

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

A Hutchinson,<sup>1\*</sup> JE Coster,<sup>1</sup> KL Cooper,<sup>1</sup>  
A McIntosh,<sup>1</sup> SJ Walters,<sup>2</sup> PA Bath,<sup>3</sup>  
M Pearson,<sup>4</sup> TA Young,<sup>5</sup> K Rantell,<sup>2</sup>  
Mj Campbell<sup>2</sup> and J Ratcliffe<sup>5</sup>

<sup>1</sup>Section of Public Health, ScHARR, University of Sheffield, UK

<sup>2</sup>Section of Health Services Research, ScHARR, University of Sheffield, UK

<sup>3</sup>Department of Information Studies, University of Sheffield, UK

<sup>4</sup>Clinical Effectiveness and Evaluation Unit, Royal College of Physicians,  
London, UK

<sup>5</sup>Section of Health Economics and Decision Sciences,  
ScHARR, University of Sheffield, UK

\*Corresponding author



## *Executive summary*

*Health Technology Assessment* 2010; Vol. 14: No. 10  
DOI: 10.3310/hta14100

**Health Technology Assessment**  
**NIHR HTA programme**  
[www.hta.ac.uk](http://www.hta.ac.uk)





## 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.

## Publication

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).

# 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](http://www.prepress-projects.co.uk)), on behalf of NETSCC, HTA.

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

MR