‘Failure to rescue’ as a measure of quality of hospital care: the limitations of secondary diagnosis coding in English hospital data

Martin McKee, James Coles and Philip James

Summary

Although it is widely recognized that quality of care varies between hospitals, a robust and valid measure of outcome that can be used in comparisons has proven elusive. One measure that has recently been proposed by US researchers is the ‘failure to rescue’ (FTR) rate. This is based on the assumption that, whereas complications may reflect both patient severity and health care factors, the ability to save patients once complications arise is much more closely related to the quality of health care. We describe an evaluation of FTR in a national sample of English hospitals using hospital episode data. We found that the rate of secondary diagnosis recording in England is about one-tenth that in the United States. The FTR rate would be highly sensitive to variations in the completeness of coding of secondary diagnoses. Unless coding is of uniformly high quality, any attempt to compare severity adjusted outcomes will be potentially unreliable.

Keywords: hospitals; quality of care; outcomes; administrative data; United Kingdom

Introduction

It has been recognized for over 100 years that the outcome of care varies between hospitals.1 By the 1990s, after many years of only intermittent attention, it had risen high on the political agenda with a growing demand for information that would inform patients and those purchasing care on their behalf about where to seek treatment. This has now led to publication of measures of outcome by hospital or physician, most frequently in the form of death rates. Examples include the publication of survival following myocardial infarction, by the Scottish Health Department,2 and following cardiac surgery, by the health authorities in New York.3

These efforts have been criticized on several counts. The criticisms can be divided broadly into two areas. The first are technical concerns,4,5 such as the instability arising from analysis of small numbers or the dangers of confounding, with different hospitals treating patients of differing levels of severity but where the analyst has limited capacity to adjust for this differing risk. The second are managerial concerns, such as how to determine what should be done in response to the information obtained and the scope for opportunistic behaviour.6

The issue of risk adjustment is especially problematic. A recent study of outcome following gastrointestinal haemorrhage showed how rankings of hospitals changed considerably once information on co-morbidity, extracted from routine data, was incorporated, but then changed further once information obtained at endoscopy was added.7 On the basis of experience with data from non-randomized evaluative studies, it is likely to be almost impossible to adjust adequately for differences in risk of mortality using routinely available hospital discharge data.8

Silber et al. have proposed a possible means of overcoming the problem of risk adjustment. In brief, their approach is based on the principle that the probability of a patient dying in hospital is a function of both their clinical condition, or pre-admission severity, and of the quality of care provided. Traditionally, differences in the clinical condition have been adjusted for by means of information on co-morbidity, as far as is possible with the data available. It is then assumed that residual variation between either clinicians or hospitals reflects difference in quality of care. Silber et al. contend that an alternative approach can be developed based on a separation of the development of complications from whether or not the patient survives them. They argue that the development of complications is influenced considerably by the patient’s condition on admission, which is outside the control of the clinical team. In contrast, survival following development of a complication is a valid measure of the quality of care provided.

To operationalize this they have coined the term ‘failure to rescue’ as a measure of quality of hospital care: the limitations of secondary diagnosis coding in English hospital data

Martin McKee, Professor of European Public Health
CASPE Research, King’s Fund, 11–13 Cavendish Square, London W1M 0AN.
James Coles, Director, Research
CHKS Ltd, 1 Arden Court, Arden Road, Alcester B49 6HN.
Philip James, Director
Address correspondence to Professor McKee.

© Faculty of Public Health Medicine 1999
rescue’ (FTR), which they define as death of a patient who has developed one or more of a specified list of complications. They surmise that low FTR rates will be associated with good clinical care. They note the importance of differentiating, for example, hospitals that admit patients with a high rate of complications, but are very good at managing them, from others whose patients have a low frequency of complications, but that have a poor record of managing them. Both hospitals, whose case-mix may differ simply because of admission thresholds, might do well on simple measures of mortality. FTR thus obviates the need to take other aspects of case-mix into account.

In studies undertaken in the United States, differences in complication rates have been shown to be poorly correlated with differences in mortality. Furthermore, differences in complication rates are associated with patient severity but this is not the case for FTR, which is, instead, associated with hospital characteristics, such as the percentage of board certified anaesthetists. From these findings, Silber et al. argued that FTR offers a possible solution to some of the problems associated with either mortality or complication rates as measures of outcome. Furthermore, they found that it performed at least as well as existing severity measures based on condition on admission. They have not, however, examined whether their measure is of value when comparing individual hospitals.

Comparisons of hospital performance are acquiring much greater prominence in the United Kingdom, not least because of public disquiet in the light of events in Bristol. The government has published a new performance management framework in which it states: ‘As part of the Performance Framework, we will develop and publish sophisticated measures of clinical quality on a specialty by specialty named hospital basis’ (emphasis added). It continues, ‘We need to ensure we are comparing like with like. So, over time, figures will need to be “risk adjusted” to standardise for factors such as age, severity, case-mix, and concurrent illness.’ It is widely accepted that existing measures of case-mix used in the United Kingdom have limited capacity to adjust adequately for severity. The government accepts that further work is needed but that it is only ‘... over time, [that] figures will need to be “risk adjusted” ’ (emphasis added).

This study, which is part of a pilot study for an examination of links between hospital organizational factors and clinical outcome, has examined whether FTR could be used as a measure of quality of care in English NHS hospitals.

If such a measure is to be used routinely, it must be valid when derived from routinely collected data. Such validation would require a detailed examination of case notes or, preferably, prospectively collected clinical data. However, the cost of such an exercise would be high, and it is prudent first to see whether it is possible to gain some idea of its possible performance by means of an analysis of existing, routinely collected data. As there is no ‘gold standard’ for quality of care using routinely collected data, the method chosen must be indirect. We have developed a pragmatic approach that seeks to describe how FTR behaves using English data and then to explore construct validity by identifying criteria that, if such a measure were to perform as intended, should, in general, be met. If, on the basis of such an evaluation, FTR was seen to hold the potential to differentiate good and poor quality care in England, a further, more detailed evaluation would be required, but if it failed to meet these initial criteria, such an evaluation would be unnecessary.

Methods

Hospital discharge data, in the format based on the English NHS Contract Minimum Data Set (CMDS), have been collected by CHKS Ltd from hospitals in the United Kingdom, as part of a continuing programme of quality assurance. Data are supplied by the hospitals to CHKS, which inspects, cleans and removes any inconsistencies with reference to the hospital management.

Initially, the two index procedures selected by Silber et al. were examined: trans-urethral resection of the prostate (TURP) and cholecystectomy. However, it immediately became clear that, in a particular hospital, the numbers involved were too small for meaningful analysis. Consequently, all general surgical and urological episodes in which an operative procedure was recorded were examined.

Death rates and complication rates were calculated as the number of deaths (or complications) in relation to all finished consultant episodes. FTR rates were calculated from deaths in episodes in which complications were recorded.

ICD-9 codes corresponding to each of the adverse occurrences listed by Silber et al. were identified (see Table 1). Data relate to the calendar years 1996 and 1997. Analyses were undertaken using Excel and SPSS packages. In regression analyses, hospitals or years with fewer than five deaths were excluded (this excluded urology in 25 hospitals in 1996 data, and 28 hospitals in 1997).

The first question was whether secondary diagnoses are recorded equally in English and US data. Clearly, if the level of recording differs markedly, the scope for transferability is limited.

Second, do FTR rates increase with age to the same extent as do mortality and complication rates? Complications, and thus mortality, are strongly age dependent, whereas the ability to rescue a patient after the development of a complication might be expected to be less so. This information is also necessary to decide whether one must adjust for age as, in the US study, patients under 65 years of age were excluded.

A third issue is whether FTR rates correlate with complication rates. It would be expected that they would do so only weakly, on the basis of the premise that the development of complications is a function of the patient’s pre-admission condition as well as the quality of the care they receive, whereas the FTR rate reflects only the latter. In contrast, if they correlate closely it might suggest that there was a bias towards greater
coding of complications among patients who died rather than those of equal severity who survived.

Fourth, how do the three measures behave from year to year? Do hospitals with high FTR rates in one year also have high rates the next? How does this compare with crude mortality or with complication rates?

Fifth, do hospitals with high or low FTR rates have identifiable features, such as those in an inner city location? These hospitals have faced particular pressures in recent years, with high staff turnovers and dependence on agency staff, which might be expected to lead to a poorer quality of care. Do teaching hospitals differ on any of the measures?

Results

Data were available for the two years, on 954 536 general surgery episodes and 430 005 urology episodes, from 73 and 66 hospitals, respectively. The mean number of general surgical episodes with a surgical procedure recorded per hospital per year was 6643, of which 41.6 per cent (SD 11.6) had complications recorded and 17.6 per cent (SD 5.8) died. The corresponding figures for urology were 3259, of which 18.9 per cent of cases (SD 7.9) had complications recorded and 5.7 per cent (SD 3.4) died. The rates per thousand episodes of each of the designated complications are shown in Table 1, first, for all ages for both specialties and, second, as the US study was limited to patients 65 and over, for patients in both specialties combined and in the same age group in data sets from both countries. In England, considering all ages, complications are infrequent, affecting less than 1 per cent of episodes. When the same age groups are compared, frequencies in England are generally much less than in the United States, with the exception of atrial fibrillation and flutter.

The distribution of the three measures is shown in Table 2. Table 3 shows the effect of age on the three measures. As expected, the measures of death rates and complication rates increase markedly with increasing age. The FTR rate shows no consistent effect.

The correlation between the three measures is shown in Table 4. FTR tends to be only weakly associated with death

| Table 1 Overall complication rates in each specialty per 1000 episodes |
|--------------------------|------------------|------------------|------------------|
| Complication             | Urology          | General surgery  | Surgery and urology - age 65 and over | USA |
| Pulmonary embolism (I26.x) | 0.28             | 0.67             | 0.98             | 1.8 |
| Second-degree heart block (I44.1) | 0.04         | 0.02             | 0.05             | 1.3 |
| Third-degree block (I44.2)     | 0.09             | 0.06             | 0.16             | 0.5 |
| Cardiac arrest (I46.x)        | 0.28             | 0.93             | 1.45             | 10.9 |
| Ventricular tachycardia (I47.2) | 0.04           | 0.04             | 0.08             | 0.5 |
| Atrial fibrillation and flutter (I48.x) | 5.29         | 6.58             | 13.71            | 13.6 |
| Heart failure (I50.x)         | 3.38             | 4.70             | 9.69             | 21.1 |
| Pleural effusion (J90.x)      | 0.37             | 1.87             | 2.06             | 25.8 |
| Pneumothorax (J93.x)          | 0.11             | 0.28             | 0.24             | 1.7 |
| Renal dysfunction (N99.0)     | 0.09             | 0.28             | 0.41             | 21.8 |
| Wound infection (T81.4)       | 1.07             | 3.61             | 3.88             | 7.2 |
| Pneumonia (J13.x, J14.x, J15.x, J16.x, J18.x) | 0.65      | 2.75             | 4.21             | 27.6 |
| Stroke (I61.x, I62.x, I63.x, I64.x) | 0.12  | 0.36             | 0.56             | 2.7 |

Source of US data: calculated from data in Ref. 10.
rates (urology in 1997 is an exception) and is inversely associated with complication rates. The pattern is reasonably consistent in both years and for both specialties. The consistency of the measures over time is shown in Table 5. Complication rates were most strongly correlated, followed by death rates and then by FTR rates.

The distribution of various characteristics of hospitals when ranked in order of FTR rates is examined in Table 6. For general surgery, teaching hospitals and those in large cities (recognizing that these characteristics are highly correlated) constituted a greater proportion of hospitals in the category with the highest FTR scores than in that with the lowest scores. However, the mean FTR rates did not differ significantly between hospitals in towns and small cities, outer areas of large cities, and inner cities (11.0 per cent, 11.8 per cent and 12.0 per cent, respectively) (Fig. 1) or between teaching and non-teaching hospitals (11.2 per cent and 11.9 per cent, respectively).

**Discussion**

The challenge of finding a way in which routinely available data might be used to differentiate hospitals according to the quality of care they provide is tantalizing. ‘Failure to rescue’ initially seemed to offer a way forward. Conceptually, it is extremely attractive, as it overcomes many of the known limitations of severity adjustment systems. However, in this evaluation using English hospital data, this promise has been less than adequately fulfilled. Recorded complication rates are substantially lower than those in the United States. This could be due to a less severely ill population of patients in England or to differences in coding. That the latter is important is suggested by examining the pattern of specific complications. It is noteworthy that atrial fibrillation and flutter, which is much more clear cut than many other complications and which requires a specific, often long-term, treatment, is as common in England as in the United States. In contrast, the difference is especially large for those complications where there might be expected to be most discretion about diagnosis and that are more transient, such as renal dysfunction, which, in the United States, will often include transient fluid and electrolyte disturbance. Similarly, it is conceivable that ‘pleural effusion’ in the United States...
include many patients who have some transient filling of the
costo-phrenic angle on a chest X-ray.

In the United States there are powerful incentives to record
factors that enhance the apparent severity of the patient, as was
seen when New York began to publish risk-adjusted mortality
rates by surgeon for cardiac surgery and the recorded rates
of certain complications increased dramatically.6

The observation that death rates are poorly correlated with
FTR rates is consistent with the argument that death rates are
a function of pre-admission severity and quality of care. How-
ever, the inverse relationship between complications and FTR
requires further consideration. One possible explanation is
that this indicates a systematic difference in secondary diag-
nosis coding, with those hospitals recording low complication
rates only coding complications in patients who die.

The observation that complication rates were more closely
correlated from one year to the next than were death or FTR
rates suggests a systematic bias in relation to depth of coding
between hospitals.

Any system designed to compare hospital performance on
the basis of mortality faces major difficulties because of the
small numbers involved. These would be reduced even further
if only a few procedures were included. FTR, by focusing on
complications rather than a procedure, offers the potential to
overcome this difficulty.

Unfortunately, these results suggest that FTR is unable to
fulfil its full potential in the United Kingdom at present.

However, some findings demand further examination. In
particular, the apparent variation in the intensity of coding of
secondary diacoses among hospitals tends to swamp any
variation in FTR rates. This is consistent with findings from
an earlier study that examined severity adjusted rankings of
English hospitals according to surgical death rates, in which
some hospitals consistently failed to code complications and
co-morbidities, so that comparisons were substantially
distorted.4

There are, however, some encouraging findings. The ap-
aparent independence of FTR from age is important, suggesting
that it is relatively independent of pre-admission severity,
and there may be some scope for using it as a screening method
in association with simple checks on the depth of secondary
diagnosis coding, possibly narrowing the range of complica-
tions that are considered.

In recent years, the National Health Service has poured
resources into the process of collecting information. There has,
however, been much less emphasis on how this information
might be used, with many of the intrinsic definitional prob-
lems not being resolved.11 The emphasis on publication of
measures of performance throughout the public sector in the
United Kingdom has brought to the fore the issue of case-
mix or, as it is popularly referred to, measures of ‘added value’.
In judging hospital care, this can work only if there is some
degree of consistency in the coding of secondary diagnoses.
If this is to be achieved, it will require a major managerial
effort. Furthermore, such efforts face the intrinsic problem
that recording complications is especially susceptible to a
wide range of biases, such as a hospital providing low-quality
care failing even to diagnose complications, as well as the
scope for opportunistnic behaviour if such data are to feed into
a performance management framework. None the less, we
believe that the concept on which FTR is based is sound, and
if the quality of secondary coding could be improved, it
would justify further evaluation, although perhaps in the
context of prospectively collected clinical databases, with
agreed definitions, rather than the existing hospital episode
data system.

Although the findings of this study are largely negative,
we believe that they are of interest because they highlight,
once again, the importance of the quality of coding of
secondary diagnoses, and not just the easier to measure
indicators such as the frequency of records with no primary
diagnosis.

Table 6 Characteristics of hospitals ranked by FTR rates (values are percentages)

<table>
<thead>
<tr>
<th></th>
<th>Urology</th>
<th>General surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low FTR (10 lowest in each year)</td>
<td>High FTR (10 highest in each year)</td>
</tr>
<tr>
<td></td>
<td>Large city</td>
<td>North</td>
</tr>
<tr>
<td>Low FTR (10 lowest in each year)</td>
<td>33.3</td>
<td>40.0</td>
</tr>
<tr>
<td>High FTR (10 highest in each year)</td>
<td>36.4</td>
<td>36.4</td>
</tr>
</tbody>
</table>

Figure 1 FTR rates in hospitals in different settings (general surgery, 1997).
Acknowledgement

This study was funded by a grant from the Nuffield Provincial Hospitals Trust.

References


Accepted on 29 June 1999