White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/96911/

Version: Accepted Version

Article:
Hutchinson, A., Coster, J.E., Cooper, K.L. et al. (2013) A structured judgement method to enhance mortality case note review: development and evaluation. BMJ Quality and Safety, 22 (12). pp. 1032-1040. ISSN: 2044-5415

https://doi.org/10.1136/bmjqs-2013-001839

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

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

A structured judgement method to enhance mortality case note review: development and evaluation

Allen Hutchinson,[1] Joanne E Coster,[1] Katy L Cooper,[1] Michael Pearson,[2] Aileen McIntosh,[1] Peter A Bath[3]

[1]Section of Public Health, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

[2]Department of Clinical Evaluation, University of Liverpool, Liverpool, UK

[3]Information School, University of Sheffield, Sheffield, UK

Correspondence to:

Professor Allen Hutchinson,

Section of Public Health, School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent St., Sheffield S1 4DA, UK;

allen.hutchinson@sheffield.ac.uk

Keywords

Chart review methodologies, Quality improvement methodologies, Healthcare quality improvement, Patient safety, Qualitative research

ABSTRACT

Background Case note review remains a prime means of retrospectively assessing quality of care. This study examines a new implicit judgement method, combining structured reviewer comments with quality of care scores, to assess care of people who die in hospital.

Methods Using 1566 case notes from 20 English hospitals, 40 physicians each reviewed 30–40 case notes, writing structured judgement-based comments on care provided within three phases of care, and on care overall, and scoring quality of care from 1 (unsatisfactory) to 6 (very best care). Quality of care comments on 119 people who died (7.6% of the cohort) were analysed independently by two researchers to investigate how well reviewers provided

structured short judgement notes on quality of care, together with appropriate care scores. Consistency between explanatory textual data and related scores was explored, using overall care score to group cases.

Results Physician reviewers made informative, clinical judgement-based comments across all phases of care and usually provided a coherent quality of care score relating to each phase. The majority of comments (83%) were explicit judgements. About a fifth of patients were considered to have received less than satisfactory care, often experiencing a series of adverse events.

Conclusions A combination of implicit judgement, explicit explanatory comment and related quality of care scores can be used effectively to review the spectrum of care provided for people who die in hospital. The method can be used to quickly evaluate deaths so that lessons can be learned about both poor and high quality care.

BACKGROUND

Hospital death rates are a matter of public concern in the United Kingdom and have been the subject of both country-wide data analysis and local intensive reviews, one of which has resulted in a major public debate.[1,2] Concerns about hospital deaths in well-developed health systems, especially when linked to the occurrence of adverse events,[3] have also been expressed internationally. This has resulted in a number of rigorous epidemiological studies of adverse event frequency, for example in Australia, Canada and Sweden.[4-6] More recently there have been large studies of hospital deaths, together with associated events, which have examined whether some hospital deaths might have been preventable.[7,8] On a day to day level, however, there remains a need for rigorous methods to enable

clinical teams to retrospectively assess quality of care in a timely manner and, thus, to identify when deaths were inevitable, or whether they might have been prevented with better care. This could assist, for example, in the discussions on care that currently take place in hospital Morbidity and Mortality meetings.

Internationally, case note review remains a prime means of retrospectively assessing quality of care,[3-8] despite the known methodological and practical challenges of this review method.[9-11] Two principal review methods are used:- explicit criterion-based methods, and implicit (sometimes called holistic) methods which are based on clinical judgement.

Criterion-based methods, usually using frameworks of pre-determined criteria to identify elements of care which are either met or not met, are useful for large-scale audits of care or for screening case notes using criterion-based trigger tools.[9]

Implicit review methods are based on clinical judgement, and are probably more effective for identifying and recording the detail and nuance of care (both unsatisfactory and good).[12] Thus, implicit review methods are probably more appropriate for detailed exploration of the care for people who die in hospital. However, unstructured implicit review formats have been criticised for low inter-rater reliability (high variability) and for potential reviewer bias,[9-11,13] whereas structured implicit review limits the variability and creates specific frameworks so that reviewers are able to make, justify and organise statements on care.[14]

Initial models of structured implicit review methods were actually a fusion of implicit judgements of quality of care which were required of the reviewer in order for them to complete a set of explicit review criteria (for example, a criterion such as 'no appropriate nursing interventions carried out').[14] A framework such as this was used by Pearson et al to monitor nursing care quality.[15] More recently, Hogan et al used this approach in a study of the frequency of adverse event rates and preventable deaths in English hospitals, where a judgement-based, structured, explicit one-to-five scale was used by reviewers to rate quality of care, from very poor to excellent.[8] In a study of adverse event frequency and preventability upon 8400 patient records in the Netherlands, Zegers and colleagues used two six-point scales on which reviewers used their judgement to record whether injury was caused by health care management or the disease process and to assess the degree of preventability.[7,16]

However, this form of judgement-based structured implicit review only provides a scale-based quantitative result and there is no means of understanding how or why the reviewer judgement was made. Thus the method is useful for large scale monitoring or epidemiological studies of adverse events, but has rather less value for more detailed review at the ward or hospital level of why an event occurred.

To increase the value of structured implicit review in the context of reviewing the whole spectrum of care quality, rather than focussing only on adverse event rates, we designed and tested a structured care review method, drawing on the initial work of Kahn and colleagues.[14] This required reviewers to make implicit clinical judgements and to write explicit comments to support judgement-based quality of care scores.[9] In the developmental stage of the study, multi-professional groups of reviewers independently reviewed the same records, first using a quantitative and then using a qualitative review process. For each case the review process was undertaken for three phases of care

(admission, initial management and later management), followed by an overall judgement of the care provided for the patient. For each phase of care, and for care overall, reviewers, both physicians and nurses, were asked to rate quality of care on a one (unsatisfactory) to six (excellent) scale. This was similar to a four-stage, phase of care approach, together with overall care quality, subsequently used by Hogan et al to provide a framework on which to rate quality of care.[8]

There was moderate inter-rater reliability of these judgement-based scores when two or three physicians, working separately, used structured implicit review on the same set of case notes (Intra-Class Correlation Coefficient [ICC] 0.52). Physician reviewers tended to make more explicit written judgements on the quality of care provided than did nurse reviewers, who more often made commentaries about the process/pathway of care.[13]

Subsequently we asked 40 physician reviewers to undertake this enhanced form of structured implicit review to examine the quality of care provided for 1566 people with either Chronic Obstructive Pulmonary Disease [COPD] or heart failure as their main diagnosis. There was no oversampling of deaths and each set of case notes was reviewed only once. There were two reviewers (one for COPD cases and one for heart failure cases) for each of twenty randomly-selected large hospitals in England and each reviewer judged between 30-40 consecutively selected sets of case notes and associated clinical records in their own hospital. Reviewers were either senior respiratory or cardiology physicians in training. Our initial quantitative analysis, reported elsewhere, examined the range of phase of care scores and overall care scores for each of the 20 hospitals and the relationship of the care scores to broader quality of care markers.[9]

Here we report a new qualitative and quantitative analysis of the commentaries written by the reviewers to support their judgement scores of care provided for the 119 cases who died in hospital (8% of the total within the cohort of 1566 cases). The purpose of the analysis was to explore whether physician reviewers can consistently provide short, structured, judgement-based comments on quality of care that they can also justify with an appropriate care score. The consistency between the explanatory textual data and the related scores is explored with a view to considering whether this structured method, combining implicit judgements supported by explanatory comments, together with quality of care scores, can be used for routine mortality case note review.

METHODS

Hospital and reviewer selection

Acute care hospitals in England were first grouped into quartiles using mortality data. Equal numbers of hospitals from the upper and lower quartiles were then randomly selected (20 in total). Each randomly selected hospital had to provide two reviewers, who were all volunteers and specialists in training. Each was initially approached by specialists in their own hospitals and initial research team contact with the specialists was made through the Royal College of Physicians.

Reviewer training

All reviewers received training in the review methods and in data recording prior to data collection. A full-day training session comprised a description of the methods, discussion about the need to be as explicit as possible about the judgement commentaries and a session of reviewing a set of case notes in pairs with tutors. Finally, all of the reviewers judged the care from the same set of anonymised case notes and then commented on their findings in a managed small group discussion, which again emphasised the need to be explicit in their judgements. Data was collected via an electronic form which enabled direct entry by reviewers of both comments and scores for all relevant care phases and care overall. This enabled reviewers to structure their commentaries and the data collection program was also demonstrated during the training day.

Finally, reviewers were provided with a set of national clinical practice guidelines relevant to their clinical specialty. Regular contact was maintained between the study team and the reviewers, who could ask for advice during the review period, using a telephone helpline.

Data collection

Each set of case notes was reviewed by a single physician reviewer. Quality of care was assessed in three phases – admission, initial management and later management and also for care overall. For each phase of care and for care overall, reviewers wrote short textual comments on the quality of care provided and were encouraged to be explicit in their comments on care. They also gave the care a score from one to six for each phase and for overall care, based on the criteria in Table 1.

Table 1: Care score criteria

Score 1	Unsatisfactory: 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
Score 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
Score 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
Score 4	This was satisfactory care, only falling short of current best practice in more than two minor areas
Score 5	This was good care, which only fell short of current best practice in one or two minor areas
Score 6	Very best care: this was excellent care and met current best practice

Analysis methods

Of the 1566 cases reviewed, 119 cases had died during their hospital admission.To explore the type and content of written comments by the reviewers on each of the 119 cases, a textual analysis framework, developed during the study prior to this analysis and previously reported,[9] was applied to all of the phase and overall care comments. Two authors (AH, JC) reviewed and categorised the comments independently and any differences in categorisation were resolved through discussion.

Comments were categorised into three groups (see Box 1). All comments in categories B (implicit judgement comments) and C (explicit judgement comments) were subsequently classified by the two study analysts as indicating good quality of care (positive comments) or as indicating poor quality of care (negative comments). These two categories of comment for each case were then grouped by their related overall quality of care scores, which were then used to classify each case into one of six groups, from unsatisfactory care (score one) to very best care (score six). Examples of the detailed textual analysis are presented in the results in Tables 2 and 3.

The association between the quality scores for care overall for the group of 119 people who died was compared with the distribution of scores for the 1447 cases who survived, using the Chi-squared test. The association between the comment category and type and their relationship one to another were explored across overall care scores using the Chi-squared test. Chi-squared tests were undertaken using Microsoft Excel and p-values were calculated using Graphpad software (http://graphpad.com/).

Box 1: Reviewer comment categories

Category A. Little or no comment about care and/or little or no judgement, including, for example, a description of what was in the case note or a description of what happened to the patient (not the care they received).

Note: Category A did not contribute to the analysis presented here, since this analysis was concerned with judgements, rather than descriptive reports.

Category B. Limited comment about quality of care and/or implied judgement. This category included an implied judgment and/or a description of the care delivered (not just a description of a patient pathway) and/or a description of an omission of care.

Category C. Comments about care with explicit judgements and views. This category included explicit judgements of care delivered, questioning or queries about the care delivered, explanations or justification of care delivered, alternative options or justification of care that should have been delivered or concerns about care.

Table 2: Reviewer commentary on care judged unsatisfactory overall

Overall care score 1	Reviewer comments	Comment type (Pos or Neg) / category
Admission phase score 1	Poor history documentation	Neg/B
	Poor examination documentation	Neg/B
	Initial investigations requested CXR, ECG, bloods but no comment made re these	Neg/B
	No ABGs and patient was tachypnoeic and hypoxic	Neg/B
	No O2 (not documented)	Neg/B
	No GTN	Neg/B
	Pitiful dose of frusemide (furosemide) (20 mg IV)	Neg/C
	Extremely poor management	Neg/C
Initial management phase score 1	Medical team made no attempt to adequately treat the heart failure	Neg/C
	No comment on the CXR	Neg/B
	No ABGs	Neg/B
	CPAP started without ABGs	Neg/B
	Did record a resuscitation status	Pos/B
	Documentation very poor, for example, no reference to the fact that she was so unwell or whether they thought it likely that she would die	Neg/C
	No discussion with the family or relatives	Neg/C
Later management	See previous	
Overall care score 1	All aspects of this case were very poor. History, examination, medical management, documentation	Neg/C
	If this lady was clearly dying and had multiple co-morbidities, they should have documented this, made the lady comfortable and called the family in	Neg/C

ABGs, arterial blood gases; CPAP, continuous positive airway pressure; CXR, chest radiograph; ECG, electrocardiograph; GTN, glyceryl trinitrate; IV, intravenous; Neg, negative; Pos, positive.

Table 3: Reviewer commentary on care judged short of best practice

Overall care score 3	Reviewer comments	Comment type (Pos or Neg) / category
Admission phase score 4	pH 7.436 Good history taken of COPD symptoms and normal functional status,	Pos/C
	alternative diagnosis of PE and CCF not excluded in a patient with risk factors for both	Neg/B
	Clinical cardiovascular exam not thorough (no mention of JVP, pedal oedema, chest expansion, sputum characteristics)	Neg/C
Initial management phase score 4	Patient received appropriate treatment for COPD (ie, steroids, antibiotics and nebulizers),	Pos/C
	however the CXR result was never recorded ?looked at	Neg/B
Later management phase score 2	Although the patient was recorded to be clinically improving 2/7 post admission and team were considering early discharge, his ABG was not improving and patient's SOB+tachypnoea attributed to anxiety, pt (patient) gradually deteriorated	Neg/B
	Patient changed to inhalers too soon	Neg/C
	Seen appropriately by respiratory team, frusemide (furosemide) and aminophylline infusion appropriately suggested	Pos/C
	Nursing staff inappropriately withheld oral medications as they thought he was nil by mouth	Neg/C

Overall care score 3	Reviewer comments	Comment type (Pos or Neg) / category
	Developed severe type 2 respiratory failure but no decision on resus status made until patient very unwell. This needed to be made by on call team	Neg/B
	Earlier referral to ITU and I.v aminophylline may have changed outcome	Neg/C
	Good chest physio(therapy) input	Pos/C
Overall care score 3	Patient appropriately treated initially with nebs, antibiotics and steroids	Pos/C
	however patient's treatment plan not escalated until he was in severe type 2 respiratory failure	Neg/B
	NIV/ITU not considered in this patient ?why-he had no other co-morbidities and no previous hospital admissions	Neg/C
	Resus decision made inappropriately by on call team when patient very unwell	Neg/C

ABG, arterial blood gases; CCF, congestive cardiac failure; COPD, chronic obstructive pulmonary disease; CXR, chest radiograph; ITU, intensive therapy unit; I.v, intravenous; JVP, jugular venous pressure; nebs, nebuliser; NIV, non-invasive intubation; PE, pulmonary embolus; pt, patient; resus, resuscitation; SOB, shortness of breath.

RESULTS

The overall quality of care scores for the patients who died are compared, in Table 4, with the scores for all patients who survived. The proportions of cases in which care fell short of good practice are relatively similar across the two groups of cases, although there are a higher proportion of 'satisfactory' cases and a somewhat lower proportion of the 'good' cases among those people who died than in the survivor group. There were no statistically significant differences between the two groups (χ^2=9.800; degrees of freedom (df)=5; p=0.0811).

Table 4 Quality of care overall: score comparisons between people who died and those that survived

Quality of care	Care fell short of good practice			Satisfactory care	Good or better care		Total numbers of reviews
Quality of care scores	**1 (unsatisfactory)**	**2**	**3**	**4**	**5 (good)**	**6 (very best care)**	
People who died (%)	5 (4.3)	3 (2.6)	13 (11.1)	32 (27.4)	40 (34.1)	24 (20.5)	117* (100%)
People who survived (%)	59 (4.1)	97 (6.7)	134 (9.3)	288 (19.9)	640 (44.2)	229 (15.8)	1447* (100%)

*Two cases from the group of 119 people are not included in this analysis due to incomplete data. Both had phase scores of five or six with no negative comments but for each the overall care score was missing, so they could not be grouped by overall care score. (χ^2=9.800; degrees of freedom (df)=5; p=0.0811).

Relationship of positive and negative comments to overall care scores

Table 5 summarises the relationship between the overall care score for each case and the types of comment (whether positive or negative judgements) provided by the reviewers for each of the phases and for overall care. There was a significant association between the total number of positive and negative comments and the overall scores (χ^2=205.50; df=5; p<0.0001).

In the care score range unsatisfactory (one) to falling short of best practice (three), the proportion of negative comments outweighs the positive comments. When the care is rated from satisfactory (four) to very best care (six) the positive comments increasingly outweigh the negative. Generally, the positive to negative ratio of comments for each phase remains stable across each overall group score band. So where the overall score is three or less, across each of the phases there are more negative comments than there are positive comments, and the reverse is true for the summary of the higher scores, indicating that the reviewer judgements are generally consistent with the overall score that was given. The ratios of positive to negative comments ranges between 0.28 for overall care score one to 21.17 for those cases grouped by overall care score six.

Table 5 Numbers of positive and negative comments per overall score

Overall score (N cases)	Admission phase comment type and nos.		Early management phase comment type and nos.		Later management phase comment type and nos.		Overall care comment type and nos.		Total number of comments		Positive/ negative comment ratio
	Pos	Neg.	Pos	Neg.	Pos	Neg.	Pos	Neg.	Pos	Neg.	
1 (5)	6	15	2	18	3	5	2	8	13	48	0.28
2 (3)	4	6	2	9	0	5	0	6	6	26	0.23
3 (13)	11	19	5	16	6	11	7	14	29	60	0.48
4 (32)	39	27	27	24	22	12	24	21	112	84	1.33
5 (40)	49	15	49	12	29	2	47	9	174	38	4.58
6 (24)	36	4	40	1	25	1	29	0	127	6	21.17
Overall (117)	145	86	125	80	85	36	109	58	464	262	1.78

(χ^2=205.50; df=5; p<0.0001)

There are fewer comments in total in the later phases of care because some patients died early in the course of the admission. There is also some indication in the textual commentaries that a number of reviewers felt most of what needed to be said had already been said in the earlier phase of care comments for a particular case, and so did not need to be repeated.

In general, the phase of care comments were more detailed than the overall care comments. Occasionally, however, reviewers gave an unexpectedly high score related to a qualitative judgement that suggested a lower quality of care had occurred (see, for example, the case in Box 4).

Categorisation of comments: implicit and explicit judgements about care quality

Table 6 summarises the numbers of comments grouped by category (category B – implicit judgements of care, category C – explicit judgements of care) and comment type (positive or negative) for each overall care score.

Table 6 Comments by type and category and overall score

Overall score (N cases)	Admission phase comment type, category and number				Early management phase comment type, category and number				Later management phase comment type, category and number				Overall care comment type, category and number			
	Pos		Neg.		Pos		Neg.		Pos		Neg.		Pos		Neg.	
	B	C	B	C	B	C	B	C	B	C	B	C	B	C	B	C
1 (5)	2	4	11	4	0	2	7	11	2	1	1	4	0	2	0	8
2 (3)	3	1	4	2	0	2	4	5	0	0	0	5	0	0	0	6
3 (13)	2	9	8	11	1	4	7	9	2	4	1	10	1	6	1	13
4 (32)	1	38	2	25	2	25	6	18	0	22	1	11	0	24	1	20
5 (40)	9	40	6	9	2	47	2	10	6	23	0	2	3	44	1	8
6 (24)	9	27	4	0	7	33	0	1	3	22	0	1	3	26	0	0
Total (117)	26	119	35	51	12	113	26	54	13	72	3	33	7	102	3	55

Results in Table 7 show that, overall, there were more than four times as many explicit comments (judgements) as there were implicit comments. For the lower overall care scores (one to three) there tended to be a rather higher ratio of implicit (B) judgements than there were for the higher care scores, although the implicit judgements were always in the minority. This tend is confirmed by a significant statistical association between the total number of implicit/explicit judgements of care and the overall care score (χ^2=48.37; df=5; p<0.0001). Thus the pattern of more explicit comments than implicit comments was seen for all quality of care scores, from one (poor care) to six (best care), indicating that reviewers were on the whole prepared to make explicit judgements where care was poor as well as where care was good.

These results suggest that the reviewers were on the whole prepared to make the type of judgements and explicit comments asked of them during training and which would be valuable in a quality of care review.

Table 7 Comparison between implicit/explicit and positive/negative comments

Overall score (N cases)	Ratio of explicit (C) to implicit (B) comments								
	Total positive comments			Total negative comments			All comments		
	B	C	Ratio C to B	B	C	Ratio C to B	B	C	Ratio C to B
1 (5)	4	9	2.25	19	27	1.42	23	36	1.57
2 (3)	3	3	1.00	8	18	2.25	11	21	1.91
3 (13)	6	23	3.83	17	43	2.53	23	66	2.87
4 (32)	3	109	36.33	10	74	7.40	13	183	14.08
5 (40)	20	154	7.70	9	29	3.22	29	183	6.31
6 (24)	22	108	4.91	4	2	0.50	26	110	4.23

Total (117)	58	406	7.00	67	193	2.88	125	599	4.79

χ^2=48.37; df=5; p<0.0001

Content and nature of comments

Study of the individual comments showed that a number of B category comments contained concise technical summaries in addition to implicit judgements on the quality of care. Many of the C category comments were incisive clinical observations with a strong view of the quality of care, especially when the reviewer considered that the care was poor. Comments across the range of overall scores often included consideration of the broader, non-technical, processes of care (for example, communication with relatives), as well as technical aspects of care.

Of the 21 case reviews with low overall scores (scores of one, two or three), 15 were accompanied by an explicit clinically-relevant judgement that justified the low score. Some related to cases where care was generally poor throughout the inpatient episode, while others related to cases where a specific aspect of care was of concern. In two of the cases, incorrect diagnosis was the main problem, while in 12 cases there was concern about suboptimal management. There were usually multiple smaller events that were additive, rather than one main adverse event, which only occurred in one of the 12 cases. Two of the 15 cases were considered to have such poor record keeping as to be a threat to the care of the patient.

Tables 2 and 3 provide examples demonstrating the range, type and category of comments made by reviewers in two cases. All of the comments are as written by the reviewers and the scores given for each phase of care are included. Reviewers were able both to comment on the technical aspects of care and to take a holistic view of the overall care plan.

Table 2 is also used to demonstrate how the categorisation of the comments was applied in the analysis. For example:

- Although the reviewer explicitly grades the documentation as poor in the admission, this is only a description of the documentation without any explanation and therefore is categorised as a B level comment. In the initial management phase, however, there is a judgement (very poor documentation) together with an explanation, which rates a C category.

- When the reviewer implies in the initial management phase that it was poor practice not to take an arterial blood gas sample ('No ABGs (Arterial Blood Gas) and patient was tachypnoeic and hypoxic') there is no explicit statement that this was unsatisfactory (and it is thus a B category comment).

- A judgement on the therapy ('pitiful dose of frusemide (furosemide) (20mg Intravenous)) is a C category comment.

- When commenting on the technical aspects of care, the reviewer could also be explicit about how the care should have been managed overall, in the context of the patient's illness. This is an explicit, category C, judgement.

The case in Table 2 also illustrates a pattern that is of a group or 'constellation' of events which of themselves may not cause severe harm but which, taken together, can lead to harm to the patient. This pattern was also found in the main study among some of the patients who survived.[17]

Although there are usually more negative comments than there are positive comments when overall care scores are low, as shown in Table 5, the case in Table 3 shows examples of how positive and negative comments can be juxtaposed in each phase. In retrospect, this case also brings the question of whether the overall score of three was the most appropriate – it might be argued from the level of the comment that the case could have been given a lower overall care score of two (see, for example, the comments on later management).

Comments on good care tended to be more global than those for unsatisfactory care but may also be quite explicit. Cases which demonstrate this and also how a single adverse event may change the reviewer's overall consideration of the case are included as additional material (see online supplementary tables S7 and S8).

Some of the reviewers in this study were more 'explanatory' than others so that, in some cases, the number of comments may reflect individual style rather than the strength of the comment. For example, comments such as 'good care' or 'unclear treatment' are short explicit judgements without further detail, while other reviewers are more extensively explicit.

Of the 63 case reviews (54% of the total number of mortality reviews) that scored most highly (5 or 6), there were 52 accompanied by a short explicit comment in the overall care section indicating that all key aspects of care had been good or excellent (e.g. 'well looked after') and in 16 of the 63 reviews there were comments about the inevitable outcome of the case despite the good care received.

DISCUSSION

In this study we have shown that physician reviewers are able to use structured review to, make implicit quality and safety judgements, write explicit short care commentaries and give coherent matching quality of care scores. Quantitative scores and qualitative comments matched well, indicating that physician reviewers can appropriately score the quality of care on a rating scale.

These physician reviewers could identify and explain both technical and non-technical aspects of care, and could rank these aspects of care using a set of 'benchmark' scores, ranging from very good care to very unsatisfactory care. For people with complex illnesses, the outcome is not always survival. However, structured explicit judgments can show how high quality care was provided – even if the patient has not survived. For example, there were a number of instances where explicit comments were made about the quality of non-technical care such as way information was provided to patients and their relatives. Conversely, when poor care occurs, the method can identify the points at which care fails to meet expected standards, and when the situation can be, or is, rescued. It is interesting to note that in Table 1 the proportions of those who died and had less than satisfactory care (about 20% of the cases), were similar to those who survived and had poor care.

During the training session, reviewers were encouraged to be as direct as possible in their commentaries, and in the results overall (Tables 3 and 4) there were many more explicit comments than there were implicit comments. Nevertheless when poor care was being described, while explicit comments predominated, there was a noteworthy proportion of implicit, B level, comments. Sometimes these B level comments were about documentation (which was not in the C category) or concerned missed tests which the reviewer listed and didn't specifically make a judgement upon – for example, No ABGs (Arterial Blood Gases), see Box 3. It may be that in this case the reviewer felt that the result said it all and that an explicit comment was superfluous. On the other hand, it could also be that some reviewers might have felt uncomfortable about making direct comments about very poor care.

With the hindsight of these results, and when undertaking reviews such as this in health service settings, training should include discussion of an initial sample of commentaries and scores with each reviewer to assist in maximising the number of explicit comments. Of course, training might identify some reviewers who do not feel able to make explicit comments and so would not be suitable for this type of review.

The phase of care structure also contributes to an understanding of how care may vary, and at what point. Interestingly, a phase of care approach has also been used by Shannon and colleagues in a review of cardiac surgical care,[18] albeit in a rather more structured system with distinct changes in physical settings. In the context of assessing whether death was a preventable outcome, Hogan et al used a four-phase model to identify adverse incidents – initial assessment, treatment plan, ongoing monitoring and preparation for discharge.[8] Under the conditions of a service review a three-phase model might be easier to manage but either a three or four phase approach would be appropriate.

Qualitative comments from the reviewers were useful in that they could succinctly identify what was done badly in poor cases. Such short explicit judgements could support a wider, more detailed, service review to assess what could be improved in a particular setting or condition. Furthermore, since this structured review method assesses both process and outcome of care, this mixed type of review, using qualitative comments with scores, might be a useful addition to review measures which only assess outcomes or are criterion based. This mixed qualitative and criterion based method is published in detail elsewhere.[9]

In this study, assessments of the quality and safety of the care provided showed that, for over 80% of the patients who died, care was rated at least satisfactory and, for approximately half of the cases, care was judged to be of high quality. The processes of care described enable a qualitative judgement to be associated with an objective score that is explicable to, and understandable by, a wide range of people and would be understood by the public too. However, having graded a case as poor or not, there is the added advantage that the structured comments also provide the reasoning behind the judgement in a format to which clinical teams and individuals should be able to respond in a review process.

Limitations

In this study the 40 reviewers were all volunteers who undertook the work in their own hospitals. Although there might be concerns about the impartiality of using internal review teams, results have shown that reviewers can make incisive short notes (commentaries) about quality of care, and can critically review care provided in their own hospitals.

Internal review teams have also been used in other settings. Sharek et al commented on the strong performance of hospital-based internal review teams, albeit when using more structured, criterion-based, trigger tools to identify adverse events.[19]

Although it could be argued that two reviewers per case might enhance the quality and depth of a case note review, there is some evidence to suggest that this use of a more intensive resource does not necessarily improve the review process. While we were able to show in our development study that there was reasonable coherence of quantitative care scores and criterion-based scores between physician reviewers,[9,13] other work by Hofer and colleagues found that multiple reviewing of the same set of case notes did not enhance the results.[20]

Finally, it is important to recognise that there are limits to the extent to which the quantitative analysis of the reviews can be used. For example, averaging phase scores across each case, to determine whether phase score averages are similar to the overall care score, is not appropriate. An example of this can be found in the online resource in Box 6 where care was judged excellent until moments before the patient died. The value of this current study is that the context and the basis for any quantitative score can be found in the phase of care comments associated with each score.

Conclusions

This method is a refinement on both global implicit judgement and structured implicit judgement used upon a set of case notes, because it is able to provide information on aspects of each phase of care, enabling more detailed, yet still brief, comments to show explicitly how care may vary or be consistent with expected standards. For example, this method could be used to identify whether care has led to a preventable death, or to identify good quality of care even though the overall outcome is failure to survive. Thus, although the study did not explicitly seek to judge a death as preventable, as did Hogan and colleagues,[8] review training could straightforwardly include an explicit judgement commentary about whether a death was preventable or was not preventable (as some of the study reviewers actually did).

Results also show how explicit written judgements and quality of care scoring can be used together and thus may offer a range of case note review methods for use under differing circumstances, together with opportunities for providing training and assessment of 'reviewer quality'.

Structured judgement review provides the framework for a quality of care review that can be used by clinical leaders and quality managers to identify potential priority areas for evaluation. For example, scoring allows for a screening of the overall care quality for a case overall, or can identify issues in a particular phase of care, say at admission or initial management. Explicit comments allow exploration of particular aspects of care, for instance where good treatment plans might be inadequately implemented. For these purposes it is not necessary to analyse whether comments are implicit or explicit. The data collection framework is straightforward, has been previously published and is easily available.[9]

Who should act as the reviewers? Because of the complexity of illness often presented in hospital settings, studies of adverse events have used experienced generalists with some specialist support.[8] This structured implicit review method could be used in a similar way

either with in-hospital teams or by visiting teams from other hospitals. We do not know whether the review results would be better when undertaken by experienced specialists than the reviewers in our study. However, our results have shown that this form of review can be undertaken by specialists at a senior level in a training programme – so increasing the pool of trained senior reviewers in a hospital - and thus the method offers the opportunity for early review of the care of people who die in hospital so that, where necessary, timely quality improvement lessons can be learnt.

ABBREVIATIONS

COPD: Chronic Obstructive Pulmonary Disease

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.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Allen Hutchinson took the lead in the conception and design of the study, took the lead on the analysis of the qualitative mortality review data, was principal author of all of the drafts of this paper and has given approval for this version of the paper to be published.

Joanne Coster contributed to the conception and design of the study, undertook data collection and analysis of the mortality review data, contributed to all of the drafts of this paper and has given approval for this version of the paper to be published.

Katy Cooper made contributions to the design of the study, undertook data collection and contributed to analysis of the mortality review data, contributed to all of the drafts of this paper and has given approval for this version of the paper to be published.

Michael Pearson contributed to the conception and design of the study overall, contributed to the interpretation of the mortality review data, contributed to all of the drafts of this paper and has given approval for this version of the paper to be published.

Aileen McIntosh contributed to the conception and design of the study overall, took a lead on the qualitative analysis framework, contributed to all of the drafts of this paper and has given approval for this version of the paper to be published.

Peter Bath contributed to the conception and design of study and of the qualitative analysis framework, undertook the statistical analysis for the quantitative analysis, contributed to all of the drafts of this paper and has given approval for this version of the paper to be published.

No other authors contributed to this analysis and authorship.

Allen Hutchinson acts as Guarantor for this article.

ACKNOWLEGEMENTS

This project was funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) Programme (project number RM03/JH08/AH) and was published in full in Health Technology Assessment 2010,**14(10):**1-170.

The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HTA programme, NIHR, NHS or the Department of Health.

REFERENCES

1. The Dr Foster Hospital Guide 2009: How Safe Is Your Hospital? London: Dr Foster Ltd 2009.
2. The Stationary Office: Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. London: The Stationary Office 2013. ISBN 9780102981476.
3. Thomas EJ, Brennan TA. Incidence and types of preventable adverse events in elderly patients: population based review of medical records. BMJ 2000;**320**:741-4.
4. Wilson RM, Runciman WB, Gibberd RW et al. The Quality in Australian Health Care Study. Med J Aust 1995;**163**:458-71.
5. Baker GR, Norton PG, Flintoft V et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ 2004;**170**:1678-86.

6. Soop M, Fryksmark U, Koster M et al. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. Int J Qual Health Care 2009;**21**:285-91.
7. Zegers M, de Bruijne MC, Wagner C et al. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. Qual Saf Health Care 2009;**18**:297-302.
8. Hogan H, Healey F, Neale G et al. Preventable deaths in care in English acute hospitals: a retrospective case record study. BMJ Qual Saf 2012. doi:10.1136/bmjqs-2012-001159.
9. Hutchinson A, Coster JE, Cooper KL et al. Comparison of case note review methods for evaluating quality and safety in health care. Health Technol Assess 2010;**14**:1-170.
10. Lilford R, Edwards A, Girling A, et al. Inter-rater reliability of case-note audit: a systematic review. J Health Serv Res Policy 2007;**12**:173-80.
11. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. JAMA 2001;**286**:415-20.
12. Mohammed MA, Mant J, Bentham L et al. 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:**102-106.
13. Hutchinson A, Coster JE, Cooper KL et al. Assessing quality of care from hospital case notes: comparison of reliability of two methods. Qual Saf Health Care 2010 Dec, **19(6)**:e2.
14. Kahn KL, Rubenstein LV, Sherwood MJ et al. Structured Implicit Review for Physician Implicit Measurement of Quality of Care: Development of the Form and Guidelines for Its Use. California: Rand Corporation 1989.
15. Pearson ML, Lee JL, Chang BL et al. Structured implicit review: a new method for monitoring nursing care quality. Med Care 2000;**38**:1074-1091.

16. Zegers M, de Bruijne MC, Wagner C, et al. Design of a retrospective patient record on the occurrence of adverse events among patients in Dutch hospitals. BMC Health Serv Res 2007;7:27 doi:10.1186/1472-6963-7-27
17. Hutchinson A, McIntosh A, Coster JE et al. When is an event an event? Data from the quality and safety continuum. In: From safe design to safe practice. Edited by Hignett S, Norris B, Catchpole K et al. Cambridge: The Ergonomics Society 2008. ISBN 978-0-9554225-2-2
18. Shannon FL, Fazzalari FL, Theurer PF et al. A method to evaluate cardiac surgery mortality: phase of care mortality analysis. Ann Thoracic Surg 2012;**93**:36-43.
19. Sharek PJ, Parry G, Goldmann D et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalised patients. Health Serv Res 2011;**46**:654-678.
20. Hofer TP, Bernstein SJ, DeMonner S et al. Discussion between Reviewers Does Not Improve Reliability of Peer Review of Hospital Quality. Med Care 2000;**38**:152-61.

ADDITIONAL MATERIAL

Table S7 and Table S8 provide additional examples of reviewer scores and comments, together with the assessment of each comment type and category. Table S7 is a case where the care was judged to be good. Overall there were fewer comments than those cases where care was judged to be unsatisfactory. In the later management section of this case there is a description of care included which, though not a judgement and therefore not placed in a category, illuminates the case and its management for the reader.

Table S7

Overall care score 5		Comment type and category
Admission Phase score 5	Good initial assessment. Heart failure identified and confirmed radiologically and appropriate treatment commenced. Important differential diagnoses considered and also treated for chest infection.	Pos/C Pos/C Pos/C
Early management Phase score	Regularly reviewed as he failed to improve and significant effort made to optimise his medication.	Pos/B Pos/C

missing		
Later management Phase score 5	Ceiling of treatment considered, documented and communicated to family. When it became apparent that his situation was hopeless, active treatment was withdrawn. Again, communication with the family was good and documentation was clear.	Pos/C Pos/C
Overall care Score 5	Well looked after. A decent attempt to treat her from a difficult starting point. Good multidisciplinary involvement and good communication with the patient and family throughout.	Pos/C Pos/C Pos/C Pos/C

Table S8 is an example of how the impact of one adverse event may impact on the overall score, even when care is good in other components of care. Just at the point when the patient died, an inappropriate action occurred which the reviewer regarded as very unsatisfactory, although it did not contribute to the patient's death. The reviewer gave high care scores for three phases but then deemed the overall care poor because of the adverse event at point of death.

Table S8

Overall care score 1		Comment type (Pos or Neg) and category
Admission Phase score 5	excellent history taking and clinical examination. Oedema of feet not recorded though is noted in subsequent ward examinations.	Pos/C Neg/ B
Early management Phase score 6	appropriate medical management. Decision taken to manage conservatively as end stage heart failure	Pos/C
Later management Phase score 6	discussion with family about end stage heart failure. Decommissioned ICD as was appropriate	Pos/C
Overall care Score 1	Excellent management as deemed appropriate in this terminal situation. DNR form done by team. Despite this when patient was in periarrest situation, patient was given 1 DC shock. This did not seem appropriate at all as DNR decision was discussed with the family.	Pos/ C Neg/ C