion score and hospital mortality group

	Lower-mortality group scores			Higher-mortality group scores			Mean difference	95% CI		p-value
	n	Mean	SD	n	Mean	SD		Lower	Upper	
Total criterion score ^a	438	10.3	1.7	447	10.1	1.8	-0.2	-0.9	0.6	0.637
Holistic: admission phase	434	4.7	1.2	441	4.5	1.4	-0.2	-0.6	0.2	0.321
Holistic: initial management phase	435	4.8	1.2	438	4.7	1.2	-0.1	-0.4	0.2	0.649
Holistic pre-discharge phase	423	4.8	1.1	423	4.7	1.2	-0.1	-0.4	0.2	0.542
Holistic: overall quality-of-care rating	428	4.5	1.3	444	4.3	1.3	-0.2	-0.6	0.3	0.419

a Six-point scale.
Mean difference is the mean score for case notes of patients in the lower-mortality group hospitals minus mean score for the higher-mortality group.
The p-values and CIs were adjusted for clustering by reviewer.

TABLE 20 Relationship between the heart failure mean holistic scale scores, total criterion score and hospital mortality group

	Lower-mortality group scores			Higher-mortality group scores			Mean difference	95% CI		p-value
	n	Mean	SD	n	Mean	SD		Lower	Upper	
Total criterion score ^a	393	8.0	1.4	309	7.8	1.5	-0.2	-0.7	0.2	0.312
Holistic: admission phase	393	4.7	0.9	309	4.6	1.2	-0.1	-0.6	0.3	0.501
Holistic: initial management phase	382	4.6	1.1	304	4.5	1.3	-0.1	-0.6	0.3	0.541
Holistic pre-discharge phase	384	4.7	1.0	300	4.3	1.4	-0.4	-1.0	0.1	0.122
Holistic: overall quality-of-care rating	385	4.6	1.1	308	4.2	1.4	-0.4	-0.9	0.1	0.115

a Six-point scale.
Mean difference is the mean score for case notes of patients in the lower-mortality group hospitals minus mean score for the higher-mortality group.
The p-values and CIs were adjusted for clustering by reviewer.

TABLE 21 COPD – mean holistic overall score by mortality group

	Mean (SD)	Median	Range
Higher-mortality hospital group (n = 10)	4.3 (0.62)	4.4	2.9–5.0
Lower-mortality hospital group (n = 10)	4.5 (0.56)	4.5	3.7–5.5
There were no statistically significant differences in mean scores per mortality group ($t=-0.83, p=0.418$).			

TABLE 22 Heart failure – mean holistic overall score by mortality group

	Mean (SD)	Median	Range
Higher-mortality hospital group (n = 10)	4.2 (0.58)	4.1	3.1–5.1
Lower-mortality hospital group (n = 10)	4.6 (0.38)	4.6	3.9–5.1
There were no statistically significant differences in mean scores per mortality group ($t=-1.75, p=0.097$).			

TABLE 23 COPD – mean holistic phase score by mortality group

	Mean (SD)	Median	Range
Higher-mortality hospital group (n = 10)	13.4 (1.65)	13.6	10.2–15.2
Lower-mortality hospital group (n = 10)	13.9 (1.30)	14.0	11.6–16.1
There were no statistically significant differences in mean scores per mortality group ($t=-0.77, p=0.453$).			

TABLE 24 Heart failure – mean holistic phase score by mortality group

	Mean (SD)	Median	Range
Higher-mortality hospital group (n = 10)	13.0 (1.83)	13.4	10.3–15.3
Lower-mortality hospital group (n = 10)	13.8 (1.44)	13.8	11.5–15.8
There were no statistically significant differences in mean scores per mortality group ($t=-1.10, p=0.288$).			

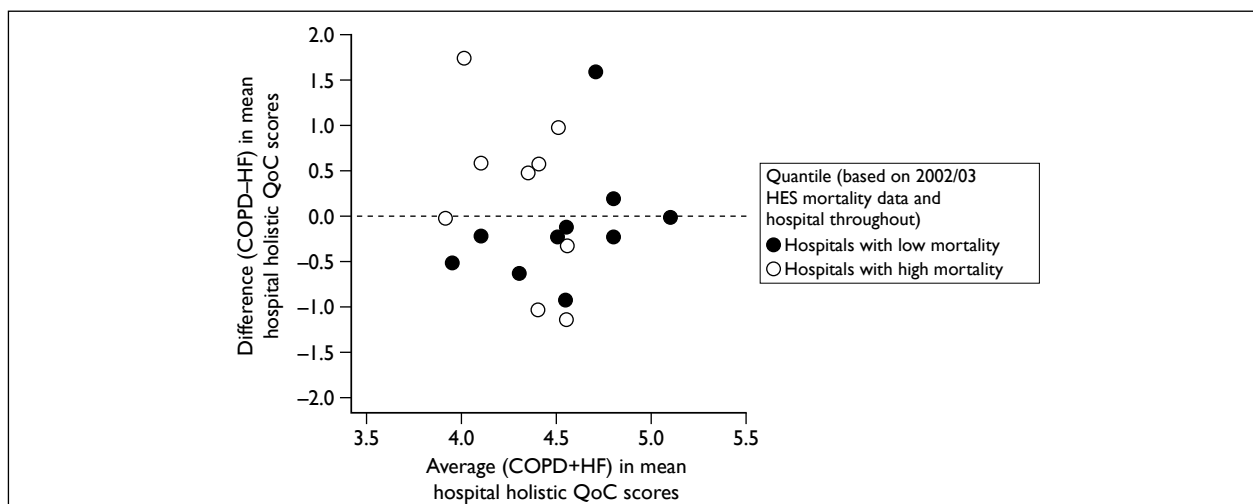


FIGURE 15 Relationship between mean quality-of-care scores for COPD and heart failure within each hospital, grouped by mortality level (note: one hospital in the higher-mortality hospital group was an outlier in all of the analyses and was therefore excluded from the Bland–Altman plot analysis). HF, heart failure.

Correlations between holistic scale scores and hospital-level outcome variables

Pearson's correlation coefficient was used for analysis of continuous data. Spearman's correlation coefficient was used for analysis of ordered categorical data (e.g. HCC star rating). Levels of strength of correlation are assumed as in *Box 10*, above.

Rather than using statistical significance to identify important results, correlations were identified as important if the correlation coefficient was greater than 0.4; that is, at least a moderate relationship existed between the two variables under investigation. Very weak and weak relationships ($r \leq 0.4$) were regarded as having no correlation.

There were only a limited number of moderate to very strong correlations between the quality-of-care scores for the upper- and lower-mortality hospital groups and the outcome and risk-adjustment variables selected for the study. A positive correlation indicates that, as criterion or overall holistic score increases, the variable levels also increase. Thus, for example, a positive correlation for NHS staff survey question 22e (care of patient/service user is top priority) demonstrates an increase in COPD or heart failure criterion score, as there is an increase in the mean score for staff who agree that care of patient/service user is top priority in the hospital.

An example of a negative correlation would be an increase in COPD or heart failure score that correlates with a decrease in the number

of incidents reported to the NPSA per 100 bed-days (increased reporting is thought to indicate a positive safety culture⁵³).

Tables 25–28 present the moderate to very strong correlations only. The results of all of the holistic data correlations, including those without statistically significant results, are presented in Appendices 7–10. Pearson correlation coefficients are used unless otherwise indicated in the table. Each table is accompanied by scatter plots fitted with regression lines to demonstrate the strength of the correlation.

COPD mean holistic overall score (1 = unsatisfactory to 6 = very best care)

Table 25 shows that there were only two moderate negative correlations. Thus, as the COPD scale score increased (an indication of better quality of care), the frequency decreased at which NHS staff said they were encouraged to report errors (Figure 16), and also when staff said that care of the patient/service user was a high priority (Figure 17). These correlations are in an unexpected direction, as better quality of care might have been expected to relate to both a positive safety culture and to high priority for patient care.

Heart failure mean holistic overall score (1 = unsatisfactory to 6 = very best care)

A total of four variables were significantly correlated with heart failure mean score: quality of services, existing national targets, new national targets, and NHS staff survey question 24a (know how to report concerns) (Table 26).

TABLE 25 Correlation between COPD holistic mean overall score and outcome variables

Variable	Correlation coefficient	p-value	Relationship
NHS staff survey Q27b: Encouraged to report errors (mean)	-0.490	0.028	Moderate
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	-0.503	0.024	Moderate

TABLE 26 Correlation between heart failure holistic mean overall score and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Quality of services	0.651 ^a	0.002	Strong
Existing national targets	0.765 ^a	<0.001	Strong
New national targets	0.453 ^a	0.045	Moderate
NHS staff survey Q24a: Know how to report concerns (% yes)	0.509	0.022	Moderate

a Spearman's rank correlation used.

TABLE 27 Correlation between COPD holistic mean phase score and outcome variables

Variable	Correlation coefficient	p-value	Relationship
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	-0.454	0.044	Moderate

TABLE 28 Correlation between heart failure holistic mean phase score and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Heart failure mean age	-0.552	0.014	Moderate
Quality of services	0.486 ^a	0.030	Moderate
Percentage of patients with acute MI receiving thrombolysis	0.463	0.046	Moderate
Meets existing national targets	0.691 ^a	0.001	Strong
NHS staff survey Q27e: Hospital takes action to ensure does not happen again (mean)	0.470	0.037	Moderate
NHS staff survey Q24a: Know how to report (% yes)	0.546	0.013	Moderate
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	0.470	0.037	Moderate

a Spearman's rank correlation used.

There is a strong positive correlation between quality of services (fair, good or excellent) and heart failure mean score: those trusts with a fair score had lower mean heart failure scores than trusts with a good or excellent score (Figure 18). Existing national targets (partially met, almost met,

fully met) were also strongly positively correlated with heart failure mean scores where an increase in mean scores indicated an increased likelihood of trusts meeting existing national targets (Figure 19). There was a moderate positive correlation between new national targets (weak, fair, good,

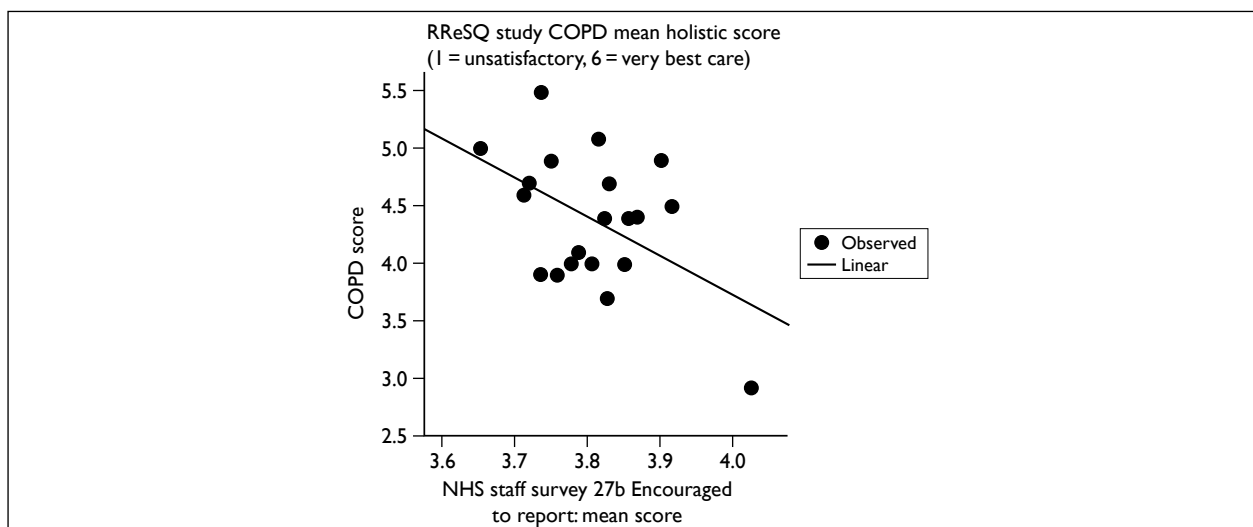


FIGURE 16 Scatter plot (with regression line) for COPD score by NHS staff survey Q27b: Encouraged to report errors (mean score). RReSQ, Records Review for Safety and Quality (study).

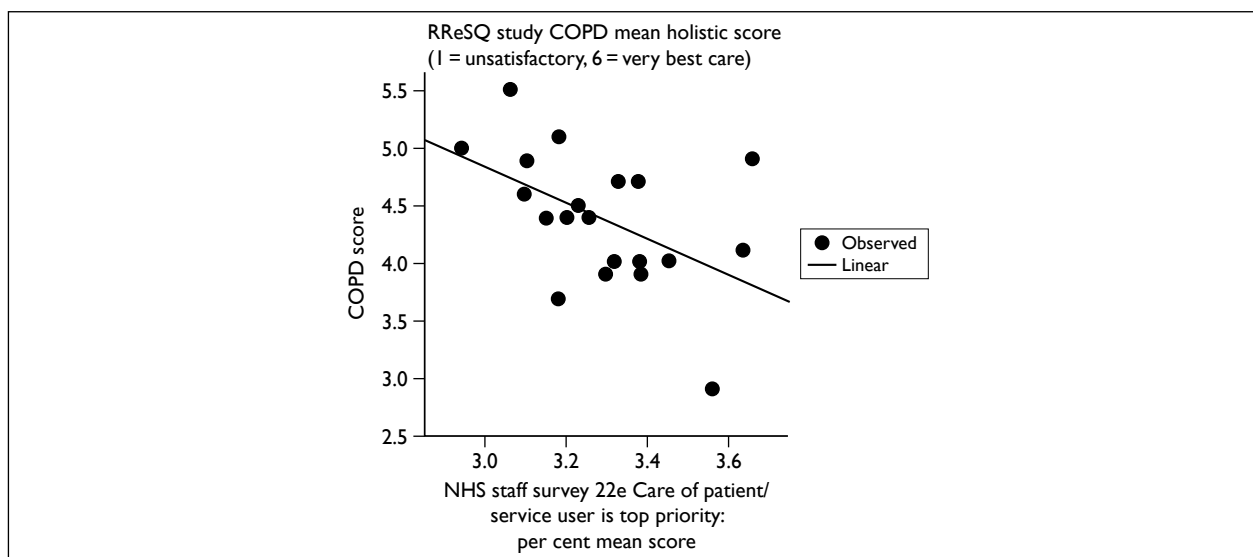


FIGURE 17 Scatter plot (with regression line) for COPD score by NHS staff survey Q22e: Care of patient/service user is top priority (mean score). RReSQ, Records Review for Safety and Quality (study).

excellent) and heart failure – mean heart failure scores increased with increased levels of excellence (Figure 20).

Finally, there was a moderate positive correlation between NHS staff survey question 24a and heart failure mean score: the higher the percentage of responders who knew how to report incidents, the higher was the mean heart failure score (Figure 21).

COPD mean holistic phase score

Only NHS staff survey question 22e (care of patient/service user is top priority) was significantly correlated with COPD mean holistic score (Table 27). As with the overall score analysis above (Table 25), the correlation between the two variables was negative, and as COPD mean holistic score increases the mean score for care of patient/service users being a top priority decreased – a lower score indicates more disagreement with the statement (Figure 22).

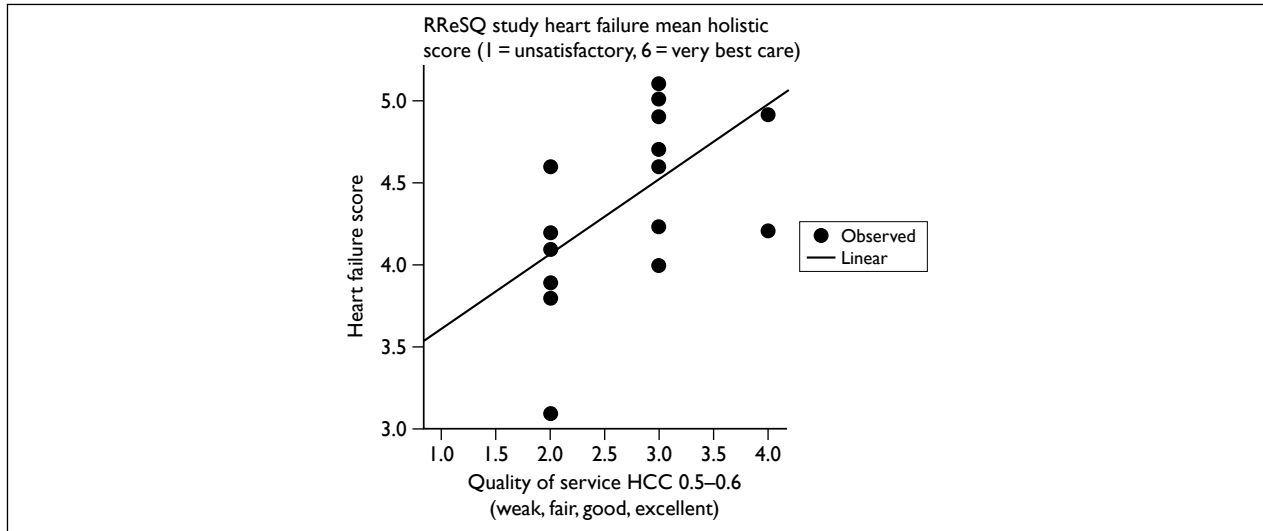


FIGURE 18 Scatter plot (with regression line) for heart failure score by quality of service. RReSQ, Records Review for Safety and Quality (study).

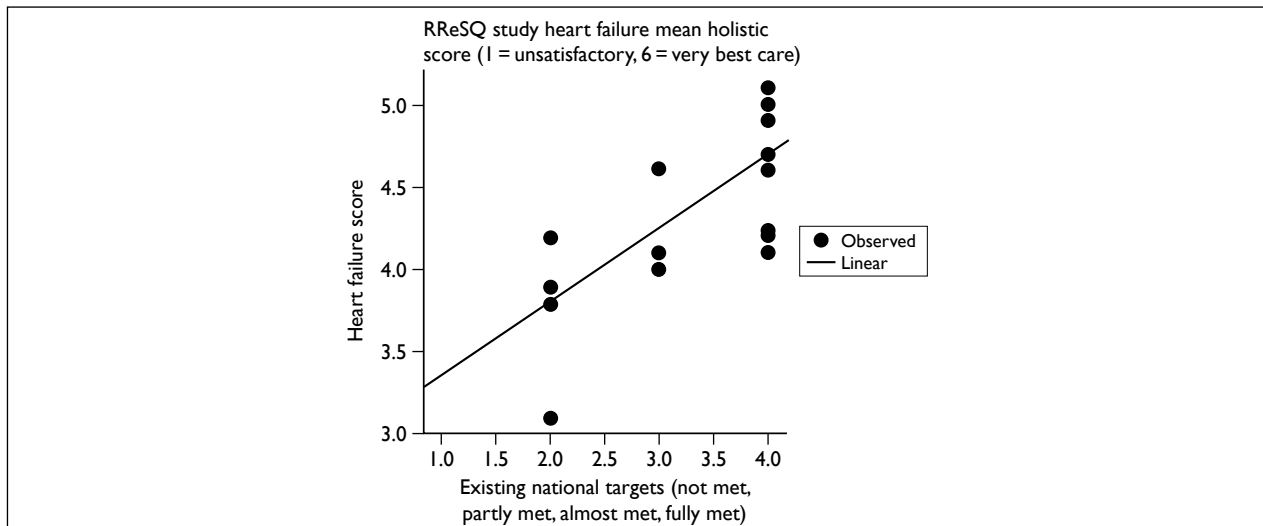


FIGURE 19 Scatter plot (with regression line) for heart failure score by existing national targets. RReSQ, Records Review for Safety and Quality (study).

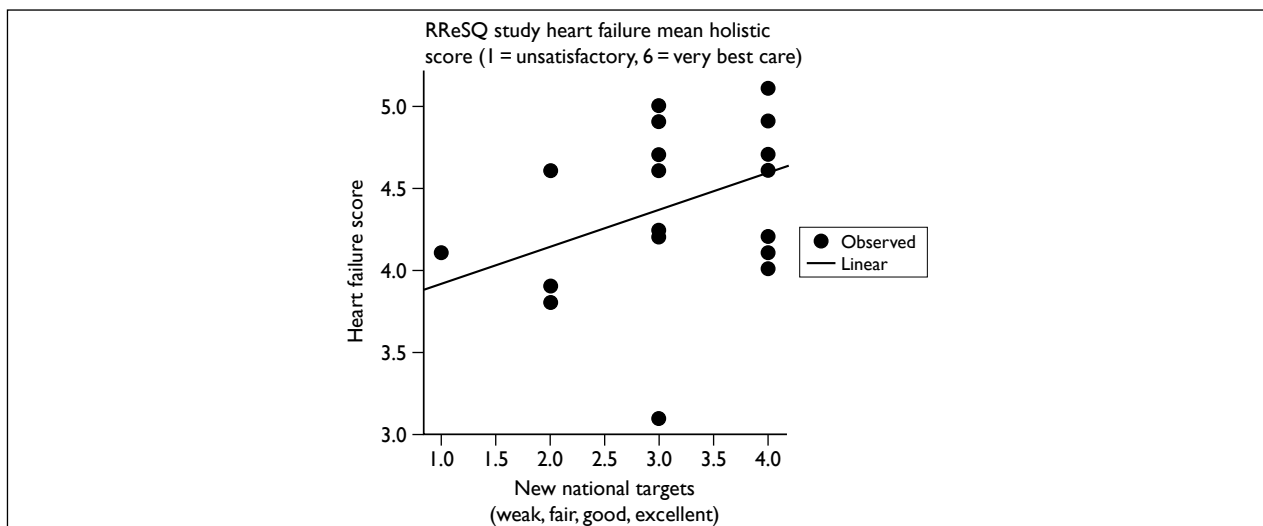


FIGURE 20 Scatter plot (with regression line) for heart failure score by new national targets. RReSQ, Records Review for Safety and Quality (study).

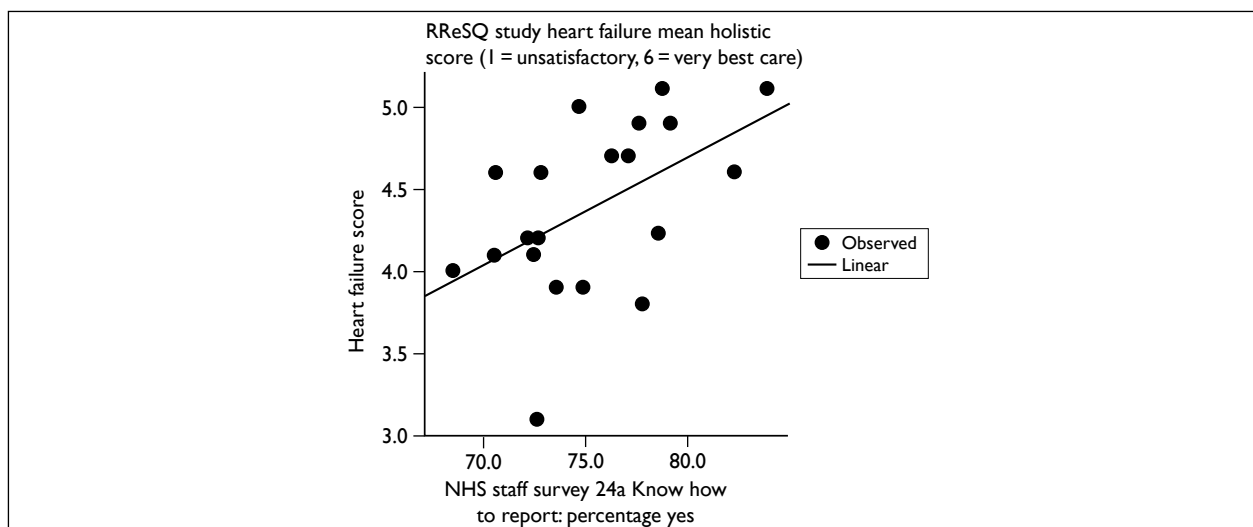


FIGURE 21 Scatter plot (with regression line) for heart failure score by NHS staff survey Q24a: Know how to report (% yes). RReSQ, Records Review for Safety and Quality (study).

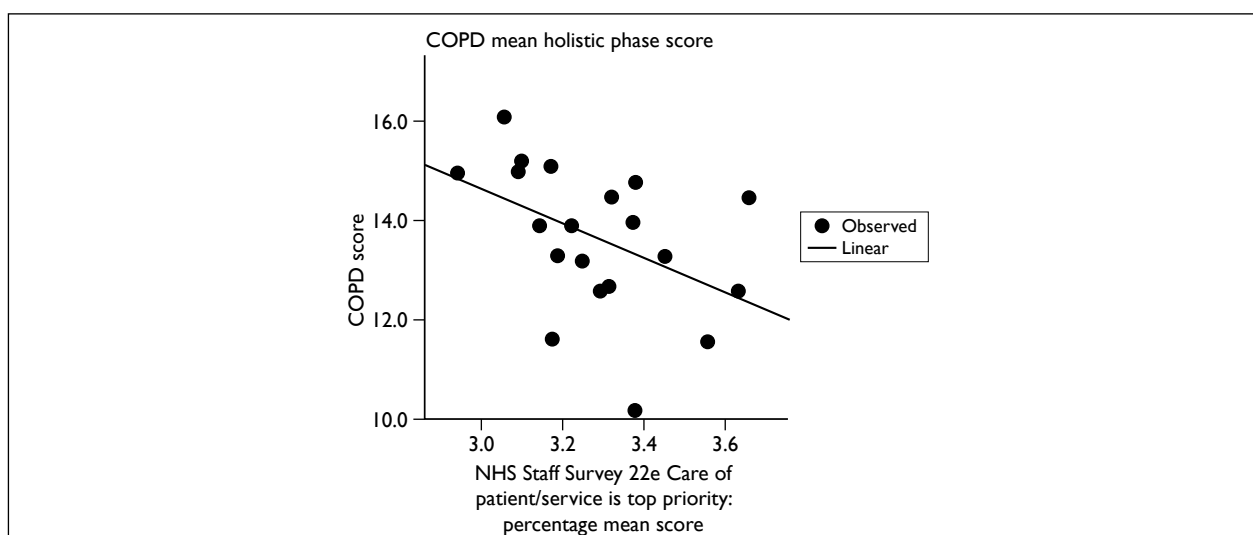


FIGURE 22 Scatter plot (with regression line) for COPD holistic score by NHS staff survey Q22e: Care of patients/service users is top priority (mean score).

Heart failure mean holistic phase score

Seven variables were significantly correlated with heart failure mean holistic score:

1. heart failure mean age [Hospital Episode Statistics (HES) data]
2. quality of services
3. percentage of patients with acute MI receiving thrombolysis
4. existing national targets
5. NHS staff survey question 27e (trust takes action to ensure never happens again)
6. NHS staff survey question 24a (know how to report incidents)

7. NHS staff survey question 22e (care of patient/service user is top priority).

Mean age is negatively correlated with score: as mean age increases mean holistic score decreases (Figure 23). This may be explained by older people having more complex problems, although reviewers in the phase one study showed that quality of care could be scored highly even when a patient did not survive.

Quality of service (fair, good, excellent) is positively correlated with mean score; thus, as heart failure mean phase score increases, the more likely a hospital is to be of good or excellent status (Figure

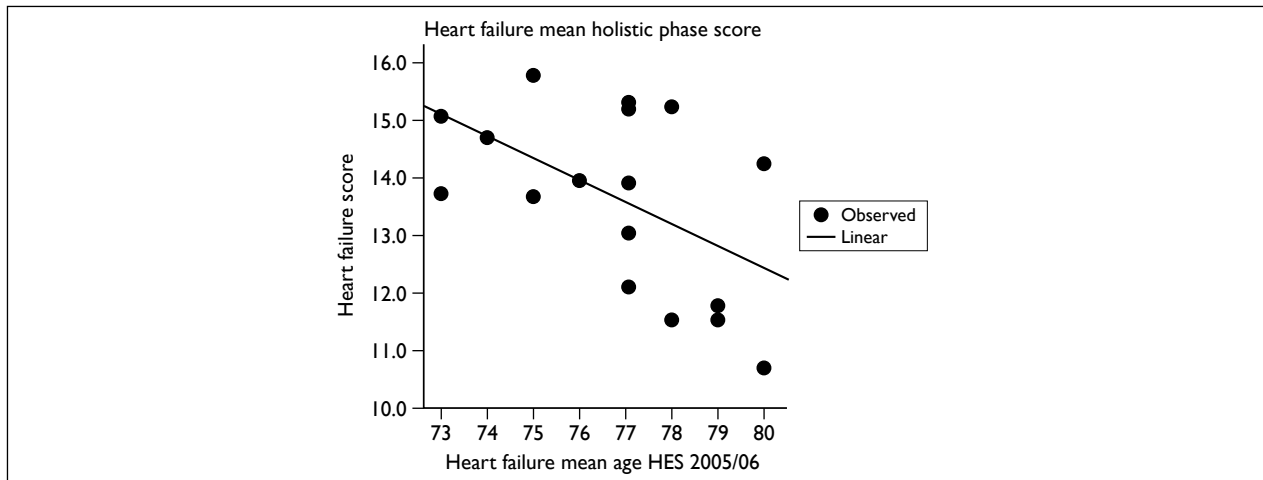


FIGURE 23 Scatter plot (with regression line) for heart failure holistic score by mean patient age.

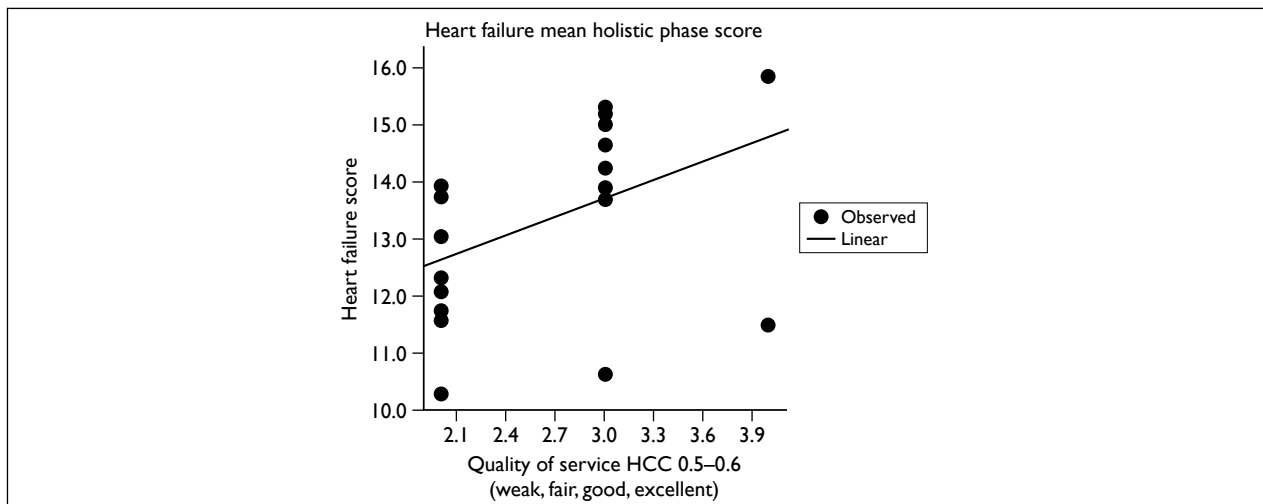


FIGURE 24 Scatter plot (with regression line) for heart failure holistic score by quality of services.

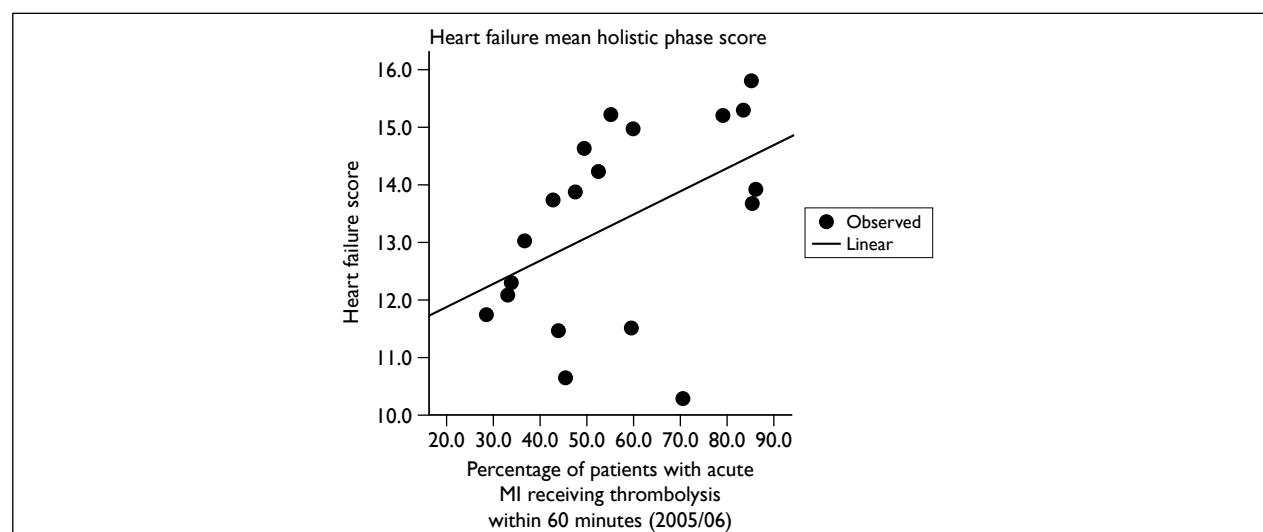


FIGURE 25 Scatter plot (with regression line) for heart failure holistic score by percentage of patients with acute MI receiving thrombolysis.

24). There is a moderate positive correlation between the percentage of patients with acute MI receiving thrombolysis and heart failure mean holistic phase score (an important association in heart failure care) (Figure 25). A strong positive correlation exists between existing national targets (partly met, almost met, fully met) and heart failure mean holistic phase score, mean scores increase as the level of meeting targets increases (Figure 26).

Additionally, there were moderate positive correlations between mean score for NHS staff

survey questions 27e and 22e and heart failure mean holistic phase scores – the higher the heart failure score, the stronger the respondent agreed with the statement that the hospital takes action to ensure that it does not happen again or care of patient/service user is top priority (Figure 27). Finally, there was a moderate positive correlation between NHS staff survey question 24a and heart failure mean holistic phase score, where the higher the percentage of responders agreeing (% yes) they know how to report incidents results, the higher the heart failure mean holistic score (Figure 28).

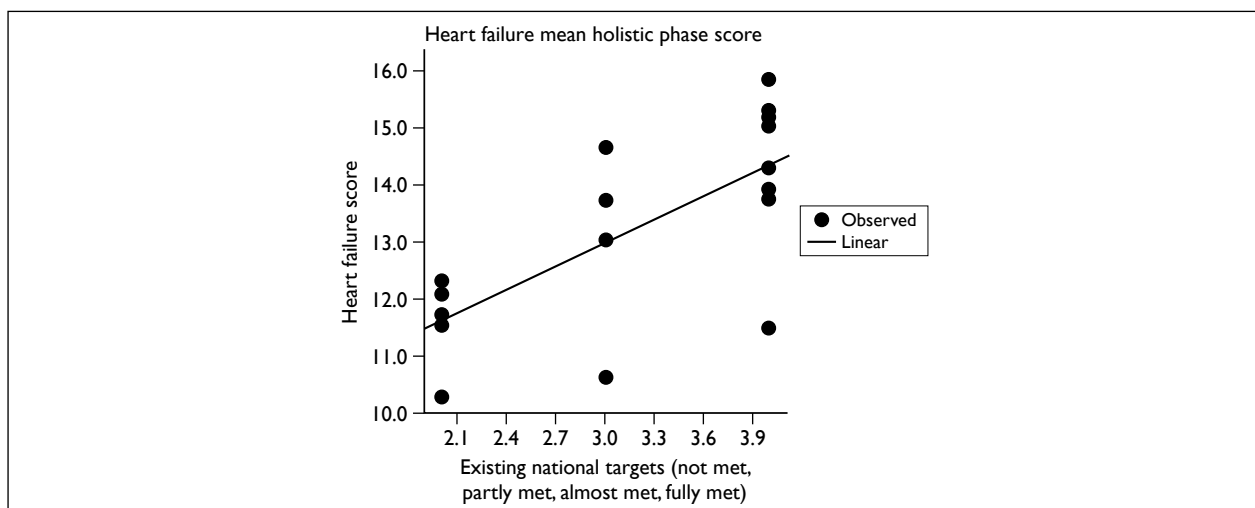


FIGURE 26 Scatter plot (with regression line) for heart failure holistic score by existing national targets.

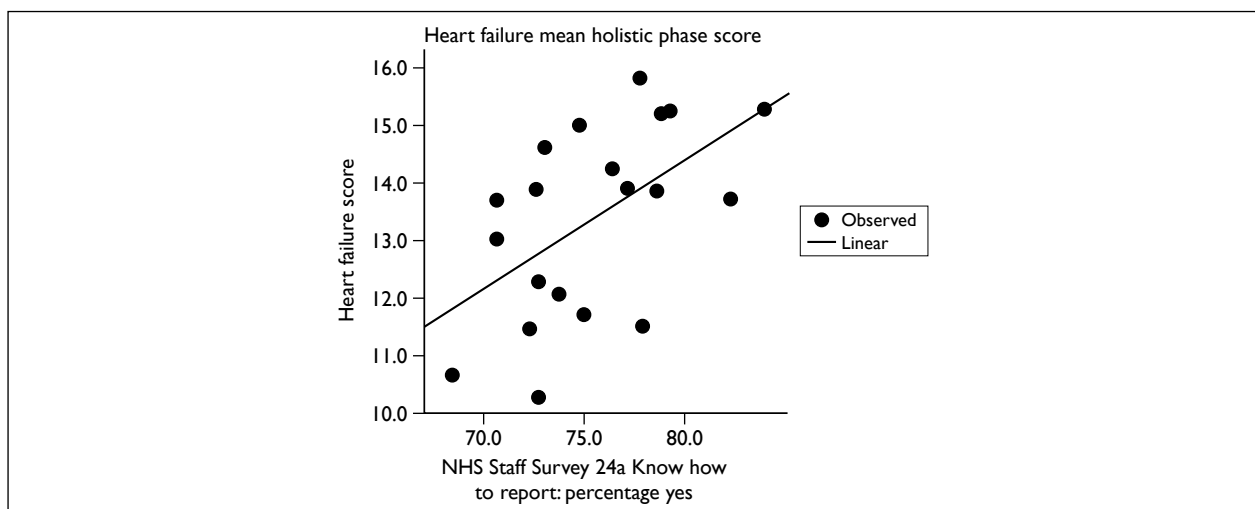


FIGURE 27 Scatter plot (with regression line) for heart failure holistic score by NHS staff survey Q24a: Know how to report (% yes).

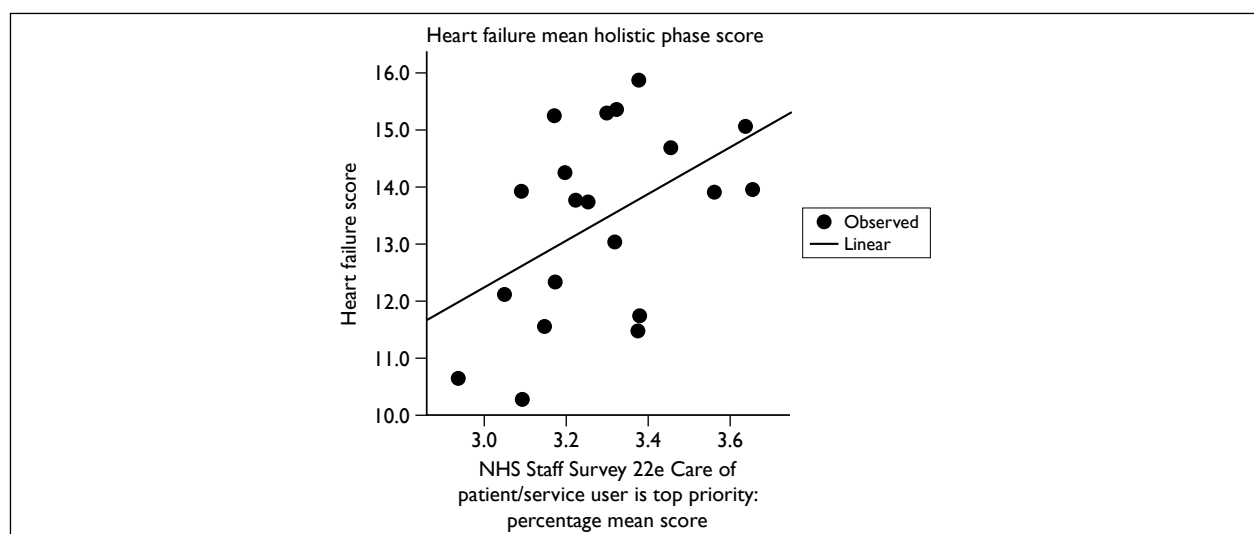


FIGURE 28 Scatter plot (with regression line) for heart failure holistic score by NHS staff survey Q22e: Care of patients/service users is top priority (mean score).

TABLE 29 COPD mean criterion-based score (out of 13) by mortality group

	Mean (SD)	Median	Range
Higher-mortality group hospitals (n = 10)	10.1 (1.10)	10.4	7.9–11.5
Lower-mortality group hospitals (n = 10)	10.2 (0.73)	10.7	8.4–11.1

There were no statistically significant differences in mean scores per mortality group ($t = -0.21, p = 0.840$).

TABLE 30 Heart failure mean criterion-based score (out of 11) by mortality group

	Mean (SD)	Median	Range
Higher-mortality group hospitals (n = 10)	7.8 (0.41)	7.9	7.0–8.4
Lower-mortality group hospitals (n = 10)	8.0 (0.50)	7.9	7.5–9.3

There were no statistically significant differences in mean scores per mortality group ($t = -0.83, p = 0.417$).

Correlations between criterion-based scores and hospital-level outcome variables

Tables 29 and 30 show the relationships between criterion scores and mortality grouping. There were no statistically significant differences in mean scores per mortality group. These findings are similar to the holistic quality-of-care data (Tables 21–24).

Correlations between criterion-based scores and outcome variables

Correlations at the moderate to very strong level are shown in Tables 31 and 32, with accompanying

scatter plots. The results of all correlations are presented in Appendices 11 and 12.

COPD mean criterion score (out of 13)

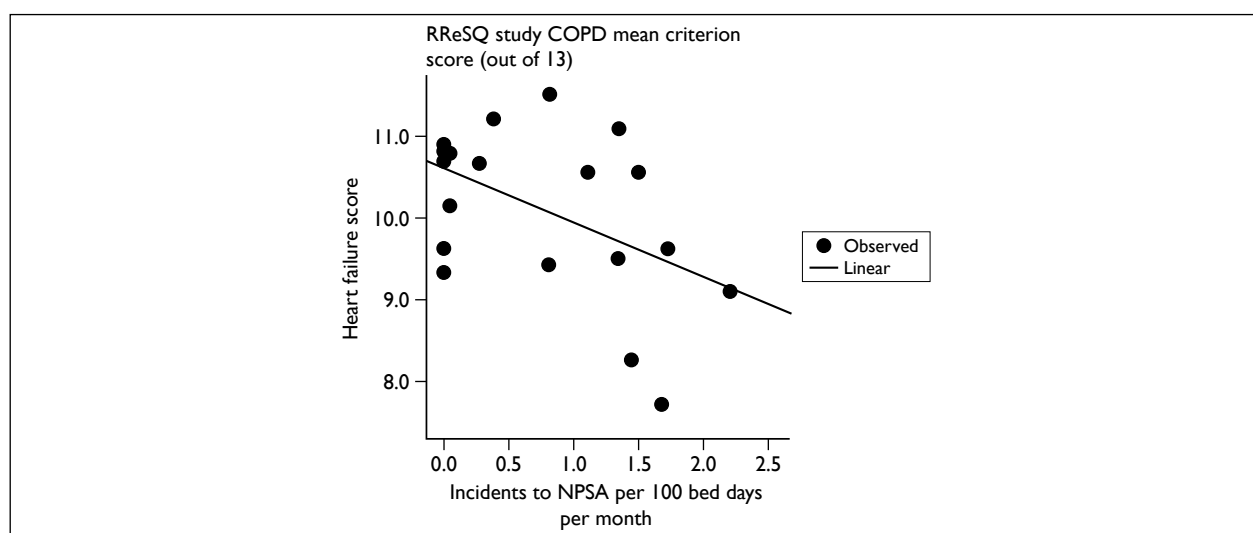
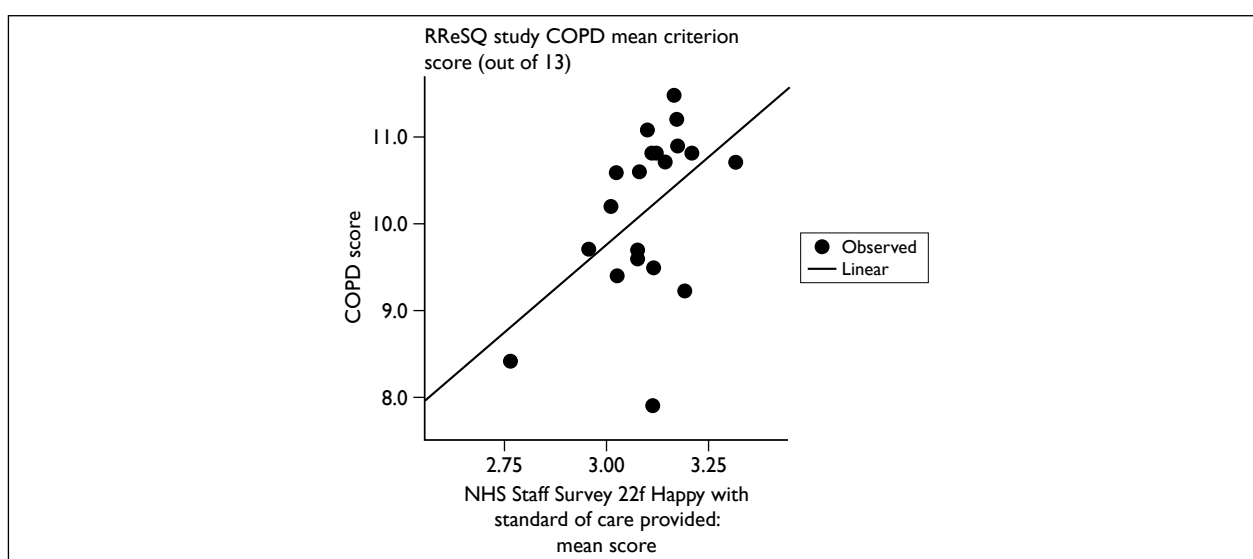
Two variables were significantly correlated with COPD mean criteria score: incidents reported to the NPSA per 100 bed-days and NHS staff survey question 22f (happy with standard of care provided) (Table 31). A moderate negative correlation was observed with NPSA bed-days, as the number of incidents per 100 bed-days decreased the mean criteria score increased (Figure 29). Again this is an unexpected direction for the correlation, as increased reporting is a

TABLE 31 Correlations between COPD mean criterion score and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Incidents reported to NPSA per 100 bed-days	-0.489	0.029	Moderate
NHS staff survey Q22f: Happy with standard of care provided (mean)	0.475	0.034	Moderate

TABLE 32 Correlations between heart failure mean criterion score and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Incidents reported to NPSA per 100 bed-days	0.528	0.017	Moderate
NHS staff survey Q24b: System for reporting (% no)	0.620	0.004	Strong

**FIGURE 29** Scatter plot (with regression line) for COPD criterion score by incidents reported to NPSA per 100 bed-days. RReSQ, Records Review for Safety and Quality (study).**FIGURE 30** Scatter plot (with regression line) for COPD algorithmic score by NHS staff survey Q22f: Happy with standard of care provided (mean score). RReSQ, Records Review for Safety and Quality (study).

positive safety culture marker. A moderate positive correlation was observed with NHS staff survey 22f (mean score), where the higher the score, the happier the staff with their standard of care; thus, as the mean response to 22f increased, mean COPD score increased (Figure 30).

Heart failure mean criteria score (out of 11)

Two variables were significantly correlated with heart failure mean criteria score: incidents reported to the NPSA per 100 bed-days and NHS staff survey question 24b (knowing the system for reporting errors) (Table 32). In a reversal of results

from the COPD criterion analysis there was a significant positive correlation between incidents reported to the NPSA and heart failure mean score, as the number of incidents increased the mean heart failure score increased (Figure 31). There was also a strong positive correlation between the percentage of responders to the NHS staff survey stating they did not know how to report errors confidentially and mean heart failure score. That is, as mean heart failure score increased, the percentage of responders reporting that they did not know how to report concerns also increased (Figure 32).

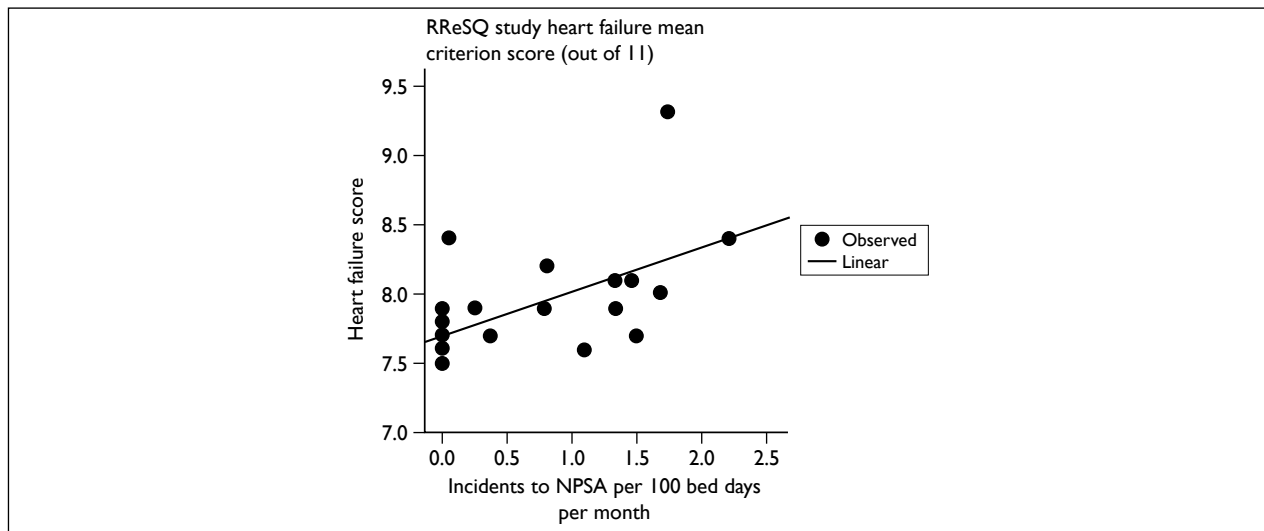


FIGURE 31 Scatter plot (with regression line) for heart failure criterion score by NPSA incidents bed-days. RReSQ, Records Review for Safety and Quality (study).

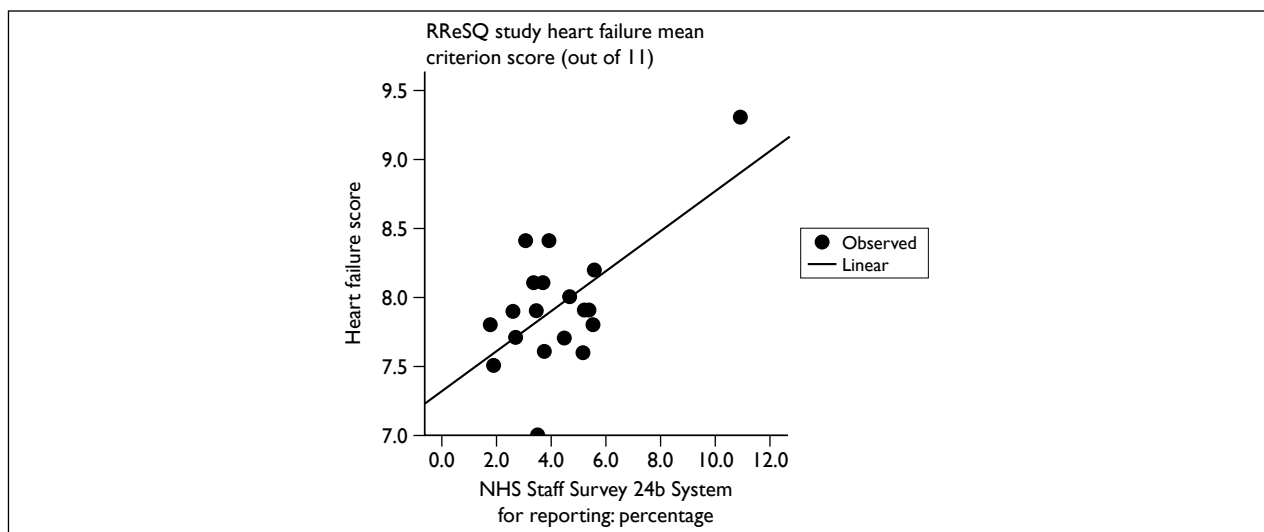


FIGURE 32 Scatter plot (with regression line) for heart failure criterion score by NHS staff survey Q24b: System for reporting (% no). RReSQ, Records Review for Safety and Quality (study).

Outliers in the analysis

It was agreed to identify which hospitals were outliers in each analysis to establish whether it was the same hospital each time. The following outliers were identified (hospital scores presented in parentheses):

- COPD mean holistic overall score – hospital RC9 (2.9)
- Heart failure mean holistic overall score – hospital RN1 (3.1)
- COPD mean holistic phase score – hospital RQW (10.2)
- Heart failure mean holistic phase score – *no outlier identified*
- COPD mean criterion-based score – *no outlier identified*
- Heart failure mean criterion-based score – hospital RNH (9.3), hospital RQW (7.0).

Hospital RNH (from the heart failure criterion score analysis) was the most noticeable outlier in all analyses and a decision was taken to rerun the analysis to see if there were any changes in the correlation results.

There were some minor differences in results following the exclusion of hospital RNH. Incidents reported to the NPSA remained moderately correlated with heart failure mean score: the higher the number of incidents, the higher the mean heart failure score. The percentage of heart failure patients dying in hospital within 28 days moderately correlated with heart failure mean score: the lower the percentage of patients who die within 28 days, the higher the mean score. Finally, NHS staff survey question 24b [system for reporting error (% replying that they did not know of the system)] is no longer strongly correlated with mean heart failure score.

Regression analysis

A priori, one of the objectives of this analysis was to fit multiple regression models to COPD and heart failure scores to establish any predictors. With a sample of 20 hospitals the number of variables in a multiple regression model should be restricted to two, as anything over two would result in overfitting a model. Furthermore, consideration should be taken of categorical variables such as existing national targets, where each level of response (not met, partly met, almost met, fully met) counts as one variable in the regression analysis – thus a categorical variable with three or more categories, which is the case for all categorical variables in this data set, would result in overfitting.

It was possible to fit a multiple regression model for COPD criterion score, where NPSA incidents and NHS staff survey question 22f (happy with standard of care provided) were significant predictors of COPD score at the 10% significance level. A one-unit increase in the number of incidents reported to NPSA per 100 bed-days resulted in a 0.5 decrease in COPD criterion score, and a one-unit increase in the staff who were happy with standard of care resulted in a 3.14 change in COPD score. Multiple regression models were fitted using the forwards stepwise method, by which the most significant variable at the univariate stage is selected first.

When the regression model was fitted to the COPD criterion score, this model included only the two variables that were significantly correlated with the score. When included in a regression model together, neither 'incidents per 100 bed-days' nor 'NHS staff survey question 22f' were significant at the 5% significance level (*Table 33*).

A multiple regression model was also fitted to heart failure criterion score, where NPSA incidents and NHS staff survey question 22f were significant predictors of heart failure score at the 10% significance level. A one-unit increase in the number of incidents reported to NPSA per 100 bed-days resulted in a 0.21 increase in heart failure criterion score, and a one-unit increase in the percentage of staff reporting 'no' to system for reporting errors resulted in a 0.11 change in heart failure score.

When a regression model was fitted to the heart failure criterion score, this model included only the two variables that were significantly correlated with criterion score. When included in a regression model together, incidents per 100 bed-days was no longer significant when included with NHS staff survey question 24b (percentage of staff reporting they did not know of the system) (*Table 34*).

For the remaining four analyses it was not possible to fit a multiple regression model, either because no two variables were significant in the same model or due to problems with overfitting for categorical variables.

Discussion

This part of the study set out to explore how much of the variance in outcome for two important clinical conditions (COPD and heart failure) can be

TABLE 33 COPD criterion score regression with outcome variables

	Regression coefficient (standard error)	t-test	p-value
Constant	0.82 (5.48)	0.15	0.883
Incidents reported to NPSA per 100 bed-days	-0.50 (0.26)	-1.89	0.076
NHS staff survey Q22f: Happy with standard of care provided (mean)	3.14 (1.75)	1.79	0.091

TABLE 34 Heart failure criterion score regression with outcome variables

	Regression coefficient (standard error)	t-test	p-value
Constant	7.30 (0.19)	38.98	<0.001
Incidents reported to NPSA per 100 bed-days	0.21 (0.12)	1.77	0.095
NHS staff survey Q24b: System for reporting – per cent reporting they did not know	0.11 (0.04)	2.56	0.020

attributed to differences in quality of clinical care in acute hospitals. The main findings of the study were:

- Within-hospital and between-hospital variation in quality of care can be identified from a combination of holistic scale scores and criterion-based review.
- While there were trends towards hospitals having lower mortality also having higher quality-of-care scores, none of these differences was statistically significant.
- Although there were some correlations between quality-of-care scores and hospital-level outcome data, there was no clear relationship between the quality of care and hospital-level outcomes for the two indicator conditions in this study. This may reflect the complexity of the process–outcome relationship at the patient group level.
- Available hospital-level outcome indicator data are probably insufficiently sensitive to reflect the quality of care recorded in patient case notes. Furthermore, the nuances of patient care may mean that high-quality care may be given even when the patient's outcome appears poor, and vice versa. These findings may be pointing to process measures as being more useful than outcome measures when reviewing the case notes of people who have chronic disease or multiple conditions.

The study explored a complex methodological and clinical question, and the extent to which it

can be investigated is dependent on a number of factors, which, in this study in particular, are related to the available outcome measures and the availability of process of care measures. Choice of outcome measure can be critical in exploring the relationship between the process and outcome of care. For example, the death of a patient (outcome) in the terminal stages of a chronic illness may not be influenced even by the highest quality of care (process).

Previous studies have already indicated that the relationship between process and outcome is difficult to assess, and the systematic review by Pitches *et al.*³⁸ found that only about one-half of the 51 correlations between process measures and outcome indicators in the 36 studies were positive.

Process (or quality-of-care) measures in this study were derived from a multimethod case note review process that was refined during phase one of the study. For each case, reviewers produced a synthesis of their perspective on the quality of care provided, rather than from direct observation. Only the quantitative data from holistic scale scores and criterion-based scores could be used for the analysis, with the holistic textual data being used to validate the reviewer's holistic scale scores being given to overall and phases of care.

Condition-specific mortality and proxy outcomes were derived from a range of hospital-level data. These measures were selected from a much more extensive list through a process of group discussion,

drawing also on the research team's experience in another study that examined the influence of hospital-level process and outcome measures on incident reporting in acute hospitals.⁵³ Resource constraints and the difficulties of obtaining research governance and patient consent meant that, although desirable, it was not possible to obtain patient-level outcome data for the 1565 cases reviewed.

In seeking to explore the relationship between the quality of care for the two tracer conditions in hospitals and overall quality of care and quality markers across hospital institutions, we found only a limited number of positive associations. Correlations with continuous outcome data, where positive (*Table 17*), mainly reflected the selection of mortality to group the hospitals.

In exploring the relationship between quality-of-care scores and hospital mortality group, the findings are consistent for both conditions. Across all phases of care and overall care using holistic review, and total criterion score, there was a trend towards higher mean scores for the lower-mortality hospital groups (*Tables 19 and 20*). However, when account was taken of clustering by reviewer, there were no statistically significant differences in scores between the two groups of hospitals. These results are supported by the finding that there are no statistical differences for the mean holistic scale scores between the higher- and lower-mortality hospitals (*Tables 21–24*).

If the quality of care is good for one condition in a hospital, as measured by case note review, should we expect that it will also be good for another condition? From each hospital there were approximately 40 reviews available for the quality of care in COPD and in heart failure (80 reviews in total). It might be hypothesised that a hospital from the lower-mortality group might be expected to have higher quality-of-care scores and that the levels of quality of care for the two conditions within each hospital bear some relationship. We did not find this – in general, scores were only slightly higher in the lower-mortality group than in the higher-mortality hospital group. Analysis (*Figure 15*) shows that there is little association within hospitals, and we are unable to show that better quality care results in better quality hospital-level outcomes.

We explored the relationships between holistic and criterion-based quality-of-care scores on the one hand and hospital-level process and

outcome indicators on the other. These indicators were drawn from a number of sources, and their relationship with the outcome of quality of care for the two conditions took different forms. Some, such as the HCC indicators, were related to the general external reference measures. Others, particularly from the NHS staff survey, are known to be related to safety culture and incident reporting.⁵³

There are a number of positive correlations in the expected direction for holistic scale score and mean phase score data, although this is mainly for heart failure – there are only limited correlations for COPD holistic data. However, there were also a number of negative correlations for which it is difficult to find an explanation. For example, as COPD mean holistic score increases, the mean score for care of patient/service users being a top priority decreased. It is possible that some of these associations are chance findings because of the number of correlations being undertaken. However, we limited this effect by ensuring that all statistical tests were two-sided, with a significance level of $p \leq 0.05$, and weak and very weak associations were excluded from consideration in the analysis.

Why is the correlation between recorded process and outcome so apparently poor? We have already indicated that the methods may not be sophisticated enough to measure the associations. It may be that the use of hospital-level measures that are not condition specific, and that reflect the organisation as a whole, are too abstract for the purpose of assessing care for people with such conditions such as COPD and heart failure.

However, there may be other important confounding issues that relate to the meaning of quality of care and to the relationship between process and outcome when quality of care is measured.

When Mohammed and colleagues²⁷ explored the care of people with stroke, they found, as we did, that both criterion and holistic methods are valuable in reviewing care, and others have suggested similar approaches. They also found that clinical practice issues, such as the impact of advance directives, made it difficult to assess quality unless detailed information at the patient level was used to understand outcomes. In our study it became clear that reviewers could be very critical of quality of care and were willing to make explicit judgements about clinical practice. Crucially, though, reviewers were able to say that what might be regarded as a poor outcome (e.g. a patient died)

can be accompanied by very high-quality care. For a very ill patient, with little or no chance of surviving, high-quality palliative care may be both appropriate and in the patient's and family's best interests. Alternatively, patients may survive very poor care. Again, choice of outcome measure is critical.

Nevertheless, although individual level data may increase the likelihood of defining process–outcome relationships, this still remains a methodological challenge. When Gibbs and colleagues⁷ used patient-level outcomes as well as hospital-level outcomes to investigate the relationship between process and outcome among surgical patients, they found that people who were more severely ill, or who died or had complications, had higher quality-of-care ratings than those with a lower predicted outcome risk. Furthermore, Pitches *et al.*³⁸ have shown that numbers of studies find uncertain relationships between process and risk-adjusted outcomes of care. The results of this study seem to be following a similar trend.

Limitations

Because of a lack of proven methodologies for chronic disease care, and because of the difficulty of accessing individual outcome data, it was not possible to undertake a full risk-adjusted analysis. Whereas Daley *et al.*³⁷ were able to access a number of possible predictors for surgical care, these are mainly unavailable for chronic disease management. In this study, the availability of the outcome measures was limited to hospital-level data and it has not been possible to capture more individual-level data. This is clearly a potential limitation. However, even though risk adjustment measures have been available for the assessment of the effect of process of care on surgical outcomes, a number of studies have found it difficult to

show positive associations between risk-adjusted mortality and good quality of care.^{7,37,54,55}

Reviewers are not perfect. For instance, some may have a more positive view on cases than they should have, or be too inexperienced to identify flaws in care. Nevertheless, we have no evidence that they glossed over errors, as they did identify about 20% of cases that fell into the unsatisfactory range and about 4% that were identified as adverse incidents or near misses. The ability to judge the quality of care provided is also dependent on the quality of recording in the case notes, and there may not be enough data in the records to be able to make an adequate judgement on quality of care – although reviewers indicated that about 85% of the case notes were in a satisfactory or better condition.

Although the two groups of hospitals in this study had considerable differences in mortality rates between the lower-mortality hospitals and the higher-mortality hospitals, a whole range of factors might account for those differences, such as case mix and age. Indeed, in this study, there is some indication that quality-of-care scores went down as age went up in the group of heart failure cases, but this may be because of higher levels of mortality risk before admission. We do not know enough about these cases to model this risk, although we did randomly select the 10 hospitals in each group.

We did not account for any measurement error in the process variables and the number of measurements per hospital. As the correlation of true scores is equal to the correlation of observed scores divided by the square root of the product of their reliabilities, it is possible that measurement error may dilute the magnitude of the correlation coefficients. This means that the correlations observed between the process and outcome variables may be lower than the true population correlation.

Chapter 4

Overall conclusions of the research

Implications for reviewing quality of care

To return to our practical research question, what do our results tell us about which method of review would be best used for which purposes and by which professional groups? We have found that all three professional groups perform well when using criterion-based review. If this type of review is chosen, perhaps for large-scale clinical audit to inform service development, the decision on which professional groups should undertake reviews using this method might depend mainly on cost and availability of staff. The data on resource use show that the doctors are considerably more expensive at cost per review, because they have the highest salary levels, although their review times are similar to the nursing/clinical group. However, medically trained reviewers may have a place when using structured review methods to identify variations in care, such as adverse events, as clinical training might be of advantage by enhancing watchfulness. Furthermore, review of small numbers of cases has relatively little cost impact, and for criterion-based review of limited numbers of cases, for example in an investigation of quality of care, staff with nursing or medical training may add value to an evaluation.

The decision on who should undertake structured holistic review is more complex. It is a method that might best be used when more is required than just the sum of the results of collecting a set of review criteria. While all groups can use the method of holistic-scale scoring, the overall results conceal wide ranges of agreement, sometimes close to random for phase-of-care results (*Table 7*). Particularly for the more technical phases of care, such as investigations, admissions and initial management, these results suggest that the three groups of staff are interpreting the recorded care differently when they each review the same record. This probably reflects their background knowledge of the clinical situation and of how the care is delivered. Even when considerable training in the review method has been provided, it is unrealistic to expect the non-clinical audit staff to fully appreciate the details of the medical care, let alone when that care has or has not deviated from best practice. Results suggest that nursing-trained

reviewers sometimes identify different problems from those found by physician reviewers. It is possible that extended training and selection of staff might reduce this difference, for instance by selection of specialist consultant nurses or of very experienced doctors. Selection of person skills according to task might provide the best outcome for the more difficult reviews.

Moreover, reviews of cases of serious unintended incidents or of poor-outcome cases might benefit from structured reviews by pairs of reviewers – one with a nursing background and one with a medical background. Our results have shown that these two groups of reviewers offer different types of results, with nurses tending towards care process issues and doctors offering judgements on more technical interventions. If reviews were supported by effective training, including the enabling of staff to make explicit judgements on care, joint mixed professional reviewing, perhaps using more senior doctors and consultant-level specialist nurses, might offer a wider range of insights than if case records were reviewed by two professionals from the same background. Whereas Weingart *et al.*⁶ wondered whether the differences in holistic review results from physicians and nurses reviewers could be problematic, the differences we found in our study could be put to a positive advantage.

Textual data provides much finer-grained information than do scores or criterion-based review, even when it is provided in short phrases and sentences. Full analysis of the textual data in a clinical setting, rather than in a research project, is likely to be costly and difficult to do when undertaking large-scale audits or quality and safety reviews. However, the increasing practice of undertaking smaller scale reviews, for example where there are a small series of cases with poor outcomes that require detailed review, is identifying a need for structured reviews that would benefit from a combination of data, such as is provided by criterion-based scores, holistic scoring and structured textual commentary.

Overall, the results of this study suggest that there may be significant gains to be made in clinical audit and evaluation through better understanding

of the products of the different methods of review and of the value in training and selection of reviewers.

Reviewing using a mixed method approach – how can it be used in future?

Because of the research aims of the first part of the study, we separated out the review methods into two distinct sections – holistic and criterion based. In the second part of the study, those methods were combined to provide holistic scale scores, criterion-based scores and textual data about the quality of care of each of the phases and of overall care. Here, the textual data was used only to validate the holistic scale score data. Results suggest that when reviewing or auditing small groups of cases, for example when there are concerns about the outcome of interventions, mixed holistic and criterion-based review, which also captures textual data, may prove a powerful model. In using these mixed methods, careful attention will be required to aspects such as the selection and training of reviewers, including recognition of the problems associated with inter-rater reliability and bias.

Relationships between quality and outcome of care

No strong relationships were found between quality of care, as measured by case note review, and outcome of care at the hospital level. Other authors have found a similar lack of direct association. The finding that reviewers considered quality of care as a broad concept, where patients who fared poorly overall because of their underlying condition might nevertheless have high-quality care, reflects a similar finding by Gibbs *et al.*,⁷ when examining

patient-level outcomes; in their study, patients with poor outcomes or with a higher predicted risk of mortality or morbidity had higher quality-of-care ratings than those with a low predicted risk of adverse outcome. Mohammed and colleagues²⁷ found that a combination of holistic data and criterion-based data was required to understand the influence of care on outcomes, and Pitches *et al.*³⁸ concluded from their systematic review of the literature that ‘the general notion that hospitals with higher risk-adjusted mortality have poorer quality of care is neither consistent nor reliable’.

Lilford and colleagues⁵⁶ have recently provided a useful critique of risk-adjusted outcomes in the assessment of health-care quality in which they question the value of using outcome measures to evaluate the quality of care. We are unable to confirm their proposition to use process measures rather than outcome measures with our results, but would suggest that for the purposes of reviewing quality of care, process measurement that allows for an integration of both criterion-based and holistic review can provide a sound basis for decision-making.

Meanwhile, the real challenge of outcomes is to define what is measurable – and there are actually few measures that are validated.⁵⁷ Where attempts have been made to develop outcome measures, there has been some tendency for health professionals to dismiss outcomes as too difficult to use. But patients are concerned with the ultimate outcome of their therapy and thus we cannot ignore the need for appropriate measures, and risk adjustment methods, for chronic disease management. In the meantime, however, given the lack of agreement on specific outcome measures, we suggest that process measures are a reasonable proxy.

Chapter 5

Future research agenda

Senior clinical staff are increasingly called upon to assess quality of care from case notes under conditions in which there is cause for concern. This research has identified three aspects of case note review that could be used to support quality review and that could benefit from further research.

1. Research to assess the inter-rater reliability among experienced physician reviewers, including the effect of selection and training. This research should take account of other work from North America that has shown the potential for major discrepancies in inter-rater reliability.
2. Reviewer recruitment appears to be difficult in the UK, perhaps because this form of clinical review has been treated in an ad hoc fashion to date. A qualitative study could be undertaken to understand the possible barriers and factors that might enhance recruitment and training of clinician reviewers.
3. Doctors and nurses may view quality differently, but there has been no UK research (and little international research) to explore whether experienced physicians and experienced specialist nurses review in a similar or complimentary manner (possibly enhancing the overall scope and quality of reviews).
4. Research on the reliability of structured holistic assessment of the quality and safety of care using scales remains unusual in the literature. Evaluation of the sensitivity and internal consistency of these methods would be of value.
5. The extent to which review criteria are reproducible remains a research question.

There is some evidence from US and UK studies using the RAND appropriateness method that panels constructing review criteria from the same clinical guideline have only moderate levels of agreement. The extensive need for review criteria generated by national clinical guidelines suggests that further research could be useful before these criteria become a significant part of the new quality-improvement programmes.

There is an important research agenda relating to linkage between process and outcome data for chronic disease care, and in relation to case note review. Although the potential for data retrieval from electronic records is considerable (information might be gathered from data mining and natural language programming), this agenda relates to paper-based records, which will remain the main data source in hospital care over the next 5 years.

6. There is a need for further research to explore risk-adjusted outcome measure methodologies for chronic diseases. Methods of risk prediction should be developed in a fashion that enables the production of (at least some) measures across the spectrum of chronic disease, allowing better methods of outcome comparison.
7. There is a continued need for validated condition-specific outcome measures for a range of chronic diseases, of a type that can be used in health services evaluation – that is, comprising a minimum data set of items that might, in future evaluations, be collectable through electronic records systems.



Acknowledgements

This study was undertaken by two partner organisations: the University of Sheffield – the School of Health and Related Research (SchARR) and the Department of Information Studies – and the Royal College of Physicians Clinical Evaluation and Effectiveness Unit (CEEU). An important feature of the study was the contact made with clinical teams in over 30 hospitals in England, many of whom went on to contribute to the study. CEEU staff took a major part in enabling the Sheffield team to contact the hospitals and their specialist teams, as well as providing expertise in case note review and quality-of-care methodology.

Karen Beck provided administrative support at the University of Sheffield throughout the project and her contribution was exceptionally helpful. In particular her substantial contribution to the development of data collection software was invaluable.

Jon Nicholl provided valuable methodological advice in the early part of the study.

We especially wish to acknowledge the enthusiasm and assistance of reviewers and all of their colleagues in the study hospitals, without whom this study could not have taken place.

We thank the external referees for their contributions to shaping the future research agenda.

Contribution of authors

Allen Hutchinson was Principal Investigator, principal author of the application, responsible for

the overall management of the project and main author of the report.

Joanne E Coster (née Dean) was project manager, contributed to the application and made a major contribution to the fieldwork and analysis.

Katy L Cooper contributed to the project management and made a major contribution to the fieldwork and analysis.

Aileen McIntosh was lead qualitative researcher and made a major contribution to the qualitative analysis and a significant contribution to the project development.

Stephen J Walters was lead statistical adviser and senior statistical analyst.

Peter A Bath made senior contributions to project development and analysis.

Michael Pearson was lead clinical adviser and contributed expertise in quality assessment methods and analysis.

Tracey A Young was a statistical analyst and took the lead in developing and analysing the outcomes component of the study.

Khadija Rantell was a statistical analyst and contributed to the management of the data sets in the first stage of the study.

Michael J Campbell provided senior and specialist statistical advice to the project design and analysis.

Julie Ratcliffe provided health economics advice and analysis.



References

1. Codman EA. A Study in Hospital Efficiency. *As demonstrated by the case report of the first five years of a private hospital*. Boston, MA Thomas Todd; 1918–1920.
2. Rubenstein LV, Kahn KL, Reimisch EJ, Sherwood MJ, Rogers WH, Kambers C. Changes in quality of care for five diseases measured by implicit review, 1981–1986. *JAMA* 1990;**264**:1974–79.
3. Leape LL, Brennan TA, Laird N, Cawthors AG, Localio AR, Barnes BA, *et al.* The nature of adverse events in hospitalised patients: results of Harvard Medical Practice Study II. *NEJM* 1991;**324**:377–84.
4. Runciman WB, Webb RK, Helps SC, Thomas EJ, Sexton EJ, Studdert DM, *et al.* A comparison of iatrogenic injury studies in Australia and the USA II reviewer behaviour and quality of care. *Int J Qual Health Care* 2000;**12**:379–88.
5. Thomas EJ, Studdert DM, Burstein HR, Orav JE, Zeena TBS, Williams EJ, *et al.* Incidences and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;**38**:261–71.
6. Weingart SN, Davis RB, Palmer RH, Cahalane M, Hamel MB, Mukamal K, *et al.* Discrepancies between explicit and implicit review: physician and nurse assessments of complication and quality. *Health Serv Res* 2002;**32**:483–98.
7. Gibbs J, Clark K, Khuri S, Henderson W, Hur K, Daley J. Validating risk adjusted surgical outcomes: chart review of process of care. *Int J Qual Health Care* 2001;**13**:187–96.
8. Ashton C, Kuykendall D, Johnson ML, Wray N. An empirical assessment of the validity of explicit and implicit process of care criteria for quality assessment. *Med Care* 1999;**37**:798–808.
9. Thomas EJ, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. *Ann Int Med* 2002;**136**:812–16.
10. Hofer TP, Asch SM, Hayward RA, Rubenstein LV, Hogan MM, Adams J, *et al.* Profiling quality of care: is there a role for peer review? *BMC Health Serv Res* 2004;**4**:9. URL: www.biomedcentral.com/1472-6963/4/9.
11. Hayward RA, Hofer TPE. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA* 2001;**286**:415–20.
12. Lilford R, Edwards A, Girling A, Hofer T, Di Tanna GL, Petty J, *et al.* Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Policy*, 2007;**12**:173–80.
13. Hulka BS, Romm FJ, Parkerson GR, Russell IT, Clapp NE, Johnson FS. Peer review in ambulatory care: use of explicit criteria and implicit judgements. *Med Care* 1979;**17**(Suppl.):1–73.
14. Hayward RA, McMahon LF, Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? *Ann Intern Med* 1993;**118**:550–6.
15. Fischhoff B. Hindsight ≠ foresight: The effect of outcome knowledge on judgement under uncertainty. *J Exp Psychol* 1975;**1**:288–99.
16. Lilford RJ, Mohammed MA, Brauholtz D, Hofer TP. The measurement of active errors: methodology issues. *Qual Saf Health Care* 2003;**12**(Suppl. 1):8–12.
17. Localio RA, Weaver SL, Landis R, Lawthers AG, Brennan TA, Hebert L, *et al.* Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. *Ann Intern Med* 1996;**125**:457–64.
18. Agency for Health Care Policy and Research. *Using clinical practice guidelines to evaluate quality of care*. Vol. 2. Rockville, MD: US Department of Health and Human Services, Public Health Service; 1995.
19. Hadorn DC, Baker DW, Kamberg CJ, Brook RH. Practice guidelines. Phase II of the AHCPR – sponsored heart failure guideline: translating practice recommendations into review criteria. *J Qual Improvement* 1996;**22**:265–76.
20. The North of England Study of Standards and Performance in General Practice. Medical audit in general practice. I: Effects on doctors' clinical behaviour for common childhood conditions. *BMJ* 1992;**304**:1480–4.
21. Hutchinson A, McIntosh A, Anderson J, Gilbert C, Field R. Developing primary care review criteria from evidenced-based guidelines: coronary heart disease as a model. *BJGP* 2003;**53**:691–6.

22. National Institute for Clinical Excellence. *Chronic heart failure: management of chronic heart failure in adults in primary and secondary care. Clinical Guideline 5*. London: National Institute for Clinical Excellence; 2003.
23. Rudd AG, Lowe D, Irwin P, Rutledge Z, Pearson M. Intercollegiate Stroke Working Party. National stroke audit: a tool for change? *Qual Health Care* 2001;**10**:141–51.
24. Gompertz P, Dennis M, Hopkins A, Ebrahim S. Development and reliability of the stroke audit form. UK Stroke Audit Group. *Age aging*. 1994;**23**:378–83.
25. Gompertz PH, Irwin P, Morris R, Lowe D, Rutledge Z, Rudd AG, *et al*. Reliability and validity of the Intercollegiate Stroke Audit Package. *J Eval Clin Prac* 2001;**7**:1–11.
26. Camacho LA, Rubin HR. Assessment of the validity and reliability of three systems of medical record screening for quality of care assessment. *Med Care* 1998;**36**:748–51.
27. Mohammed MA, Mant J, Bentham L, Stevens A, Hussain S. Process and mortality of stroke patients with and without do not resuscitate order in the West Midlands, UK. *Int J Qual Health Care*, 2006;**18**:102–6.
28. Rubenstein LR, Kahn KL, Harris ER, Sherwood MJ, Rodgers WH, Brook RH. *Structured implicit review of the medical record: a method for measuring the quality of in-hospital medical care and a summary of quality changes following implementation of the Medicare prospective payments system*. Santa Monica, CA: RAND; 1991.
29. Pearson M, Lee JL, Chang BL, Elliott M, Kahn KL, Rubenstein LV. Structured implicit review: a new method for monitoring nursing care quality. *Med Care* 2000;**38**:1074–91.
30. Keeler EB, Rubenstein KLK, Draper D, Harrison ER, McGinty MJ, Rogers WH, *et al*. Health Programme of RAND. *JAMA* 1992;**268**:1702–8.
31. National Institute for Clinical Excellence. *Chronic obstructive pulmonary disease. Management of chronic obstructive pulmonary disease in adults in primary and secondary care*. London: National Institute for Clinical Excellence; 2004.
32. van Belle, Gerald. *Statistical rules of thumb*. 1st edn. Wiley Series in Probability and Statistics. London: Wiley InterScience; 2002.
33. Hospital Episode Statistics Online: Data on hospital providers. URL: www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=212 (accessed 20 November 2006).
34. Royal College of Physicians and British Thoracic Society. *Report of the 2003 National COPD Audit*. London: Royal College of Physicians, 2004.
35. Fleiss JL. *Statistical methods for rates and proportions*. 2nd edn. New York, NY: Wiley; 1981.
36. Dale, JR. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* 1986;**42**:909–17.
37. Daley J, Khuri SF, Henderson W, Hur K, Gibbs J, Barbour G, *et al*. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs surgical risk study. *J Am Coll Surg* 1997;**185**:315–27.
38. Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Services Res* 2007;**7**:91. URL: www.biomedcentral.com/1472-6963/7/91.
39. Glaser B, Strauss A. *Discovery of grounded theory: strategies for qualitative research*. London: Weidenfeld & Nicolson; 1967.
40. NHS Employers. URL: www.nhsemployers.org/restricted/downloads/download.asp?ref=363&hash=0bb7dc2313394a93337d3adf51cf6c3f&itemplate=e_aboutus_3col_aboutus-2028.
41. Wardle TD, Burnham R, Greig E, Preston S, Harris RA, Borrill Z, *et al*. A confidential study of deaths after emergency medical admission: issues relating to quality of care. *Clin Med* 2003;**3**:425–34.
42. Potter J, Peel P, Mian S, Lowe D, Irwin P, Pearson M, *et al*. National audit of continence care for older people: management of faecal incontinence. *Age Ageing* 2007;**36**:268–73.
43. Hutchinson A, McIntosh A, Coster JE, Cooper KL, Bath PA, Walters SJ, *et al*. When is an event an event? Data from the quality and safety continuum. In Hignett S, Norris B, Catchpole K, Hutchinson A, Tapley S, editors. *From safe design to safe practice*. Cambridge: The Ergonomics Society; 2008.
44. Armitage P, Berry G, Matthews, JNS. *Statistical methods in medical research*. 4th edn. Oxford: Blackwell Science; 2002.
45. Luck J, Peabody JW, Dressellhaus TR, Lee M, Glassman P. How well does chart abstraction measure quality? A prospective comparison of standardised patients with the medical record. *Am J Med* 2000;**108**:642–9.

46. Healthcare Commission. URL: www.healthcarecommission.org.uk (accessed 30 November 2008).
47. Intercollegiate Stroke Audit Working Party. *National sentinel stroke audit*. London: Royal College of Physicians of London; 2007.
48. McNaughton H, McPherson K, Taylor W, Weatherall M. Relationship between process and outcome in stroke care. *Stroke* 2003;**34**:713–17.
49. Wilson B, Thornton JG, Hewison J, Lilford RJ, Watt I, Brauholtz D, *et al.* The Leeds University maternity audit project. *Int J Qual Health Care* 2002;**14**:175–81.
50. Dr Foster. *The hospital guide: how good is my hospital, 2005*. URL: www.drfooster.co.uk/hospitalreport/pdfs/howGood.pdf (accessed 20 November 2006).
51. Healthcare Commission NHS Staff Survey. www.healthcarecommission.org.uk/nationalfindings/surveys.cfm (accessed 20 November 2006).
52. National Patient Safety Agency. *Quarterly national reporting and learning system data summary, autumn 2006*. URL: www.npsa.nhs.uk/site/media/documents/1953_NRLS_Data.pdf (accessed 20 November 2006).
53. Hutchinson A, Young TA, Cooper KL, McIntosh A, Karnon JD, Scobie S, *et al.* Trends in healthcare incident reporting and relationship to safety and quality data in acute hospitals: results from the National Reporting and Learning System. *Qual Saf Health Care* 2009;**18**:5–10.
54. Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;**335**:648–50.
55. Dubois RW, Rogers WH, Moxley JH, *et al.* Hospital inpatient mortality: is it a predictor of quality? *N Engl J Med* 1987;**317**:1674–80.
56. Thomas JW, Holloway JJ, Guire KE. Validating risk-adjusted mortality as an indicator of quality of care. *Inquiry* 1993;**30**:6–22.
57. Pearson M, Goldacre M, Coles J, Amess M, Cleary R, Fletcher J, *et al.* *Health outcome indicators: asthma report of working group to the Department of Health*. Oxford: National Centre for Health Outcomes Development; 1999.

Appendix I

COPD review criteria

NICE definition of an exacerbation of COPD

ICD-10: J42, J43, J44. 'An exacerbation is a sustained worsening of the patient's symptoms from their usual stable state, which is beyond normal day-to-day variations, and is acute in onset. Commonly reported symptoms are worsening breathlessness, cough, increased sputum production and change in sputum colour.'³²

Key data

1. Audit record number	<input style="width: 200px; height: 20px;" type="text"/>							
2. Hospital number	<input style="width: 200px; height: 20px;" type="text"/>							
3. Date of birth	<input style="width: 200px; height: 20px;" type="text"/>							
4. Gender	male <input type="checkbox"/> female <input type="checkbox"/>							
5. First part of patient's postcode	<input style="width: 200px; height: 20px;" type="text"/>							
6. Date of this admission to hospital	<input style="width: 200px; height: 20px;" type="text"/>							
7. Did the patient die during this admission	yes <input type="checkbox"/> no <input type="checkbox"/>							
If yes, was the recorded cause of death <table border="0" style="margin-left: 100px;"> <tr> <td>COPD or complications of COPD</td> <td><input type="checkbox"/></td> </tr> <tr> <td>other cause(s)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>not recorded</td> <td><input type="checkbox"/></td> </tr> </table>	COPD or complications of COPD	<input type="checkbox"/>	other cause(s)	<input type="checkbox"/>	not recorded	<input type="checkbox"/>		
COPD or complications of COPD	<input type="checkbox"/>							
other cause(s)	<input type="checkbox"/>							
not recorded	<input type="checkbox"/>							
8. Date of discharge from hospital (or death if applicable)	<input style="width: 200px; height: 20px;" type="text"/>							
9. Was the patient accepted by an early discharge (or hospital at home) scheme?	yes <input type="checkbox"/> no <input type="checkbox"/> not applicable <input type="checkbox"/>							
10. Prior to this admission, has the patient previously been admitted to hospital for COPD or accepted onto an early discharge scheme?	yes <input type="checkbox"/> no <input type="checkbox"/>							

History/patient characteristics

11. What is the patient's smoking status?

- current smoker
- ex smoker (stopped more than 3 months)
- life long non-smoker
- not recorded

If current or ex-smoker,

how many cigarettes smoked per day?

or

pack years?

don't know

12. Does the patient have any significant co-morbidities?

Please tick all that apply

- none
- heart disease
- hypertension
- stroke
- locomotor problems
- neurological problems
- diabetes
- visual impairment
- depression/anxiety
- other

13. What are the patient's social circumstances?

- lives alone, no support
- lives alone with social service support
- lives with spouse or close relative
- lives in nursing/residential home
- lives in warden controlled (sheltered) housing

Admission

14. At admission, was there a record of

level of breathlessness	increased	<input type="checkbox"/>
	not increased	<input type="checkbox"/>
	not recorded	<input type="checkbox"/>
level of sputum	increased	<input type="checkbox"/>
	not increased	<input type="checkbox"/>
	not recorded	<input type="checkbox"/>
changes in colour of sputum	changed	<input type="checkbox"/>
	not changed	<input type="checkbox"/>
	not recorded	<input type="checkbox"/>
sputum colour	white or grey	<input type="checkbox"/>
	yellow or green	<input type="checkbox"/>
	no sputum	<input type="checkbox"/>
	not recorded	<input type="checkbox"/>

15. Was the patient's dyspnoea rating (e.g. on the MRC dyspnoea scale) recorded

yes	<input type="checkbox"/>
no	<input type="checkbox"/>

16. What is the patient's performance status?

normal activity	<input type="checkbox"/>
strenuous activity limited	<input type="checkbox"/>
limited activity but self care <input type="checkbox"/>	
limited self care	<input type="checkbox"/>
bed/chair bound, no self care <input type="checkbox"/>	
not known	<input type="checkbox"/>

17. Was a chest X-ray taken?

yes	<input type="checkbox"/>
no	<input type="checkbox"/>

If yes, is the X-ray report in the notes?

yes	<input type="checkbox"/>
no	<input type="checkbox"/>

18. Was respiratory rate measured

yes
no

If yes, what was the first reading
after admission

per minute

19. Were blood gases taken?

yes
no

If yes, what was the first recorded (after admission) value for?

pH or: H⁺ (mmol/l) not recorded

PCO₂ (kPa) or: mmHg not recorded

PO₂ (kPa) or: mmHg not recorded

Was level of O₂ to be given stipulated in notes/on chart?

yes
no

20. Was an ECG performed?

yes
no

21. Was urea recorded?

yes
no

If yes, what was the first recorded (after admission) value?

mmol/l

not recorded

22. Was serum albumin recorded?

yes
no

If yes, what was the first recorded (after admission) value?

mmol/l

not recorded

23. Was there a record of medications being taken at time of admission?

yes

no

If yes, were there 5 or more medications recorded?

yes

no

24. What was the patient's temperature at admission?

not recorded

25. Is there a spirometry reading in the notes for this admission?

Yes

No

If yes what is the FEV₁ level
(if more than one, give most recent)

not recorded

26. Is there a record of peripheral oedema?

yes - present

yes – not present

not recorded

If present, was it?

leg/ ankles

sacral

not recorded

Initial management

27. Was a course of antibiotics prescribed?

yes
no

28. Were nebulised bronchodilators prescribed?

yes
no

29. Did the patient receive systemic corticosteroids?

yes
no

30. How many sets of arterial blood gases results are in the records for this stay?

31. Did the patient have a pH less than 7.35 at any time during this stay?

yes
no

If yes, did they receive ventilatory support?

respiratory stimulant (e.g. doxapram)
non-invasive
invasive
none

If the patient had a pH of less than 7.35 and did not receive ventilatory support, is it noted why not?

patient refused
no facilities
not appropriate
failed
other (please state)

not recorded

Pre-discharge phase

-
32. Was oximetry (O₂ saturation levels) undertaken, after acute phase but prior to discharge?
- | | | |
|--|--------------|--------------------------|
| | yes | <input type="checkbox"/> |
| | no | <input type="checkbox"/> |
| | not recorded | <input type="checkbox"/> |
- If yes, what were the results? %
- | | | |
|--|--------------|--------------------------|
| | not recorded | <input type="checkbox"/> |
|--|--------------|--------------------------|
-
33. If a current smoker, was help toward smoking cessation given
- | | | |
|--|---|--------------------------|
| | referred to smoking cessation programme | <input type="checkbox"/> |
| | advice given and recorded | <input type="checkbox"/> |
| | nothing recorded | <input type="checkbox"/> |
| | not applicable (because non-smoker) | <input type="checkbox"/> |
-
34. Was there an assessment of the patient's home circumstances and their ability to cope?
- | | | |
|--|--------------|--------------------------|
| | yes | <input type="checkbox"/> |
| | no | <input type="checkbox"/> |
| | not recorded | <input type="checkbox"/> |
-
35. Where was the patient discharged to?
- | | | |
|--|---|--------------------------|
| | own home – independent of help | <input type="checkbox"/> |
| | own home – with additional social support | <input type="checkbox"/> |
| | sheltered housing or living with relative | <input type="checkbox"/> |
| | nursing or residential care | <input type="checkbox"/> |
| | not applicable – died in hospital | <input type="checkbox"/> |
-
36. Is there a letter to the patient's primary care team?
- | | | |
|--|--------------|--------------------------|
| | yes | <input type="checkbox"/> |
| | no | <input type="checkbox"/> |
| | not recorded | <input type="checkbox"/> |
- If yes, did the letter include a clear list of the patient's medication?
- | | | |
|--|--------------|--------------------------|
| | yes | <input type="checkbox"/> |
| | no | <input type="checkbox"/> |
| | not recorded | <input type="checkbox"/> |
-
37. Which type of consultant was the patient under at time of discharge?
- | | | |
|--|---------------------------|--------------------------|
| | respiratory physician | <input type="checkbox"/> |
| | care of elderly physician | <input type="checkbox"/> |
| | general physician | <input type="checkbox"/> |
| | other | <input type="checkbox"/> |
| | not recorded | <input type="checkbox"/> |

Appendix 2

Heart failure review criteria

NICE definition of an exacerbation of heart failure (heart failure due to left ventricular systolic dysfunction)

ICD-10: I50.0, I50.1, I50.9, I11.0. 'An exacerbation of heart failure is a sustained worsening of the patient's symptoms from their usual stable state, which is beyond normal day-to-day variations, and is acute in onset. Commonly reported symptoms are worsening breathlessness, tiredness and swelling of the feet and/or ankles.'²³

Key data

1. Audit record number

2. Hospital number

3. Date of birth

4. Gender

male
female

5. First part of patient's postcode

6. Date of admission to hospital

7. Did the patient die during this admission

yes
no

If yes, was the recorded cause of death

heart failure or complications of heart failure
other cause(s)
not recorded

8. Date of discharge from hospital (or death if applicable)

Patient history

9. What is the patient's smoking status
- | | | |
|--|--|--------------------------|
| | current smoker | <input type="checkbox"/> |
| | ex smoker (stopped more than 3 months) | <input type="checkbox"/> |
| | life long non-smoker | <input type="checkbox"/> |
| | not recorded | <input type="checkbox"/> |

If current or ex-smoker,
how many cigarettes smoked per day?
or
pack years?

don't know

10. Is there a record of the patient's weekly alcohol intake?

units per week

not recorded

11. Does the patient have any significant co-morbidities?

Please tick all that apply

- | | |
|---------------------|--------------------------|
| none | <input type="checkbox"/> |
| hypertension | <input type="checkbox"/> |
| stroke | <input type="checkbox"/> |
| respiratory disease | <input type="checkbox"/> |
| locomotor problems | <input type="checkbox"/> |
| neurological | <input type="checkbox"/> |
| problems | <input type="checkbox"/> |
| visual impairment | <input type="checkbox"/> |
| diabetes | <input type="checkbox"/> |
| depression/anxiety | <input type="checkbox"/> |
| other | <input type="checkbox"/> |

18. Is there a record for any of the following used to assess functional capacity?

Tick all that apply:

- | | |
|---|--------------------------|
| breathlessness scale (eg NYHA scale) | <input type="checkbox"/> |
| objective measure of O ₂ capacity | |
| treadmill test without O ₂ consumption test | <input type="checkbox"/> |
| cardiopulmonary exercise test with O ₂ consumption | <input type="checkbox"/> |
| none recorded | <input type="checkbox"/> |

19. Is there a record of raised jugular venous pressure
(also called jugular venous distension)

- | | |
|--------------------|--------------------------|
| yes – elevated | <input type="checkbox"/> |
| yes – not elevated | <input type="checkbox"/> |
| not recorded | <input type="checkbox"/> |

20. Is there a record of peripheral oedema?

- | | |
|-------------------|--------------------------|
| yes – present | <input type="checkbox"/> |
| yes – not present | <input type="checkbox"/> |
| not recorded | <input type="checkbox"/> |

If present, was it?

- | | |
|--------------------|--------------------------|
| leg/swollen ankles | <input type="checkbox"/> |
| sacral | <input type="checkbox"/> |
| not recorded | <input type="checkbox"/> |

21. What was first blood pressure (on arrival)?

mmHg

- | | |
|--------------|--------------------------|
| not recorded | <input type="checkbox"/> |
|--------------|--------------------------|

Initial management

22. For each of the following medications, is it recorded if the patient was prescribed

diuretic	prior to admission	<input type="checkbox"/>
	prescribed during stay	<input type="checkbox"/>
	on discharge	<input type="checkbox"/>
	not prescribed	<input type="checkbox"/>
	not recorded	<input type="checkbox"/>
ACE inhibitor	prior to admission	<input type="checkbox"/>
	prescribed during stay	<input type="checkbox"/>
	on discharge	<input type="checkbox"/>
	not prescribed	<input type="checkbox"/>
	not recorded	<input type="checkbox"/>
beta blocker	prior to admission	<input type="checkbox"/>
	prescribed during stay	<input type="checkbox"/>
	on discharge	<input type="checkbox"/>
	not prescribed	<input type="checkbox"/>
	not recorded	<input type="checkbox"/>

23. Is it recorded that the patient received oxygen during this stay?

yes
no

24. Is there a record of the patient's weight in the notes?

yes
no

25. Is there a record of the patient's blood pressure (daily) in the notes?

yes
no

26. Is there a record of the patient's electrolytes in the notes?

yes
no

Pre-discharge phase

27. Pre-discharge assessments

Was blood pressure taken prior to discharge (within 48 hours of discharge)?

- yes
- no
- not recorded

Was the level of breathlessness prior to discharge (within 48 hours of discharge) documented?

- breathless at rest
- breathless on minor exertion
- breathless on strenuous exercise
- not recorded

Was patient's weight prior to discharge (within 48 hours of discharge) recorded?

- yes
- no

Was serum creatinine recorded within 48 hours of discharge

- yes
- no

28. Is there a past record of/ or plan to refer for an exercise/rehabilitation programme

- yes
- no

29. If a current smoker, was help toward smoking cessation given

- referred to smoking cessation programme
- advice given and recorded
- nothing recorded
- not applicable (because non-smoker)

30. Was there an assessment of the patient's home circumstances and their ability to cope?

- yes
- no

31. Where was the patient discharged to?

- own home – independent of help
- own home – with additional social support
- sheltered housing or living with relative
- nursing or residential care
- not applicable – died in hospital

32. Is there a letter to the patient's primary care team?

- yes
- no

If yes, did the letter include a clear list of the patient's medication?

- yes
- no
- not recorded

33. Which type of consultant was the patient under at time of discharge?

- cardiologist
- care of elderly
- physician
- general physician
- other
- not recorded

Appendix 3

Validity of review criterion questionnaire (COPD)

Key data

No.	Criterion	Essential	Desirable	Non-essential	Comments
1	Audit record number (RReSQ Study Reference Number)				
2	Hospital number				
3	Date of birth				
4	Gender				
5	First part of patient's postcode				
6	Date of this admission to hospital (dd/mm/yyyy)				
7	<i>Did the patient die during this admission?</i> Yes No				
	<i>If yes, was the recorded cause of death:</i> COPD or complications of COPD Other cause(s) Not recorded				
8	<i>Date of discharge from hospital (or death if applicable):</i> Discharge Death				
9	<i>Was the patient accepted by an early discharge (or hospital at home) scheme?</i> Yes No Not applicable				
10	<i>Prior to this admission, has the patient previously been admitted to hospital for COPD or accepted on to an early discharge scheme?</i> Yes No				

History/patient characteristics

No.	Criterion	Essential	Desirable	Non-essential	Comments
11	<p><i>What is the patient's smoking status?</i></p> <p>Current smoker Ex-smoker (stopped more than 3 months) Lifelong non-smoker Not recorded</p> <p><i>If current or ex-cigarette smoker:</i></p> <p><i>How many cigarettes smoked per day?</i> <i>Or pack-years?</i> Don't know</p>				
12	<p><i>Does the patient have any comorbidities?</i></p> <p>Please tick all that apply:</p> <ul style="list-style-type: none"> – None – Heart disease – Hypertension – Stroke – Locomotor problems – Neurological problems – Diabetes – Visual impairment – Depression/anxiety – Other 				
13	<p><i>What are the patient's social circumstances?</i></p> <p>Lives alone, no support Lives alone with social service support Lives with spouse, close relative or carer Lives in nursing/residential home Lives in warden controlled (sheltered) housing Not known</p>				

Admission

No.	Criterion	Essential	Desirable	Non-essential	Comments
14	<p>At admission, was there a record of:</p> <p>Level of breathlessness:</p> <ul style="list-style-type: none"> – Increased – Not increased – Not recorded <p>Level of sputum:</p> <ul style="list-style-type: none"> – Increased – Not increased – Not recorded <p>Changes in colour of sputum:</p> <ul style="list-style-type: none"> – Changed – Not changed – Not recorded <p>Sputum colour:</p> <ul style="list-style-type: none"> – White or grey – Yellow or green – No sputum – Not recorded 				
15	<p>Was the patient's dyspnoea rating (e.g. on the MRC dyspnoea scale) recorded?</p> <p>Yes</p> <p>No</p>				
16	<p>What is the patient's performance status, prior to admission?</p> <p>Normal activity</p> <p>Strenuous activity limited</p> <p>Limited activity but self care</p> <p>Limited self care</p> <p>Bed/chair bound, no self care</p> <p>Not known</p>				
17	<p>Was a chest X-ray taken within 24 hours?</p> <p>Yes</p> <p>No</p> <p>If yes, is the X-ray report in the notes?</p> <p>Yes</p> <p>No</p>				
18	<p>Was respiratory rate measured within 24 hours?</p> <p>Yes</p> <p>No</p> <p>If yes, what was the first reading after admission (per minute)?</p>				

No.	Criterion	Essential	Desirable	Non-essential	Comments
19	<p>Were blood gases taken within 24 hours?</p> <p>Yes No</p> <p>If yes, what was the first recorded (after admission) value for?</p> <p>pH or H⁺ (mmol/l) Not recorded</p> <p>Pco₂ (kPa or mmHg) Not recorded</p> <p>Po₂ (kPa) or mmHg Not recorded</p> <p>Was level of O₂ to be given stipulated in notes/on chart?</p> <p>Yes No</p>				
20	<p>Was an ECG performed?</p> <p>Yes No</p>				
21	<p>Was urea recorded?</p> <p>Yes No</p> <p>If yes, what was the first recorded (after admission) value?</p> <p>mmol/l Not recorded</p>				
22	<p>Was serum albumin recorded?</p> <p>Yes No</p> <p>If yes, what was the first recorded (after admission) value?</p> <p>mmol/l Not recorded</p>				
23	<p>Was there a record of medications being taken at time of admission (within 24 hours)?</p> <p>Yes No</p> <p>If yes, were there five or more medications recorded?</p> <p>Yes no</p>				
24	<p>What was the patient's temperature at admission? (°C)</p> <p>Not recorded</p>				

No.	Criterion	Essential	Desirable	Non-essential	Comments
25	<p><i>Is there a spirometry reading in the notes for this admission?</i></p> <p>Yes No</p> <p><i>If yes, what is the FEV₁ level (if more than one, give most recent)</i></p> <p>Not recorded</p>				
26	<p><i>Is there a record of peripheral oedema?</i></p> <p>Yes – present Yes – not present not recorded</p> <p><i>If peripheral oedema was present, was it:</i></p> <p>In leg/ankles? Sacral? Not recorded</p>				

Initial management

No.	Criterion	Essential	Desirable	Non-essential	Comments
27	<p><i>Was a course of antibiotics prescribed?</i></p> <p>Yes No</p>				
28	<p><i>Were nebulised bronchodilators prescribed?</i></p> <p>Yes No</p>				
29	<p><i>Did the patient receive systemic corticosteroids?</i></p> <p>Yes No</p>				
30	<p>How many sets of arterial blood gases results are in the records for this stay?</p>				
31	<p><i>Did the patient have a pH less than 7.35 at any time during this stay?</i></p> <p>Yes No</p>				
	<p><i>If yes, did they receive ventilatory support?</i></p> <p>Please tick all that apply:</p> <ul style="list-style-type: none"> – Respiratory stimulant (e.g. doxapram) – Non-invasive – Invasive – None 				
	<p><i>If the patient had a pH of less than 7.35 and did not receive ventilatory support, is it noted? If not, why not?</i></p> <p>Patient refused No facilities Not appropriate Failed Other Not recorded</p>				

Pre-discharge

No.	Criterion	Essential	Desirable	Non-essential	Comments
32	Was oximetry (O_2 saturation levels) undertaken, after acute phase but prior to discharge (within 48 hours of discharge)? Yes No				
	If yes, what were the results? Not recorded				
33	If a current smoker, was help toward smoking cessation given? Referred to smoking cessation programme Advice given and recorded? Nothing recorded? Not applicable (because non-smoker)				
34	Was there an assessment of the patient's home circumstances and their ability to cope? Yes No				
35	Where was the patient discharged to? Own home – independent of help Own home – with additional social support Sheltered housing or living with relative/carer Nursing or residential care Other hospital Not applicable – died in hospital Not recorded				
36	Is there a letter to the patient's primary care team? Yes No				
	If yes, did the letter include a clear list of the patient's medication? Yes No				
37	Which type of consultant was the patient under at time of discharge? Respiratory physician Care of elderly physician General physician Other Not recorded				
38	Time taken to complete (in hours and minutes)				

Appendix 4

A holistic review data collection page

Microsoft Access - [COPDHolistic : Form]

File Edit Insert Records Window Help

Type a question for help

COPD Holistic Study AUDIT RECORD NUMBER 25

NICE Definition Key data History/Patient Characteristics Admission Initial management Pre-discharge phase Overall assessment Default

1. Audit record number: 25

2. Hospital number:* 123-456-789

3. Date of birth: Day (dd):* 21 Month (mm):* 12 Year (yyyy): 1946

4. Gender: Male 5. First part of patient's postcode:* 55

6. Date of this admission to hospital: 27/11/2004

7. Date of discharge from hospital (or death if applicable): 29/11/2004 Death/Discharge: Discharge

8. Prior to this admission, has the patient previously been admitted to hospital for COPD or accepted onto an early discharge scheme? Yes

< Previous Page Next Page > Delete Record Close Form Save Record < Go to Previous Record Go to Next Record >

Logged In As: Royal Bournemouth Hospital (S714R)

Record: 1 of 1

start Pegasus Mail Software Manual v4 ~... HeartFailureHolistic : ... COPDHolistic : Form 14:40

Appendix 5

Reviewer training scenarios to assist in recognising variation in care quality using holistic review

The scenarios used stroke care as exemplars in order not to influence the views of the reviewers on what they might perceive as appropriate care for either of the two tracer conditions – COPD or heart failure.

Mrs X – scenario 1

Mrs X, a 78-year-old lively lady, has a dizzy turn and for an hour loses the use of her right arm and feels weakness in her right leg. She has had a transient ischaemic attack (TIA) and goes to her GP, very concerned. He gets her an outpatient appointment at her local hospital for two months' time. A week after she has seen him she suffers a completed stroke (involving her right side and speech) and is taken to hospital, where she is admitted to a general medical ward. It is a Friday. She is seen by a consultant on the following Tuesday. She has a CT scan on the Thursday. She develops pneumonia on the Friday, which is treated. She is referred the following week to a geriatrician. She is eventually sent home very disabled, no information is sent to her GP, and there proves to be a long wait for community rehabilitation, so after 2 weeks she is sent to a nursing home where she dies after 3 months.

Quality comments

This lady showed early signs of a pending stroke, which could have been prevented.

- No early diagnosis is made and no rapid referral to a specialist service is provided.
- No aspirin is prescribed to prevent the stroke.
- No specialist care in a stroke unit.
- There is a delay in providing a CT scan.
- Dysphagia (difficulty swallowing) is not picked up by early screening, leading to pneumonia.
- She does not receive early therapy input to begin rehabilitation and mobilisation.
- The communication systems between hospital and community fail.
- The community services are unable to provide the rehabilitation she needs.

Mrs X – scenario 2

Mrs X, a 78-year-old lively lady, has a dizzy turn and for an hour loses the use of her right arm and feels weakness in her right leg. She has had a TIA and goes to her GP, very concerned. He gets her an appointment for the TIA clinic in 2 days' time and starts her on aspirin. At the TIA one-stop clinic she is duplex scanned and referred for early carotid endarterectomy, which she has within 10 days. She returns home on appropriate secondary prevention, having received dietary advice to eat more healthily. She was a lifelong smoker so has been given cessation advice and nicotine replacement therapy. She continues with her active life, with regular follow-up visits to her GP.

Quality comments

Evidence shows that the danger of stroke is high after a TIA.

- She receives appropriate rapid referral to a specialist service.
- Secondary prevention with aspirin is commenced immediately.
- Early investigations are carried out.
- Appropriate surgical treatment is provided quickly.
- Secondary prevention and lifestyle advice are provided prior to discharge.
- Her health is monitored regularly by her GP.

Mr Y – scenario 3

Mr Y is a 60-year-old man. He has a rapid-onset stroke on a Saturday and is taken to his local hospital. Because the hospital CT scanner does not work over the weekend, he is booked for a scan later in the week but this is cancelled for emergency admissions and somehow never gets rebooked. There is no acute stroke unit and, with the rehabilitation stroke unit being full, he is sent to the acute medical admissions ward. There he is not screened for dysphagia so his problems with swallowing are not picked up and no protocol begins. He also develops pneumonia after a week.

Because of his incontinence he is catheterised in the acute medical unit and the catheter remains in situ when he is sent to the general medical ward and somehow gets left there. He is visited by the physiotherapist for his pneumonia and begins mobilising for half an hour per day when his pneumonia has resolved. Mobilising is also made more difficult by his catheter, the fact that he feels so weak from not eating properly and following the pneumonia. Other complications arise from his catheter. He spends many weeks in hospital and eventually is sent home, catheter still in situ because he has persisting incontinence from poor bladder tone that he developed from prolonged catheterisation. He remains weak and increasingly disabled. No rehabilitation continues at home and after a year he suffers a second stroke from which he does not recover. The family feel he had poor care and are making a complaint through the HCC.

Quality comments

- A CT scan is never carried out.
- He does not receive specialist care in a stroke unit.
- His dysphagia (difficulty swallowing) is not picked up, leading to pneumonia and malnutrition.
- Prolonged, possibly unnecessary, use of a catheter leads to complications.
- No early supported discharge from hospital is offered.
- No rehabilitation is provided on his return to primary care.

Mr Y – scenario 4

Mr Y is a 60-year-old man. He has a rapid-onset stroke on a Saturday and is taken to his local hospital. There, despite it being a weekend, he has a CT scan, from which ischaemic stroke is confirmed. He is admitted directly to the acute stroke unit. He is attached to monitoring systems (oxygen levels and ECG) by specialist nurses, who take regular observations of his vital signs (blood sugars, temperature, etc.). His clinical assessment on arrival in the unit, which includes a swallow

screen, identifies that he has dysphagia. He is designated 'nil by mouth' and tube feeding is begun by the nurses, using the protocol agreed with the stroke team. The speech and language therapist sees him on the Monday to begin treatment and his dietary regime comes under the guidance of the specialist dietitian. He is incontinent of urine in the first 48 hours, which is conservatively managed, without catheterisation, and resolves spontaneously. The physiotherapist begins his early mobilisation regime, which the nurses practice with him. An occupational therapy referral is made and therapy begins within 1 week of referral. After a fortnight the secondary prevention regime is established (aspirin having been started on the second day), using a protocol based on the national clinical guidelines for stroke. Because Mr Y is anxious to get home by this stage he has been referred to the specialist stroke Early Supported Discharge team and, once it is clear that he can safely get himself out of bed, he goes home under their care for continuing rehabilitation at home.

Quality comments

- A CT scan is carried out early, confirming the diagnosis.
- Secondary prevention with aspirin is commenced early.
- Hospital care is provided in a specialised stroke unit.
- He receives regular monitoring of physiological indicators while in hospital.
- Dysphagia is identified early and managed, reducing the risk of pneumonia and malnutrition.
- Urinary incontinence is managed without the need for a catheter (this is associated with better clinical outcome).
- Mobilisation therapy is begun early and monitored (this is one of the key features of stroke units associated with better clinical outcome).
- Early supported discharge by a specialist team is deemed suitable and is provided (this is proven to result in the same outcomes as hospital rehabilitation and is liked by patients).

Appendix 6

Record review for safety and quality study

A report of the record reviews undertaken by reviewer XXXX. A collaborative project by the Royal College of Physicians and the University of Sheffield.

Foreword

Medical record review has become a standard means of assessing variations in quality of care. This is despite uncertainty about which methods of record review are most effective and reliable. The aim of this audit was to assess which are the most effective and appropriate methods of reviewing quality of care from medical records. Further work is currently being undertaken to test out the conclusions and to assess whether it is possible to demonstrate a linkage between quality of care and outcomes.

All possible safeguards to preserve the quality of the data collected have been made by the University of Sheffield. Nevertheless it is important to interpret your results in this report using your knowledge of your own service and any difficulties you experienced in collecting your audit data that may have affected your own outcomes.

We are grateful to everyone who has helped with the project and appreciate the very considerable amount of time and effort that has gone into obtaining local data. We very much hope that this information will be useful for local audit purposes.

Introduction

Nine hospitals took part in this first phase of the audit. Overall, 1484 textual record reviews and 1400 criterion-based record reviews were returned to the study team. Records were reviewed from two specialties: COPD and heart failure. Reviewers were either nurses, non-clinical audit staff, SpRs or other clinical staff.

This report presents the results of the audit of COPD records undertaken by reviewer 5732. Textual reviews were undertaken on 38 out of 50 records, and criterion-based reviews were undertaken on 36 out of the same 50 records. Reviews are of patients admitted with an exacerbation of COPD, and who had a primary diagnosis of COPD, during the time period 1 September 2004 to 28 February 2005.

Methods

- *Textual review* – quality of care is assessed using the reviewer's own professional opinion.
- *Criterion-based review* – quality of care is assessed using a set of specific criteria.

A phases-of-care approach was adopted for both review methods. For textual review, reviewers were asked to comment on the care received by the patient in the admission, initial management and pre-discharge phases. They were then asked to make a final overall comment. Reviewers were also asked to rate each the quality of care in phase on a six-point scale and to rate the quality of the overall care on a 10-point scale.

For criterion-based review, criteria were grouped under the phase-of-care headings used in the holistic review. Reviewers were asked to answer the questions using information from the patient record.

Training for data collectors

Two training days were held for data collectors. The training days were both held in London (Royal College of Pathologists and BMA House).

The training provided an introduction to the review methods and familiarised the reviewers with the materials to be used while undertaking the reviews (data collection software and review help notes). There were also sessions on recognising care-quality variance when using the two review methods. For these sessions, examples were used from stroke care. This was so that we did not influence reviewers' perceptions of care quality for the two audit conditions. When reviewers were unable to attend a training day, two of the project team visited their hospital and provided training on-site. Reviewers who were unable to be present during the site visit were trained via telephone. Where possible, these reviewers were also assigned a 'buddy' (someone at their hospital who had undertaken the training in a face-to-face setting).

- Fifteen reviewers were trained at the training days.
- Ten reviewers were trained during a site visit.
- Fourteen reviewers were provided with telephone training.

The project team were available throughout the data collection period to answer queries and provide support and advice.

Participants

Nine hospitals took part in the COPD audit and eight of those hospitals also took part in the heart failure audit. Hospitals were randomly selected to participate in the audit, and consultants at each hospital were approached to provide their approval for the audit to take place and to assist in finding staff to review records.

The COPD audit involved the following reviewers:

Staff type	Number of reviewers
Doctor	6
Nurse	6
Non-clinical audit staff	6
Clinical other (e.g. physio or pharmacist)	2
Total	20

The heart failure audit involved the following reviewers:

Staff type	Number of reviewers
Doctor	10
Nurse	5
Non-clinical audit staff	3
Clinical other (e.g. physio or pharmacist)	1
Total	19

Data return

Reviewers were asked to review 50 heart failure or COPD records, using each of the review methods (resulting in 100 record reviews). If all reviewers had returned all reviews, this would have resulted in a total of 3900 reviews.

However, not all reviewers were able to return the full amount of reviews. This was for a variety of reasons, for example staff changing jobs. This was particularly a problem for the SpR reviewers, some of whom rotated to a post in a different hospital during the audit period. Also, there were some difficulties in recruiting reviewers in some hospitals. This meant that these reviewers started the audit later than other reviewers and, as such, had less time to complete all the reviews. Due to work pressures, one hospital site was unable to return any heart failure reviews.

Percentage of data returned

Condition	Review type	Total number of reviews returned	Reviews returned (%)
COPD	Textual	901	90
COPD	Criterion based	834	83
Heart failure	Textual	581	61
Heart failure	Criterion based	563	59
Total		2879	74

Textual data

We are using the textual data from the holistic review to investigate whether different staff types (e.g. audit staff, nurses and specialist registrars) make different types of comment or comment on different issues when asked to review quality of care from patient records.

Each reviewer's comments have been coded according to the type of comment, i.e. whether the comment is a judgement of the care provided or a description of the care provided. Comments have also been coded according to whether the comment relates to the patient records or patient care. We will use this information to determine which type of reviewer (SpR, nurse, non-clinical audit staff) provide the most useful types of comments about quality of care.

In some hospitals, different types of staff have reviewed the same records. This is so that we can compare the comments to gain an understanding of the types of comments made by different staff. All of the analysis has been anonymised and each reviewer is only identifiable by their reviewer ID.

This analysis is ongoing. We hope to publish the results of the study findings and will provide you with details of any publications.

Results

The following results relate to a review of 38 holistic and 36 criterion-based patient records by reviewer 5732.

Key data

Gender

	<i>n</i>	Per cent
Female	18	50
Male	18	50
Total	36	100

Number of patients accepted on to an early discharge scheme

	<i>n</i>	Per cent
Not accepted onto an early discharge scheme	36	100
Not applicable	0	0
Accepted on to an early discharge scheme	0	0
Total	0	0

Number of patients discharged or died

	<i>n</i>	Per cent
Discharge	35	97.2
	0	0
Died from COPD or complications of COPD		
Not recorded	1	2.8
Total	36	100

Number of patients with previous admissions for COPD

	<i>n</i>	Per cent
No previous admissions for COPD	19	52.8
Previous admissions for COPD	17	47.2
Total	36	100

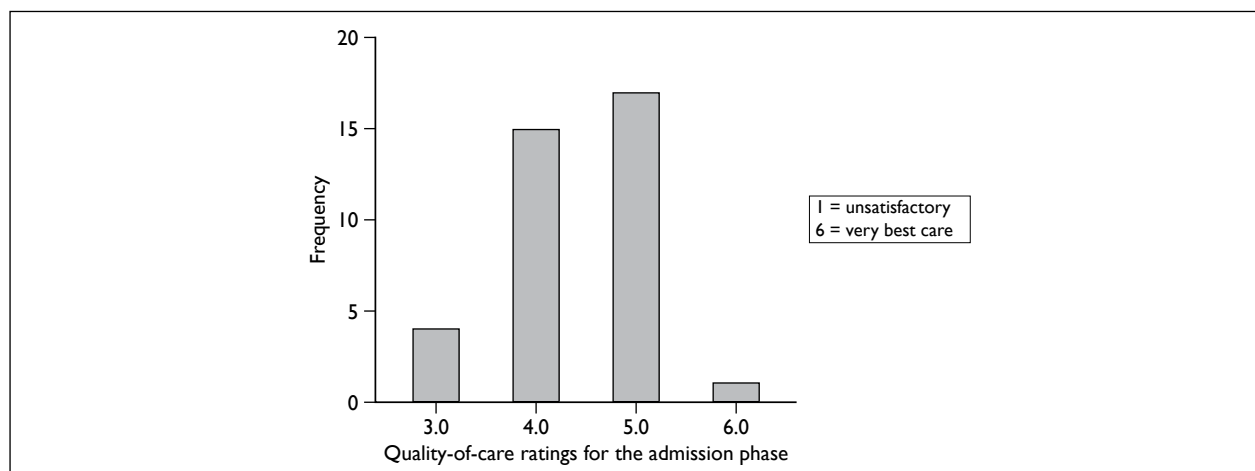
History and patient characteristics**Smoking status**

	<i>n</i>	Per cent
Current smoker	15	41.7
Ex-smoker (stopped more than 3 months)	18	50.0
Lifelong non-smoker	2	5.6
Not recorded	1	2.8
Total	36	100

Social circumstances

	<i>n</i>	Per cent
Lives alone with social service support	1	2.8
Lives alone, no support	11	30.6
Lives in nursing/residential home	3	8.3
Lives in warden controlled (sheltered) housing	1	2.8
Lives with spouse, close relative or carer	20	55.6
Total	36	100

Admission phase



Please note, some of these results are subjective and are the opinions of the individual reviewers.

Rating scale results: quality-of-care ratings for the admission phase

Quality-of-care rating for the admission phase – all reviewers from site 439

Reviewer	Mean quality-of-care rating	SD	Median	Range
5731	4.2	1.08	5.0	1.0–6.0
5732	4.4	0.72	4.0	3.0–6.0
5833	4.1	1.26	5.0	1.0–6.0
5834	5.0	0.99	5.0	2.0–6.0

Quality-of-care rating for the admission phase – all hospitals

Hospital	Mean	SD	Median	Range
Hospital 203	4.8	0.7	5.0	2.0–6.0
Hospital 211	4.9	0.6	5.0	1.0–6.0
Hospital 260	4.3	0.6	4.0	3.0–5.0
Hospital 271	4.5	1.2	5.0	1.0–6.0
Hospital 415	4.1	1.0	4.0	1.0–6.0
Hospital 420	4.0	1.4	4.0	2.0–6.0
Hospital 439	4.5	1.1	5.0	1.0–6.0
Hospital 441	5.4	1.0	6.0	2.0–6.0
Hospital 452	4.0	1.4	4.0	1.0–6.0

Criterion-based review – admission phase

Level of breathlessness

	<i>n</i>	Per cent
Increased	36	100
Not increased	0	0
Total	36	100

Level of sputum

	<i>n</i>	Per cent
Increased	3	8.3
Not increased	7	19.4
Not recorded	26	72.2
Total	36	100

Changes in sputum colour

	<i>n</i>	Per cent
Changed	3	8.3
Not changed	5	13.9
Not recorded	28	77.8
Total	36	100

Sputum colour

	<i>n</i>	Per cent
No sputum	9	25.0
Not recorded	6	16.7
White or grey	10	27.8
Yellow or green	11	30.6
Total	36	100

Was the dyspnoea rating recorded?

	<i>n</i>	Per cent
No	12	33.3
Yes	24	66.7
Total	36	100

Performance status

	<i>n</i>	Per cent
Limited activity but self care	7	19.4
Limited self care	4	11.1
Normal activity	9	25.0
Not known	13	36.1
Strenuous activity limited	3	8.3
Total	36	100

Chest X-ray within 24 hours?

	<i>n</i>	Per cent
No	2	5.6
Yes	34	94.4
Total	36	100

If yes, is X-ray report in notes?

	<i>n</i>	Per cent
Missing	2	5.6
No	0	0.0
Yes	34	94.4
Total	36	100

Respiratory rate within 24 hours?

	<i>n</i>	Per cent
Yes	2	5.6
No	34	94.4
Total	36	100

If yes, first reading is:

	<i>n</i>	Per cent
≤12	2	5.6
13	1	2.8
14	1	2.8
15	2	5.6
16	2	5.6
19	2	5.6
20	4	11.1
22	1	2.8
23	1	2.8
24	5	13.9
26	2	5.6
28	4	11.1
30	3	8.3
32	1	2.8
34	1	2.8
36	1	2.8
40	3	8.3
Total	36	100

Blood gases within 24 hours?

	<i>n</i>	Per cent
No	8	22.2
Yes	28	77.8
Total	36	100

ECG performed?

	<i>n</i>	Per cent
No	8	22.2
Yes	28	77.8
Total	36	100

Urea recorded?

	<i>n</i>	Per cent
No	3	8.3
Yes	33	91.7
Total	36	100

Serum albumin recorded?

	<i>n</i>	Per cent
No	8	22.2
Yes	28	77.8
Total	36	100

Record of medications at admission?

	<i>n</i>	Per cent
No	8	22.2
Yes	28	77.8
Total	36	100

If peripheral oedema present, was it:

	<i>n</i>	Per cent
No peripheral oedema/not recorded	20	55.6
Leg/ankles	15	41.7
Sacral	0	0.0
Not recorded	1	2.8
Total	36	100

Temperature at admission

°C	<i>n</i>	Per cent
34.5	1	2.8
35	1	2.8
35.4	1	2.8
35.5	1	2.8
35.7	1	2.8
36	1	2.8
36.1	3	8.3
36.2	2	5.6
36.3	2	5.6
36.4	3	8.3
36.5	2	5.6
36.7	3	8.3
36.8	1	2.8
36.9	2	5.6
37	4	11.1
37.1	1	2.8
37.2	2	5.6
37.4	1	2.8
37.5	1	2.8
37.8	2	5.6
38.2	1	2.8
Total	36	100

Spirometry reading this admission?

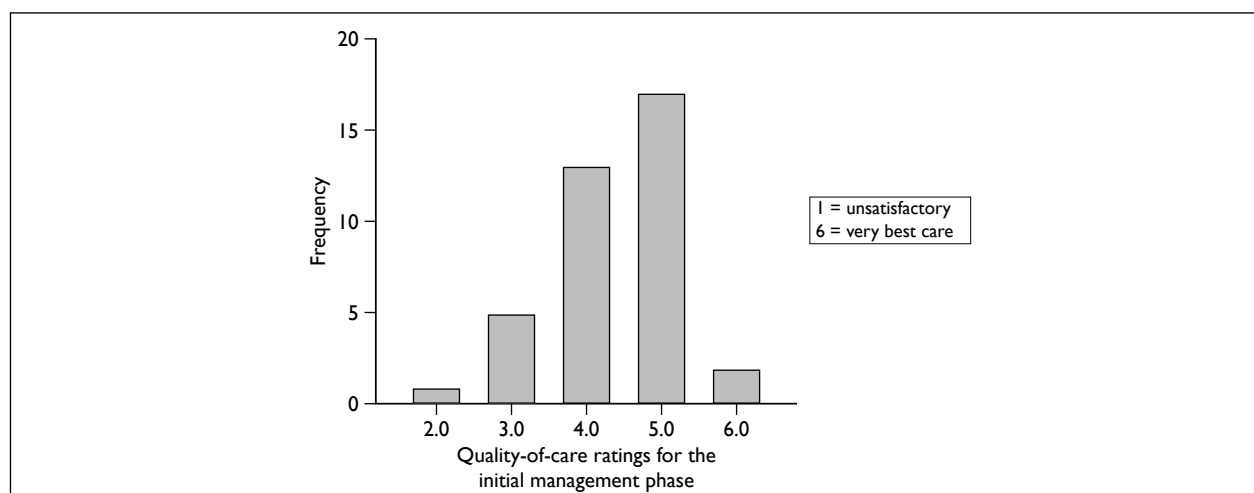
	<i>n</i>	Per cent
No	24	66.7
Yes	12	33.3
Total	36	100

Record of peripheral oedema?

	<i>n</i>	Per cent
Not recorded	5	13.9
Yes – not present	15	41.7
Yes – present	16	44.4
Total	36	100

Initial management phase

Rating scale results – quality-of-care ratings for the initial management phase



Please note, some of these results are subjective and are the opinions of the individual reviewers.

Quality-of-care rating for the initial management phase – all reviewers from site 439

Reviewer	Mean quality-of-care rating	SD	Median	Range
5731	4.5	0.92	5.0	2.0–6.0
5732	4.4	0.88	5.0	2.0–6.0
5833	4.4	1.33	5.0	1.0–6.0
5834	5.2	0.83	5.0	3.0–6.0

Quality-of-care rating for the initial management phase – all hospitals

Hospital	Mean	SD	Median	Range
Hospital 203	4.9	0.8	5.0	2.0–6.0
Hospital 211	4.9	0.8	5.0	1.0–6.0
Hospital 260	4.5	0.5	4.0	3.0–5.0
Hospital 271	4.6	1.1	5.0	1.0–6.0
Hospital 415	4.2	1.0	4.0	1.0–6.0
Hospital 420	3.8	1.5	3.0	1.0–6.0
Hospital 439	4.6	1.1	5.0	1.0–6.0
Hospital 441	5.3	0.9	6.0	2.0–6.0
Hospital 452	4.0	1.3	4.0	1.0–6.0

Criterion-based review: initial management phase

Were antibiotics prescribed?

	<i>n</i>	Per cent
No	5	13.9
Yes	31	86.1
Total	36	100

Were nebulised bronchodilators prescribed?

	<i>n</i>	Per cent
No	3	8.3
Yes	33	91.7
Total	36	100

Did patient receive systemic corticosteroids?

	<i>n</i>	Per cent
No	2	5.6
Yes	34	94.4
Total	36	100

Number of arterial blood gas results

	<i>n</i>	Per cent
0	7	19.4
1	18	50.0
2	2	5.6
3	4	11.1
4	2	5.6
5	1	2.8
6	2	5.6
Total	36	100

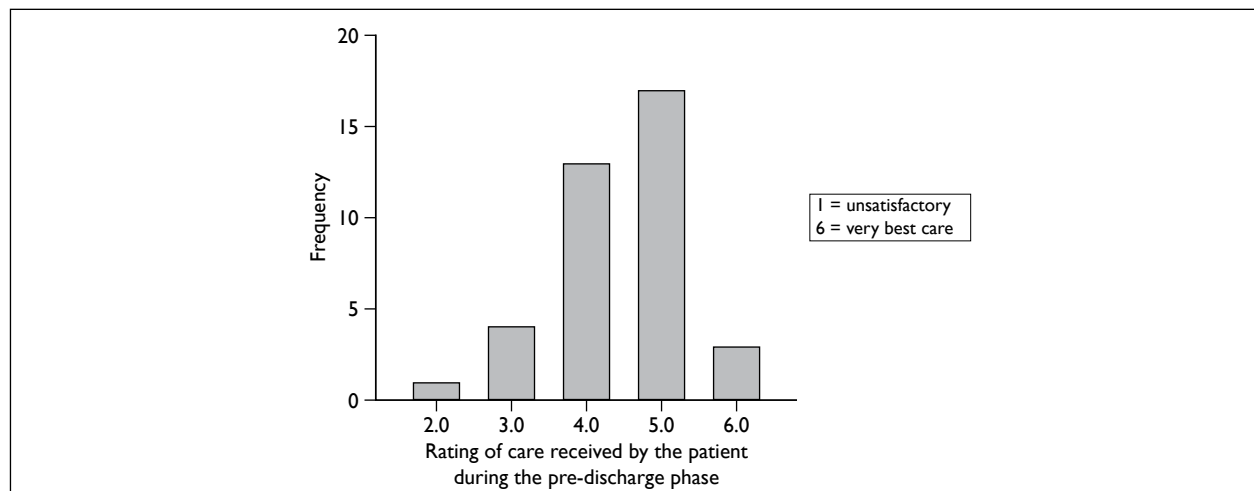
pH less than 7.35 at any time?

	<i>n</i>	Per cent
Missing data	6	16.7
No	17	47.2
Yes	13	36.1
Total	36	100

Pre-discharge phase

Rating scale results – quality-of-care ratings for the pre-discharge phase

Please note, some of these results are subjective and are the opinions of the individual reviewers.



Quality-of-care rating for the pre-discharge phase – all reviewers from site 439

Reviewer	Mean quality-of-care rating	SD	Median	Range
5731	4.5	0.85	5.0	2.0–5.0
5732	4.4	0.89	5.0	2.0–6.0
5833	4.4	1.14	5.0	1.0–6.0
5834	4.6	1.23	5.0	1.0–6.0

Quality-of-care rating for the pre-discharge phase – all hospitals

Hospital	Mean	SD	Median	Range
Hospital 203	4.74	0.76	5.0	1–6
Hospital 211	4.8	0.84	5.0	1–5
Hospital 260	4.3	0.57	4.0	3–5
Hospital 271	4.5	1.36	5.0	1–6
Hospital 415	4.2	1.16	4.0	1–6
Hospital 420	3.4	1.67	3.0	1–6
Hospital 439	4.5	1.14	5.0	1–6
Hospital 441	4.8	0.98	5.0	2–6
Hospital 452	3.6	1.53	3.0	1–6

Criterion-based audit – pre-discharge phase

Oximetry within 48 hours of discharge?

	<i>n</i>	Per cent
Yes	13	36.1
No	23	63.9
Total	36	100

Assessment of home circumstances?

	<i>n</i>	Per cent
No	30	83.3
Yes	6	16.7
Total	36	100

Where was patient discharged to?

	<i>n</i>	Per cent
Not applicable – died in hospital	1	2.8
Nursing or residential care	1	2.8
Other hospital	1	2.8
Own home – independent of help	25	69.4
Own home – with additional social support	4	11.1
Sheltered housing or living with relative	4	11.1
Total	36	100

Discharge letter to primary care team?

	<i>n</i>	Per cent
No	4	11.1
Yes	32	88.9
Total	36	100

If yes, is there a clear list of medications?

	<i>n</i>	Per cent
No discharge letter	4	11.1
No	1	2.8
Yes	31	86.1
Total	36	100

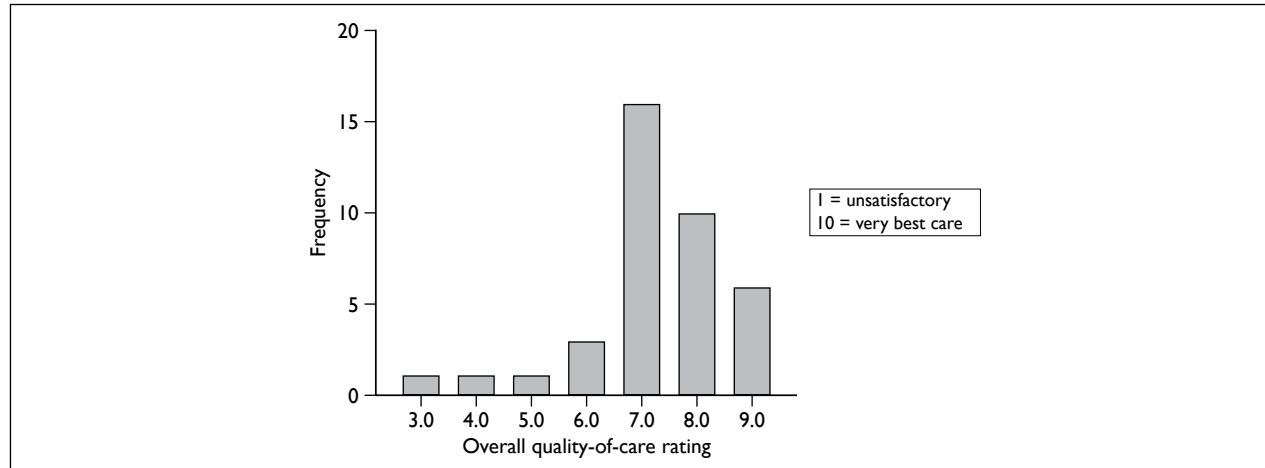
Type of consultant at discharge?

	<i>n</i>	Per cent
Care of elderly physician	5	13.9
General physician	9	25.0
Other	2	5.6
Respiratory physician	20	55.6
Total	36	100

Overall care

Rating scale results: quality-of-care ratings for the overall care

Please note, some of these results are subjective and are the opinions of the individual reviewers.



Quality-of-care rating overall – all reviewers from site 439

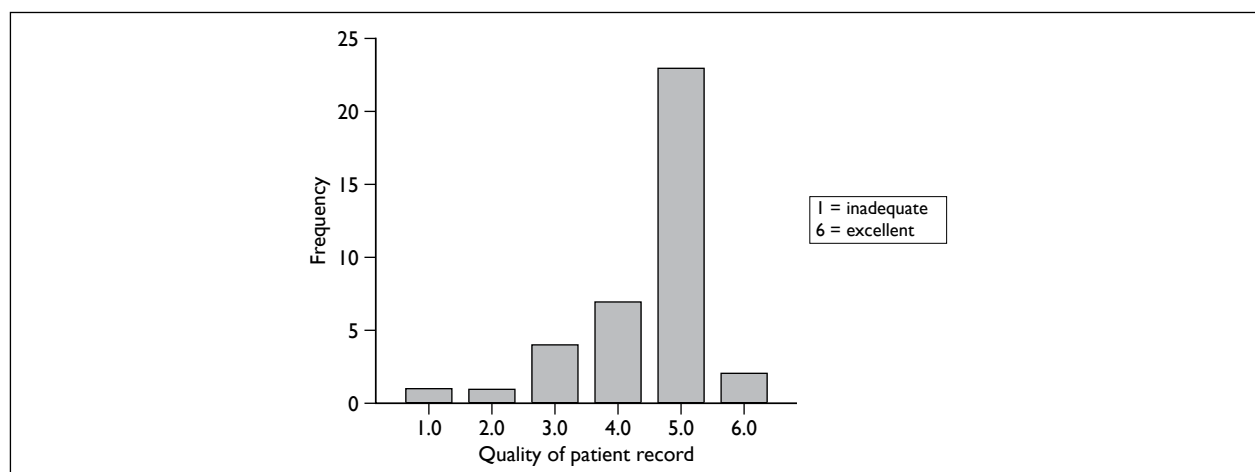
Reviewer	Mean quality-of-care rating	SD	Median	Range
5731	7.4	1.90	8.0	2.0–10.0
5732	7.3	1.31	7.0	3.0–9.0
5833	7.4	2.21	8.0	2.0–10.0
5834	7.9	1.54	8.0	3.0–10.0

Overall quality-of-care rating – all hospitals

Hospital	Mean	SD	Median	Range
Hospital 203	8.0	1.26	8.0	3–10
Hospital 211	8.2	1.46	9.0	2–9
Hospital 260	7.2	1.00	7.0	4–9
Hospital 271	7.6	1.90	8.0	1–10
Hospital 415	7.3	1.50	8.0	1–10
Hospital 420	5.9	2.40	6.0	2–10
Hospital 439	7.5	1.79	8.0	2–10
Hospital 441	8.3	1.25	8.0	6–10
Hospital 452	5.7	2.09	6.0	1–10

Patient records

Rating scale results – ratings for the quality of patient records.



Appendix 7

COPD – correlations between holistic mean overall scale scores^a and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Percentage of patients with COPD who die in hospital within 28 days	-0.295	0.207	Weak
HSMR from Dr Foster (3-year mortality)	-0.157	0.508	Very weak
HSMR from Dr Foster (1-year mortality)	-0.072	0.763	Very weak
Incidents to NPSA per 100 bed-days	-0.345	0.136	Weak
SMR for deaths in low-mortality HRGs	0.049	0.838	Very weak
COPD finished consultant episodes	0.125	0.610	Very weak
COPD bed-days	0.102	0.677	Very weak
COPD mean length of stay	-0.167	0.495	Very weak
COPD mean age	0.011	0.963	Very weak
Star rating (0 worst to 3 best)	-0.077 ^b	0.746	Very weak
Use of resources HCC	-0.184 ^b	0.437	Very weak
Patient's experience	0.222	0.347	Weak
Quality of services	-0.118 ^b	0.620	Very weak
Percentage of patients with acute MI receiving thrombolysis	0.185	0.448	Very weak
Existing national targets	-0.072 ^b	0.764	Very weak
New national targets	0.105 ^b	0.658	Very weak
NHS staff survey Q25a: Seen errors in the past month (% yes)	-0.328	0.158	Weak
NHS staff survey Q27b: Encouraged to report errors (mean)	-0.490	0.028	Moderate
NHS staff survey Q27e: Trust takes action to ensure does not happen again (mean)	-0.131	0.581	Very weak
NHS staff survey Q24a: Know how to report (% yes)	-0.230	0.329	Weak
NHS staff survey Q24b: System for reporting (% yes)	-0.243	0.303	Weak
NHS staff survey Q24b: System for reporting (% no)	-0.154	0.516	Very weak
NHS staff survey Q24b: System for reporting (% don't know)	0.294	0.208	Very weak
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	-0.503	0.024	Moderate
NHS staff survey Q22f: Happy with standard of care provided (mean)	-0.282	0.228	Weak

HRGs, Healthcare Resource Groups.
 a Scale: 1 = unsatisfactory to 6 = very best care.
 b Spearman's rank correlation used.
 Correlations at the <0.05 significance level are shown in bold text.

Appendix 8

Heart failure – correlations between holistic mean overall scale scores^a and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Percentage of patients with heart failure who die in hospital within 28 days	-0.334	0.149	Weak
HSMR from Dr Foster (3-year mortality)	-0.183	0.439	Very weak
HSMR from Dr Foster (1-year mortality)	-0.308	0.186	Weak
Incidents to NPSA per 100 bed-days	-0.228	0.335	Weak
SMR for deaths in low-mortality HRGs	-0.093	0.696	Very weak
Heart failure finished consultant episodes	-0.174	0.477	Very weak
Heart failure bed-days	-0.237	0.329	Weak
Heart failure mean length of stay	0.064	0.795	Very weak
Heart failure mean age	-0.445	0.056	Moderate
Star rating (0 worst to 3 best)	0.240 ^a	0.309	Weak
Use of resources HCC	0.345 ^a	0.136	Weak
Patient's experience	-0.365	0.114	Weak
Quality of services	0.651^b	0.002	Strong
Percentage of patients with acute MI receiving thrombolysis	0.350	0.142	Weak
Existing national targets	0.765^b	<0.001	Strong
New national targets	0.453^b	0.045	Moderate
NHS staff survey Q25a: Seen errors in the past month (% yes)	-0.261	0.267	Weak
NHS staff survey Q27b: Encouraged to report errors (mean)	0.308	0.187	Weak
NHS staff survey Q27e: Trust takes action to ensure does not happen again (mean)	0.430	0.059	Moderate
NHS staff survey Q24a: Know how to report (% yes)	0.509	0.022	Moderate
NHS staff survey Q24b: System for reporting (% yes)	0.264	0.261	Weak
NHS staff survey Q24b: System for reporting (% no)	0.126	0.598	Very weak
NHS staff survey Q24b: System for reporting (% don't know)	-0.306	0.189	Weak
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	0.442	0.051	Moderate
NHS staff survey Q22f: Happy with standard of care provided (mean)	0.078	0.744	Very weak

HRGs, Healthcare Resource Groups.
^a Scale: 1 = unsatisfactory to 6 = very best care.
^b Spearman's rank correlation used.
Correlations at the <0.05 significance level are shown in bold text.

Appendix 9

COPD – correlations between holistic mean phase scale scores and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Percentage of patients with COPD who die in hospital within 28 days	-0.290	0.215	Weak
HSMR from Dr Foster (3-year mortality)	-0.135	0.569	Very weak
HSMR from Dr Foster (1-year mortality)	-0.095	0.691	Very weak
Incidents to NPSA per 100 bed-days	-0.067	0.778	Very weak
SMR for deaths in low-mortality HRGs	0.070	0.771	Very weak
COPD finished consultant episodes	0.422	0.072	Moderate
COPD bed-days	0.387	0.102	Weak
COPD mean length of stay	-0.146	0.550	Very weak
COPD mean age	0.108	0.658	Very weak
Star rating (0 worst to 3 best)	0.035 ^a	0.882	Very weak
Use of resources HCC	-0.054 ^a	0.821	Very weak
Patient's experience	-0.365	0.114	Weak
Quality of services	-0.083	0.729	Very weak
Percentage of patients with acute MI receiving thrombolysis	0.350	0.142	Weak
Existing national targets	-0.054 ^a	0.821	Very weak
New national targets	0.073 ^a	0.759	Very weak
NHS staff survey Q25a: Seen errors in the past month (% yes)	-0.236	0.316	Weak
NHS staff survey Q27b: Encouraged to report errors (mean)	-0.330	0.156	Weak
NHS staff survey Q27e: Trust takes action to ensure does not happen again (mean)	0.038	0.872	Very weak
NHS staff survey Q24a: Know how to report (% yes)	-0.248	0.292	Weak
NHS staff survey Q24b: System for reporting (% yes)	-0.203	0.390	Weak
NHS staff survey Q24b: System for reporting (% no)	-0.103	0.664	Very weak
NHS staff survey Q24b: System for reporting (% don't know)	0.238	0.312	Weak
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	-0.454	0.044	Moderate
NHS staff survey Q22f: Happy with standard of care provided (mean)	-0.268	0.252	Weak

HRGs, Healthcare Resource Groups.
 a Spearman's rank correlation used.
 Correlations at the <0.05 significance level are shown in bold text.

Appendix 10

Heart failure – correlations between holistic mean overall scale scores and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Percentage of patients with heart failure who die in hospital within 28 days	-0.274	0.242	Weak
HSMR from Dr Foster (3-year mortality)	-0.139	0.560	Very weak
HSMR from Dr Foster (1-year mortality)	-0.272	0.245	Weak
Incidents to NPSA per 100 bed-days	-0.215	0.362	Weak
SMR for deaths in low-mortality HRGs	-0.064	0.789	Very weak
Heart failure finished consultant episodes	-0.147	0.549	Very weak
Heart failure bed-days	-0.226	0.353	Weak
Heart failure mean length of stay	0.029	0.905	Very weak
Heart failure mean age	-0.552	0.014	Moderate
Star rating (0 worst to 3 best)	0.222 ^a	0.347	Weak
Use of resources HCC	0.248 ^a	0.292	Weak
Patient's experience	-0.292	0.211	Weak
Quality of services	0.486^a	0.030	Moderate
Percentage of patients with acute MI receiving thrombolysis	0.463	0.046	Moderate
Existing national targets	0.691^a	0.001	Strong
New national targets	0.226 ^a	0.338	Weak
NHS staff survey Q25a: Seen errors in the past month (% yes)	-0.212	0.369	Weak
NHS staff survey Q27b: Encouraged to report errors (mean)	0.331	0.154	Weak
NHS staff survey Q27e: Trust takes action to ensure does not happen again (mean)	0.470	0.037	Moderate
NHS staff survey Q24a: Know how to report (% yes)	0.546	0.013	Moderate
NHS staff survey Q24b: System for reporting (% yes)	0.286	0.222	Weak
NHS staff survey Q24b: System for reporting (% no)	0.076	0.750	Very weak
NHS staff survey Q24b: System for reporting (% don't know)	-0.313	0.180	Weak
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	0.470	0.037	Moderate
NHS staff survey Q22f: Happy with standard of care provided (mean)	0.009	0.972	Very weak

HRGs, Healthcare Resource Groups.
^a Spearman's rank correlation used.
 Correlations at the <0.05 significance level are shown in bold text.

Appendix I I

COPD – correlations between holistic mean criterion scores and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Percentage of patients with COPD who die in hospital within 28 days	-0.297	0.203	Weak
HSMR from Dr Foster (3-year mortality)	-0.247	0.295	Weak
HSMR from Dr Foster (1-year mortality)	-0.111	0.640	Very weak
Incidents to NPSA per 100 bed-days	-0.489	0.029	Moderate
SMR for deaths in low-mortality HRGs	0.308	0.187	Weak
COPD finished consultant episodes	-0.022	0.928	Very weak
COPD bed-days	-0.007	0.979	Very weak
COPD mean length of stay	0.118	0.629	Very weak
COPD mean age	0.125	0.611	Very weak
Star rating (0 worst to 3 best)	0.019 ^a	0.936	Very weak
Use of resources HCC	-0.311 ^a	0.182	Weak
Patient's experience	0.101	0.672	Very weak
Quality of services	-0.188 ^a	0.427	Very weak
Percentage of patients with acute MI receiving thrombolysis	0.042	0.866	Very weak
Existing national targets	-0.207 ^a	0.381	Weak
New national targets	-0.049 ^a	0.838	Very weak
NHS staff survey Q25a: Seen errors in the past month (% yes)	0.069	0.772	Very weak
NHS staff survey Q27b: Encouraged to report errors (mean)	-0.238	0.312	Weak
NHS staff survey Q27e: Trust takes action to ensure does not happen again (mean)	-0.084	0.726	Very weak
NHS staff survey Q24a: Know how to report (% yes)	-0.263	0.263	Very weak
NHS staff survey Q24b: System for reporting (% yes)	0.093	0.698	Very weak
NHS staff survey Q24b: System for reporting (% no)	-0.131	0.582	Very weak
NHS staff survey Q24b: System for reporting (% don't know)	-0.052	0.827	Very weak
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	-0.139	0.558	Very weak
NHS staff survey Q22f: Happy with standard of care provided (mean)	0.475	0.034	Moderate

HRGs, Healthcare Resource Groups.
^a Spearman's rank correlation used.
 Correlations at the <0.05 significance level are shown in bold text.

Appendix 12

Heart failure – correlations between mean criterion scores and outcome variables

Variable	Correlation coefficient	p-value	Relationship
Percentage of patients with heart failure who die in hospital within 28 days	-0.357	0.122	Weak
HSMR from Dr Foster (3-year mortality)	0.014	0.952	Very weak
HSMR from Dr Foster (1-year mortality)	-0.218	0.357	Weak
Incidents to NPSA per 100 bed-days	0.528	0.017	Moderate
SMR for deaths in low-mortality HRGs	0.280	0.232	Weak
Heart failure finished consultant episodes	-0.255	0.291	Weak
Heart failure bed-days	-0.259	0.284	Weak
Heart failure mean length of stay	0.017	0.944	Very weak
Heart failure mean age	-0.325	0.172	Weak
Star rating (0 worst to 3 best)	-0.254 ^a	0.174	Weak
Use of resources HCC	0.029 ^a	0.279	Very weak
Patient's experience	-0.377	0.903	Weak
Quality of services	-0.033 ^a	0.891	Very weak
Percentage of patients with acute MI receiving thrombolysis	-0.102	0.679	Very weak
Existing national targets	-0.147 ^a	0.535	Very weak
New national targets	0.155 ^a	0.515	Very weak
NHS staff survey Q25a: Seen errors in the past month (% yes)	0.192	0.416	Very weak
NHS staff survey Q27b: Encouraged to report errors (mean)	0.137	0.564	Very weak
NHS staff survey Q27e: Trust takes action to ensure does not happen again (mean)	-0.092	0.698	Very weak
NHS staff survey Q24a: Know how to report (% yes)	-0.142	0.551	Very weak
NHS staff survey Q24b: System for reporting (% yes)	-0.145	0.543	Very weak
NHS staff survey Q24b: System for reporting (% no)	0.620	0.004	Strong
NHS staff survey Q24b: System for reporting (% don't know)	-0.051	0.833	Very weak
NHS staff survey Q22e: Care of patient/service user is top priority (mean)	0.336	0.147	Weak
NHS staff survey Q22f: Happy with standard of care provided (mean)	-0.291	0.214	Weak

HRGs, Healthcare Resource Groups.
^a Spearman's rank correlation used.
Correlations at the <0.05 significance level are shown in bold text.

Appendix 13

Comparison of holistic and criterion-based review methods using structured clinical records in stroke care

Background

UK stroke care tends to have structured medical records for hospital inpatients, with prospective completion of patient records based on structured phases of care, in some units. This, it is hypothesised, changes the type and quality of data collected in the medical record and thus may more accurately enable conformance with, and assessment of, good care standards.

It is currently unknown whether the use of structured medical records has any effect on the quality of information available for peer review, or whether it might differentially influence the quality of information that is captured by explicit or implicit review methods. It may also be that there is a higher level of inter-rater reliability to be found between, and within, types of reviewers when using structured, prospective record-keeping compared with that found in the main research project.

Major national sentinel audit projects in the UK have already used review criteria-based, explicit review methods to explore quality variance in stroke care.¹ These were undertaken by teams of nurses or physicians who had been trained in records review methods. Substantial variations in organisation and clinical care have been identified across the 8200 cases included in the national stroke audit. The audit did not use a holistic approach, which has been hypothesised as an alternative means of identifying quality variation, especially in complex cases.²

Study questions

This small adjunct study seeks to answer two related questions. First, what are the similarities and differences in peer-review information captured by explicit (review criteria-based) methods and implicit (holistic) methods from structured (stroke care) clinical records? Second, are there differences in the type of information recorded by clinical audit staff (including nurses) and by doctors, using the two types of review methods?

This study is nested within the main medical records study, which addresses similar research questions, but in which the records are not structured. Although the stroke care study is small, the overall results of the stroke care study can therefore be contrasted with the non-structured record results in the phase one study to begin to explore whether there are differences between the type of information that can be extracted from the two different types of records.

Secondary study questions

- Does structured prospective medical record-keeping in stroke care influence the type, extent and quality of data recorded in clinical audit review (compared with the type and quality of data found in unstructured record-keeping in the phase one study)?
- When using explicit (review criteria-based) clinical audit review methods, does the use of structured recording in stroke care change the proportions of recorded criteria compared with the proportions recorded in unstructured records for COPD and heart failure care?
- For both explicit (review criteria-based) and implicit (holistic) clinical audit case note review methods, does reliability improve between and within reviewer types when structured recording is used in comparison to unstructured recording?

Methods

The overall research approach was to investigate the impact of structured prospective record-keeping on the reliability and completeness of holistic and criterion-based case note review methods. Quality of care was assessed using a combined holistic and algorithmic method, as used in the phase two outcomes study, by one nursing-trained reviewer and two medical reviewers. The same case notes were reviewed by each reviewer.

Stroke

Stroke is recognised as the third biggest cause of death in the UK. It is also the largest single cause

of severe disability in older people. In excess of 110,000 people in England each year will suffer from a stroke, which incurs NHS costs of over £2.8B per year.³ All hospitals that care for patients with stroke were required to have a specialist stroke service by 2004, as set out by the National Service Framework for Older People.⁴ To support our choice of stroke for this adjunct study, we took into account the availability of an evidence-based guideline produced by the RCP, together with the existence of the National Sentinel Stroke Audit.¹ The National Sentinel Stroke Audit Criteria provided the basis for developing a set of review criteria for safety and quality assessment for stroke management for this study (see also the section on criterion development in Chapter 2).

Number of case notes for review

Each reviewer was asked to review 40 stroke care records, as in the phase one study.

Selection and recruitment of hospitals

Only one hospital participated in this small study. This hospital was chosen because of the study team's close links with the stroke care staff, whose input was crucial to the development of the review criteria and the running of the study. A second hospital was also approached, but, although they were keen to participate, they did not have staff available to take part.

Numbers and types of review and reviewers

Two doctors in training (SpRs) and one clinical audit nurse were recruited to review 40 records of patients admitted to hospital for care for an acute stroke.

Holistic review data capture

Holistic review data was collected using the same methods as in the phase one study. Reviewers were asked to provide a textual comment on the quality of care and also to rate the quality of care on a six-point rating scale (1 = unsatisfactory, 6 = very best care). This was done for each of three phases of care (admission, initial management and pre-discharge) and for care overall. Care overall was rated on a 10-point rating scale (1 = unsatisfactory, 10 = very best care).

Assessing the quality of recording in the case notes

Evaluation of quality of care through case note review is critically dependent on the quality of recording in the case notes, together with that in associated data sources, such as computerised

pathology and radiology results. In order to assess the quality and completeness of the records under review, reviewers were asked to assess the quality of each record using a six-point rating scale (1 = inadequate, 6 = excellent), as per the main study.

Review criteria development for stroke care

The basis for the development of the review criteria was an already established criterion-based audit data set within the National Sentinel Stroke Audit, an organisational and clinical audit comprising 94 criteria. We developed a shorter version of the clinical component of the national audit data set through discussions with stroke care staff, using the same approach as the phase one study.

For the stroke study criterion-based questions, predefined answer options were provided for all questions. Usually, the options were 'yes', 'no' and 'not recorded'. Reviewers were instructed to answer 'yes' if the care was provided or the result of a test was in the patient record, and 'no' if the care was not provided for a valid reason. Examples of valid reasons were provided for each question and included things such as if the patient died, or was unconscious or was receiving palliative care. Reviewers were instructed to answer 'not recorded' if the information they were looking for was missing from the record. Where information is missing from the record it is presumed that care was not provided.

Developing data capture tools

Data collection materials were developed in Microsoft ACCESS[®] and were designed to be easily used by the reviewers. The software had a facility to easily export formatted data to the study team.

Staff training

Where possible, we intended that staff reviewing records for the stroke care study should attend one of the training days for the main study, although only the nursing-trained clinical reviewer was able to attend the training day. The two doctors were unable to attend due to clinical commitments, so one attended a one-to-one training session with the study project manager, while the other was trained via telephone by the project manager.

Analysis methods

Holistic scale score analysis

Summary statistics for the holistic quality-of-care ratings were calculated for each reviewer, for the phases of care and overall data. Box-and-whisker

plots comparing the quality-of-care ratings for the phases of care and overall quality of care for each reviewer were also produced.

Measuring reliability between reviewer pairs

The reliability between the reviewers overall quality-of-care ratings was assessed by calculating intraclass correlation coefficients (ICCs) in SPSS. ICCs were calculated for each reviewer pair (doctor 1 versus doctor 2, doctor 1 versus clinical audit 1, and doctor 2 versus clinical audit 1), as well as a combined ICC for all three reviewers.

Criterion-based review

The score for each criterion was summed to create quality-of-care scores for each phase of care and overall. Summary statistics for the criterion-based quality-of-care score were calculated, as were box-and-whisker plots comparing the three reviewers' data.

Measuring reliability between reviewer pairs

Intraclass correlation coefficients were used to assess the inter-rater reliability of the criterion-based quality-of-care scores. As with the holistic reliability analysis, ICCs were calculated for each reviewer pair and for all three reviewers.

Ethics review

The ethics review was the same as that for the main studies.

Health-care governance

Although the main study had already received ethical approval, further discussions were held with the clinical effectiveness manager of the study hospital to determine whether this small adjunct study was research, clinical audit or service review.⁵

Ethical principles were considered, following the decision that this work would be undertaken as service review and conducted in line with local governance procedures.

Results

Holistic quality-of-care rating scale

Completeness of data capture

Each of the doctor reviewers reviewed 37 out of the 40 records and the clinical reviewer reviewed 40 out of 40 records. Only the 37 reviews that were reviewed by each of the three reviewers were included in the analysis. The level of completeness of holistic review was assessed by calculating the amount of missing data for each of the quality-of-care rating scales. The amount of missing data for each phase for each reviewer is presented in *Table 35* which shows that the completeness level of the holistic rating scale is high. One reviewer (1733, Clinical) had no missing data. The amount of missing data for this section of the review is small. The most missing data was recorded by Doctor 1 for the Initial Management Phase, however, 8.8% missing data equates to 3 instances of missing data.

Quality of care

Table 36 presents the results of the quality-of-care analysis of stroke care using a holistic rating scale, where 1 = unsatisfactory care and 6 = very best care. On the whole the reviewer mean/median results were similar for each phase of care. Doctor 2 tends to rate the quality of care lower than the clinical reviewer and doctor 1, particularly in the pre-discharge section. The clinical reviewer uses only a limited section of the rating scale (between 3 and 5), whereas the other two reviewers (doctors) tend to use all of the available scale.

TABLE 35 Holistic review completeness of data collection

	Clinical reviewer (nurse) (%)	Doctor 1 reviewer (%)	Doctor 2 reviewer (%)
Phase of care (37 reviews each)			
Admission phase missing data	0	0	2.7
Initial management phase missing data	0	8.8	2.7
Pre-discharge phase missing data	0	2.7	0
Overall missing data	0	0	2.7

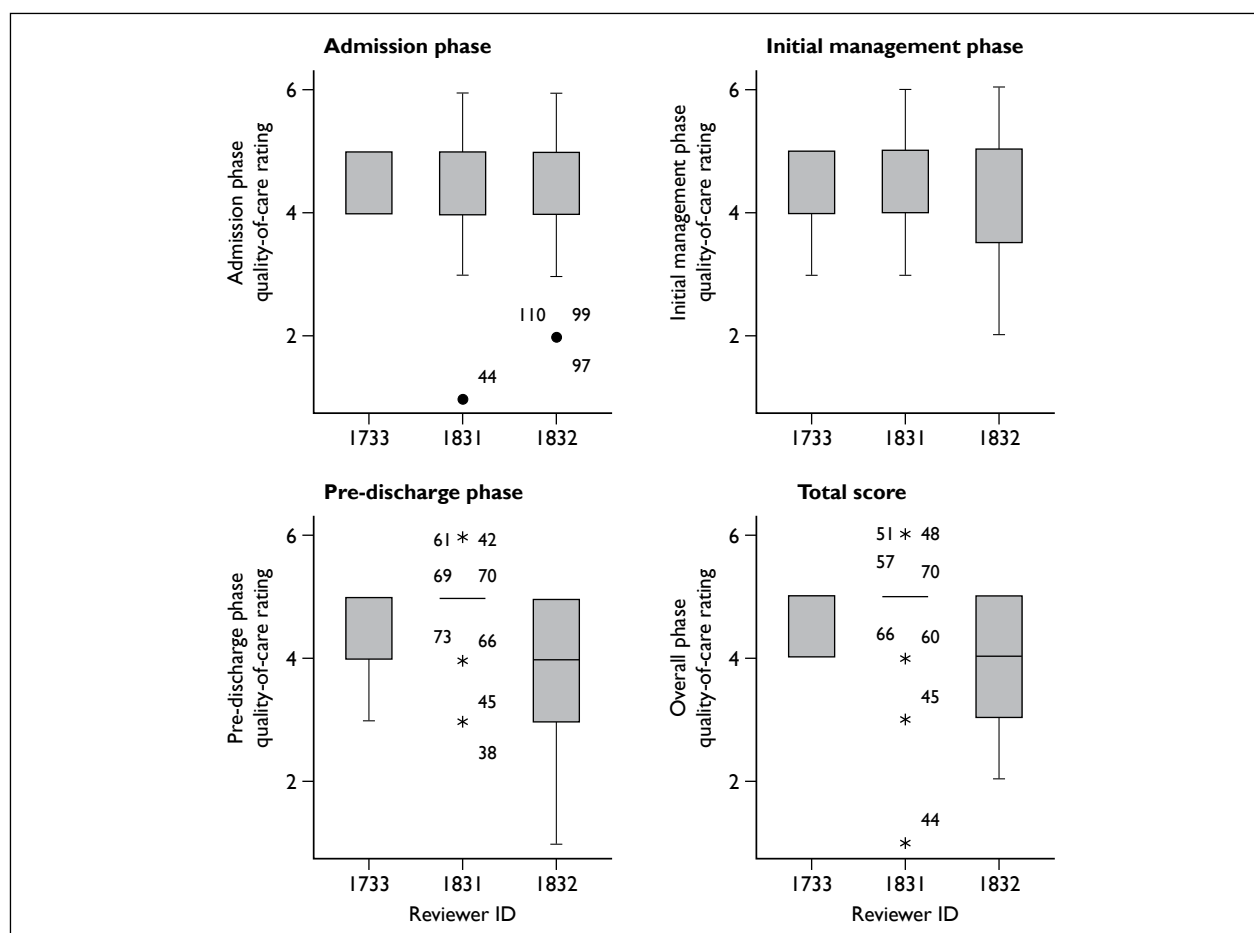


FIGURE 33 Holistic phase quality-of-care rating.

Box-and-whisker plots of holistic scale score data (Figure 33)

The box-and-whisker plots compare the median quality-of-care scale rating and the interquartile range, for each reviewer, for each phase of care.

Reliability between reviewers

Table 37 presents the results of the stroke care holistic review inter-rater reliability analysis and shows the level of agreement between reviewers for the overall quality-of-care ratings.

Comparison with results from the main study

The pair of doctor reviewers achieves the highest reliability. This is in line with findings from the main study, which found that pairs of doctor reviewers achieved the highest reliability for holistic review. The phase one reliability study also found that there was low reliability between different-staff-type reviewer pairs, for example ‘doctor and nurse’ for holistic review (see Table 7, p. 21). This finding is supported by the analysis of the stroke data as

the different-staff-type reviewer pairs achieve lower ICCs here.

The amount of comparison that can be undertaken between the stroke reviewer reliability analysis and the main study reviewer pair reliability analysis is limited, due to the small number of stroke reviewers taking part. The doctor reviewer pair ICC is similar to that of the COPD doctor reviewer pair in the phase one reliability study (0.328 versus 0.33, respectively – Table 37). In this much smaller study on stroke care the different-staff-type reviewer pair reliability comparisons are generally much lower than those presented in Table 6a (p. 20) in the phase one study.

Criterion-based review

Completeness of data capture

Each reviewer completed 37 reviews (of the same patient records). Where a reviewer did not select one of the predefined answer options for each criterion, this was classed as missing data. The results in Table 38 show that there are very

TABLE 36 Holistic scale score results

		Clinical reviewer (nurse)	Doctor 1 reviewer	Doctor 2 reviewer
Phase of care				
Admission phase quality-of-care rating	Mean	4.5	4.7	4.4
	(SD)	(0.5)	(0.9)	(1.3)
	Median	4.0	5.0	5.0
	(IQR)	(4.0–5.0)	(4.0–5.0)	(4.0–5.0)
	Min.–max.	4.0–5.0	1.0–6.0	2.0–6.0
Initial management phase quality-of-care rating	Mean	4.5	4.88	4.3
	(SD)	(0.6)	(0.81)	(1.1)
	Median	5.00	5.00	5.0
	(IQR)	(4.00–5.00)	(4.00–5.00)	(3.3–5.0)
	Min.–max.	3.0–5.0	3.0–6.0	2.0–6.0
Pre-discharge phase quality-of-care rating	Mean	4.5	5.0	3.8
	(SD)	(0.6)	(0.7)	(1.1)
	Median	5.0	5.0	4.0
	(IQR)	(4.0–5.0)	(5.0–5.0)	(3.0–5.0)
	Min.–max.	3.0–5.0	3.0–6.0	1.0–5.0
Overall phase quality-of- care rating	Mean	4.4	4.8	4.0
	(SD)	(0.5)	(0.9)	(0.9)
	Median	4.0	5.0	4.0
	(IQR)	(4.0–5.0)	(4.5–5.0)	(3.0–5.0)
	Min.–max.	4.0–5.0	1.0–6.0	2.0–5.0

IQR, interquartile range.

TABLE 37 Inter-rater reliability of holistic overall quality-of-care ratings

Reviewers	ICC	Significance
Doctor 1 vs doctor 2	0.328	0.022
Doctor 1 vs clinical audit 1	–0.285	0.959
Doctor 2 vs clinical audit 1	0.047	0.389
All staff (doctor 1, doctor 2, clinical audit 1)	0.077	0.210

low missing data rates for all reviewers for the criterion-based data collection.

Criterion-based quality-of-care scores

Quality-of-care scores were assigned to the criterion in the admission, initial management and pre-discharge phases. The method used was similar to that in the phase one study, whereby each time a criterion was ‘met’ or ‘done’ or the reviewer selected ‘not done’ for a valid reason option, a score of 1 was given. If a reviewer selected ‘not recorded’, the review item did not receive a score

as this option presumes the care was not provided. The mean and median quality-of-care scores are presented in *Table 39*.

Box-and-whisker plots of criterion-based total quality-of-care scores (Figure 34)

The box-and-whisker plots compare the median criterion-based quality-of-care scores and the interquartile range, for each reviewer for each phase of care.

TABLE 38 Criterion-based data completeness rates

	Clinical reviewer (nurse)	Doctor 1 reviewer	Doctor 2 reviewer	Total
Total number of reviews	37	37	37	111
Total number of data items available (sum of admission, initial management and pre-discharge phases)	1406	1406	1406	4218
Total number of data items missing	0	19	7	26
Percentage of data missing from each reviewer	0	0.01	0.004	0.006

TABLE 39 Criterion-based quality-of-care scores

		Clinical reviewer (nurse)	Doctor 1 reviewer	Doctor 2 reviewer
Phase of care				
Admission phase (out of 11)	Mean	9.2	10.5	7.8
	(SD)	(1.2)	(0.6)	(2.3)
	Median	9.0	11.0	8.0
	(IQR)	(9.0–10.0)	(10.0–11.0)	(6.0–10.0)
	Min.–max.	6.0–11.0	9.0–11.0	3.0–11.0
Initial management phase (out of 6)	Mean	5.2	5.8	4.6
	(SD)	(0.9)	(0.4)	(1.2)
	Median	5.0	6.0	5.0
	(IQR)	(5.0–6.0)	(6.0–6.0)	(4.0–5.0)
	Min.–max.	2.0–6.0	5.0–6.0	1.0–6.0
Pre-discharge phase (out of 16)	Mean	13.5	15.2	10.1
	(SD)	(1.3)	(1.9)	(2.8)
	Median	13.0	16.0	10.0
	(IQR)	(13.0–14.5)	(15.0–16.0)	(8.0–12.0)
	Min.–max.	11.0–16.0	5.0–16.0	4.0–16.0
Total score (sum of all phases, max. 33)	Mean	28.0	31.2	22.6
	(SD)	(2.3)	(2.0)	(4.5)
	Median	29.0	32.0	22.6
	(IQR)	(27.0–29.0)	(31.0–32.0)	(20.8–25.9)
	Min.–max.	22.0–33.0	22.0–33.0	11.4–33.0
IQR, interquartile range.				

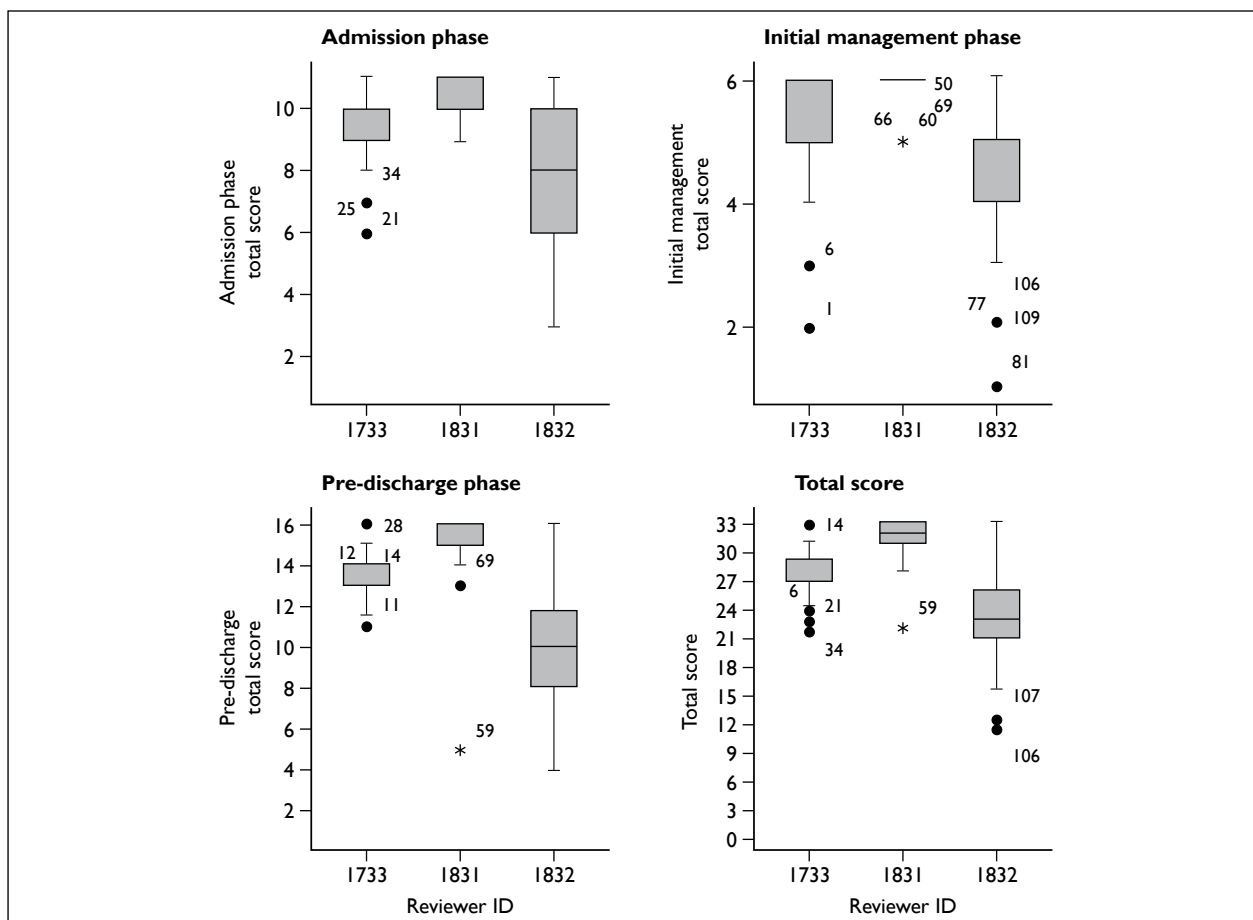


FIGURE 34 Criterion-based quality-of-care scores.

Reliability between reviewers

Table 40 presents the results of the stroke criterion-based review reliability analysis. The quality-of-care scores assigned to the criterion-based data was used to calculate ICCs for each reviewer pair.

Comparison with results from the phase one study

As with the holistic data, the amount of comparison that can be undertaken between the stroke reviewer reliability analysis and the main study reliability analysis is limited, due to the small number of stroke reviewers taking part.

On the whole, the reliability results for the stroke data are much lower than those of the phase one study, for which the reliability results were 0.88 (range 0.64–0.96) for the pairs of doctor reviewers (Table 9). The stroke review reliability for the doctor pairs is much less at only 0.031, suggesting that the reviewers are not completing the data collection form in the same way. However, reliability results for doctors in the phase one study did vary quite widely between individual pairs.

Time taken to complete reviews

Data presented in Table 41 show that each stroke care record review took approximately 1 hour.

The length of hospital stay for patients with stroke tends to be long, and the records associated with the care are large. From our sample, the mean length of stay for each patient was 33.7 days, but this ranged from 2 days to 314 days. Also, the method used was a joint holistic and criterion-based method, with scale scores, textual data and review of criteria, so the length of time taken to perform each review is probably not unreasonable.

Quality of records

The quality of the case notes reviewed for this study was rated on a rating scale (1 = poor, 6 = excellent). Case notes received similar quality ratings from the clinical audit and doctor 1 reviewers, whereas doctor 2 tended to assign a lower-quality rating than the other two reviewers (Table 42).

In the phase one study, the mean quality-of-case-notes ratings for COPD and heart failure were 4.3

TABLE 40 Inter-rater reliability between criterion-based review quality-of-care scores

Reviewers	ICC	Significance
Doctor 1 vs doctor 2	0.031	0.426
Doctor 1 vs clinical audit 1	0.126	0.226
Doctor 2 vs clinical audit 1	0.384	0.009
All staff (doctor 1, doctor 2, clinical audit 1)	0.199	0.022

TABLE 41 Summary statistics for time taken to review records (minutes)

	Clinical reviewer (nurse)	Doctor 1 reviewer	Doctor 2 reviewer
Mean	53.78	70.14	66.57
Median	50.00	70.00	60.00
SD	12.985	12.047	26.031
Minimum	40	45	30
Maximum	110	90	150

TABLE 42 Quality of record ratings

Quality of records rating	Clinical reviewer (nurse)	Doctor 1 reviewer	Doctor 2 reviewer
Mean	4.38	4.70	3.33
(SD)	(0.5)	(0.661)	(0.717)
Median	4.0	5.00	3.00
(IQR)	(4.0–5.0)	(4.0–5.0)	(3.0–4.0)
Min.–max.	3–5	3–6	1–4

IQR, interquartile range.

(SD 1.2) and 4.7 (SD 0.9), respectively. The stroke-care case notes received similar ratings.

Conclusions

The size of this adjunct study was limited by resources and, subsequently, by access, thus reducing its generalisability. For the criterion-based component of the review there is some indication that the reviewers were able to capture a more complete data set than the 39 reviewers in the main study. This may have been due to the quality of the recording, and perhaps the structured nature of the case notes, although it is also possible that the reviewers were more skilled at the task than were the main study reviewers.

The inter-rater reliability results were poor for the holistic reviews, more so than the main study, although even the main study showed that there were considerable differences between reviewers in their inter- and intra-rater reliability. There may be a number of reasons for this poor level of agreement, including the general level of difficulty of providing holistic reviews of case notes belonging to patients who had prolonged hospital stays. Under such conditions, it may be that holistic reviewing requires a very high level of training and experience to be able to identify variations in care from the mass of available information, perhaps supported by electronic means of screening, such as might be possible by using trigger methodologies (Appendix 14) based on a condition-specific set of review criteria.

References

1. Intercollegiate Stroke Audit Working Party. *National Sentinel Stroke Audit*. London: Royal College of Physicians of London; 2007.
2. Mohammed MA, Mant J, Bentham L, Stevens A, Hussain S. Process and mortality of stroke patients with and without do not resuscitate order in the West Midlands, UK. *Int J Qual Healthcare* 2006;**18**:102–6.
3. National Audit Office. URL: www.nao.org.uk/pn/05-06/0506452.htm
4. Department of Health. *National Service Framework for Older People*. London: Department of Health; 2001.
5. Mawson S, Gerrish K, Schofield J, Debbage S, Somers A. A pragmatic governance framework for differentiating between research, audit and service review activities. *Clin Manag* 2007;**15**:29–35.

Appendix 14

The place of trigger tool methodology in case note review for quality and safety

Context of the analysis

The initial study proposal indicated that it would be valuable to undertake an exploration of electronic trigger tool methods for assessing safety and quality in the two study conditions, acute exacerbations of COPD and heart failure, in contrasting trigger tools with paper-based holistic and criterion review.

Because this trigger tool study required research ethics approval it could only practically be undertaken in a hospital local to the study team, given the extent of the resource commitments of the main studies. It proved to be the case that none of the local hospitals had sufficient electronic records systems to support even a small study. In the most likely setting for the research there were a total of seven separate paper-based record systems. The research commissioners therefore agreed that a short review of trigger tool methods in the context of paper-based records would be an appropriate alternative.

Trigger tool methods

Health-care trigger tools were first described by Classen and colleagues¹ as an electronic screening tool for identifying markers or 'sentinels'² for possible adverse drug events (ADEs). Since this original prototype was developed in the context of a hospital with an electronic record system, it was possible to develop data-searching techniques that scanned for a drug, test or procedure that is usually associated with the management of an ADE. If a marker was found then a full review of a medical record could be undertaken to determine whether there had indeed been an avoidable event.

It should be noted here, however, that the term 'ADE' includes both preventable adverse events and adverse drug reactions that may be unforeseeable, even under circumstances of the very best care. So the sensitivity of the original electronic trigger tool system for identifying safety events was limited by the choice of tracer – in this case an ADE.

Resar *et al.*² point out that this initial trigger tool system had the benefit of much reducing the staff time that might otherwise be needed to screen all case notes (e.g. under circumstances when routine screening is identified by an organisation as a priority or in a search for ADEs). Routine automated screening could also be carried out within a very short time of an event, conceivably within a short enough timeframe for the patient still to be under active management of an event.

The initial trigger tool concept applied to electronic records and therefore posed a potentially cost-effective and timely method for screening large numbers of case notes, of a range of patients and conditions. For example, Jha *et al.*³ reviewed care over 21,964 patient-days and compared the results of voluntary reporting of ADEs with chart review (398 ADEs) and computer monitoring using a trigger tool [2620 alerts, of which 275 (10.5%) were ADEs]. Little commentary was made on the low specificity of the computerised alert screening method.

Rozich *et al.*⁴ subsequently used Classen's¹ original ideas to develop and test a more broadly based ADE trigger tool of 24 criteria, which could be used across a wide range of hospital types, from community to tertiary hospitals.

Trigger tools are essentially composed of a set of review criteria that are designed to identify possible active incidents or errors, such as the patient being given the wrong medication or a failure by the clinical team to respond to deteriorating vital signs. In this sense, although the evidence base on which they are built may be different, trigger tools are similar to evidence-based review criteria derived from clinical practice guidelines. Review criteria that are guideline based are more usually focused on positive acts than 'incident-based' trigger tools. For instance, the lack of a guideline-based review criterion in the case notes may indicate that an event has not happened – possibly linked to a 'failure-to-provide' event. For example, the failure

to record that a measure of glucose level has been taken in a person with diabetes might indicate a more general failure to actively manage the case.

Initially, trigger tools were used either in near 'real time' or, retrospectively, from paper-based or electronic records, to identify possible preventable events in medication safety. Trigger technology was subsequently broadened to identify possible harm in services such as paediatric intensive care,⁵ where situation-specific trigger criteria have been developed to screen for some of the more frequently occurring, preventable, safety events. In the initial validation study of this tool, the most commonly identified adverse events were health-care-associated infections, catheter infiltrates and unplanned extubations requiring reintubation. Trigger tools for adult intensive care have included such criteria as abrupt falls in haemoglobin level – indicative of severe bleeding – or the occurrence of a case of pneumonia in a person who is already a patient in a hospital.⁶

Recent developments have taken a different approach in looking globally for adverse events across whole hospital inpatient systems,⁷ and providing a measure for comparing one hospital's results with another, using denominators such as:

- adverse events per 1000 patient-days, or
- adverse events per 100 admissions, or
- percentage of admissions with an adverse event.

More controversially, the Institute of Healthcare Improvement (IHI)⁸ has developed an outpatient care trigger tool that 'bands together multiple episodes of care across a continuum',⁸ using triggers data from malpractice claims to categorise outpatient-care-related adverse events. The tool comprises 11 criteria 'to provide "clues" to the possibility of adverse events in a patient record'.⁸

Strengths and weaknesses of trigger tool methods

Although the value of using the electronic screening trigger programs has been recognised, it is also apparent that many hospital record systems are still paper based, so that more recent versions of trigger tools have been directed towards supporting the screening of case notes by trained reviewers, using what is effectively a criterion-based, explicit approach. Just like review criteria, trigger tools bring structure to a review, being used as a framework for screening case notes or

electronic records and identifying pointers to potential adverse events, which are then explored through full, holistic case note review.

Although these more broadly based service reviews are now being more widely promoted by the IHI and through projects in the UK NHS,⁸ rather less is currently being said about the limitations of the method. These limitations can be seen as:

1. development effort required for the criteria for the trigger tool
2. requirement for case note review when trigger criteria are found
3. validity and interpretation of the results.

Development resource

Extensive effort is required to develop a set of review criteria that have some evidence base, face validity and reproducibility. Although Resar and colleagues² do not indicate how much effort was required to develop the four trigger programs they outline in their 2003 review article, they indicate in another article that many person-hours were required to develop the IHI adverse drug event tool.⁷

To create trigger tools for (types of) adverse events, it would be necessary first to create an initial list of possible adverse events for a clinical condition or care setting – this has been the approach in the IHI global adverse events tool.⁷ Once an initial list of trigger criteria is developed then validation and reproducibility testing adds a further burden. For instance, development of the IHI medication tool⁴ was undertaken in 86 hospitals and was based on a review of 2837 records. This is a highly resource-intensive process and it is unlikely that this level of funding will be available often, particularly in the UK, so that trigger tools are likely to be limited in number, and, for the foreseeable future, essentially of North American origin.

Use of case note review

During a screening review using trigger tools, a positive finding of any one criterion requires that a full case note review must be undertaken. Resar and colleagues² point out that 'the reviewer must review the use of the trigger in the context of the care documented'. For example, in a medication review an event that appears to be an ADE may be an adverse drug reaction (unpredictable and probably not preventable) rather than an adverse event.

One of the key limitations of trigger technology is that any adverse events not identified by a trigger would be missed⁹ unless a general screen was carried out, which defeats the efficiency purpose. Moreover, it could be argued that if it takes around 15–20 minutes⁴ to manually scan to identify one in a set of adverse review criteria, this is about the same length of time that an experienced reviewer might take to undertake a structured implicit review. The IHI trigger tool study⁴ used full retrospective case note review with 23 (together with one open) review criteria. Instead of being a screening tool, therefore, the trigger criteria could be seen part of a mixed explicit/implicit case note review methodology.

Validity and interpretation of the results

When trigger tools are developed using rigorous methods and with extensive validation, there is undoubtedly a role for such review methods when they are used to review sets of case notes within an institution, and, in combination with full implicit review, they can also be used to explore safety and quality between institutions. However, Brown and colleagues¹⁰ point out a number of methodological limitations when trigger results alone are used as screening methods. Under such circumstances, the arguments comprise concerns over both sensitivity and specificity.

The authors identify the problem of a lack of a 'gold standard' for identifying the actual level of events occurring (even observation is not an accurate measure), so that the sensitivity of a specific trigger tool – for example for measuring the rate of ADEs in a particular population – may be higher than reported but is not as high as some other (possibly more expensive) methods, such as full holistic review. Trigger tools can, of course, be used denominator free or used with a denominator such as 100 bed-days, just the same as in criterion-based clinical audit.

If the specificity of the criteria in a trigger tool is high then only a narrow range of events may be identified. Conversely, Brown and colleagues¹⁰ point out that if specificity is low then there will be many false-positives and resource inefficient review. Use of tools with low levels of specificity might yield biased data in comparisons between organisations.

Is there a role for trigger tools in case note review of care for conditions such as COPD and heart failure?

The application of trigger tools to care for chronic conditions is certainly possible, in somewhat similar manner to the production of evidence-based review criteria. In the evaluation of safety, trigger tools have some advantage in that they can be developed to directly identify possible poor care, unlike the more usual review criteria that usually indicate possible gaps in care. However, the extent of the development work required to identify the range of indicators needed to trace possible flaws in care might be even greater than that needed for evidence-based review criteria to evaluate the process of care delivery.

It is likely, therefore, that only limited trigger tool sets of criteria will be available in the foreseeable future and that these will be more likely to be applied to specific instances, such as medication events, or to complex and event prone settings, such as intensive care units.

Nevertheless, where trigger tool criteria do exist it may be worth exploring their use as electronic medical records become commonplace in hospitals. While methodological limitations will remain, and care will be required in interpreting data from trigger tools that are used to provide 'comparable' data, research should be undertaken on the utility of a combined method of trigger tool screening with structured holistic review of identified records.

References

1. Classen D, Pestotnik SL, Evans RS, Burke JP. Computerised surveillance of adverse drug events in hospital patients. *JAMA* 1991;**266**:2847–51.
2. Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Health Care* 2003;**12**: ii39-ii45.
3. Jha AK, Kuperman GJ, Teich JM, Leape LL, Shea B, Rittenberg MA, *et al.* Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *JAMIA* 1998;**5**:305–14.
4. Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. *Qual Saf Health Care* 2003;**12**:194–200.

5. Sharek PJ, Horbar JD, Mason W, Bisarya H, Thurm CW, Suresh G, *et al.* Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. *Pediatrics* 2006;**118**:1332–40.
6. Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. *Joint Comm J Qual Improv* 2006;**32**:585–90.
7. Griffin FA, Resar RK. *IHI global trigger tool for measuring adverse events*. IHI Innovation Series white paper. Cambridge, MA: Institute of Healthcare Improvement; 2007.
8. Institute of Healthcare Improvement. URL: www.IHI.org (accessed 13 March 2008).
9. Dean B. Adverse drug events: what's the truth? *Qual Saf Health Care* 2003;**12**:165–6.
10. Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, *et al.* *An epistemology of patient safety research: a framework for study design and interpretation*. Birmingham: MRC Patient Safety Methodology Network; 2007.



Health Technology Assessment reports published to date

Volume 1, 1997

No. 1

Home parenteral nutrition: a systematic review.

By Richards DM, Deeks JJ, Sheldon TA, Shaffer JL.

No. 2

Diagnosis, management and screening of early localised prostate cancer.

A review by Selley S, Donovan J, Faulkner A, Coast J, Gillatt D.

No. 3

The diagnosis, management, treatment and costs of prostate cancer in England and Wales.

A review by Chamberlain J, Melia J, Moss S, Brown J.

No. 4

Screening for fragile X syndrome.

A review by Murray J, Cuckle H, Taylor G, Hewison J.

No. 5

A review of near patient testing in primary care.

By Hobbs FDR, Delaney BC, Fitzmaurice DA, Wilson S, Hyde CJ, Thorpe GH, *et al.*

No. 6

Systematic review of outpatient services for chronic pain control.

By McQuay HJ, Moore RA, Eccleston C, Morley S, de C Williams AC.

No. 7

Neonatal screening for inborn errors of metabolism: cost, yield and outcome.

A review by Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, *et al.*

No. 8

Preschool vision screening.

A review by Snowdon SK, Stewart-Brown SL.

No. 9

Implications of socio-cultural contexts for the ethics of clinical trials.

A review by Ashcroft RE, Chadwick DW, Clark SRL, Edwards RHT, Frith L, Hutton JL.

No. 10

A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment.

By Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S.

No. 11

Newborn screening for inborn errors of metabolism: a systematic review.

By Seymour CA, Thomason MJ, Chalmers RA, Addison GM, Bain MD, Cockburn F, *et al.*

No. 12

Routine preoperative testing: a systematic review of the evidence.

By Munro J, Booth A, Nicholl J.

No. 13

Systematic review of the effectiveness of laxatives in the elderly.

By Petticrew M, Watt I, Sheldon T.

No. 14

When and how to assess fast-changing technologies: a comparative study of medical applications of four generic technologies.

A review by Mowatt G, Bower DJ, Brebner JA, Cairns JA, Grant AM, McKee L.

Volume 2, 1998

No. 1

Antenatal screening for Down's syndrome.

A review by Wald NJ, Kennard A, Hackshaw A, McGuire A.

No. 2

Screening for ovarian cancer: a systematic review.

By Bell R, Petticrew M, Luengo S, Sheldon TA.

No. 3

Consensus development methods, and their use in clinical guideline development.

A review by Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, *et al.*

No. 4

A cost-utility analysis of interferon beta for multiple sclerosis.

By Parkin D, McNamee P, Jacoby A, Miller P, Thomas S, Bates D.

No. 5

Effectiveness and efficiency of methods of dialysis therapy for end-stage renal disease: systematic reviews.

By MacLeod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, *et al.*

No. 6

Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model.

By Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G.

No. 7

Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomised controlled trials.

By Song F, Glenny AM.

No. 8

Bone marrow and peripheral blood stem cell transplantation for malignancy.

A review by Johnson PWM, Simnett SJ, Sweetenham JW, Morgan GJ, Stewart LA.

No. 9

Screening for speech and language delay: a systematic review of the literature.

By Law J, Boyle J, Harris F, Harkness A, Nye C.

No. 10

Resource allocation for chronic stable angina: a systematic review of effectiveness, costs and cost-effectiveness of alternative interventions.

By Sculpher MJ, Petticrew M, Kelland JL, Elliott RA, Holdright DR, Buxton MJ.

No. 11

Detection, adherence and control of hypertension for the prevention of stroke: a systematic review.

By Ebrahim S.

No. 12

Postoperative analgesia and vomiting, with special reference to day-case surgery: a systematic review.

By McQuay HJ, Moore RA.

No. 13

Choosing between randomised and nonrandomised studies: a systematic review.

By Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C.

No. 14

Evaluating patient-based outcome measures for use in clinical trials.

A review by Fitzpatrick R, Davey C, Buxton MJ, Jones DR.

No. 15

Ethical issues in the design and conduct of randomised controlled trials.

A review by Edwards SJL, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J.

No. 16

Qualitative research methods in health technology assessment: a review of the literature.

By Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P.

No. 17

The costs and benefits of paramedic skills in pre-hospital trauma care.

By Nicholl J, Hughes S, Dixon S, Turner J, Yates D.

No. 18

Systematic review of endoscopic ultrasound in gastro-oesophageal cancer.

By Harris KM, Kelly S, Berry E, Hutton J, Roderick P, Cullingworth J, *et al.*

No. 19

Systematic reviews of trials and other studies.

By Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F.

No. 20

Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses.

A review by Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, *et al.*

Volume 3, 1999

No. 1

Informed decision making: an annotated bibliography and systematic review.

By Bekker H, Thornton JG, Airey CM, Connelly JB, Hewison J, Robinson MB, *et al.*

No. 2

Handling uncertainty when performing economic evaluation of healthcare interventions.

A review by Briggs AH, Gray AM.

No. 3

The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review.

By Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H.

No. 4

A randomised controlled trial of different approaches to universal antenatal HIV testing: uptake and acceptability. Annex: Antenatal HIV testing – assessment of a routine voluntary approach.

By Simpson WM, Johnstone FD, Boyd FM, Goldberg DJ, Hart GJ, Gormley SM, *et al.*

No. 5

Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review.

By Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ.

No. 6

Assessing the costs of healthcare technologies in clinical trials.

A review by Johnston K, Buxton MJ, Jones DR, Fitzpatrick R.

No. 7

Cooperatives and their primary care emergency centres: organisation and impact.

By Hallam L, Henthorne K.

No. 8

Screening for cystic fibrosis.

A review by Murray J, Cuckle H, Taylor G, Littlewood J, Hewison J.

No. 9

A review of the use of health status measures in economic evaluation.

By Brazier J, Deverill M, Green C, Harper R, Booth A.

No. 10

Methods for the analysis of quality-of-life and survival data in health technology assessment.

A review by Billingham LJ, Abrams KR, Jones DR.

No. 11

Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analysis.

By Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, Anionwu EN.

No. 12

Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses.

A review by Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al.*

No. 13

'Early warning systems' for identifying new healthcare technologies.

By Robert G, Stevens A, Gabbay J.

No. 14

A systematic review of the role of human papillomavirus testing within a cervical screening programme.

By Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A, *et al.*

No. 15

Near patient testing in diabetes clinics: appraising the costs and outcomes.

By Grieve R, Beech R, Vincent J, Mazurkiewicz J.

No. 16

Positron emission tomography: establishing priorities for health technology assessment.

A review by Robert G, Milne R.

No. 17 (Pt 1)

The debridement of chronic wounds: a systematic review.

By Bradley M, Cullum N, Sheldon T.

No. 17 (Pt 2)

Systematic reviews of wound care management: (2) Dressings and topical agents used in the healing of chronic wounds.

By Bradley M, Cullum N, Nelson EA, Petticrew M, Sheldon T, Torgerson D.

No. 18

A systematic literature review of spiral and electron beam computed tomography: with particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease.

By Berry E, Kelly S, Hutton J, Harris KM, Roderick P, Boyce JC, *et al.*

No. 19

What role for statins? A review and economic model.

By Ebrahim S, Davey Smith G, McCabe C, Payne N, Pickin M, Sheldon TA, *et al.*

No. 20

Factors that limit the quality, number and progress of randomised controlled trials.

A review by Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, *et al.*

No. 21

Antimicrobial prophylaxis in total hip replacement: a systematic review.

By Glenny AM, Song F.

No. 22

Health promoting schools and health promotion in schools: two systematic reviews.

By Lister-Sharp D, Chapman S, Stewart-Brown S, Sowden A.

No. 23

Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee.

A review by Lord J, Victor C, Littlejohns P, Ross FM, Axford JS.

Volume 4, 2000**No. 1**

The estimation of marginal time preference in a UK-wide sample (TEMPUS) project.

A review by Cairns JA, van der Pol MM.

No. 2

Geriatric rehabilitation following fractures in older people: a systematic review.

By Cameron I, Crotty M, Currie C, Finnegan T, Gillespie L, Gillespie W, *et al.*

No. 3

Screening for sickle cell disease and thalassaemia: a systematic review with supplementary research.

By Davies SC, Cronin E, Gill M, Greengross P, Hickman M, Normand C.

No. 4

Community provision of hearing aids and related audiology services.

A review by Reeves DJ, Alborz A, Hickson FS, Bamford JM.

No. 5

False-negative results in screening programmes: systematic review of impact and implications.

By Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K.

No. 6

Costs and benefits of community postnatal support workers: a randomised controlled trial.

By Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A.

No. 7

Implantable contraceptives (subdermal implants and hormonally impregnated intrauterine systems) versus other forms of reversible contraceptives: two systematic reviews to assess relative effectiveness, acceptability, tolerability and cost-effectiveness.

By French RS, Cowan FM, Mansour DJA, Morris S, Procter T, Hughes D, *et al.*

No. 8

An introduction to statistical methods for health technology assessment.

A review by White SJ, Ashby D, Brown PJ.

No. 9

Disease-modifying drugs for multiple sclerosis: a rapid and systematic review.

By Clegg A, Bryant J, Milne R.

No. 10

Publication and related biases.

A review by Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ.

No. 11

Cost and outcome implications of the organisation of vascular services.

By Michaels J, Brazier J, Palfreyman S, Shackley P, Slack R.

No. 12

Monitoring blood glucose control in diabetes mellitus: a systematic review.

By Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R.

No. 13

The effectiveness of domiciliary health visiting: a systematic review of international studies and a selective review of the British literature.

By Elkan R, Kendrick D, Hewitt M, Robinson JJA, Tolley K, Blair M, *et al.*

No. 14

The determinants of screening uptake and interventions for increasing uptake: a systematic review.

By Jepson R, Clegg A, Forbes C, Lewis R, Sowden A, Kleijnen J.

No. 15

The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth.

A rapid review by Song F, O'Meara S, Wilson P, Golder S, Kleijnen J.

No. 16

Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost-effectiveness and women's views.

By Bricker L, Garcia J, Henderson J, Mugford M, Neilson J, Roberts T, *et al.*

No. 17

A rapid and systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer.

By Lister-Sharp D, McDonagh MS, Khan KS, Kleijnen J.

No. 18

Liquid-based cytology in cervical screening: a rapid and systematic review.

By Payne N, Chilcott J, McGoogan E.

No. 19

Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care.

By King M, Sibbald B, Ward E, Bower P, Lloyd M, Gabbay M, *et al.*

No. 20

Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography?

By Kerry S, Hilton S, Patel S, Dundas D, Rink E, Lord J.

No. 21

Systematic reviews of wound care management: (3) antimicrobial agents for chronic wounds; (4) diabetic foot ulceration.

By O'Meara S, Cullum N, Majid M, Sheldon T.

No. 22

Using routine data to complement and enhance the results of randomised controlled trials.

By Lewsey JD, Leyland AH, Murray GD, Boddy FA.

No. 23

Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review.

By Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C.

No. 24

Outcome measures for adult critical care: a systematic review.

By Hayes JA, Black NA, Jenkinson C, Young JD, Rowan KM, Daly K, *et al.*

No. 25

A systematic review to evaluate the effectiveness of interventions to promote the initiation of breastfeeding.

By Fairbank L, O'Meara S, Renfrew MJ, Woolridge M, Sowden AJ, Lister-Sharp D.

No. 26

Implantable cardioverter defibrillators: arrhythmias. A rapid and systematic review.

By Parkes J, Bryant J, Milne R.

No. 27

Treatments for fatigue in multiple sclerosis: a rapid and systematic review.

By Brañas P, Jordan R, Fry-Smith A, Burls A, Hyde C.

No. 28

Early asthma prophylaxis, natural history, skeletal development and economy (EASE): a pilot randomised controlled trial.

By Baxter-Jones ADG, Helms PJ, Russell G, Grant A, Ross S, Cairns JA, *et al.*

No. 29

Screening for hypercholesterolaemia versus case finding for familial hypercholesterolaemia: a systematic review and cost-effectiveness analysis.

By Marks D, Wonderling D, Thorogood M, Lambert H, Humphries SE, Neil HAW.

No. 30

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists in the medical management of unstable angina.

By McDonagh MS, Bachmann LM, Golder S, Kleijnen J, ter Riet G.

No. 31

A randomised controlled trial of prehospital intravenous fluid replacement therapy in serious trauma.

By Turner J, Nicholl J, Webber L, Cox H, Dixon S, Yates D.

No. 32

Intrathecal pumps for giving opioids in chronic pain: a systematic review.

By Williams JE, Louw G, Towler G.

No. 33

Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review.

By Shepherd J, Waugh N, Hewitson P.

No. 34

A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies.

By MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS.

No. 35

Intravascular ultrasound-guided interventions in coronary artery disease: a systematic literature review, with decision-analytic modelling, of outcomes and cost-effectiveness.

By Berry E, Kelly S, Hutton J, Lindsay HSJ, Blaxill JM, Evans JA, *et al.*

No. 36

A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression.

By Simpson S, Corney R, Fitzgerald P, Beecham J.

No. 37

Systematic review of treatments for atopic eczema.

By Hoare C, Li Wan Po A, Williams H.

No. 38

Bayesian methods in health technology assessment: a review.

By Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR.

No. 39

The management of dyspepsia: a systematic review.

By Delaney B, Moayyedi P, Deeks J, Innes M, Soo S, Barton P, *et al.*

No. 40

A systematic review of treatments for severe psoriasis.

By Griffiths CEM, Clark CM, Chalmers RJG, Li Wan Po A, Williams HC.

Volume 5, 2001

No. 1

Clinical and cost-effectiveness of donepezil, rivastigmine and galantamine for Alzheimer's disease: a rapid and systematic review.

By Clegg A, Bryant J, Nicholson T, McIntyre L, De Broe S, Gerard K, *et al.*

No. 2

The clinical effectiveness and cost-effectiveness of riluzole for motor neurone disease: a rapid and systematic review.

By Stewart A, Sandercock J, Bryan S, Hyde C, Barton PM, Fry-Smith A, *et al.*

No. 3

Equity and the economic evaluation of healthcare.

By Sassi F, Archard L, Le Grand J.

No. 4

Quality-of-life measures in chronic diseases of childhood.

By Eiser C, Morse R.

No. 5

Eliciting public preferences for healthcare: a systematic review of techniques.

By Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al.*

No. 6

General health status measures for people with cognitive impairment: learning disability and acquired brain injury.

By Riemsma RP, Forbes CA, Glanville JM, Eastwood AJ, Kleijnen J.

No. 7

An assessment of screening strategies for fragile X syndrome in the UK.

By Pembrey ME, Barnicoat AJ, Carmichael B, Bobrow M, Turner G.

No. 8

Issues in methodological research: perspectives from researchers and commissioners.

By Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, *et al.*

No. 9

Systematic reviews of wound care management: (5) beds; (6) compression; (7) laser therapy, therapeutic ultrasound, electrotherapy and electromagnetic therapy.

By Cullum N, Nelson EA, Flemming K, Sheldon T.

No. 10

Effects of educational and psychosocial interventions for adolescents with diabetes mellitus: a systematic review.

By Hampson SE, Skinner TC, Hart J, Storey L, Gage H, Foxcroft D, *et al.*

No. 11

Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees: a rapid and systematic review.

By Jobanputra P, Parry D, Fry-Smith A, Burls A.

No. 12

Statistical assessment of the learning curves of health technologies.

By Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT.

No. 13

The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review.

By Dinnes J, Cave C, Huang S, Major K, Milne R.

No. 14

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of debriding agents in treating surgical wounds healing by secondary intention.

By Lewis R, Whiting P, ter Riet G, O'Meara S, Glanville J.

No. 15

Home treatment for mental health problems: a systematic review.

By Burns T, Knapp M, Catty J, Healey A, Henderson J, Watt H, *et al.*

No. 16

How to develop cost-conscious guidelines.

By Eccles M, Mason J.

No. 17

The role of specialist nurses in multiple sclerosis: a rapid and systematic review.

By De Broe S, Christopher F, Waugh N.

No. 18

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of orlistat in the management of obesity.

By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 19

The clinical effectiveness and cost-effectiveness of pioglitazone for type 2 diabetes mellitus: a rapid and systematic review.

By Chilcott J, Wight J, Lloyd Jones M, Tappenden P.

No. 20

Extended scope of nursing practice: a multicentre randomised controlled trial of appropriately trained nurses and preregistration house officers in preoperative assessment in elective general surgery.

By Kinley H, Czoski-Murray C, George S, McCabe C, Primrose J, Reilly C, *et al.*

No. 21

Systematic reviews of the effectiveness of day care for people with severe mental disorders: (1) Acute day hospital versus admission; (2) Vocational rehabilitation; (3) Day hospital versus outpatient care.

By Marshall M, Crowther R, Almaraz-Serrano A, Creed F, Sledge W, Kluiters H, *et al.*

No. 22

The measurement and monitoring of surgical adverse events.

By Bruce J, Russell EM, Mollison J, Krukowski ZH.

No. 23

Action research: a systematic review and guidance for assessment.

By Waterman H, Tillen D, Dickson R, de Koning K.

No. 24

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer.

By Ward S, Morris E, Bansback N, Calvert N, Crellin A, Forman D, *et al.*

No. 25

A rapid and systematic review of the evidence for the clinical effectiveness and cost-effectiveness of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer.

By Lloyd Jones M, Hummel S, Bansback N, Orr B, Seymour M.

No. 26

Comparison of the effectiveness of inhaler devices in asthma and chronic obstructive airways disease: a systematic review of the literature.

By Brocklebank D, Ram F, Wright J, Barry P, Cates C, Davies L, *et al.*

No. 27

The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint.

By Bryan S, Weatherburn G, Bungay H, Hatrick C, Salas C, Parry D, *et al.*

No. 28

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of topotecan for ovarian cancer.

By Forbes C, Shirran L, Bagnall A-M, Duffy S, ter Riet G.

No. 29

Superseded by a report published in a later volume.

No. 30

The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial.

By Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M.

No. 31

Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients.

By McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, *et al.*

No. 32

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer.

By Clegg A, Scott DA, Sidhu M, Hewitson P, Waugh N.

No. 33

Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives.

By Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G.

No. 34

Depot antipsychotic medication in the treatment of patients with schizophrenia: (1) Meta-review; (2) Patient and nurse attitudes.

By David AS, Adams C.

No. 35

A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression.

By Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, *et al.*

No. 36

Cost analysis of child health surveillance.

By Sanderson D, Wright D, Acton C, Duree D.

Volume 6, 2002**No. 1**

A study of the methods used to select review criteria for clinical audit.

By Hearnshaw H, Harker R, Cheater F, Baker R, Grimshaw G.

No. 2

Fludarabine as second-line therapy for B cell chronic lymphocytic leukaemia: a technology assessment.

By Hyde C, Wake B, Bryan S, Barton P, Fry-Smith A, Davenport C, *et al.*

No. 3

Rituximab as third-line treatment for refractory or recurrent Stage III or IV follicular non-Hodgkin's lymphoma: a systematic review and economic evaluation.

By Wake B, Hyde C, Bryan S, Barton P, Song F, Fry-Smith A, *et al.*

No. 4

A systematic review of discharge arrangements for older people.

By Parker SG, Peet SM, McPherson A, Cannaby AM, Baker R, Wilson A, *et al.*

No. 5

The clinical effectiveness and cost-effectiveness of inhaler devices used in the routine management of chronic asthma in older children: a systematic review and economic evaluation.

By Peters J, Stevenson M, Beverley C, Lim J, Smith S.

No. 6

The clinical effectiveness and cost-effectiveness of sibutramine in the management of obesity: a technology assessment.

By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 7

The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review.

By Berry E, Kelly S, Westwood ME, Davies LM, Gough MJ, Bamford JM, *et al.*

No. 8

Promoting physical activity in South Asian Muslim women through 'exercise on prescription'.

By Carroll B, Ali N, Azam N.

No. 9

Zanamivir for the treatment of influenza in adults: a systematic review and economic evaluation.

By Burls A, Clark W, Stewart T, Preston C, Bryan S, Jefferson T, *et al.*

No. 10

A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models.

By Richards RG, Sampson FC, Beard SM, Tappenden P.

No. 11

Screening for gestational diabetes: a systematic review and economic evaluation.

By Scott DA, Loveman E, McIntyre L, Waugh N.

No. 12

The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: a systematic review and economic evaluation.

By Clegg AJ, Colquitt J, Sidhu MK, Royle P, Loveman E, Walker A.

No. 13

The clinical effectiveness of trastuzumab for breast cancer: a systematic review.

By Lewis R, Bagnall A-M, Forbes C, Shirran E, Duffy S, Kleijnen J, *et al.*

No. 14

The clinical effectiveness and cost-effectiveness of vinorelbine for breast cancer: a systematic review and economic evaluation.

By Lewis R, Bagnall A-M, King S, Woolcott N, Forbes C, Shirran L, *et al.*

No. 15

A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease.

By Vale L, Wyness L, McCormack K, McKenzie L, Brazzelli M, Stearns SC.

No. 16

The clinical effectiveness and cost-effectiveness of bupropion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation.

By Woolcott NF, Jones L, Forbes CA, Mather LC, Sowden AJ, Song FJ, *et al.*

No. 17

A systematic review of effectiveness and economic evaluation of new drug treatments for juvenile idiopathic arthritis: etanercept.

By Cummins C, Connock M, Fry-Smith A, Burls A.

No. 18

Clinical effectiveness and cost-effectiveness of growth hormone in children: a systematic review and economic evaluation.

By Bryant J, Cave C, Mihaylova B, Chase D, McIntyre L, Gerard K, *et al.*

No. 19

Clinical effectiveness and cost-effectiveness of growth hormone in adults in relation to impact on quality of life: a systematic review and economic evaluation.

By Bryant J, Loveman E, Chase D, Mihaylova B, Cave C, Gerard K, *et al.*

No. 20

Clinical medication review by a pharmacist of patients on repeat prescriptions in general practice: a randomised controlled trial.

By Zermansky AG, Petty DR, Raynor DK, Lowe CJ, Freemantle N, Vail A.

No. 21

The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation.

By Jobanputra P, Barton P, Bryan S, Burls A.

No. 22

A systematic review and economic evaluation of computerised cognitive behaviour therapy for depression and anxiety.

By Kaltenthaler E, Shackley P, Stevens K, Beverley C, Parry G, Chilcott J.

No. 23

A systematic review and economic evaluation of pegylated liposomal doxorubicin hydrochloride for ovarian cancer.

By Forbes C, Wilby J, Richardson G, Sculpher M, Mather L, Reimsma R.

No. 24

A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change.

By Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, *et al.*

No. 25

A systematic review update of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists.

By Robinson M, Ginnelly L, Sculpher M, Jones L, Riemsma R, Palmer S, *et al.*

No. 26

A systematic review of the effectiveness, cost-effectiveness and barriers to implementation of thrombolytic and neuroprotective therapy for acute ischaemic stroke in the NHS.

By Sandercock P, Berge E, Dennis M, Forbes J, Hand P, Kwan J, *et al.*

No. 27

A randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care in a bronchiectasis clinic.

By Caine N, Sharples LD, Hollingworth W, French J, Keogan M, Exley A, *et al.*

No. 28

Clinical effectiveness and cost – consequences of selective serotonin reuptake inhibitors in the treatment of sex offenders.

By Adi Y, Ashcroft D, Browne K, Beech A, Fry-Smith A, Hyde C.

No. 29

Treatment of established osteoporosis: a systematic review and cost-utility analysis.

By Kanis JA, Brazier JE, Stevenson M, Calvert NW, Lloyd Jones M.

No. 30

Which anaesthetic agents are cost-effective in day surgery? Literature review, national survey of practice and randomised controlled trial.

By Elliott RA Payne K, Moore JK, Davies LM, Harper NJN, St Leger AS, *et al.*

No. 31

Screening for hepatitis C among injecting drug users and in genitourinary medicine clinics: systematic reviews of effectiveness, modelling study and national survey of current practice.

By Stein K, Dalziel K, Walker A, McIntyre L, Jenkins B, Horne J, *et al.*

No. 32

The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature.

By Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, *et al.*

No. 33

The effectiveness and cost-effectiveness of imatinib in chronic myeloid leukaemia: a systematic review.

By Garside R, Round A, Dalziel K, Stein K, Royle R.

No. 34

A comparative study of hypertonic saline, daily and alternate-day rhDNase in children with cystic fibrosis.

By Suri R, Wallis C, Bush A, Thompson S, Normand C, Flather M, *et al.*

No. 35

A systematic review of the costs and effectiveness of different models of paediatric home care.

By Parker G, Bhakta P, Lovett CA, Paisley S, Olsen R, Turner D, *et al.*

Volume 7, 2003

No. 1

How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study.

By Egger M, Jüni P, Bartlett C, Hohenstein F, Sterne J.

No. 2

Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of home versus hospital or satellite unit haemodialysis for people with end-stage renal failure.

By Mowatt G, Vale L, Perez J, Wyness L, Fraser C, MacLeod A, *et al.*

No. 3

Systematic review and economic evaluation of the effectiveness of infliximab for the treatment of Crohn's disease.

By Clark W, Raftery J, Barton P, Song F, Fry-Smith A, Burls A.

No. 4

A review of the clinical effectiveness and cost-effectiveness of routine anti-D prophylaxis for pregnant women who are rhesus negative.

By Chilcott J, Lloyd Jones M, Wight J, Forman K, Wray J, Beverley C, *et al.*

No. 5

Systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma.

By Riley RD, Burchill SA, Abrams KR, Heney D, Lambert PC, Jones DR, *et al.*

No. 6

The cost-effectiveness of screening for *Helicobacter pylori* to reduce mortality and morbidity from gastric cancer and peptic ulcer disease: a discrete-event simulation model.

By Roderick P, Davies R, Raftery J, Crabbe D, Pearce R, Bhandari P, *et al.*

No. 7

The clinical effectiveness and cost-effectiveness of routine dental checks: a systematic review and economic evaluation.

By Davenport C, Elley K, Salas C, Taylor-Weetman CL, Fry-Smith A, Bryan S, *et al.*

No. 8

A multicentre randomised controlled trial assessing the costs and benefits of using structured information and analysis of women's preferences in the management of menorrhagia.

By Kennedy ADM, Sculpher MJ, Coulter A, Dwyer N, Rees M, Horsley S, *et al.*

No. 9

Clinical effectiveness and cost-utility of photodynamic therapy for wet age-related macular degeneration: a systematic review and economic evaluation.

By Meads C, Salas C, Roberts T, Moore D, Fry-Smith A, Hyde C.

No. 10

Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities.

By Grimshaw GM, Szczepura A, Hultén M, MacDonald F, Nevin NC, Sutton F, *et al.*

No. 11

First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS).

By Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM.

No. 12

The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation.

By Calvert N, Hind D, McWilliams RG, Thomas SM, Beverley C, Davidson A.

No. 13

A systematic review of atypical antipsychotics in schizophrenia.

By Bagnall A-M, Jones L, Lewis R, Ginnelly L, Glanville J, Torgerson D, *et al.*

No. 14

Prostate Testing for Cancer and Treatment (ProtecT) feasibility study.

By Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, *et al.*

No. 15

Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation.

By Boland A, Dundar Y, Bagust A, Haycox A, Hill R, Mujica Mota R, *et al.*

No. 16

Screening for fragile X syndrome: a literature review and modelling.

By Song FJ, Barton P, Sleightholme V, Yao GL, Fry-Smith A.

No. 17

Systematic review of endoscopic sinus surgery for nasal polyps.

By Dalziel K, Stein K, Round A, Garside R, Royle P.

No. 18

Towards efficient guidelines: how to monitor guideline use in primary care.

By Hutchinson A, McIntosh A, Cox S, Gilbert C.

No. 19

Effectiveness and cost-effectiveness of acute hospital-based spinal cord injuries services: systematic review.

By Bagnall A-M, Jones L, Richardson G, Duffy S, Riemsma R.

No. 20

Prioritisation of health technology assessment. The PATHS model: methods and case studies.

By Townsend J, Buxton M, Harper G.

No. 21

Systematic review of the clinical effectiveness and cost-effectiveness of tension-free vaginal tape for treatment of urinary stress incontinence.

By Cody J, Wyness L, Wallace S, Glazener C, Kilonzo M, Stearns S, *et al.*

No. 22

The clinical and cost-effectiveness of patient education models for diabetes: a systematic review and economic evaluation.

By Loveman E, Cave C, Green C, Royle P, Dunn N, Waugh N.

No. 23

The role of modelling in prioritising and planning clinical trials.

By Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P.

No. 24

Cost-benefit evaluation of routine influenza immunisation in people 65-74 years of age.

By Allsup S, Gosney M, Haycox A, Regan M.

No. 25

The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heart-beating and non-heart-beating donors.

By Wight J, Chilcott J, Holmes M, Brewer N.

No. 26

Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment.

By Williams JG, Cheung WY, Cohen DR, Hutchings HA, Longo MF, Russell IT.

No. 27

Evaluating non-randomised intervention studies.

By Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, *et al.*

No. 28

A randomised controlled trial to assess the impact of a package comprising a patient-orientated, evidence-based self-help guidebook and patient-centred consultations on disease management and satisfaction in inflammatory bowel disease.

By Kennedy A, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, *et al.*

No. 29

The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review.

By Dinnes J, Loveman E, McIntyre L, Waugh N.

No. 30

The value of digital imaging in diabetic retinopathy.

By Sharp PF, Olson J, Strachan F, Hipwell J, Ludbrook A, O'Donnell M, *et al.*

No. 31

Lowering blood pressure to prevent myocardial infarction and stroke: a new preventive strategy.

By Law M, Wald N, Morris J.

No. 32

Clinical and cost-effectiveness of capecitabine and tegafur with uracil for the treatment of metastatic colorectal cancer: systematic review and economic evaluation.

By Ward S, Kaltenthaler E, Cowan J, Brewer N.

No. 33

Clinical and cost-effectiveness of new and emerging technologies for early localised prostate cancer: a systematic review.

By Hummel S, Paisley S, Morgan A, Currie E, Brewer N.

No. 34

Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system.

By Royle P, Waugh N.

No. 35

Systematic review and economic decision modelling for the prevention and treatment of influenza A and B.

By Turner D, Wailoo A, Nicholson K, Cooper N, Sutton A, Abrams K.

No. 36

A randomised controlled trial to evaluate the clinical and cost-effectiveness of Hickman line insertions in adult cancer patients by nurses.

By Boland A, Haycox A, Bagust A, Fitzsimmons L.

No. 37

Redesigning postnatal care: a randomised controlled trial of protocol-based midwifery-led care focused on individual women's physical and psychological health needs.

By MacArthur C, Winter HR, Bick DE, Lilford RJ, Lancashire RJ, Knowles H, *et al.*

No. 38

Estimating implied rates of discount in healthcare decision-making.

By West RR, McNabb R, Thompson AGH, Sheldon TA, Grimley Evans J.

No. 39

Systematic review of isolation policies in the hospital management of methicillin-resistant *Staphylococcus aureus*: a review of the literature with epidemiological and economic modelling.

By Cooper BS, Stone SP, Kibbler CC, Cookson BD, Roberts JA, Medley GF, *et al.*

No. 40

Treatments for spasticity and pain in multiple sclerosis: a systematic review.

By Beard S, Hunn A, Wight J.

No. 41

The inclusion of reports of randomised trials published in languages other than English in systematic reviews.

By Moher D, Pham B, Lawson ML, Klassen TP.

No. 42

The impact of screening on future health-promoting behaviours and health beliefs: a systematic review.

By Bankhead CR, Brett J, Bukach C, Webster P, Stewart-Brown S, Munafo M, *et al.*

Volume 8, 2004

No. 1

What is the best imaging strategy for acute stroke?

By Wardlaw JM, Keir SL, Seymour J, Lewis S, Sandercock PAG, Dennis MS, *et al.*

No. 2

Systematic review and modelling of the investigation of acute and chronic chest pain presenting in primary care.

By Mant J, McManus RJ, Oakes RAL, Delaney BC, Barton PM, Deeks JJ, *et al.*

No. 3

The effectiveness and cost-effectiveness of microwave and thermal balloon endometrial ablation for heavy menstrual bleeding: a systematic review and economic modelling.

By Garside R, Stein K, Wyatt K, Round A, Price A.

No. 4

A systematic review of the role of bisphosphonates in metastatic disease.

By Ross JR, Saunders Y, Edmonds PM, Patel S, Wonderling D, Normand C, *et al.*

No. 5

Systematic review of the clinical effectiveness and cost-effectiveness of capecitabine (Xeloda®) for locally advanced and/or metastatic breast cancer.

By Jones L, Hawkins N, Westwood M, Wright K, Richardson G, Riemsma R.

No. 6

Effectiveness and efficiency of guideline dissemination and implementation strategies.

By Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, *et al.*

No. 7

Clinical effectiveness and costs of the Sugarbaker procedure for the treatment of pseudomyxoma peritonei.

By Bryant J, Clegg AJ, Sidhu MK, Brodin H, Royle P, Davidson P.

No. 8

Psychological treatment for insomnia in the regulation of long-term hypnotic drug use.

By Morgan K, Dixon S, Mathers N, Thompson J, Tomeny M.

No. 9

Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome.

By Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ.

No. 10

A systematic review and economic evaluation of magnetic resonance cholangiopancreatography compared with diagnostic endoscopic retrograde cholangiopancreatography.

By Kaltenthaler E, Bravo Vergel Y, Chilcott J, Thomas S, Blakeborough T, Walters SJ, *et al.*

No. 11

The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis.

By Barton P, Jobanputra P, Wilson J, Bryan S, Burls A.

No. 12

Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review.

By Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S.

No. 13

Clinical effectiveness and cost-effectiveness of pioglitazone and rosiglitazone in the treatment of type 2 diabetes: a systematic review and economic evaluation.

By Czoski-Murray C, Warren E, Chilcott J, Beverley C, Psyllaki MA, Cowan J.

No. 14

Routine examination of the newborn: the EMREN study. Evaluation of an extension of the midwife role including a randomised controlled trial of appropriately trained midwives and paediatric senior house officers.

By Townsend J, Wolke D, Hayes J, Davé S, Rogers C, Bloomfield L, *et al.*

No. 15

Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach.

By Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, *et al.*

No. 16

A multi-centre randomised controlled trial of minimally invasive direct coronary bypass grafting versus percutaneous transluminal coronary angioplasty with stenting for proximal stenosis of the left anterior descending coronary artery.

By Reeves BC, Angelini GD, Bryan AJ, Taylor FC, Cripps T, Spyt TJ, *et al.*

No. 17

Does early magnetic resonance imaging influence management or improve outcome in patients referred to secondary care with low back pain? A pragmatic randomised controlled trial.

By Gilbert FJ, Grant AM, Gillan MGC, Vale L, Scott NW, Campbell MK, *et al.*

No. 18

The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic analysis.

By Clark W, Jobanputra P, Barton P, Burls A.

No. 19

A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder.

By Bridle C, Palmer S, Bagnall A-M, Darba J, Duffy S, Sculpher M, *et al.*

No. 20

Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis.

By Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N.

No. 21

Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement.

By Avenell A, Broom J, Brown TJ, Poobalan A, Aucott L, Stearns SC, *et al.*

No. 22

Autoantibody testing in children with newly diagnosed type 1 diabetes mellitus.

By Dretzke J, Cummins C, Sandercock J, Fry-Smith A, Barrett T, Burls A.

No. 23

Clinical effectiveness and cost-effectiveness of prehospital intravenous fluids in trauma patients.

By Dretzke J, Sandercock J, Bayliss S, Burls A.

No. 24

Newer hypnotic drugs for the short-term management of insomnia: a systematic review and economic evaluation.

By Dündar Y, Boland A, Strobl J, Dodd S, Haycox A, Bagust A, *et al.*

No. 25

Development and validation of methods for assessing the quality of diagnostic accuracy studies.

By Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J.

No. 26

EVALUATE hysterectomy trial: a multicentre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy.

By Garry R, Fountain J, Brown J, Manca A, Mason S, Sculpher M, *et al.*

No. 27

Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon- β and glatiramer acetate for multiple sclerosis.

By Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C.

No. 28

Effectiveness and cost-effectiveness of imatinib for first-line treatment of chronic myeloid leukaemia in chronic phase: a systematic review and economic analysis.

By Dalziel K, Round A, Stein K, Garside R, Price A.

No. 29

VenUS I: a randomised controlled trial of two types of bandage for treating venous leg ulcers.

By Iglesias C, Nelson EA, Cullum NA, Torgerson DJ, on behalf of the VenUS Team.

No. 30

Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction.

By Mowatt G, Vale L, Brazzelli M, Hernandez R, Murray A, Scott N, *et al.*

No. 31

A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme.

By Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S.

No. 32

The Social Support and Family Health Study: a randomised controlled trial and economic evaluation of two alternative forms of postnatal support for mothers living in disadvantaged inner-city areas.

By Wiggins M, Oakley A, Roberts I, Turner H, Rajan L, Austerberry H, *et al.*

No. 33

Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review.

By Green JM, Hewison J, Bekker HL, Bryant, Cuckle HS.

No. 34

Evaluation of abnormal uterine bleeding: comparison of three outpatient procedures within cohorts defined by age and menopausal status.

By Critchley HOD, Warner P, Lee AJ, Brechin S, Guise J, Graham B.

No. 35

Coronary artery stents: a rapid systematic review and economic evaluation.

By Hill R, Bagust A, Bakhai A, Dickson R, Dündar Y, Haycox A, *et al.*

No. 36

Review of guidelines for good practice in decision-analytic modelling in health technology assessment.

By Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al.*

No. 37

Rituximab (MabThera®) for aggressive non-Hodgkin's lymphoma: systematic review and economic evaluation.

By Knight C, Hind D, Brewer N, Abbott V.

No. 38

Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation.

By Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, *et al.*

No. 39

Pegylated interferon α -2a and -2b in combination with ribavirin in the treatment of chronic hepatitis C: a systematic review and economic evaluation.

By Shepherd J, Brodin H, Cave C, Waugh N, Price A, Gabbay J.

No. 40

Clopidogrel used in combination with aspirin compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation.

By Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher M, *et al.*

No. 41

Provision, uptake and cost of cardiac rehabilitation programmes: improving services to under-represented groups.

By Beswick AD, Rees K, Griebisch I, Taylor FC, Burke M, West RR, *et al.*

No. 42

Involving South Asian patients in clinical trials.

By Hussain-Gambles M, Leese B, Atkin K, Brown J, Mason S, Tovey P.

No. 43

Clinical and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes.

By Colquitt JL, Green C, Sidhu MK, Hartwell D, Waugh N.

No. 44

Identification and assessment of ongoing trials in health technology assessment reviews.

By Song FJ, Fry-Smith A, Davenport C, Bayliss S, Adi Y, Wilson JS, *et al.*

No. 45

Systematic review and economic evaluation of a long-acting insulin analogue, insulin glargine

By Warren E, Weatherley-Jones E, Chilcott J, Beverley C.

No. 46

Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis.

By McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR, *et al.*

No. 47

Clinical and cost-effectiveness of once-daily versus more frequent use of same potency topical corticosteroids for atopic eczema: a systematic review and economic evaluation.

By Green C, Colquitt JL, Kirby J, Davidson P, Payne E.

No. 48

Acupuncture of chronic headache disorders in primary care: randomised controlled trial and economic analysis.

By Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, *et al.*

No. 49

Generalisability in economic evaluation studies in healthcare: a review and case studies.

By Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al.*

No. 50

Virtual outreach: a randomised controlled trial and economic evaluation of joint teleconferenced medical consultations.

By Wallace P, Barber J, Clayton W, Currell R, Fleming K, Garner P, *et al.*

Volume 9, 2005

No. 1

Randomised controlled multiple treatment comparison to provide a cost-effectiveness rationale for the selection of antimicrobial therapy in acne.

By Ozolins M, Eady EA, Avery A, Cunliffe WJ, O'Neill C, Simpson NB, *et al.*

No. 2

Do the findings of case series studies vary significantly according to methodological characteristics?

By Dalziel K, Round A, Stein K, Garside R, Castelnovo E, Payne L.

No. 3

Improving the referral process for familial breast cancer genetic counselling: findings of three randomised controlled trials of two interventions.

By Wilson BJ, Torrance N, Mollison J, Wordsworth S, Gray JR, Haites NE, *et al.*

No. 4

Randomised evaluation of alternative electrosurgical modalities to treat bladder outflow obstruction in men with benign prostatic hyperplasia.

By Fowler C, McAllister W, Plail R, Karim O, Yang Q.

No. 5

A pragmatic randomised controlled trial of the cost-effectiveness of palliative therapies for patients with inoperable oesophageal cancer.

By Shenfine J, McNamee P, Steen N, Bond J, Griffin SM.

No. 6

Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography.

By Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H.

No. 7

Issues in data monitoring and interim analysis of trials.

By Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, *et al.*

No. 8

Lay public's understanding of equipoise and randomisation in randomised controlled trials.

By Robinson EJ, Kerr CEP, Stevens AJ, Lilford RJ, Braunholtz DA, Edwards SJ, *et al.*

No. 9

Clinical and cost-effectiveness of electroconvulsive therapy for depressive illness, schizophrenia, catatonia and mania: systematic reviews and economic modelling studies.

By Greenhalgh J, Knight C, Hind D, Beverley C, Walters S.

No. 10

Measurement of health-related quality of life for people with dementia: development of a new instrument (DEM-QOL) and an evaluation of current methodology.

By Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P, *et al.*

No. 11

Clinical effectiveness and cost-effectiveness of drotrecogin alfa (activated) (Xigris®) for the treatment of severe sepsis in adults: a systematic review and economic evaluation.

By Green C, Dinnes J, Takeda A, Shepherd J, Hartwell D, Cave C, *et al.*

No. 12

A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy.

By Dinnes J, Deeks J, Kirby J, Roderick P.

No. 13

Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK.

By Willis BH, Barton P, Pearmain P, Bryan S, Hyde C.

No. 14

Laparoscopic surgery for inguinal hernia repair: systematic review of effectiveness and economic evaluation.

By McCormack K, Wake B, Perez J, Fraser C, Cook J, McIntosh E, *et al.*

No. 15

Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation.

By Wilby J, Kainth A, Hawkins N, Epstein D, McIntosh H, McDaid C, *et al.*

No. 16

A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine.

By Peveler R, Kendrick T, Buxton M, Longworth L, Baldwin D, Moore M, *et al.*

No. 17

Clinical effectiveness and cost-effectiveness of immediate angioplasty for acute myocardial infarction: systematic review and economic evaluation.

By Hartwell D, Colquitt J, Loveman E, Clegg AJ, Brodin H, Waugh N, *et al.*

No. 18

A randomised controlled comparison of alternative strategies in stroke care.

By Kalra L, Evans A, Perez I, Knapp M, Swift C, Donaldson N.

No. 19

The investigation and analysis of critical incidents and adverse events in healthcare.

By Woloshynowych M, Rogers S, Taylor-Adams S, Vincent C.

No. 20

Potential use of routine databases in health technology assessment.

By Raftery J, Roderick P, Stevens A.

No. 21

Clinical and cost-effectiveness of newer immunosuppressive regimens in renal transplantation: a systematic review and modelling study.

By Woodroffe R, Yao GL, Meads C, Bayliss S, Ready A, Raftery J, *et al.*

No. 22

A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis.

By Stevenson M, Lloyd Jones M, De Nigris E, Brewer N, Davis S, Oakley J.

No. 23

A systematic review to examine the impact of psycho-educational interventions on health outcomes and costs in adults and children with difficult asthma.

By Smith JR, Muggford M, Holland R, Candy B, Noble MJ, Harrison BDW, *et al.*

No. 24

An evaluation of the costs, effectiveness and quality of renal replacement therapy provision in renal satellite units in England and Wales.

By Roderick P, Nicholson T, Armitage A, Mehta R, Mullee M, Gerard K, *et al.*

No. 25

Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation.

By Wilson J, Connock M, Song F, Yao G, Fry-Smith A, Raftery J, *et al.*

No. 26

Indirect comparisons of competing interventions.

By Glenny AM, Altman DG, Song F, Sakarovich C, Deeks JJ, D'Amico R, *et al.*

No. 27

Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling.

By Robinson M, Palmer S, Sculpher M, Philips Z, Ginnelly L, Bowens A, *et al.*

No. 28

Outcomes of electrically stimulated gracilis neosphincter surgery.

By Tillin T, Chambers M, Feldman R.

No. 29

The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation.

By Garside R, Stein K, Castelnovo E, Pitt M, Ashcroft D, Dimmock P, *et al.*

No. 30

Systematic review on urine albumin testing for early detection of diabetic complications.

By Newman DJ, Mattock MB, Dawnay ABS, Kerry S, McGuire A, Yaqoob M, *et al.*

No. 31

Randomised controlled trial of the cost-effectiveness of water-based therapy for lower limb osteoarthritis.

By Cochrane T, Davey RC, Matthes Edwards SM.

No. 32

Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain.

By Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M, *et al.*

No. 33

Cost-effectiveness and safety of epidural steroids in the management of sciatica.

By Price C, Arden N, Cogan L, Rogers P.

No. 34

The British Rheumatoid Outcome Study Group (BROSG) randomised controlled trial to compare the effectiveness and cost-effectiveness of aggressive versus symptomatic therapy in established rheumatoid arthritis.

By Symmons D, Tricker K, Roberts C, Davies L, Dawes P, Scott DL.

No. 35

Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials.

By King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, *et al.*

No. 36

The clinical and cost-effectiveness of implantable cardioverter defibrillators: a systematic review.

By Bryant J, Brodin H, Loveman E, Payne E, Clegg A.

No. 37

A trial of problem-solving by community mental health nurses for anxiety, depression and life difficulties among general practice patients. The CPN-GP study.

By Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, *et al.*

No. 38

The causes and effects of socio-demographic exclusions from clinical trials.

By Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, *et al.*

No. 39

Is hydrotherapy cost-effective? A randomised controlled trial of combined hydrotherapy programmes compared with physiotherapy land techniques in children with juvenile idiopathic arthritis.

By Epps H, Ginnelly L, Utley M, Southwood T, Gallivan S, Sculpher M, *et al.*

No. 40

A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study.

By Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, *et al.*

No. 41

Displaced intracapsular hip fractures in fit, older people: a randomised comparison of reduction and fixation, bipolar hemiarthroplasty and total hip arthroplasty.

By Keating JF, Grant A, Masson M, Scott NW, Forbes JF.

No. 42

Long-term outcome of cognitive behaviour therapy clinical trials in central Scotland.

By Durham RC, Chambers JA, Power KG, Sharp DM, Macdonald RR, Major KA, *et al.*

No. 43

The effectiveness and cost-effectiveness of dual-chamber pacemakers compared with single-chamber pacemakers for bradycardia due to atrioventricular block or sick sinus syndrome: systematic review and economic evaluation.

By Castelnovo E, Stein K, Pitt M, Garside R, Payne E.

No. 44

Newborn screening for congenital heart defects: a systematic review and cost-effectiveness analysis.

By Knowles R, Griebisch I, Dezateux C, Brown J, Bull C, Wren C.

No. 45

The clinical and cost-effectiveness of left ventricular assist devices for end-stage heart failure: a systematic review and economic evaluation.

By Clegg AJ, Scott DA, Loveman E, Colquitt J, Hutchinson J, Royle P, *et al.*

No. 46

The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma.

By Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D.

No. 47

Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: systematic review and economic evaluation.

By Clar C, Cummins E, McIntyre L, Thomas S, Lamb J, Bain L, *et al.*

No. 48

Systematic review of effectiveness of different treatments for childhood retinoblastoma.

By McDaid C, Hartley S, Bagnall A-M, Ritchie G, Light K, Riemsma R.

No. 49

Towards evidence-based guidelines for the prevention of venous thromboembolism: systematic reviews of mechanical methods, oral anticoagulation, dextran and regional anaesthesia as thromboprophylaxis.

By Roderick P, Ferris G, Wilson K, Halls H, Jackson D, Collins R, *et al.*

No. 50

The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder, in children.

By Dretzke J, Frew E, Davenport C, Barlow J, Stewart-Brown S, Sandercock J, *et al.*

Volume 10, 2006

No. 1

The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease.

By Loveman E, Green C, Kirby J, Takeda A, Picot J, Payne E, *et al.*

No. 2

FOOD: a multicentre randomised trial evaluating feeding policies in patients admitted to hospital with a recent stroke.

By Dennis M, Lewis S, Cranswick G, Forbes J.

No. 3

The clinical effectiveness and cost-effectiveness of computed tomography screening for lung cancer: systematic reviews.

By Black C, Bagust A, Boland A, Walker S, McLeod C, De Verteuil R, *et al.*

No. 4

A systematic review of the effectiveness and cost-effectiveness of neuroimaging assessments used to visualise the seizure focus in people with refractory epilepsy being considered for surgery.

By Whiting P, Gupta R, Burch J, Mujica Mota RE, Wright K, Marson A, *et al.*

No. 5

Comparison of conference abstracts and presentations with full-text articles in the health technology assessments of rapidly evolving technologies.

By Dundar Y, Dodd S, Dickson R, Walley T, Haycox A, Williamson PR.

No. 6

Systematic review and evaluation of methods of assessing urinary incontinence.

By Martin JL, Williams KS, Abrams KR, Turner DA, Sutton AJ, Chapple C, *et al.*

No. 7

The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy. A systematic review.

By Connock M, Frew E, Evans B-W, Bryan S, Cummins C, Fry-Smith A, *et al.*

No. 8

Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling.

By Garside R, Pitt M, Somerville M, Stein K, Price A, Gilbert N.

No. 9

Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer: a systematic review and economic evaluation.

By Main C, Bojke L, Griffin S, Norman G, Barbieri M, Mather L, *et al.*

No. 10

Evaluation of molecular techniques in prediction and diagnosis of cytomegalovirus disease in immunocompromised patients.

By Szczepura A, Westmoreland D, Vinogradova Y, Fox J, Clark M.

No. 11

Screening for thrombophilia in high-risk situations: systematic review and cost-effectiveness analysis. The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) study.

By Wu O, Robertson L, Twaddle S, Lowe GDO, Clark P, Greaves M, *et al.*

No. 12

A series of systematic reviews to inform a decision analysis for sampling and treating infected diabetic foot ulcers.

By Nelson EA, O'Meara S, Craig D, Iglesias C, Golder S, Dalton J, *et al.*

No. 13

Randomised clinical trial, observational study and assessment of cost-effectiveness of the treatment of varicose veins (REACTIV trial).

By Michaels JA, Campbell WB, Brazier JE, MacIntyre JB, Palfreyman SJ, Ratcliffe J, *et al.*

No. 14

The cost-effectiveness of screening for oral cancer in primary care.

By Speight PM, Palmer S, Moles DR, Downer MC, Smith DH, Henriksson M, *et al.*

No. 15

Measurement of the clinical and cost-effectiveness of non-invasive diagnostic testing strategies for deep vein thrombosis.

By Goodacre S, Sampson F, Stevenson M, Wailoo A, Sutton A, Thomas S, *et al.*

No. 16

Systematic review of the effectiveness and cost-effectiveness of HealOzone® for the treatment of occlusal pit/fissure caries and root caries.

By Brazzelli M, McKenzie L, Fielding S, Fraser C, Clarkson J, Kilonzo M, *et al.*

No. 17

Randomised controlled trials of conventional antipsychotic versus new atypical drugs, and new atypical drugs versus clozapine, in people with schizophrenia responding poorly to, or intolerant of, current drug treatment.

By Lewis SW, Davies L, Jones PB, Barnes TRE, Murray RM, Kerwin R, *et al.*

No. 18

Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation.

By Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, *et al.*

No. 19

Cognitive behavioural therapy in addition to antispasmodic therapy for irritable bowel syndrome in primary care: randomised controlled trial.

By Kennedy TM, Chalder T, McCrone P, Darnley S, Knapp M, Jones RH, *et al.*

No. 20

A systematic review of the clinical effectiveness and cost-effectiveness of enzyme replacement therapies for Fabry's disease and mucopolysaccharidosis type I.

By Connock M, Juarez-Garcia A, Frew E, Mans A, Dretzke J, Fry-Smith A, *et al.*

No. 21

Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation.

By Wright M, Grieve R, Roberts J, Main J, Thomas HC, on behalf of the UK Mild Hepatitis C Trial Investigators.

No. 22

Pressure relieving support surfaces: a randomised evaluation.

By Nixon J, Nelson EA, Cranny G, Iglesias CP, Hawkins K, Cullum NA, *et al.*

No. 23

A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents.

By King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G, *et al.*

No. 24

The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review.

By Connock M, Burls A, Frew E, Fry-Smith A, Juarez-Garcia A, McCabe C, *et al.*

No. 25

Effectiveness and cost-effectiveness of salicylic acid and cryotherapy for cutaneous warts. An economic decision model.

By Thomas KS, Keogh-Brown MR, Chalmers JR, Fordham RJ, Holland RC, Armstrong SJ, *et al.*

No. 26

A systematic literature review of the effectiveness of non-pharmacological interventions to prevent wandering in dementia and evaluation of the ethical implications and acceptability of their use.

By Robinson L, Hutchings D, Corner L, Beyer F, Dickinson H, Vanoli A, *et al.*

No. 27

A review of the evidence on the effects and costs of implantable cardioverter defibrillator therapy in different patient groups, and modelling of cost-effectiveness and cost-utility for these groups in a UK context.

By Buxton M, Caine N, Chase D, Connelly D, Grace A, Jackson C, *et al.*

No. 28

Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: a systematic review and economic evaluation.

By Shepherd J, Jones J, Takeda A, Davidson P, Price A.

No. 29

An evaluation of the clinical and cost-effectiveness of pulmonary artery catheters in patient management in intensive care: a systematic review and a randomised controlled trial.

By Harvey S, Stevens K, Harrison D, Young D, Brampton W, McCabe C, *et al.*

No. 30

Accurate, practical and cost-effective assessment of carotid stenosis in the UK.

By Wardlaw JM, Chappell FM, Stevenson M, De Nigris E, Thomas S, Gillard J, *et al.*

No. 31

Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.

By Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, *et al.*

No. 32

The cost-effectiveness of testing for hepatitis C in former injecting drug users.

By Castelnovo E, Thompson-Coon J, Pitt M, Cramp M, Siebert U, Price A, *et al.*

No. 33

Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation.

By Kaltenthaler E, Brazier J, De Nigris E, Tumor I, Ferriter M, Beverley C, *et al.*

No. 34

Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy.

By Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, *et al.*

No. 35

Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation.

By Brazier J, Tumor I, Holmes M, Ferriter M, Parry G, Dent-Brown K, *et al.*

No. 36

Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model.

By Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J, *et al.*

No. 37

Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme.

By O'Dowd H, Gladwell P, Rogers CA, Hollinghurst S, Gregory A.

No. 38

A comparison of the cost-effectiveness of five strategies for the prevention of nonsteroidal anti-inflammatory drug-induced gastrointestinal toxicity: a systematic review with economic modelling.

By Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, *et al.*

No. 39

The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review.

By Waugh N, Black C, Walker S, McIntyre L, Cummins E, Hillis G.

No. 40

What are the clinical outcome and cost-effectiveness of endoscopy undertaken by nurses when compared with doctors? A Multi-Institution Nurse Endoscopy Trial (MINuET).

By Williams J, Russell I, Durai D, Cheung W-Y, Farrin A, Bloor K, *et al.*

No. 41

The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: systematic review and economic evaluation.

By Pandor A, Eggington S, Paisley S, Tappenden P, Sutcliffe P.

No. 42

A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness.

By Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, *et al.*

No. 43

Telemedicine in dermatology: a randomised controlled trial.

By Bowns IR, Collins K, Walters SJ, McDonagh AJG.

No. 44

Cost-effectiveness of cell salvage and alternative methods of minimising perioperative allogeneic blood transfusion: a systematic review and economic model.

By Davies L, Brown TJ, Haynes S, Payne K, Elliott RA, McCollum C.

No. 45

Clinical effectiveness and cost-effectiveness of laparoscopic surgery for colorectal cancer: systematic reviews and economic evaluation.

By Murray A, Lourenco T, de Verteuil R, Hernandez R, Fraser C, McKinley A, *et al.*

No. 46

Etanercept and efalizumab for the treatment of psoriasis: a systematic review.

By Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Bravo Vergel Y, *et al.*

No. 47

Systematic reviews of clinical decision tools for acute abdominal pain.

By Liu JLY, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, *et al.*

No. 48

Evaluation of the ventricular assist device programme in the UK.

By Sharples L, Buxton M, Caine N, Cafferty F, Demiris N, Dyer M, *et al.*

No. 49

A systematic review and economic model of the clinical and cost-effectiveness of immunosuppressive therapy for renal transplantation in children.

By Yao G, Albon E, Adi Y, Milford D, Bayliss S, Ready A, *et al.*

No. 50

Amniocentesis results: investigation of anxiety. The ARIA trial.

By Hewison J, Nixon J, Fountain J, Cocks K, Jones C, Mason G, *et al.*

Volume 11, 2007

No. 1

Pemetrexed disodium for the treatment of malignant pleural mesothelioma: a systematic review and economic evaluation.

By Dundar Y, Bagust A, Dickson R, Dodd S, Green J, Haycox A, *et al.*

No. 2

A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer.

By Collins R, Fenwick E, Trowman R, Perard R, Norman G, Light K, *et al.*

No. 3

A systematic review of rapid diagnostic tests for the detection of tuberculosis infection.

By Dinnes J, Deeks J, Kunst H, Gibson A, Cummins E, Waugh N, *et al.*

No. 4

The clinical effectiveness and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women.

By Stevenson M, Davis S, Lloyd-Jones M, Beverley C.

No. 5

A systematic review of quantitative and qualitative research on the role and effectiveness of written information available to patients about individual medicines.

By Raynor DK, Blenkinsopp A, Knapp P, Grime J, Nicolson DJ, Pollock K, *et al.*

No. 6

Oral naltrexone as a treatment for relapse prevention in formerly opioid-dependent drug users: a systematic review and economic evaluation.

By Adi Y, Juarez-Garcia A, Wang D, Jowett S, Frew E, Day E, *et al.*

No. 7

Glucocorticoid-induced osteoporosis: a systematic review and cost-utility analysis.

By Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M.

No. 8

Epidemiological, social, diagnostic and economic evaluation of population screening for genital chlamydial infection.

By Low N, McCarthy A, Macleod J, Salisbury C, Campbell R, Roberts TE, *et al.*

No. 9

Methadone and buprenorphine for the management of opioid dependence: a systematic review and economic evaluation.

By Connock M, Juarez-Garcia A, Jowett S, Frew E, Liu Z, Taylor RJ, *et al.*

No. 10

Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only.

By Isaacs AJ, Critchley JA, See Tai S, Buckingham K, Westley D, Harridge SDR, *et al.*

No. 11

Interferon alfa (pegylated and non-pegylated) and ribavirin for the treatment of mild chronic hepatitis C: a systematic review and economic evaluation.

By Shepherd J, Jones J, Hartwell D, Davidson P, Price A, Waugh N.

No. 12

Systematic review and economic evaluation of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer.

By Tappenden P, Jones R, Paisley S, Carroll C.

No. 13

A systematic review and economic evaluation of epoetin alfa, epoetin beta and darbepoetin alfa in anaemia associated with cancer, especially that attributable to cancer treatment.

By Wilson J, Yao GL, Raftery J, Bohlius J, Brunskill S, Sandercock J, *et al.*

No. 14

A systematic review and economic evaluation of statins for the prevention of coronary events.

By Ward S, Lloyd Jones M, Pandor A, Holmes M, Ara R, Ryan A, *et al.*

No. 15

A systematic review of the effectiveness and cost-effectiveness of different models of community-based respite care for frail older people and their carers.

By Mason A, Weatherly H, Spilsbury K, Arksey H, Golder S, Adamson J, *et al.*

No. 16

Additional therapy for young children with spastic cerebral palsy: a randomised controlled trial.

By Weindling AM, Cunningham CC, Glenn SM, Edwards RT, Reeves DJ.

No. 17

Screening for type 2 diabetes: literature review and economic modelling.

By Waugh N, Scotland G, McNamee P, Gillett M, Brennan A, Goyder E, *et al.*

No. 18

The effectiveness and cost-effectiveness of cinacalcet for secondary hyperparathyroidism in end-stage renal disease patients on dialysis: a systematic review and economic evaluation.

By Garside R, Pitt M, Anderson R, Mealing S, Roome C, Snaith A, *et al.*

No. 19

The clinical effectiveness and cost-effectiveness of gemcitabine for metastatic breast cancer: a systematic review and economic evaluation.

By Takeda AL, Jones J, Loveman E, Tan SC, Clegg AJ.

No. 20

A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease.

By Collins R, Cranny G, Burch J, Aguiar-Ibáñez R, Craig D, Wright K, *et al.*

No. 21

The clinical effectiveness and cost-effectiveness of treatments for children with idiopathic steroid-resistant nephrotic syndrome: a systematic review.

By Colquitt JL, Kirby J, Green C, Cooper K, Trompeter RS.

No. 22

A systematic review of the routine monitoring of growth in children of primary school age to identify growth-related conditions.

By Fayer D, Nixon J, Hartley S, Rithalia A, Butler G, Rudolf M, *et al.*

No. 23

Systematic review of the effectiveness of preventing and treating *Staphylococcus aureus* carriage in reducing peritoneal catheter-related infections.

By McCormack K, Rabindranath K, Kilonzo M, Vale L, Fraser C, McIntyre L, *et al.*

No. 24

The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis.

By McLoughlin DM, Mogg A, Eranti S, Pluck G, Purvis R, Edwards D, *et al.*

No. 25

A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment.

By Boyle J, McCartney E, Forbes J, O'Hare A.

No. 26

Hormonal therapies for early breast cancer: systematic review and economic evaluation.

By Hind D, Ward S, De Nigris E, Simpson E, Carroll C, Wyld L.

No. 27

Cardioprotection against the toxic effects of anthracyclines given to children with cancer: a systematic review.

By Bryant J, Picot J, Levitt G, Sullivan I, Baxter L, Clegg A.

No. 28

Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation.

By McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, *et al.*

No. 29

Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analyses.

By Colbourn T, Asseburg C, Bojke L, Philips Z, Claxton K, Ades AE, *et al.*

No. 30

Clinical effectiveness and cost-effectiveness of bone morphogenetic proteins in the non-healing of fractures and spinal fusion: a systematic review.

By Garrison KR, Donell S, Ryder J, Shemilt I, Mugford M, Harvey I, *et al.*

No. 31

A randomised controlled trial of postoperative radiotherapy following breast-conserving surgery in a minimum-risk older population. The PRIME trial.

By Prescott RJ, Kunkler IH, Williams LJ, King CC, Jack W, van der Pol M, *et al.*

No. 32

Current practice, accuracy, effectiveness and cost-effectiveness of the school entry hearing screen.

By Bamford J, Fortnum H, Bristow K, Smith J, Vamvakas G, Davies L, *et al.*

No. 33

The clinical effectiveness and cost-effectiveness of inhaled insulin in diabetes mellitus: a systematic review and economic evaluation.

By Black C, Cummins E, Royle P, Philip S, Waugh N.

No. 34

Surveillance of cirrhosis for hepatocellular carcinoma: systematic review and economic analysis.

By Thompson Coon J, Rogers G, Hewson P, Wright D, Anderson R, Cramp M, *et al.*

No. 35

The Birmingham Rehabilitation Uptake Maximisation Study (BRUM). Homebased compared with hospital-based cardiac rehabilitation in a multi-ethnic population: cost-effectiveness and patient adherence.

By Jolly K, Taylor R, Lip GYH, Greenfield S, Raftery J, Mant J, *et al.*

No. 36

A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food.

By Abubakar I, Irvine L, Aldus CF, Wyatt GM, Fordham R, Schelenz S, *et al.*

No. 37

A randomised controlled trial examining the longer-term outcomes of standard versus new antiepileptic drugs. The SANAD trial.

By Marson AG, Appleton R, Baker GA, Chadwick DW, Doughty J, Eaton B, *et al.*

No. 38

Clinical effectiveness and cost-effectiveness of different models of managing long-term oral anti-coagulation therapy: a systematic review and economic modelling.

By Connock M, Stevens C, Fry-Smith A, Jowett S, Fitzmaurice D, Moore D, *et al.*

No. 39

A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder.

By Soares-Weiser K, Bravo Vergel Y, Beynon S, Dunn G, Barbieri M, Duffy S, *et al.*

No. 40

Taxanes for the adjuvant treatment of early breast cancer: systematic review and economic evaluation.

By Ward S, Simpson E, Davis S, Hind D, Rees A, Wilkinson A.

No. 41

The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation.

By Burr JM, Mowatt G, Hernández R, Siddiqui MAR, Cook J, Lourenco T, *et al.*

No. 42

Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models.

By Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I.

No. 43

Contamination in trials of educational interventions.

By Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, *et al.*

No. 44

Overview of the clinical effectiveness of positron emission tomography imaging in selected cancers.

By Facey K, Bradbury I, Laking G, Payne E.

No. 45

The effectiveness and cost-effectiveness of carmustine implants and temozolomide for the treatment of newly diagnosed high-grade glioma: a systematic review and economic evaluation.

By Garside R, Pitt M, Anderson R, Rogers G, Dyer M, Mealing S, *et al.*

No. 46

Drug-eluting stents: a systematic review and economic evaluation.

By Hill RA, Boland A, Dickson R, Dünder Y, Haycox A, McLeod C, *et al.*

No. 47

The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model.

By Fox M, Mealing S, Anderson R, Dean J, Stein K, Price A, *et al.*

No. 48

Recruitment to randomised trials: strategies for trial enrolment and participation study. The STEPS study.

By Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, *et al.*

No. 49

Cost-effectiveness of functional cardiac testing in the diagnosis and management of coronary artery disease: a randomised controlled trial. The CECaT trial.

By Sharples L, Hughes V, Crean A, Dyer M, Buxton M, Goldsmith K, *et al.*

No. 50

Evaluation of diagnostic tests when there is no gold standard. A review of methods.

By Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM.

No. 51

Systematic reviews of the clinical effectiveness and cost-effectiveness of proton pump inhibitors in acute upper gastrointestinal bleeding.

By Leontiadis GI, Sreedharan A, Dorward S, Barton P, Delaney B, Howden CW, *et al.*

No. 52

A review and critique of modelling in prioritising and designing screening programmes.

By Karnon J, Goyder E, Tappenden P, McPhie S, Towers I, Brazier J, *et al.*

No. 53

An assessment of the impact of the NHS Health Technology Assessment Programme.

By Hanney S, Buxton M, Green C, Coulson D, Raftery J.

Volume 12, 2008

No. 1

A systematic review and economic model of switching from nonglycopeptide to glycopeptide antibiotic prophylaxis for surgery.

By Cranny G, Elliott R, Weatherly H, Chambers D, Hawkins N, Myers L, *et al.*

No. 2

'Cut down to quit' with nicotine replacement therapies in smoking cessation: a systematic review of effectiveness and economic analysis.

By Wang D, Connock M, Barton P, Fry-Smith A, Aveyard P, Moore D.

No. 3

A systematic review of the effectiveness of strategies for reducing fracture risk in children with juvenile idiopathic arthritis with additional data on long-term risk of fracture and cost of disease management.

By Thornton J, Ashcroft D, O'Neill T, Elliott R, Adams J, Roberts C, *et al.*

No. 4

Does befriending by trained lay workers improve psychological well-being and quality of life for carers of people with dementia, and at what cost? A randomised controlled trial.

By Charlesworth G, Shepstone L, Wilson E, Thalanany M, Mugford M, Poland F.

No. 5

A multi-centre retrospective cohort study comparing the efficacy, safety and cost-effectiveness of hysterectomy and uterine artery embolisation for the treatment of symptomatic uterine fibroids. The HOPEFUL study.

By Hirst A, Dutton S, Wu O, Briggs A, Edwards C, Waldenmaier L, *et al.*

No. 6

Methods of prediction and prevention of pre-eclampsia: systematic reviews of accuracy and effectiveness literature with economic modelling.

By Meads CA, Cnossen JS, Meher S, Juarez-Garcia A, ter Riet G, Duley L, *et al.*

No. 7

The use of economic evaluations in NHS decision-making: a review and empirical investigation.

By Williams I, McIver S, Moore D, Bryan S.

No. 8

Stapled haemorrhoidectomy (haemorrhoidopexy) for the treatment of haemorrhoids: a systematic review and economic evaluation.

By Burch J, Epstein D, Baba-Akbari A, Weatherly H, Fox D, Golder S, *et al.*

No. 9

The clinical effectiveness of diabetes education models for Type 2 diabetes: a systematic review.

By Loveman E, Frampton GK, Clegg AJ.

No. 10

Payment to healthcare professionals for patient recruitment to trials: systematic review and qualitative study.

By Raftery J, Bryant J, Powell J, Kerr C, Hawker S.

No. 11

Cyclooxygenase-2 selective non-steroidal anti-inflammatory drugs (etodolac, meloxicam, celecoxib, rofecoxib, etoricoxib, valdecoxib and lumiracoxib) for osteoarthritis and rheumatoid arthritis: a systematic review and economic evaluation.

By Chen Y-F, Jobanputra P, Barton P, Bryan S, Fry-Smith A, Harris G, *et al.*

No. 12

The clinical effectiveness and cost-effectiveness of central venous catheters treated with anti-infective agents in preventing bloodstream infections: a systematic review and economic evaluation.

By Hockenhull JC, Dwan K, Boland A, Smith G, Bagust A, Dundar Y, *et al.*

No. 13

Stepped treatment of older adults on laxatives. The STOOL trial.

By Mihaylov S, Stark C, McColl E, Steen N, Vanoli A, Rubin G, *et al.*

No. 14

A randomised controlled trial of cognitive behaviour therapy in adolescents with major depression treated by selective serotonin reuptake inhibitors. The ADAPT trial.

By Goodyer IM, Dubicka B, Wilkinson P, Kelvin R, Roberts C, Byford S, *et al.*

No. 15

The use of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer: systematic review and economic evaluation.

By Hind D, Tappenden P, Tumor I, Eggington E, Sutcliffe P, Ryan A.

No. 16

Ranibizumab and pegaptanib for the treatment of age-related macular degeneration: a systematic review and economic evaluation.

By Colquitt JL, Jones J, Tan SC, Takeda A, Clegg AJ, Price A.

No. 17

Systematic review of the clinical effectiveness and cost-effectiveness of 64-slice or higher computed tomography angiography as an alternative to invasive coronary angiography in the investigation of coronary artery disease.

By Mowatt G, Cummins E, Waugh N, Walker S, Cook J, Jia X, *et al.*

No. 18

Structural neuroimaging in psychosis: a systematic review and economic evaluation.

By Albon E, Tsourapas A, Frew E, Davenport C, Oyebo F, Bayliss S, *et al.*

No. 19

Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta₂ agonists for the treatment of chronic asthma in adults and children aged 12 years and over.

By Shepherd J, Rogers G, Anderson R, Main C, Thompson-Coon J, Hartwell D, *et al.*

No. 20

Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta₂ agonists for the treatment of chronic asthma in children under the age of 12 years.

By Main C, Shepherd J, Anderson R, Rogers G, Thompson-Coon J, Liu Z, *et al.*

No. 21

Ezetimibe for the treatment of hypercholesterolaemia: a systematic review and economic evaluation.

By Ara R, Tumur I, Pandor A, Duenas A, Williams R, Wilkinson A, *et al.*

No. 22

Topical or oral ibuprofen for chronic knee pain in older people. The TOIB study.

By Underwood M, Ashby D, Carnes D, Castelnovo E, Cross P, Harding G, *et al.*

No. 23

A prospective randomised comparison of minor surgery in primary and secondary care. The MiSTIC trial.

By George S, Pockney P, Primrose J, Smith H, Little P, Kinley H, *et al.*

No. 24

A review and critical appraisal of measures of therapist-patient interactions in mental health settings.

By Cahill J, Barkham M, Hardy G, Gilbody S, Richards D, Bower P, *et al.*

No. 25

The clinical effectiveness and cost-effectiveness of screening programmes for amblyopia and strabismus in children up to the age of 4–5 years: a systematic review and economic evaluation.

By Carlton J, Karnon J, Czoski-Murray C, Smith KJ, Marr J.

No. 26

A systematic review of the clinical effectiveness and cost-effectiveness and economic modelling of minimal incision total hip replacement approaches in the management of arthritic disease of the hip.

By de Verteuil R, Imamura M, Zhu S, Glazener C, Fraser C, Munro N, *et al.*

No. 27

A preliminary model-based assessment of the cost-utility of a screening programme for early age-related macular degeneration.

By Karnon J, Czoski-Murray C, Smith K, Brand C, Chakravarthy U, Davis S, *et al.*

No. 28

Intravenous magnesium sulphate and sotalol for prevention of atrial fibrillation after coronary artery bypass surgery: a systematic review and economic evaluation.

By Shepherd J, Jones J, Frampton GK, Tanajewski L, Turner D, Price A.

No. 29

Absorbent products for urinary/faecal incontinence: a comparative evaluation of key product categories.

By Fader M, Cottenden A, Getliffe K, Gage H, Clarke-O'Neill S, Jamieson K, *et al.*

No. 30

A systematic review of repetitive functional task practice with modelling of resource use, costs and effectiveness.

By French B, Leathley M, Sutton C, McAdam J, Thomas L, Forster A, *et al.*

No. 31

The effectiveness and cost-effectiveness of minimal access surgery amongst people with gastro-oesophageal reflux disease – a UK collaborative study. The REFLUX trial.

By Grant A, Wileman S, Ramsay C, Bojke L, Epstein D, Sculpher M, *et al.*

No. 32

Time to full publication of studies of anti-cancer medicines for breast cancer and the potential for publication bias: a short systematic review.

By Takeda A, Loveman E, Harris P, Hartwell D, Welch K.

No. 33

Performance of screening tests for child physical abuse in accident and emergency departments.

By Woodman J, Pitt M, Wentz R, Taylor B, Hodes D, Gilbert RE.

No. 34

Curative catheter ablation in atrial fibrillation and typical atrial flutter: systematic review and economic evaluation.

By Rodgers M, McKenna C, Palmer S, Chambers D, Van Hout S, Golder S, *et al.*

No. 35

Systematic review and economic modelling of effectiveness and cost utility of surgical treatments for men with benign prostatic enlargement.

By Lourenco T, Armstrong N, N'Dow J, Nabi G, Deverill M, Pickard R, *et al.*

No. 36

Immunoprophylaxis against respiratory syncytial virus (RSV) with palivizumab in children: a systematic review and economic evaluation.

By Wang D, Cummins C, Bayliss S, Sandercock J, Burls A.

Volume 13, 2009**No. 1**

Deferasirox for the treatment of iron overload associated with regular blood transfusions (transfusional haemosiderosis) in patients suffering with chronic anaemia: a systematic review and economic evaluation.

By McLeod C, Fleeman N, Kirkham J, Bagust A, Boland A, Chu P, *et al.*

No. 2

Thrombophilia testing in people with venous thromboembolism: systematic review and cost-effectiveness analysis.

By Simpson EL, Stevenson MD, Rawdin A, Papaioannou D.

No. 3

Surgical procedures and non-surgical devices for the management of non-apnoeic snoring: a systematic review of clinical effects and associated treatment costs.

By Main C, Liu Z, Welch K, Weiner G, Quentin Jones S, Stein K.

No. 4

Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis.

By McDaid C, Griffin S, Weatherly H, Durée K, van der Burgt M, van Hout S, Akers J, *et al.*

No. 5

Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review.

By Sutcliffe P, Hummel S, Simpson E, Young T, Rees A, Wilkinson A, *et al.*

No. 6

The harmful health effects of recreational ecstasy: a systematic review of observational evidence.

By Rogers G, Elston J, Garside R, Roome C, Taylor R, Younger P, *et al.*

No. 7

Systematic review of the clinical effectiveness and cost-effectiveness of oesophageal Doppler monitoring in critically ill and high-risk surgical patients.

By Mowatt G, Houston G, Hernández R, de Verteuil R, Fraser C, Cuthbertson B, *et al.*

No. 8

The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports.

By Taylor RS, Elston J.

No. 9

Controlling Hypertension and Hypotension Immediately Post Stroke (CHHIPS) – a randomised controlled trial.

By Potter J, Mistri A, Brodie F, Chernova J, Wilson E, Jagger C, *et al.*

No. 10

Routine antenatal anti-D prophylaxis for RhD-negative women: a systematic review and economic evaluation.

By Pilgrim H, Lloyd-Jones M, Rees A.

No. 11

Amantadine, oseltamivir and zanamivir for the prophylaxis of influenza (including a review of existing guidance no. 67): a systematic review and economic evaluation.

By Tappenden P, Jackson R, Cooper K, Rees A, Simpson E, Read R, *et al.*

No. 12

Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.

By Hobart J, Cano S.

No. 13

Treatment of severe ankle sprain: a pragmatic randomised controlled trial comparing the clinical effectiveness and cost-effectiveness of three types of mechanical ankle support with tubular bandage. The CAST trial.

By Cooke MW, Marsh JL, Clark M, Nakash R, Jarvis RM, Hutton JL, *et al.*, on behalf of the CAST trial group.

No. 14

Non-occupational postexposure prophylaxis for HIV: a systematic review.

By Bryant J, Baxter L, Hird S.

No. 15

Blood glucose self-monitoring in type 2 diabetes: a randomised controlled trial.

By Farmer AJ, Wade AN, French DP, Simon J, Yudkin P, Gray A, *et al.*

No. 16

How far does screening women for domestic (partner) violence in different health-care settings meet criteria for a screening programme? Systematic reviews of nine UK National Screening Committee criteria.

By Feder G, Ramsay J, Dunne D, Rose M, Arsene C, Norman R, *et al.*

No. 17

Spinal cord stimulation for chronic pain of neuropathic or ischaemic origin: systematic review and economic evaluation.

By Simpson, EL, Duenas A, Holmes MW, Papaioannou D, Chilcott J.

No. 18

The role of magnetic resonance imaging in the identification of suspected acoustic neuroma: a systematic review of clinical and cost-effectiveness and natural history.

By Fortnum H, O'Neill C, Taylor R, Lenthall R, Nikolopoulos T, Lightfoot G, *et al.*

No. 19

Dipsticks and diagnostic algorithms in urinary tract infection: development and validation, randomised trial, economic analysis, observational cohort and qualitative study.

By Little P, Turner S, Rumsby K, Warner G, Moore M, Lowes JA, *et al.*

No. 20

Systematic review of respite care in the frail elderly.

By Shaw C, McNamara R, Abrams K, Cannings-John R, Hood K, Longo M, *et al.*

No. 21

Neuroleptics in the treatment of aggressive challenging behaviour for people with intellectual disabilities: a randomised controlled trial (NACHBID).

By Tyrer P, Oliver-Africano P, Romeo R, Knapp M, Dickens S, Bouras N, *et al.*

No. 22

Randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of selective serotonin reuptake inhibitors plus supportive care, versus supportive care alone, for mild to moderate depression with somatic symptoms in primary care: the THREAD (THREShold for AntiDepressant response) study.

By Kendrick T, Chatwin J, Dowrick C, Tylee A, Morriss R, Peveler R, *et al.*

No. 23

Diagnostic strategies using DNA testing for hereditary haemochromatosis in at-risk populations: a systematic review and economic evaluation.

By Bryant J, Cooper K, Picot J, Clegg A, Roderick P, Rosenberg W, *et al.*

No. 24

Enhanced external counterpulsation for the treatment of stable angina and heart failure: a systematic review and economic analysis.

By McKenna C, McDaid C, Suekarran S, Hawkins N, Claxton K, Light K, *et al.*

No. 25

Development of a decision support tool for primary care management of patients with abnormal liver function tests without clinically apparent liver disease: a record-linkage population cohort study and decision analysis (ALFIE).

By Donnan PT, McLernon D, Dillon JF, Ryder S, Roderick P, Sullivan F, *et al.*

No. 26

A systematic review of presumed consent systems for deceased organ donation.

By Rithalia A, McDaid C, Suekarran S, Norman G, Myers L, Sowden A.

No. 27

Paracetamol and ibuprofen for the treatment of fever in children: the PITCH randomised controlled trial.

By Hay AD, Redmond NM, Costelloe C, Montgomery AA, Fletcher M, Hollinghurst S, *et al.*

No. 28

A randomised controlled trial to compare minimally invasive glucose monitoring devices with conventional monitoring in the management of insulin-treated diabetes mellitus (MITRE).

By Newman SP, Cooke D, Casbard A, Walker S, Meredith S, Nunn A, *et al.*

No. 29

Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making.

By Andronis L, Barton P, Bryan S.

Suppl. 1

Trastuzumab for the treatment of primary breast cancer in HER2-positive women: a single technology appraisal.

By Ward S, Pilgrim H, Hind D.

Docetaxel for the adjuvant treatment of early node-positive breast cancer: a single technology appraisal.

By Chilcott J, Lloyd Jones M, Wilkinson A.

The use of paclitaxel in the management of early stage breast cancer.

By Griffin S, Dunn G, Palmer S, Macfarlane K, Brent S, Dyker A, *et al.*

Rituximab for the first-line treatment of stage III/IV follicular non-Hodgkin's lymphoma.

By Dundar Y, Bagust A, Hounsome J, McLeod C, Boland A, Davis H, *et al.*

Bortezomib for the treatment of multiple myeloma patients.

By Green C, Bryant J, Takeda A, Cooper K, Clegg A, Smith A, *et al.*

Fludarabine phosphate for the first-line treatment of chronic lymphocytic leukaemia.

By Walker S, Palmer S, Erhorn S, Brent S, Dyker A, Ferrie L, *et al.*

Erlotinib for the treatment of relapsed non-small cell lung cancer.

By McLeod C, Bagust A, Boland A, Hockenhull J, Dundar Y, Proudlove C, *et al.*

Cetuximab plus radiotherapy for the treatment of locally advanced squamous cell carcinoma of the head and neck.

By Griffin S, Walker S, Sculpher M, White S, Erhorn S, Brent S, *et al.*

Infliximab for the treatment of adults with psoriasis.

By Loveman E, Turner D, Hartwell D, Cooper K, Clegg A.

No. 30

Psychological interventions for postnatal depression: cluster randomised trial and economic evaluation. The PoNDER trial.

By Morrell CJ, Warner R, Slade P, Dixon S, Walters S, Paley G, *et al.*

No. 31

The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis.

By Rogowski R, Burch J, Palmer S, Craigs C, Golder S, Woolacott N.

No. 32

Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care.

By Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, *et al.*

No. 33

A multicentre randomised controlled trial of the use of continuous positive airway pressure and non-invasive positive pressure ventilation in the early treatment of patients presenting to the emergency department with severe acute cardiogenic pulmonary oedema: the 3CPO trial.

By Gray AJ, Goodacre S, Newby DE, Masson MA, Sampson F, Dixon S, *et al.*, on behalf of the 3CPO study investigators.

No. 34

Early high-dose lipid-lowering therapy to avoid cardiac events: a systematic review and economic evaluation.

By Ara R, Pandor A, Stevens J, Rees A, Rafia R.

No. 35

Adefovir dipivoxil and pegylated interferon alpha for the treatment of chronic hepatitis B: an updated systematic review and economic evaluation.

By Jones J, Shepherd J, Baxter L, Gospodarevskaya E, Hartwell D, Harris P, *et al.*

No. 36

Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis.

By Hewitt CE, Gilbody SM, Brealey S, Paulden M, Palmer S, Mann R, *et al.*

No. 37

A double-blind randomised placebo-controlled trial of topical intranasal corticosteroids in 4- to 11-year-old children with persistent bilateral otitis media with effusion in primary care.

By Williamson I, Bengt S, Barton S, Petrou S, Letley L, Fasey N, *et al.*

No. 38

The effectiveness and cost-effectiveness of methods of storing donated kidneys from deceased donors: a systematic review and economic model.

By Bond M, Pitt M, Akoh J, Moxham T, Hoyle M, Anderson R.

No. 39

Rehabilitation of older patients: day hospital compared with rehabilitation at home. A randomised controlled trial.

By Parker SG, Oliver P, Pennington M, Bond J, Jagger C, Enderby PM, *et al.*

No. 40

Breastfeeding promotion for infants in neonatal units: a systematic review and economic analysis

By Renfrew MJ, Craig D, Dyson L, McCormick F, Rice S, King SE, *et al.*

No. 41

The clinical effectiveness and cost-effectiveness of bariatric (weight loss) surgery for obesity: a systematic review and economic evaluation.

By Picot J, Jones J, Colquitt JL, Gospodarevskaya E, Loveman E, Baxter L, *et al.*

No. 42

Rapid testing for group B streptococcus during labour: a test accuracy study with evaluation of acceptability and cost-effectiveness.

By Daniels J, Gray J, Pattison H, Roberts T, Edwards E, Milner P, *et al.*

No. 43

Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling.

By Honest H, Forbes CA, Durée KH, Norman G, Duffy SB, Tsourapas A, *et al.*

No. 44

The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model.

By Bond M, Mealing S, Anderson R, Elston J, Weiner G, Taylor RS, *et al.*

Suppl. 2

Gemcitabine for the treatment of metastatic breast cancer.

By Jones J, Takeda A, Tan SC, Cooper K, Loveman E, Clegg A.

Varenicline in the management of smoking cessation: a single technology appraisal.

By Hind D, Tappenden P, Peters J, Kenjegalieva K.

Alteplase for the treatment of acute ischaemic stroke: a single technology appraisal.

By Lloyd Jones M, Holmes M.

Rituximab for the treatment of rheumatoid arthritis.

By Bagust A, Boland A, Hockenhull J, Fleeman N, Greenhalgh J, Dundar Y, *et al.*

Omalizumab for the treatment of severe persistent allergic asthma.

By Jones J, Shepherd J, Hartwell D, Harris P, Cooper K, Takeda A, *et al.*

Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma.

By Boland A, Bagust A, Hockenhull J, Davis H, Chu P, Dickson R.

Adalimumab for the treatment of psoriasis.

By Turner D, Picot J, Cooper K, Loveman E.

Dabigatran etexilate for the prevention of venous thromboembolism in patients undergoing elective hip and knee surgery: a single technology appraisal.

By Holmes M, C Carroll C, Papaioannou D.

Romiplostim for the treatment of chronic immune or idiopathic thrombocytopenic purpura: a single technology appraisal.

By Mowatt G, Boachie C, Crowther M, Fraser C, Hernández R, Jia X, *et al.*

Sunitinib for the treatment of gastrointestinal stromal tumours: a critique of the submission from Pfizer.

By Bond M, Hoyle M, Moxham T, Napier M, Anderson R.

No. 45

Vitamin K to prevent fractures in older women: systematic review and economic evaluation.

By Stevenson M, Lloyd-Jones M, Papaioannou D.

No. 46

The effects of biofeedback for the treatment of essential hypertension: a systematic review.

By Greenhalgh J, Dickson R, Dundar Y.

No. 47

A randomised controlled trial of the use of aciclovir and/or prednisolone for the early treatment of Bell's palsy: the BELLS study.

By Sullivan FM, Swan IRC, Donnan PT, Morrison JM, Smith BH, McKinstry B, *et al.*

Suppl. 3

Lapatinib for the treatment of HER2-overexpressing breast cancer.

By Jones J, Takeda A, Picot J, von Keyserlingk C, Clegg A.

Infliximab for the treatment of ulcerative colitis.

By Hyde C, Bryan S, Juarez-Garcia A, Andronis L, Fry-Smith A.

Rimonabant for the treatment of overweight and obese people.

By Burch J, McKenna C, Palmer S, Norman G, Glanville J, Sculpher M, *et al.*

Telbivudine for the treatment of chronic hepatitis B infection.

By Hartwell D, Jones J, Harris P, Cooper K.

Entecavir for the treatment of chronic hepatitis B infection.

By Shepherd J, Gospodarevskaya E, Frampton G, Cooper K.

Febuxostat for the treatment of hyperuricaemia in people with gout: a single technology appraisal.

By Stevenson M, Pandor A.

Rivaroxaban for the prevention of venous thromboembolism: a single technology appraisal.

By Stevenson M, Scope A, Holmes M, Rees A, Kaltenthaler E.

Cetuximab for the treatment of recurrent and/or metastatic squamous cell carcinoma of the head and neck.

By Greenhalgh J, Bagust A, Boland A, Fleeman N, McLeod C, Dundar Y, *et al.*

Mifamurtide for the treatment of osteosarcoma: a single technology appraisal.

By Pandor A, Fitzgerald P, Stevenson M, Papaioannou D.

Ustekinumab for the treatment of moderate to severe psoriasis.

By Gospodarevskaya E, Picot J, Cooper K, Loveman E, Takeda A.

No. 48

Endovascular stents for abdominal aortic aneurysms: a systematic review and economic model.

By Chambers D, Epstein D, Walker S, Fayter D, Paton F, Wright K, *et al.*

No. 49

Clinical and cost-effectiveness of epoprostenol, iloprost, bosentan, sitaxentan and sildenafil for pulmonary arterial hypertension within their licensed indications: a systematic review and economic evaluation.

By Chen Y-F, Jowett S, Barton P, Malottki K, Hyde C, Gibbs JSR, *et al.*

No. 50

Cessation of attention deficit hyperactivity disorder drugs in the young (CADDY) – a pharmacoepidemiological and qualitative study.

By Wong ICK, Asherson P, Bilbow A, Clifford S, Coghill D, R DeSoysa R, *et al.*

No. 51

ARTISTIC: a randomised trial of human papillomavirus (HPV) testing in primary cervical screening.

By Kitchener HC, Almonte M, Gilham C, Dowie R, Stoykova B, Sargent A, *et al.*

No. 52

The clinical effectiveness of glucosamine and chondroitin supplements in slowing or arresting progression of osteoarthritis of the knee: a systematic review and economic evaluation.

By Black C, Clar C, Henderson R, MacEachern C, McNamee P, Quayyum Z, *et al.*

No. 53

Randomised preference trial of medical versus surgical termination of pregnancy less than 14 weeks' gestation (TOPS).

By Robson SC, Kelly T, Howel D, Deverill M, Hewison J, Lie MLS, *et al.*

No. 54

Randomised controlled trial of the use of three dressing preparations in the management of chronic ulceration of the foot in diabetes.

By Jeffcoate WJ, Price PE, Phillips CJ, Game FL, Mudge E, Davies S, *et al.*

No. 55

VenUS II: a randomised controlled trial of larval therapy in the management of leg ulcers.

By Dumville JC, Worthy G, Soares MO, Bland JM, Cullum N, Dowson C, *et al.*

No. 56

A prospective randomised controlled trial and economic modelling of antimicrobial silver dressings versus non-adherent control dressings for venous leg ulcers: the VULCAN trial

By Michaels JA, Campbell WB, King BM, MacIntyre J, Palfreyman SJ, Shackley P, *et al.*

No. 57

Communication of carrier status information following universal newborn screening for sickle cell disorders and cystic fibrosis: qualitative study of experience and practice.

By Kai J, Ulph F, Cullinan T, Qureshi N.

No. 58

Antiviral drugs for the treatment of influenza: a systematic review and economic evaluation.

By Burch J, Paulden M, Conti S, Stock C, Corbett M, Welton NJ, *et al.*

No. 59

Development of a toolkit and glossary to aid in the adaptation of health technology assessment (HTA) reports for use in different contexts.

By Chase D, Rosten C, Turner S, Hicks N, Milne R.

No. 60

Colour vision testing for diabetic retinopathy: a systematic review of diagnostic accuracy and economic evaluation.

By Rodgers M, Hodges R, Hawkins J, Hollingworth W, Duffy S, McKibbin M, *et al.*

No. 61

Systematic review of the effectiveness and cost-effectiveness of weight management schemes for the under fives: a short report.

By Bond M, Wyatt K, Lloyd J, Welch K, Taylor R.

No. 62

Are adverse effects incorporated in economic models? An initial review of current practice.

By Craig D, McDaid C, Fonseca T, Stock C, Duffy S, Woolacott N.

Volume 14, 2010

No. 1

Multicentre randomised controlled trial examining the cost-effectiveness of contrast-enhanced high field magnetic resonance imaging in women with primary breast cancer scheduled for wide local excision (COMICE).

By Turnbull LW, Brown SR, Olivier C, Harvey I, Brown J, Drew P, *et al.*

No. 2

Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation.

By Thompson Coon J, Hoyle M, Green C, Liu Z, Welch K, Moxham T, *et al.*

No. 3

The clinical effectiveness and cost-effectiveness of testing for cytochrome P450 polymorphisms in patients with schizophrenia treated with antipsychotics: a systematic review and economic evaluation.

By Fleeman N, McLeod C, Bagust A, Beale S, Boland A, Dundar Y, *et al.*

No. 4

Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer.

By Mowatt G, Zhu S, Kilonzo M, Boachie C, Fraser C, Griffiths TRL, *et al.*

No. 5

Effectiveness and cost-effectiveness of arthroscopic lavage in the treatment of osteoarthritis of the knee: a mixed methods study of the feasibility of conducting a surgical placebo-controlled trial (the KORAL study).

By Campbell MK, Skea ZC, Sutherland AG, Cuthbertson BH, Entwistle VA, McDonald AM, *et al.*

No. 6

A randomised 2 × 2 trial of community versus hospital pulmonary rehabilitation, followed by telephone or conventional follow-up.

By Waterhouse JC, Walters SJ, Oluboyede Y, Lawson RA.

No. 7

The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13–19: a systematic review and economic evaluation.

By Shepherd J, Kavanagh J, Picot J, Cooper K, Harden A, Barnett-Page E, *et al.*

No. 8


Dissemination and publication of research findings: an updated review of related biases.

By Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, *et al.*

No. 9

The effectiveness and cost-effectiveness of biomarkers for the prioritisation of patients awaiting coronary revascularisation: a systematic review and decision model.

By Hemingway H, Henriksson M, Chen R, Damant J, Fitzpatrick N, Abrams K, *et al.*



Health Technology Assessment programme

Director,
Professor Tom Walley,
Director, NIHR HTA
programme, Professor of
Clinical Pharmacology,
University of Liverpool

Deputy Director,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NIHR HTA
programme, Professor of
Clinical Pharmacology,
University of Liverpool

Deputy Chair,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield

Dr Bob Coates,
Consultant Advisor, NETSCC,
HTA

Dr Andrew Cook,
Consultant Advisor, NETSCC,
HTA

Dr Peter Davidson,
Director of Science Support,
NETSCC, HTA

Professor Robin E Ferner,
Consultant Physician and
Director, West Midlands Centre
for Adverse Drug Reactions,
City Hospital NHS Trust,
Birmingham

Professor Paul Glasziou,
Professor of Evidence-Based
Medicine, University of Oxford

Dr Nick Hicks,
Director of NHS Support,
NETSCC, HTA

Dr Edmund Jessop,
Medical Adviser, National
Specialist, National
Commissioning Group (NCG),
Department of Health, London

Ms Lynn Kerridge,
Chief Executive Officer,
NETSCC and NETSCC, HTA

Dr Ruairidh Milne,
Director of Strategy and
Development, NETSCC

Ms Kay Pattison,
Section Head, NHS R&D
Programme, Department of
Health

Ms Pamela Young,
Specialist Programme Manager,
NETSCC, HTA

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NIHR HTA
programme, Professor of
Clinical Pharmacology,
University of Liverpool

Chair,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield

Deputy Chair,
Dr Andrew Farmer,
Senior Lecturer in General
Practice, Department of
Primary Health Care,
University of Oxford

Professor Ann Ashburn,
Professor of Rehabilitation
and Head of Research,
Southampton General Hospital

Professor Deborah Ashby,
Professor of Medical Statistics,
Queen Mary, University of
London

Professor John Cairns,
Professor of Health Economics,
London School of Hygiene and
Tropical Medicine

Professor Peter Croft,
Director of Primary Care
Sciences Research Centre, Keele
University

Professor Nicky Cullum,
Director of Centre for Evidence-
Based Nursing, University of
York

Professor Jenny Donovan,
Professor of Social Medicine,
University of Bristol

Professor Steve Halligan,
Professor of Gastrointestinal
Radiology, University College
Hospital, London

Professor Freddie Hamdy,
Professor of Urology,
University of Sheffield

Professor Allan House,
Professor of Liaison Psychiatry,
University of Leeds

Dr Martin J Landray,
Reader in Epidemiology,
Honorary Consultant Physician,
Clinical Trial Service Unit,
University of Oxford

Professor Stuart Logan,
Director of Health & Social
Care Research, The Peninsula
Medical School, Universities of
Exeter and Plymouth

Dr Rafael Perera,
Lecturer in Medical Statistics,
Department of Primary Health
Care, University of Oxford

Professor Ian Roberts,
Professor of Epidemiology &
Public Health, London School
of Hygiene and Tropical
Medicine

Professor Mark Sculpher,
Professor of Health Economics,
University of York

Professor Helen Smith,
Professor of Primary Care,
University of Brighton

Professor Kate Thomas,
Professor of Complementary &
Alternative Medicine Research,
University of Leeds

Professor David John
Torgerson,
Director of York Trials Unit,
University of York

Professor Hywel Williams,
Professor of Dermato-
Epidemiology, University of
Nottingham

Observers

Ms Kay Pattison,
Section Head, NHS R&D
Programme, Department of
Health

Dr Morven Roberts,
Clinical Trials Manager,
Medical Research Council

Diagnostic Technologies & Screening Panel

Members

Chair,
Professor Paul Glasziou,
Professor of Evidence-Based
Medicine, University of Oxford

Deputy Chair,
Dr David Elliman,
Consultant Paediatrician and
Honorary Senior Lecturer,
Great Ormond Street Hospital,
London

Professor Judith E Adams,
Consultant Radiologist,
Manchester Royal Infirmary,
Central Manchester &
Manchester Children's
University Hospitals NHS Trust,
and Professor of Diagnostic
Radiology, Imaging Science
and Biomedical Engineering,
Cancer & Imaging Sciences,
University of Manchester

Ms Jane Bates,
Consultant Ultrasound
Practitioner, Ultrasound
Department, Leeds Teaching
Hospital NHS Trust

Dr Stephanie Dancer,
Consultant Microbiologist,
Hairmyres Hospital, East
Kilbride

Professor Glyn Elwyn,
Primary Medical Care Research
Group, Swansea Clinical School,
University of Wales

Dr Ron Gray,
Consultant Clinical
Epidemiologist, Department
of Public Health, University of
Oxford

Professor Paul D Griffiths,
Professor of Radiology,
University of Sheffield

Dr Jennifer J Kurinczuk,
Consultant Clinical
Epidemiologist, National
Perinatal Epidemiology Unit,
Oxford

Dr Susanne M Ludgate,
Medical Director, Medicines &
Healthcare Products Regulatory
Agency, London

Dr Anne Mackie,
Director of Programmes, UK
National Screening Committee

Dr Michael Millar,
Consultant Senior Lecturer in
Microbiology, Barts and The
London NHS Trust, Royal
London Hospital

Mr Stephen Pilling,
Director, Centre for Outcomes,
Research & Effectiveness,
Joint Director, National
Collaborating Centre for
Mental Health, University
College London

Mrs Una Rennard,
Service User Representative

Dr Phil Shackley,
Senior Lecturer in Health
Economics, School of
Population and Health
Sciences, University of
Newcastle upon Tyne

Dr W Stuart A Smellie,
Consultant in Chemical
Pathology, Bishop Auckland
General Hospital

Dr Nicholas Summerton,
Consultant Clinical and Public
Health Advisor, NICE

Ms Dawn Talbot,
Service User Representative

Dr Graham Taylor,
Scientific Advisor, Regional
DNA Laboratory, St James's
University Hospital, Leeds

Professor Lindsay Wilson
Turnbull,
Scientific Director of the
Centre for Magnetic Resonance
Investigations and YCR
Professor of Radiology, Hull
Royal Infirmary

Observers

Dr Tim Elliott,
Team Leader, Cancer
Screening, Department of
Health

Dr Catherine Moody,
Programme Manager,
Neuroscience and Mental
Health Board

Dr Ursula Wells,
Principal Research Officer,
Department of Health

Pharmaceuticals Panel

Members

Chair,
Professor Robin Ferner,
Consultant Physician and
Director, West Midlands Centre
for Adverse Drug Reactions,
City Hospital NHS Trust,
Birmingham

Deputy Chair,
Professor Imti Choonara,
Professor in Child Health,
University of Nottingham

Mrs Nicola Carey,
Senior Research Fellow,
School of Health and Social
Care, The University of
Reading

Mr John Chapman,
Service User Representative

Dr Peter Elton,
Director of Public Health,
Bury Primary Care Trust

Dr Ben Goldacre,
Research Fellow, Division of
Psychological Medicine and
Psychiatry, King's College
London

Mrs Barbara Greggains,
Service User Representative

Dr Bill Gutteridge,
Medical Adviser, London
Strategic Health Authority

Dr Dyfrig Hughes,
Reader in Pharmacoeconomics
and Deputy Director, Centre
for Economics and Policy in
Health, IMSCaR, Bangor
University

Professor Jonathan Ledermann,
Professor of Medical Oncology
and Director of the Cancer
Research UK and University
College London Cancer Trials
Centre

Dr Yoon K Loke,
Senior Lecturer in Clinical
Pharmacology, University of
East Anglia

Professor Femi Oyeboode,
Consultant Psychiatrist
and Head of Department,
University of Birmingham

Dr Andrew Prentice,
Senior Lecturer and Consultant
Obstetrician and Gynaecologist,
The Rosie Hospital, University
of Cambridge

Dr Martin Shelly,
General Practitioner, Leeds,
and Associate Director, NHS
Clinical Governance Support
Team, Leicester

Dr Gillian Shepherd,
Director, Health and Clinical
Excellence, Merck Serono Ltd

Mrs Katrina Simister,
Assistant Director New
Medicines, National Prescribing
Centre, Liverpool

Mr David Symes,
Service User Representative

Dr Lesley Wise,
Unit Manager,
Pharmacoepidemiology
Research Unit, VRMM,
Medicines & Healthcare
Products Regulatory Agency

Observers

Ms Kay Pattison,
Section Head, NHS R&D
Programme, Department of
Health

Mr Simon Reeve,
Head of Clinical and Cost-
Effectiveness, Medicines,
Pharmacy and Industry Group,
Department of Health

Dr Heike Weber,
Programme Manager,
Medical Research Council

Dr Ursula Wells,
Principal Research Officer,
Department of Health

Therapeutic Procedures Panel

Members

<p>Chair, Dr John C Pounsford, Consultant Physician, North Bristol NHS Trust</p> <p>Deputy Chair, Professor Scott Weich, Professor of Psychiatry, Division of Health in the Community, University of Warwick, Coventry</p> <p>Professor Jane Barlow, Professor of Public Health in the Early Years, Health Sciences Research Institute, Warwick Medical School, Coventry</p> <p>Ms Maree Barnett, Acting Branch Head of Vascular Programme, Department of Health</p> <p>Mrs Val Carlill,</p>	<p>Service User Representative</p> <p>Mrs Anthea De Barton-Watson, Service User Representative</p> <p>Mr Mark Emberton, Senior Lecturer in Oncological Urology, Institute of Urology, University College Hospital, London</p> <p>Professor Steve Goodacre, Professor of Emergency Medicine, University of Sheffield</p> <p>Professor Christopher Griffiths, Professor of Primary Care, Barts and The London School of Medicine and Dentistry</p> <p>Mr Paul Hilton,</p>	<p>Consultant Gynaecologist and Urogynaecologist, Royal Victoria Infirmary, Newcastle upon Tyne</p> <p>Professor Nicholas James, Professor of Clinical Oncology, University of Birmingham, and Consultant in Clinical Oncology, Queen Elizabeth Hospital</p> <p>Dr Peter Martin, Consultant Neurologist, Addenbrooke's Hospital, Cambridge</p> <p>Dr Kate Radford,</p>	<p>Senior Lecturer (Research), Clinical Practice Research Unit, University of Central Lancashire, Preston</p> <p>Mr Jim Reece Service User Representative</p> <p>Dr Karen Roberts, Nurse Consultant, Dunston Hill Hospital Cottages</p>
--	---	--	---

Observers

<p>Dr Phillip Leech, Principal Medical Officer for Primary Care, Department of Health</p> <p>Ms Kay Pattison, Section Head, NHS R&D Programme, Department of Health</p>	<p>Dr Morven Roberts, Clinical Trials Manager, Medical Research Council</p>	<p>Professor Tom Walley, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p>	<p>Dr Ursula Wells, Principal Research Officer, Department of Health</p>
---	---	--	--

Disease Prevention Panel

Members

<p>Chair, Dr Edmund Jessop, Medical Adviser, National Specialist, National Commissioning Group (NCG), London</p> <p>Deputy Chair, Dr David Pencheon, Director, NHS Sustainable Development Unit, Cambridge</p> <p>Dr Elizabeth Fellow-Smith, Medical Director, West London Mental Health Trust, Middlesex</p>	<p>Dr John Jackson, General Practitioner, Parkway Medical Centre, Newcastle upon Tyne</p> <p>Professor Mike Kelly, Director, Centre for Public Health Excellence, NICE, London</p> <p>Dr Chris McCall, General Practitioner, The Hadleigh Practice, Corfe Mullen, Dorset</p> <p>Ms Jeanett Martin, Director of Nursing, BarnDoc Limited, Lewisham Primary Care Trust</p>	<p>Dr Julie Mytton, Locum Consultant in Public Health Medicine, Bristol Primary Care Trust</p> <p>Miss Nicky Mullany, Service User Representative</p> <p>Professor Ian Roberts, Professor of Epidemiology and Public Health, London School of Hygiene & Tropical Medicine</p> <p>Professor Ken Stein, Senior Clinical Lecturer in Public Health, University of Exeter</p>	<p>Dr Kieran Sweeney, Honorary Clinical Senior Lecturer, Peninsula College of Medicine and Dentistry, Universities of Exeter and Plymouth</p> <p>Professor Carol Tannahill, Glasgow Centre for Population Health</p> <p>Professor Margaret Thorogood, Professor of Epidemiology, University of Warwick Medical School, Coventry</p>
---	--	---	---

Observers

<p>Ms Christine McGuire, Research & Development, Department of Health</p>	<p>Dr Caroline Stone, Programme Manager, Medical Research Council</p>
---	---

Expert Advisory Network

Members

Professor Douglas Altman,
Professor of Statistics in
Medicine, Centre for Statistics
in Medicine, University of
Oxford

Professor John Bond,
Professor of Social Gerontology
& Health Services Research,
University of Newcastle upon
Tyne

Professor Andrew Bradbury,
Professor of Vascular Surgery,
Solihull Hospital, Birmingham

Mr Shaun Brogan,
Chief Executive, Ridgeway
Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Regulation
and Improvement Authority,
Belfast

Ms Tracy Bury,
Project Manager, World
Confederation for Physical
Therapy, London

Professor Iain T Cameron,
Professor of Obstetrics and
Gynaecology and Head of the
School of Medicine, University
of Southampton

Dr Christine Clark,
Medical Writer and Consultant
Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing and
Head of Research, The
Medical School, University of
Birmingham

Professor Barry Cookson,
Director, Laboratory of Hospital
Infection, Public Health
Laboratory Service, London

Dr Carl Counsell,
Clinical Senior Lecturer in
Neurology, University of
Aberdeen

Professor Howard Cuckle,
Professor of Reproductive
Epidemiology, Department
of Paediatrics, Obstetrics &
Gynaecology, University of
Leeds

Dr Katherine Darton,
Information Unit, MIND – The
Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric
Epidemiology, Institute of Child
Health, London

Mr John Dunning,
Consultant Cardiothoracic
Surgeon, Papworth Hospital
NHS Trust, Cambridge

Mr Jonathan Earnshaw,
Consultant Vascular Surgeon,
Gloucestershire Royal Hospital,
Gloucester

Professor Martin Eccles,
Professor of Clinical
Effectiveness, Centre for Health
Services Research, University of
Newcastle upon Tyne

Professor Pam Enderby,
Dean of Faculty of Medicine,
Institute of General Practice
and Primary Care, University of
Sheffield

Professor Gene Feder,
Professor of Primary Care
Research & Development,
Centre for Health Sciences,
Barts and The London School
of Medicine and Dentistry

Mr Leonard R Fenwick,
Chief Executive, Freeman
Hospital, Newcastle upon Tyne

Mrs Gillian Fletcher,
Antenatal Teacher and Tutor
and President, National
Childbirth Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine,
University of Birmingham

Mr Tam Fry,
Honorary Chairman, Child
Growth Foundation, London

Professor Fiona Gilbert,
Consultant Radiologist and
NCRN Member, University of
Aberdeen

Professor Paul Gregg,
Professor of Orthopaedic
Surgical Science, South Tees
Hospital NHS Trust

Bec Hanley,
Co-director, TwoCan Associates,
West Sussex

Dr Maryann L Hardy,
Senior Lecturer, University of
Bradford

Mrs Sharon Hart,
Healthcare Management
Consultant, Reading

Professor Robert E Hawkins,
CRC Professor and Director
of Medical Oncology, Christie
CRC Research Centre,
Christie Hospital NHS Trust,
Manchester

Professor Richard Hobbs,
Head of Department of Primary
Care & General Practice,
University of Birmingham

Professor Alan Horwich,
Dean and Section Chairman,
The Institute of Cancer
Research, London

Professor Allen Hutchinson,
Director of Public Health and
Deputy Dean of SCHARR,
University of Sheffield

Professor Peter Jones,
Professor of Psychiatry,
University of Cambridge,
Cambridge

Professor Stan Kaye,
Cancer Research UK Professor
of Medical Oncology, Royal
Marsden Hospital and Institute
of Cancer Research, Surrey

Dr Duncan Keeley,
General Practitioner (Dr Burch
& Ptnrs), The Health Centre,
Thame

Dr Donna Lamping,
Research Degrees Programme
Director and Reader in
Psychology, Health Services
Research Unit, London School
of Hygiene and Tropical
Medicine, London

Mr George Levvy,
Chief Executive, Motor
Neurone Disease Association,
Northampton

Professor James Lindesay,
Professor of Psychiatry for the
Elderly, University of Leicester

Professor Julian Little,
Professor of Human Genome
Epidemiology, University of
Ottawa

Professor Alistaire McGuire,
Professor of Health Economics,
London School of Economics

Professor Rajan Madhok,
Medical Director and Director
of Public Health, Directorate
of Clinical Strategy & Public
Health, North & East Yorkshire
& Northern Lincolnshire
Health Authority, York

Professor Alexander Markham,
Director, Molecular Medicine
Unit, St James's University
Hospital, Leeds

Dr Peter Moore,
Freelance Science Writer,
Ashtead

Dr Andrew Mortimore,
Public Health Director,
Southampton City Primary
Care Trust

Dr Sue Moss,
Associate Director, Cancer
Screening Evaluation Unit,
Institute of Cancer Research,
Sutton

Professor Miranda Mugford,
Professor of Health Economics
and Group Co-ordinator,
University of East Anglia

Professor Jim Neilson,
Head of School of Reproductive
& Developmental Medicine
and Professor of Obstetrics
and Gynaecology, University of
Liverpool

Mrs Julietta Patnick,
National Co-ordinator, NHS
Cancer Screening Programmes,
Sheffield

Professor Robert Peveler,
Professor of Liaison Psychiatry,
Royal South Hants Hospital,
Southampton

Professor Chris Price,
Director of Clinical Research,
Bayer Diagnostics Europe,
Stoke Poges

Professor William Rosenberg,
Professor of Hepatology
and Consultant Physician,
University of Southampton

Professor Peter Sandercock,
Professor of Medical Neurology,
Department of Clinical
Neurosciences, University of
Edinburgh

Dr Susan Schonfield,
Consultant in Public Health,
Hillingdon Primary Care Trust,
Middlesex

Dr Eamonn Sheridan,
Consultant in Clinical Genetics,
St James's University Hospital,
Leeds

Dr Margaret Somerville,
Director of Public Health
Learning, Peninsula Medical
School, University of Plymouth

Professor Sarah Stewart-Brown,
Professor of Public Health,
Division of Health in the
Community, University of
Warwick, Coventry

Professor Ala Szczepura,
Professor of Health Service
Research, Centre for Health
Services Studies, University of
Warwick, Coventry

Mrs Joan Webster,
Consumer Member, Southern
Derbyshire Community Health
Council

Professor Martin Whittle,
Clinical Co-director, National
Co-ordinating Centre for
Women's and Children's
Health, Lymington

Feedback

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (www.hta.ac.uk) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.