A guide to safety, quality and mortality case note review using the structured judgement method

Prof Allen Hutchinson
Professor Emeritus in Public Health
University of Sheffield

Dr Michael McCooe
Improvement Academy Leadership Fellow / Anaesthetic Registrar

Dr Emma Ryland
Improvement Academy Leadership Fellow / Geriatric Medicine Registrar
Structured Judgement Review

1 Background to the method and its strengths

Hospital Mortality Statistics are widely used to evaluate the issues within a hospital setting and yet only paint part of a picture. Quality and Safety of care cannot be properly evaluated using these numbers. Qualitative information is required and the best way of obtaining this detail is through structured case note review.

In order to provide the benefits to patient care which are commensurate with the effort put into case note review, review methods need to be standardised, yet not rigid, and usable across services, teams and specialties. There is now also recognition that mortality and safety review requires both training and guidance for reviewers.

Structured Judgement Review blends traditional, opinion-based, review methods with a standard format, requiring reviewers to make safety and quality judgements over phases of care, to make explicit written comments about care for each phase, and to score care for each phase*. The result is a relatively short but rich set of information about each case in a form that can also be aggregated to produce knowledge about clinical services and systems of care.

The object of the review method is to look for strengths and weaknesses in the caring process, to provide information about what can be learnt about the hospital systems where care goes well, and to identify points where there may be gaps, problems or difficulty in the care process.

Where care is unsatisfactory the purpose of the review process is not to point to individuals but to ask questions of the system in which people work. Just as importantly, it is also to ask questions about why care goes so well in a complex institution – and what can be learned from this. In order to ask these questions there is a need to look at the whole range of care, at holistic care approaches and the nuances of case management, as well as at the outcomes of interventions.

Structured judgement case note review can be used for a wide range of hospital-based safety and quality reviews, across services and specialties, and not only for those cases of people who die in hospital. For example, it has been used to assess care for people who have had a cardiac arrest, to review safety and quality of care prior to and during non-elective admission to intensive care settings and to review care for people admitted at different times of the week.

A very important feature of the method is that the quality and safety of care is judged and recorded whatever the outcome of the case and that good care is judged and recorded in the same detail as that care which has been problematic. Evidence shows that most care is of good or excellent quality and that there is much to be learned from analysis of high quality care.

Footnote: All clinical examples and structured judgement comments in this document are taken from hypothetical scenarios.

2 How structured judgement review works

2.1 Phases of care

The phase of care structure provides a generalised framework for the review and allows for comparisons among groups of cases at different stages of care. The principal phase descriptors are shown in Box 1. However the use of the phase structure depends much on the type of care and service being reviewed – not all phase of care headings will be used for any particular case. Thus, for example, the procedure-based review forms may not be used in some medical cases but are likely to be used in many surgical cases. In a case where a lumbar puncture is undertaken, or a pleural effusion drained, a care procedure review form might be used but not so a perioperative care form. It is up to the reviewer to judge which phase of care forms are appropriate in a particular case. [See Box 1].

Box 1 Phase of care headings

- Admission and Initial care – first 24 hours
- On-going care
- Care during a procedure
- Perioperative/procedure care
- Discharge care or End of life care; and
- Assessment of care overall

All of the review forms are readily available from the Improvement Academy email Vicky.Padgett@yahhsn.nhs.uk.

In more specialised circumstances, for example reviews of care for people who have a cardiac arrest or who are admitted to an intensive care unit, the Academy has already provided additional appropriate phase of care forms and is always prepared to discuss providing recording forms for other aspects of review.

2.2 Writing explicit judgement statements

The purpose of the reviews is to provide information, from which teams or the organisation can learn. The central part of the review process is therefore comprised of writing short, explicit, judgement statements about the perceived safety and quality of care provided in each care phase.

Explicit judgement commentaries serve two main purposes. First, they allow the reviewer to describe why and how they assess the safety and quality of care provided. Second, they provide commentary in a way that other health professionals can readily understand if they subsequently look at the completed review.

This review guide does not include a glossary of terms that reviewers might use, since this approach would inevitably be constraining or would fail to cover all eventualities in the complexities of clinical practice. Instead, reviewers are asked to use their own words in a way that explicitly states their assessment of an aspect of care and gives a short justification for why they have made the assessment.
Much of the development and training in the structured judgement method has been concerned with enabling prospective reviewers to select and use their own phrases that express that judgement in a way that others can clearly understand. Thus the method moves away from descriptions of care that do not contain a judgement of whether care was good or poor, and away from comments or words where subsequent readers have to imply what the reviewer thought, to explicit statements that use judgement words and phrases such as ‘good’ or ‘unsatisfactory’ or ‘failure’ or ‘best practice’.

Additionally, these judgement words are accompanied by statements that provide an explicit reason why a judgement is made – e.g. unsatisfactory ‘because etc’. For example, ‘Resuscitation and ceiling of care decisions made far too late in course of admission’. The purpose here is not to write long sentences but to encapsulate the clinical judgement in a few explicit statements. Some further examples of explicit judgement comments are shown in Box 2.

**Box 2 Examples of phase of care structured judgement comments**

- Continued omission to provide oxygen and respiratory support
- Team still failed to discuss potential diagnosis with patient
- Referral to ITU was too late
- There was some evidence of good management by overnight team with prompt review and intervention
- Although patient discussed with consultant once and SpR once, for 4 days only seen by SHO. This is completely unsatisfactory

Judgement comments should be made on anything the reviewer thinks important for a particular case. In general this will include the appropriateness of management plans and subsequent implementation together with the extent to which, and where, care meets good practice. In some cases, perhaps in many reviews, there may be care in a phase that has both good and poor aspects. Both should be commented on. Commentary on holistic care is just as important as that on technical care, particularly where complex ceiling of care and end-of-life care discussions might be held. Judgements should be made on how the teams have managed end-of-life decision making and to what extent patients and their relatives have been involved. Thus, for example, a judgement comment might be couched as ‘end-of-life care met recommended practice, good ceiling of care discussion with patient and family’.

Similar approaches and levels of detail are required when care is thought not to have gone well, or where aspects of care are judged only just acceptable. Then words such as ‘unsatisfactory’, ‘poor’ or ‘doesn’t meet good practice standards’ might be necessary.

Sometimes it is just not clear what has been happening during part of the process of care, where judgement words such as ‘delay’, ‘poor planning’, ‘lack of leadership’ may be used. Or if this is due to the level of documentation, comments such as ‘inadequate record keeping’ can also be used.

Overall phase of care comments are intended to bring a focus to the review by asking for an explicit, clear, judgement on what the reviewer thinks of the whole care episode, taking all aspects into consideration. It is not necessary to repeat all of what has been commented on before, though sometimes it is useful to repeat some key messages – that is a reviewer choice. Again, though, it is important to make clear and explicit what the overall judgement is and why.
Examples are offered in Box 3. Although this process may seem like a duplication of effort, a low overall care score and associated commentary is a useful screening tool in order to pick up where difficulties may be occurring in the care process.

**Box 3 Examples of overall care structured judgement comments**

- Overall, a fundamental failure to recognise the severity of this patient’s respiratory failure
- Good multidisciplinary team involvement
- Good documentation on the whole of clinical findings, investigation results management plan and discussion with other teams
- Poor practice not to be aware of DNAR status of patient especially when it has been discussed with family, clearly documented when first put in place and reviewed later on

Cause of death information should form part of the review framework. If, on review, the certified cause of death causes the reviewer some concern, this should be explicitly stated, since there may be a care governance question involved.

So, the overall message about review language is that it should be explicit and clear, in order that you, the reviewer, feels you have made the points clearly and that others who read the review will be able to understand what you have said and why.

### 2.3 Giving phase of care scores

Care scores are recorded after the judgement comments have been written [although it is clear from team training that that there is sometimes an affinity to reverse this order!] and the score is in itself the result of a judgement by the reviewer. Only one score is given per phase of care— it is not necessary to score each judgement statement.

Scores range from Excellent [Score 5] to Very Poor [Score 1] – see Box 4 – and are given for each phase of care commented on and for care overall. These scores have a number of uses.

**Box 4 Phases of care and scores**

1. Very poor care— may have led to severe harm(s) or even death
2. Poor care – may have caused moderate or minor harm(s) or led to patient/family distress
3. Adequate care
4. Good care
5. Excellent care

For the individual reviewer, scores assist in coming to a rounded judgement on the phase of care, particularly when there may be a mix of good and unsatisfactory care within a phase. The reviewer must judge what their overall decision is about the care provided for each phase and for care overall. Scoring makes this very explicit.

Note that it is not uncommon, at least when first using the method, for reviewers to decide on a score, re-read their judgement commentary and then settle on a somewhat different score. This is not a problem – it is part of the rigour of the method.
For the organisation – the multidisciplinary team, the service or the hospital - phase of care scoring provides a form of indicator or screening tool. Where patterns of scores are appearing, high or low, these are pointers for exploration and learning. Good care can provide as many surprises and valuable lessons as can poor care.

It should be acknowledged when talking about scoring that all review methods are based on opinion - on individual judgement. Research shows that where clinical judgement in case note review is concerned, there is about a 70% agreement between clinicians from the same specialty. Therefore it is not unusual, when choosing to use pairs of reviewers, for some aspects of scoring on the same case to differ between the two reviewers.

2.4 Using avoidability of death judgements and care scores in mortality review

A number of recent national mortality review studies in Canada, The Netherlands and England have included a scale that requires the reviewer to assess the likelihood of the avoidability of a death, especially where poor care has been identified. It is likely that this will proposed for use by the NHS in England so the ‘avoidability of death’ scale has therefore been included in the Academy review programme.

However, the Academy considers that the avoidability scale offers only very limited information on its own [from a score of 6 – no evidence of avoidability, to 1 - definite evidence of avoidability] and that much more information is added by including an explicit judgement comment, similar to those used in the main part of the review process.

Making an overall summary judgement on whether death is avoidable (at least to some extent) is a complex process which goes beyond judging safety and quality review by taking into account co-morbidities and estimated life expectancy.

Nevertheless, the experience with a combination of score and an explicit judgement statement suggests that reviewer comments can indeed enhance the information provided in this assessment. The avoidability scale is shown in Box 5, together with an example of an avoidability of death judgement comment.

**Box 5 Avoidability of death scale**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Definitely avoidable</td>
</tr>
<tr>
<td>2</td>
<td>Strong evidence of avoidability</td>
</tr>
<tr>
<td>3</td>
<td>Probably avoidable, more than 50 – 50</td>
</tr>
<tr>
<td>4</td>
<td>Possibly avoidable, less than 50 – 50</td>
</tr>
<tr>
<td>5</td>
<td>Slight evidence of avoidability</td>
</tr>
<tr>
<td>6</td>
<td>Definitely not avoidable</td>
</tr>
</tbody>
</table>

**Example structured judgement commentary**
Non-invasive ventilation management was sub-optimal, but ultimately it was the patient’s wish not to continue treatment. There may have been an alternative cause of breathlessness that was not fully explored or treated, which is why there may have been some avoidability.
Score 5 – slight evidence of avoidability.
2.5 Judging the quality of recording in the case notes

Case note review of course depends critically on the content and the legibility of the records. Safety of care also depends to some extent on good record keeping. Therefore, as part of the overall care assessment, the reviewer is also asked to record their judgement on the quality and legibility of the records, again using a 1-to-5 score.

2.6 Recording the data

Data are recorded on phase of care sheets together with a short descriptive data set. At the time of publication of this guide the data collection sheets are available either for completion on paper or electronically. Additionally, an intranet, Word-based, data base is also under testing which will allow the aggregation of review data and functions such as keyword searching.

2.7 When care is judged poor

In an instance where care is judged poor or lacking or if a death is scored as having strong evidence of avoidability of death then the review should feed into the Trust’s own governance structures for further input.

3 The review in practice

3.1 What is the purpose and focus of the review?

Case note review takes up expensive clinical resource so that time spent on establishing the purpose and desired outcome of the review is important. For instance, reviews have sought to evaluate care for all or some patients who come to a particular service, or have been more extensive and explored the care for the majority of people who die in hospital over a particular time period. Sometimes case selection is used to identify cases where care has been thought problematic, although generalising messages from complex cases can produce ‘solutions’ which may themselves have unintended consequences. Timely review, rather than review after a delay, provides better information.

Under most conditions, and given the constraints on reviewer availability and the need to produce usable information from the reviews, the principle of ‘less is more’ applies. A simple time-based longitudinal sample of around 40 - 50 cases will produce an extensive amount of quantitative and qualitative information on what goes right and what is not working properly.

Time spent on the analysis and information presentation outweighs the benefit of adding a few more cases to the sample. The textual information allows for themes to be developed that then allows a focus for the next improvement steps. Such an approach also has the benefit of being able to learn from, and celebrate, the cases where care has gone well.

3.2 What form should the analysis take?

There are two main areas where the information from structured judgement review may be used – at the individual patient level and at the system or service level where information is aggregated – and there is much to be learned from both of these approaches.
At the individual case level a well-executed review will provide commentary and judgements that can demonstrate how care was provided over the course of the patient’s stay in hospital. Evidence from the large national reviews of safety and quality of care indicates that the majority of people have good or excellent care while in hospital. It is important not to ignore information on good care since this may act to reinforce good practice or provide ideas on local innovations in care management.

Both good care and poor care identified by structured review in individual cases can form the basis of a group discussion on safety and quality of care since structured judgement review provides a framework round which to have that discussion. This can, for instance, be used as the basis for ‘Morbidity and Mortality’ reviews.

For groups of cases, phase of care scores can be used to identify and explore where patterns of good or less satisfactory practice are occurring. Although clinical teams often seek to explain why individual cases may have had unsatisfactory care, patterns present an opportunity to ask ‘why does this happen’ and allows a system approach to problem solving.

In aggregate, the care scores act as indicators of the reviewers’ assessments of safety and quality. For cases grouped together from a specialty or a service, a graphical representation of the scores probably works best, allowing a visualisation of where the spread of the scores lie. For example, scores may be satisfactory to good across all phases of care, or be better in aggregate in some phases than others.

Figure 1 demonstrates a spread within and between care phases, demonstrating a range from excellent to very poor. As is usually the case, there are many more good care cases than there are poor care cases. Although it is likely that review organisations will wish to explore those cases where scores are low, the information from high score cases will also be instructive and provide for valuable learning contrasts between good and poor care.

**Figure 1- Aggregation of Phase of Care Scores - Cardiac arrest care as an example**
Judgement comments will give some understanding of the reviewer’s reasons for a care score. While useful information can be gained from the commentaries in a single review, drawing multiple sets of comments together can provide much more powerful information. For a group of cases it is likely that there will be comments that focus on a number of similar topics or themes – for example good practice in a particular area, or topics such as the timeliness of an intervention or senior review.

Thus the first task is to aggregate the comments together from each care phase, then to identify comments that seem to be referring to similar issues. For example these issues might be on fluid management, or on ceiling of care discussions, or on senior review. Group these comments together as themes and give them a working title [which may change as more data are added]. Start to group the comments together when, say, 10 cases of similar type have been reviewed.

As the extent of the information expands under each theme it is quite likely that there will be a contrasting spread of comments, from critical to very positive, and the positive will often outweigh the critical. The review team is then in a position to ask the ‘why’ questions around these contrasts. Why does care work well in this situation and not in another? What was the background to cases rated poor arising on a particular Saturday night when there appeared to be problems managing deteriorating cases? And so on. An example of some review themes is shown in Box 6.

Box 6 Example of themes derived from a cardiac arrest review

| ▶ Opportunities taken and missed |
| ▶ Early Warning Score recognition |
| ▶ Senior review and case review |
| ▶ Recognition of change |
| ▶ DNACPR management |
| ▶ Documentation |
| ▶ Fluid management |

4 Some practical matters

4.1 How many reviewers should be used?

Organisations vary considerably in the way that reviews are managed, so that some review systems use only single reviewers while others are undertaken by small teams - for example a professional mix such as physicians and nurses. Some hospitals depend on only a few skilled reviewers to undertake large numbers of reviews, while others are turning towards having a larger trained reviewer group. The final decision on who does what is usually a pragmatic one relating to who has the appropriate clinical skills, time and interest.

Nevertheless, the international research evidence suggests that the use of one reviewer per case, using a standardised approach following training, is often the most satisfactory approach. This is because case note reviews are based on clinical judgement and while two reviewers usually agree on key points in a case, in general there is only about 70% agreement between professionals from the same discipline.
The extent of agreement per case goes down further the more reviewers there are per case. From a practical viewpoint, therefore, one reviewer may suffice. If two reviewers are used, for example, where one reviewer is a physician and the other is a nurse, or where a case requires additional specialist clinical review advice, they should both read and judge the whole record.

4.2 Hindsight bias – can we avoid it?

Where the outcome of a process is known before a review begins, there is strong evidence to suggest that everyone brings a bias to the review as the result of that knowledge. Mortality review does not escape this problem, since by definition the outcome is already known. All that can be done here is to remind reviewers that hindsight bias is a problem even for the best reviewers, that the case notes only ever record part of the care process and that what reviewers know now – that the patient died – was not known by the caring team, and was often not anticipated. It is likely that hindsight bias becomes less of a problem where reviews are undertaken on people who do not die during a hospital episode.

4.3 Internal versus external review – which should we choose?

While major national case note review initiatives usually depend on a group of trained external reviewers, the tradition in most hospitals has been to use internal reviewers. Both approaches have their merits. External reviewers bring new eyes to the way care is managed and are unlikely to be swayed by the internal politics of an institution.

On the other hand, although internal review can be criticised for being too close to the care teams, there is some research evidence to suggest that well trained internal reviewers can be as critical and reliable as external review teams. Furthermore, internal reviewers know something of the way the organisation is structured and the way care is delivered. They also know their way through the medical records system. Most importantly, perhaps, internal reviewers provide a continuing learning set for the organisation.¹

¹ The authors wish to thank the members of the Regional Mortality Programme Steering Group for their guidance and support in the production of this Review Guide. Thanks also to colleagues at Sheffield Teaching Hospitals Foundation Trust for permission to use Figure 1.

This guide will be reviewed from time to time. The authorship team welcomes comments on the current version and will be pleased to consider suggestions for new topics or content improvement.