Quality of care in sepsis management: development and testing of measures for improvement

Charis Marwick1, Emily Watts2, Josie Evans3 and Peter Davey3*

1Department of Infectious Diseases, Medicine and Cardiovascular Group, Ninewells Hospital, Dundee DD1 9SY, Scotland, UK; 2University of Edinburgh Medical School, Edinburgh, Scotland, UK; 3Division of Community Health Sciences, College of Medicine, Dentistry and Nursing, University of Dundee, Dundee, Scotland, UK

Received 1 February 2007; returned 5 April 2007; revised 26 May 2007; accepted 6 June 2007

Objectives: To develop and test a set of measures of quality of care in the process of sepsis management, to determine the inter-rater reliability of case-note review in assessment of these measures and to assess our current standard of care.

Methods: Five measures of process of care and one of outcome were identified from the literature review and previous experience. Failure modes and effects analysis was used by a multidisciplinary team to validate these measures and prioritize them in terms of associated risk. Forty sets of case notes were reviewed by two independent teams and the inter-rater reliability was determined using observed percentage agreement and the kappa statistic. We used the data to calculate the proportion of patients in whom we are currently meeting targets for good quality of care.

Results: The multidisciplinary team did not identify any additional areas of concern and assigned the highest risk priority to a delay of over 4 h from recognition of sepsis to antibiotic administration. The inter-rater agreement was >80% for four of the measures, but was only 62.5% for appropriateness of antibiotic therapy. Room for improvement in practice exists, for example, antibiotic administration within 4 h was not achieved in 40% of patients.

Conclusions: Four of our five measures of care are suitable for use in assessing the effect of interventions aimed at improving sepsis management, with at least moderate inter-rater reliability. Specific areas where increased clarity should improve agreement further have been identified.

Keywords: audit, outcome, bacteraemia, treatment delay, patient safety, agreement, kappa statistic

Introduction

Sepsis is estimated to cause 10 000 deaths per year in the UK and appropriate measures of quality of care are required in order to improve its management. Our study had three principal objectives: first, to identify the key processes of care that should be delivered; secondly, to assess the reliability of measures of these processes and thirdly, to establish the current standards of care in sepsis management in our hospital.

Patients and methods

Failure modes and effects analysis

Failure modes and effects analysis (FMEA) is used to assess the risk of failure and harm in processes and to identify the most important areas for process improvements.1 We asked an Infectious Diseases team to discuss five failure modes and allocate each a risk priority number.

(i) After the onset of sepsis:
(a) Was there a delay of ≥24 h in taking a blood culture?
(b) Was there a delay of ≥4 h in the patient receiving an antibiotic?

(ii) Was there a delay in medical review of ≥30 min if the Standardized Early Warning Score (SEWS) was ≥4?
(iii) Was the first dose of antibiotic therapy written for immediate administration?
(iv) Was the initial choice of antibiotic therapy appropriate according to local policy?
The risk priority number for each process was calculated by multiplying scores for three domains: likelihood of occurrence, likelihood of detection and severity of the failure mode. The maximum score for each domain was 10 so the range of possible risk priority numbers was 1–1000.

Reliability of measures of processes of care from case-note review

We used a computer program to randomly select 100 cases from 3571 blood cultures taken between 1 January and 30 April 2006, in order to identify 40 cases with sepsis (the systemic inflammatory response syndrome plus an identifiable site of infection). We measured agreement between two independent groups, each with three members of staff from the infection unit. The observed agreement, expected ‘chance’ agreement, kappa statistic, proportions of positive (\( p_{pos} \)) and negative (\( p_{neg} \)) agreements and bias-adjusted kappa were calculated for each measure with the Diagnostic and Agreement Statistics spreadsheet.\(^2\)

In addition to data about processes of care, we evaluated outcome by identifying any adverse event relevant to sepsis that occurred using the Institute for Healthcare Improvement’s Adverse Event Trigger Tool.\(^3\)

### Results

#### Failure modes and effects analysis

The FMEA meetings did not identify additional areas of concern not already covered in our five measures of quality of care. A delay in antibiotic administration was considered to be the highest risk failure mode, with a risk priority number of 280, and the lack of available nursing observations considered to be the lowest risk failure mode, with a risk priority number of 16.

#### Reliability of measures of processes of care from case-note review

Levels of agreement for objective measures, such as a time delay, were high (82.5% to 95.0%) with kappa of 0.43–0.75, representing at least moderate agreement (Table 1). For more subjective measures, such as appropriateness of antibiotics, agreement was lower (62.5% to 77.5%) with kappa as low as 0.08, representing only slight agreement (Table 2).\(^4\)

### Assessing the current standard of care

There were deficiencies in standards for all measures of quality of care (Table 2). Forty per cent of patients had an unacceptable delay (>4 h) in receiving an antibiotic, the measure with the highest risk priority number assigned by the FMEA team. Adverse events were frequent, with 15 out of 40 patients having an adverse event reported by at least one team. There was no in-patient mortality.

### Discussion

The first aim of this project was to identify and test measures of quality of care important in the management of sepsis. We

### Table 1. Inter-rater reliability for measures of quality of care

<table>
<thead>
<tr>
<th>Measure</th>
<th>Observed agreement (%)</th>
<th>Expected per cent agreement</th>
<th>Kappa (95% CI)</th>
<th>( p_{pos} )</th>
<th>( p_{neg} )</th>
<th>Bias-adjusted kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Delay in blood culture</td>
<td>38/40 (95.0)</td>
<td>90.3</td>
<td>0.48 (−0.12 to 1.08)</td>
<td>0.50</td>
<td>0.97</td>
<td>0.47</td>
</tr>
<tr>
<td>(ii) Delay in medical assessment</td>
<td>33/40 (82.5)</td>
<td>69.5</td>
<td>0.43 (0.07–0.78)</td>
<td>0.53</td>
<td>0.89</td>
<td>0.43</td>
</tr>
<tr>
<td>(iii) Delay in antibiotic administration</td>
<td>35/40 (87.5)</td>
<td>50.5</td>
<td>0.75 (0.54–0.95)</td>
<td>0.86</td>
<td>0.89</td>
<td>0.75</td>
</tr>
<tr>
<td>(iv) Antibiotic prescribed for immediate administration</td>
<td>34/40 (85.0)</td>
<td>58.0</td>
<td>0.64 (0.38–0.90)</td>
<td>0.89</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>(v) Appropriate antibiotic prescribed</td>
<td>25/40 (62.5)</td>
<td>56.5</td>
<td>0.14 (−0.14 to 0.41)</td>
<td>0.74</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td>(vi) Appropriateness of antibiotic—complete agreement</td>
<td>19/29a (65.5)</td>
<td>59.7</td>
<td>0.18 (−0.15 to 0.51)</td>
<td>0.76</td>
<td>0.38</td>
<td>0.14</td>
</tr>
<tr>
<td>(vii) Adverse event</td>
<td>34/40 (85.0)</td>
<td>57.6</td>
<td>0.60 (0.33–0.86)</td>
<td>0.72</td>
<td>0.88</td>
<td>0.60</td>
</tr>
<tr>
<td>(viii) Adverse event related to sepsis</td>
<td>6/9b (66.7)</td>
<td>54.3</td>
<td>0.27 (−0.19 to 0.73)</td>
<td>0.77</td>
<td>0.40</td>
<td>0.17</td>
</tr>
<tr>
<td>(ix) Both vii and viii (above)</td>
<td>31/40 (77.5)</td>
<td>64.5</td>
<td>0.42 (0.09–0.76)</td>
<td>0.56</td>
<td>0.87</td>
<td>0.42</td>
</tr>
</tbody>
</table>

\(^{a}\)The denominator of 29 reflects the number of cases for which there was agreement on the actual antibiotic prescribed.

\(^{b}\)The denominator of 9 reflects the number of cases for which there was agreement on the actual adverse event that occurred.

### Table 2. Meeting of targets for process measures of quality of care

<table>
<thead>
<tr>
<th>Measure</th>
<th>Target met (%)</th>
<th>Target not met (%)</th>
<th>Information not available (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delay in blood cultures?</td>
<td>37/38 (97.4)</td>
<td>1/38 (2.6)</td>
<td>0/38 (0)</td>
</tr>
<tr>
<td>Delay in medical assessment?</td>
<td>29/33 (87.9)</td>
<td>4/33 (12.1)</td>
<td>0/33 (0)</td>
</tr>
<tr>
<td>Delay in antibiotic administration?</td>
<td>20/35 (57.1)</td>
<td>14/35 (40)</td>
<td>1/35 (2.9)</td>
</tr>
<tr>
<td>Antibiotic prescribed for immediate administration?</td>
<td>25/34 (73.5)</td>
<td>9/34 (26.5)</td>
<td>0/34 (0)</td>
</tr>
<tr>
<td>Antibiotic appropriate?</td>
<td>16/19 (84.2)</td>
<td>3/19 (15.8)</td>
<td>0/19 (0)</td>
</tr>
</tbody>
</table>
identified five measures from the previous literature and no additional areas of concern were identified by our teams. The calculation of a risk priority number gives a quantitative assessment of the level of risk attributed to each failure mode (Table 1). We recognize that our multidisciplinary teams included members from Infectious Diseases and that further investigation with teams from other disciplines is required.

The second aim of the study was to test the reliability of case-note review as a method of assessment of measures of quality of care in the management of sepsis. The kappa statistic compares the observed agreement with the agreement that could be expected to occur by chance. The Cochrane Effective Practice and Organisation of Care Group define a reliable primary outcome measure as at least 90% agreement between two or more raters or kappa greater than or equal to 0.8. However, there is considerable inconsistency in the level of kappa that is used to indicate acceptable agreement, for example, a kappa of 0.4 has been advocated for indicators of the quality of care of pneumonia. Rather than selecting a single, arbitrary threshold for acceptability, it may be preferable to report agreement over the whole range of kappa from 0 to 1.

A kappa value becomes unreliable when the incidence rate for the discrete variable being measured is low, thus it is important to present the underlying observation rates in addition to the agreement and kappa statistic. The low kappa scores achieved for some of our measures (e.g. 0.47 for delay in blood culture being taken) were initially surprising because the level of agreement was high (95%). The explanation is that there were very few cases in which there was a delay of over 24 h from onset of sepsis to having blood cultured. Therefore, when asked whether or not this delay had occurred, the reviewers could have predicted the answer before even looking at the case notes, so similar agreement could have easily occurred by chance.

The most complex and subjective measure with the lowest inter-rater reliability was the appropriateness of antibiotic therapy. Previous studies have reported kappa values for inter-rater reliability for composite quality indicators that are frankly not credible, for example, ‘>0.98 in all cases’. When data abstraction is carried out by trained abstractors, the inter-rater reliability is likely to be the best possible, yet kappa values may be as low as 0.39 for composite quality indicators. The key issue is whether or not the actual agreement is sufficient for the application of the quality indicator, it is essential to report the actual levels of agreement, not just kappa.

To try and improve the accuracy of data abstraction, we used teams of three and then compared agreement across these teams. However, previous research has questioned the value of using teams of two or more people to extract data. Two independent, single reviewers may, therefore, be sufficient for data abstraction and tests of inter-rater reliability.

The main strengths of this study are that we have involved a multidisciplinary clinical team in a rigorous analysis of the reliability of key processes of care for patients with sepsis. Strengths in the analysis and reporting include the use of multiple reliability indicators and the reporting of raw data as well as per cent agreement and kappa.

An important weakness of the study was that we pre-selected patients with sepsis for detailed case-note review rather than reviewing a random sample of patients. This meant that we could not calculate the incidence of sepsis in the underlying (sample) population, which would have made our kappa values more reliable. A second weakness is that the number of case notes reviewed was too small to achieve clinically significant kappa values in subgroup analyses. The decision to analyse 40 sets of notes was based on power calculations, but these had not taken into account the possibility of requiring subgroup analyses.

Conclusions and further work

Four of five of our measures of quality of care are suitable for the assessment of quality of care with at least moderate agreement as defined by kappa. The reliability could be improved further by increasing the clarity of start points in questions regarding timing. The measure regarding appropriateness of antibiotic prescribed needs further development.

Acknowledgements

Many thanks for the time and effort in carrying out FMEA and the case-note reviews: Stuart Bell, Keith Eldridge, Catherine Houlihan, Karen Kendall, Lorraine Milton, Emily Morrison, Arlene Shaw and Claire Sloan. Also thanks to Ross Martin for providing the list of patients with blood cultures and to Simon Ogston for his instruction in statistical analysis.

Funding

This study was supported with funding from the NHS Tayside Acute Services Division Grant Scheme.

Transparency declarations

Within the last 2 years, C. M. was sponsored by Wyeth Pharmaceuticals to attend a meeting. P. D. has served on Advisory Boards for Johnson&Johnson (Global Anti-infectives) and Wyeth (UK tigecycline), received honoraria for speaking from Johnson&Johnson, Optimer, Pfizer and Wyeth and received research funding from Boehringer Ingelheim, GlaxoSmithKline and Pfizer. E. W. and J. E. have no conflicts of interest to declare.

References

Quality of care in sepsis management


