An assessment of failure to rescue derived from routine NHS data as a nursing sensitive patient safety indicator

(Report to Policy Research Programme)

Simon Jones
Alex Bottle
Peter Griffiths

November 2010

Acknowledgements

This is an independent report commissioned and funded by the Policy Research Programme in the Department of Health. The views expressed are not necessarily those of the Department. The project was undertaken by the National Nursing Research (NNRU) Unit at Kings College, London, the Dr Foster Unit at Imperial College and Dr Foster Intelligence. We would like to thank Dr Bryan Mcintosh for his substantial help in making revisions to this report.

Contact address for further information:
National Nursing Research Unit
King's College London
James Clerk Maxwell Building
57 Waterloo Road
London SE1 8WA

nnru@kcl.ac.uk

http://www.kcl.ac.uk/schools/nursing/nnru

Contents

Summary

Background

"Failure to rescue" refers to a failure to prevent a clinically important deterioration resulting from a complication of an underlying illness (e.g. cardiac arrest in a patient with acute myocardial infarction) or a complication of medical care (e.g. major haemorrhage after surgery). Rates of failure to rescue are widely recognized and used as patient safety indicators but have not been widely used in the NHS. Recently they have received renewed attention after the National Institute for Health and Clinical Excellence (NICE) released a clinical guideline for recognition of and response to acute illness in adults in hospital (Armitage et al., 2007) following reports that deteriorating patients frequently experienced delays in receiving adequate care leading to significant numbers of avoidable deaths (National Patient Safety Agency, 2007). Because of the role of nurses in early identification of deterioration, failure to rescue is potentially highly sensitive to nursing care and it has been widely advocated as a potential nursing sensitive outcome indicator (Griffiths et al 2008).

There is a variety of indicators used to determine rates of failure to rescue. Most commonly, the indicators are mortality rates amongst a subset of patients who experience a complication, which, although life threatening, is amenable to successful treatment if there is timely and appropriate intervention. Generally, the indicators are based on information collected from hospital administrative data. Because the indicators rely on the identification of a group of patients who experience particular complications, the validity of such indicators can be compromised if coding of secondary diagnoses (such as complications) is poor. The under recording of secondary diagnoses in such databases is a known issue and previously, McKee et al (1999) reported that English Hospital Episode Statistics (HES) from 1996/7 and 1997/8 were unsuitable for deriving failure to rescue measures, primarily because of low rates of coding.

In response to such limitations two alternative approaches to deriving the indicators have been proposed, both of which attempt to get around the problem of low coding of complications. The first, proposed by the originator of failure to rescue indicators, is to assume that all surgical deaths are in fact examples of failure to rescue since, in most cases, death results from a complication whether or not such a complication is recorded in the administrative data set (Silber 2007). The resulting indicator means that providers with poor coding are not 'rewarded' by excluding some deaths from the 'basket' of cases considered for failure to rescue. The second approach is predicated on the recognition that not all failures to rescue will result in a death but that complicated recoveries are longer than uncomplicated ones. If failure to rescue results in serious deterioration, which in turn leads to extended hospital stay then stays which fall well outside the norm can also be used as an (indirect) indicator of failure to rescue (Rafferty et al 2007).

This report provides a preliminary assessment to determine if it is possible to derive valid "failure to rescue" indicators for surgical patients from Hospital Episode Statistics (HES) or similar datasets based on routinely collected hospital data in England. While there is considerable interest in failure to rescue indicators for both medical and surgical patients there has been significantly more validation among surgical patients. While early definitions of failure to rescue patient safety indicators from the US Agency for Healthcare Research and Quality (AHRQ) included medical patients (Agency for Healthcare Research and Quality, 2003), they were dropped from subsequent versions due to the problem of adequately identifying failure to recue cases using administrative data sets. Furthermore, while researchers have consistently found associations between quality inputs, such as high staffing levels, and failure to rescue in surgery, evidence of such association for medical patients is generally lacking (Kane et al 2007).

Aims and Objectives

The primary purpose of this project was to assess if there was evidence of a substantive change in coding practice in England that might significantly alter conclusions about the potential to derive a valid failure to rescue indicator for surgical patients and to assess the relative methods of alternative approaches to deriving FTR indicators derived from English data.

We consider three definitions of failure to rescue for among surgical patients:

- The AHRQ definition which counts deaths among a subset of surgical patients experiencing certain complications (FTR-A)
- A revised indicator based on Silber et al (2007) which includes all surgical deaths as cases of failure to rescue (FTR-S)
- A 'long stay' based indicator which counts all patients with abnormally long (longer than the 75th percentile) hospital stays for the diagnosis / procedure as failure to rescue (FTR-L).

Summary of FTR Implementations used in this paper

Designation	Denominator	Numerator
FTR-A	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission *who experienced a complication* listed in Appendix 1	*Number of deaths* amongst *surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission *who experienced a complication* listed in Appendix 1
FTR-S	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission *who experienced a complication* listed in Appendix 1	*Number of deaths* amongst surgical patients (elective or emergency) aged between 18 and 90 who were operated on within two days of admission (*whether or not they were recorded as experiencing a complication*)
FTR-L	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission.	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission *who stay longer than the 75th percentile* for their HRG

The assessment of validity is based on the approaches undertaken by McKee et al (1999) and Silber et al (2007). Specifically it considers:

- Whether coding of secondary diagnoses has increased since the previous assessment – indicating improved potential for deriving mortality based indicators
- How rates derive from English data compare with the US – in order to determine if findings from UK data are plausible
- Whether FTR rates are associated with age, rates of complications and coding practices – in order to determine sources of bias and the need for further risk adjustment
- Whether FTR rates are stable over time – indicating that it is a relatively stable property, as would be expected if it indicates quality
- Whether FTR is associated with Hospital Standardised Mortality (HSMR) – to assess whether there are plausible associations with other recognised mortality indicators and whether FTR indicators give additional information
- Whether FTR rates are associated with hospital factors that are generally supported as linked to quality (e.g. nurse staffing levels, teaching status) – to assess the validity of the claim that the indicator reflects some dimension of underlying 'quality'.

Methods

The research used hospital discharge data from the Commissioning Data Sets (CDS) data from April 1997 to March 2009 to identify all surgical admissions (including day cases) to general acute hospitals. The failure to rescue 'basket' for FTR-A and FTR-S was based on the AHRQ definition

> *"surgical discharges age 18 years and older ... defined by specific DRGs or MS-DRGs and an ICD-9-CM code for an operating room procedure, principal procedure within 2 days of admission OR admission type of elective (ATYPE=3) with potential complications of care (e.g., pneumonia, DVT/PE, sepsis, shock/cardiac arrest, or GI hemorrhage/acute ulcer) (Agency for Healthcare Research and Quality, 2009).*

Because there is a different coding system in use in England, (ICD 10 / OPCS) we had to 'translate' the AHRQ indicator specification using automated mapping combined with direct inspection of the codebooks. For FTR-A the failure to rescue cases were those among this group of patients who died and for FTR-S it was all deaths among surgical patients (with or without a recorded complication). For FTR-L, failure to rescue cases were any patients with hospital stays longer than the 75th percentile for that patient's HRG.

We used bivariate correlations and multiple Poisson regression models to determine associations between FTR rates and a range of quality related organisational factors derived from routine NHS data. These factors included staffing variables such as the numbers of nurses and doctors per bed, teaching status and nursing staff stability. Because levels of different staff groups were highly correlated with each other, we could not include staff to bed ratios for all groups of staff in simultaneously in the regression models. Therefore, our regression models considered total numbers of professionally qualified staff (doctors and nurses) per bed and the relative numbers (ratios) of each staff group to nurses to give an indication of the effect of both staffing levels and skill mix.

Results

- Between 1997/8 and 2008/9, we found 66,100,672 surgical admissions of whom 442,462 (0.7%) died and 2,496,356 (3.8%) had an eligible complication for FTR-A or FTR-S, of whom 226,237 died (0.3%). 4,993,863 (7.6%) were classified as long stay.
- FTR rates were 9.1% (FTR-A), 17.7% (FTR-S) and 7.6% (FTR-L).
- The level of secondary diagnosis coding has increased substantially since previous reports. The median percentage of surgical admissions with at least one secondary diagnosis recorded has increased from 26% in year 1997/8 to 40% in 2008/9.
- FTR rates are age related. For example, there is a 12-fold increase in FTR-A rates from the youngest (18-39) to oldest (75+) age groups which is not reflected in US FTR rates calculated using a similar definition. FTR-A rates for England are lower in all age groups than in the US.
- FTR rates seem to be stable over time. The inter-year correlations between 2007/8 and 2008/9 were high for all FTR indicators (FTR-A 0.917, FTR-S 0.723, FTR-L 0.940).
- None of the FTR indicators are significantly correlated with coding depth (average number of secondary diagnoses coded).
- Mortality based FTR rates (FTR-A, FTR-S) were associated with a number of hospital characteristics that are plausibly associated with quality, for example in regression models:
    - Hospitals with more professionally qualified clinical staff per bed were associated with lower rates of failure to rescue
    - Higher bed occupancy was associated with higher rates of failure to rescue
- However some results ran counter to expectation, for example:
    - a more stable nursing workforce was associated with *higher* rates of failure to rescue
    - Teaching hospitals had *higher* failure to rescue (FTR-S only)
- FTR-L was also plausibly associated with quality in some respects, for example in regression models.
    - a more stable nursing workforce was associated with lower failure to rescue
    - higher bed occupancy was associated with higher levels of failure to rescue
- However, again there were some unexpected findings for example:

    - Hospitals with more professionally qualified clinical staff per bed were associated with *higher* rates of failure to rescue

- All FTR indicators showed more variation than would be expected by sampling error alone.

Conclusion

A number of our findings suggest that low levels of recording of secondary diagnoses in hospital administrative data, previously a significant reason for concluding that mortality based failure to rescue indicators could not be derived from English data, is no longer a significant obstacle. Rates of secondary coding have increased substantially over the years studied. Whereas McKee *et al.* (1999) found that failure to rescue rates showed a relatively low year on year correlation we found strong year on year correlations for all our failure to rescue rates over recent years. This suggests a degree of stability consistent with the rate reflecting an underlying characteristic of hospital performance. Whilst the quality of data has clearly improved, there is still scope for improvement amongst some trusts, although FTR rates are not significantly correlated with a trust's coding depth (average number of secondary diagnoses coded), which suggests an absence of systematic bias caused by coding practices.

While we found a number of associations between the failure to rescue indicators and presumed markers of quality, some of which have been demonstrated in previous US work, not all the relationships observed are clearly or plausibly indicative of variations in quality. One of the proposed advantages of failure to rescue over other mortality measures is that 'rescue' is within the control of the hospital whereas other factors, for example the development of a complication in the first place, are more related to patient characteristics and other factors outside the control of the hospital. This is proposed as a solution to the problem of risk adjustment. However, we found substantial over-dispersion in funnel plots suggesting that there was much more variation than would be expected by chance alone. It is possible that such over-dispersion is as a result of inadequate risk adjustment or other institutional factors independent of quality (Spiegelhalter, 2005).

Although we adjusted for some organisational and system characteristics outside the control of the hospital (for example percentage of cases with a comorbidity and level of GP coverage) it seems likely that there are other confounding variables that were not accounted for in our models. For example, we found a clear pattern of increasing rates of both FTR-A and FTR-S with increasing age and we conclude that any failure to rescue indicator would need to be risk adjusted for age and other patient level factors before coming to a final judgement. Non-risk adjusted FTR rates cannot be used to make comparisons between providers.

Although FTR-L performed well in some respects, the absence of specific issues relating to secondary coding (which this was designed to avoid) for the other indicators means that there was no clear advantage for this over the mortality based measures (FTR-A and FTR-S). The advantage is further undermined by the need for additional risk adjustment, which will likely rely on secondary coding. As an indicator, it is potentially linked to a number of factors beyond the control of the hospital, such as provision of community services and was subject to more over-dispersion that either FTR-A or FTR-S.

Because FTR-S counts all surgical deaths as part of the numerator but not the denominator it provides an active disincentive to trusts to 'game' by under coding secondary diagnoses and thus omitting them from the indicator. However, because it fails to add the additional cases to the denominator it is inherently inaccurate and will tend to exaggerate FTR-S rates in hospitals with poor coding. FTR-S showed more over-dispersion that FTR-A and we conclude that FTR-A is the most promising indicator, although further work is required.

It is notable that previous US work showed that higher nurse staffing was associated with lower levels of mortality based failure to rescue measures (FTR-A and FTR-S e.g. Silber et al 2007). Although our bivariate correlations on these measures showed a similar association, the association disappeared in our regression models. Although higher levels of clinically qualified staff were associated with lower levels of FTR-A and FTR-S, a higher nurse to doctor ratio was associated with higher rates of failure to rescue, suggesting that medical staffing might be more significant. However, the purpose of assessing these associations in this paper was primarily to validate the FTR indicator and the inconsistent relationships found raise questions as to whether the indicators are truly measuring quality in their current form. Our results do not establish any of the FTR indicators as being specifically nurse sensitive.

Implications for Policy

- **There is potential to derive mortality based failure to rescue indicators for surgical patients from routine administrative data in England**
- **Failure to rescue indicators offer some advantages over standardised mortality measures such as HSMR for surgical patients**
- **Our FTR-A indicator, based on the AHRQ definition, is worthy of further exploration as a potentially valid safety measure**
- **Unadjusted failure to rescue cannot be used to compare the quality of care between NHS trusts**
- **FTR-L may have a role as a system level indicator although its precise meaning remains unclear**
- **Failure to rescue does not appear to be a specifically nurse sensitive indicator although this conclusion should be regarded as tentative**

Implications for research

- **Further assessment is required to develop and test risk adjustment models**
- **A case note audit needs to be performed to verify the extent that cases identified by the indicator are indeed cases of 'failure to rescue'**

- **The indicator could be further validated against the successful implementation of quality improvement initiatives which focus on interventions to reduce failure to rescue (e.g. global track and trigger tools)**
- **FTR-L should be examined in patients discharged home only, where external drivers of stay may be fewer**

Introduction

"Failure to rescue" refers to a failure to prevent a clinically important deterioration resulting from a complication of an underlying illness (e.g. cardiac arrest in a patient with acute myocardial infarction) or a complication of medical care (e.g. major haemorrhage after surgery). Rates of failure to rescue are widely recognized and used as patient safety indicators but have not hitherto been used in the NHS. Recently they have received renewed attention after the National Institute for Health and Clinical Excellence (NICE) released a clinical guideline for recognition of and response to acute illness in adults in hospital (Armitage et al., 2007) following reports that deteriorating patients frequently experienced delays in receiving adequate care leading to significant numbers of avoidable deaths (National Patient Safety Agency, 2007). Because of the role of nurses in early identification of deterioration, failure to rescue is potentially highly sensitive to nursing care and it has been widely advocated as a potential nursing sensitive outcome indicator (Griffiths et al 2008).

This report provides a preliminary assessment to determine if it is possible to derive valid "failure to rescue" indicators for surgical patients from Hospital Episode Statistics (HES) or similar datasets based on routinely collected hospital data in England.

Background

The Next Stage Review of the NHS (Department of Health, 2008) placed great emphasis on the use of information on clinical performance and quality as a key tool to guide and incentivise improvement and innovation in the NHS. Subsequent to this report a range of initiatives have progressed which either support quality measurement for local clinical use (e.g. the "Measuring for Quality Improvement Programme'' and regional quality observatories) or use measures to incentivise quality (e.g. the Commissioning for Quality and Innovation payments). Subsequent developments place increased emphasis on the use of outcomes, rather than process targets.

The developments of the Next Stage Review gave further impetus to ongoing developments, which have seen increasing use of clinical performance data at the level of organisations, specialities, and clinical services. In the UK notable examples have been the publication of unit-level mortality statistics by the Society for Cardiothoracic Surgery, the publication of Hospital Standardised Mortality Ratios (HSMR) by Dr Foster and NHS surveys of patient experience (Marshall et al., 2003). In part inspired by the perceived success of these developments and in part concerned that a focus on some parts of the system would lead a decline in quality in others (Marshall et al., 2003) there has been increasing interest in developing measures which can meaningfully demonstrate the quality of nursing care, particularly in acute care settings. Several systems of quality indicators designed to focus specifically on the quality of nursing care have been developed, primarily in the USA. However, our review of these indicators suggested that there was limited evidence for the associations between so-called nurse-sensitive outcomes and variation in the quality and quantity of nursing care available to patients (Griffiths et al., 2008).

Some of the strongest evidence for associations between nursing inputs and patient outcomes relates to mortality. A number of studies have observed increases in mortality associated with lower levels of nurse staffing. A systematic review of observational studies estimated that 4.2% of deaths in acute hospitals could be associated with variations in registered nurse staffing if the relationships observed were causal (Kane et al., 2007). Kane and colleagues found the evidence for causality to be relatively strong with conclusions robust to variation in study designs and some evidence of a dose-response and temporal association. However, it is unlikely that the relationships observed are completely without confounding and so this estimate should be regarded as a maximum. Research undertaken in the UK suggests that medical staffing levels are more strongly associated with mortality than nurse staffing and that there is a strong correlation between the two variables (Jarman et al., 1999). While the evidence establishes that nursing may have a significant contribution to make to this important outcome, using mortality rates as a direct and specific indicator of nursing care quality is hard to sustain.

Our review identified an alternative measure, where there was also strong evidence for an association with nursing. *'Failure to rescue'*, originally defined as death after a complication (Silber et al., 1992), has been widely used as a nurse-sensitive outcome. It has been adopted as a valid patient safety indicator by the US Agency for Healthcare Research and Quality (AHRQ) (Agency for Healthcare Research and Quality, 2007). As with mortality, an association between low levels of nurse staffing and high levels of failure to rescue is supported by a meta-analysis of observational studies. The attributable fraction of events associated with nurse staffing (16%) in surgical units is relatively high (Kane et al., 2007).

The 'theory' behind the failure to rescue indicates identifies a clear, specific, and plausible causal contribution for nursing. While complications are predominantly a result of patient characteristics, the ability of a hospital to successfully treat (rescue) a patient given the complication is largely a result of the quality of care provided (Silber et al., 2007). Nursing's specific contribution relates to the ability to detect early signs of deterioration and take appropriate actions, including communicating well with other colleagues that ensure that they are able (and willing) to respond in a timely and appropriate manner. The potential significance of this measure is reflected in recent reports and research into responses to deteriorating patients in acute care (Luettel et al., 2007) which emphasise the complexity of response and the numerous potential points of failure including:

- not *taking* formal / informal observations
- not *recording* observations
- not *recognising* early signs of deterioration or appreciating their significance
- not *communicating* observations causing concern
- not *responding* appropriately.

The issue of 'failure to rescue' is a high priority for the NHS. In 2007 The National Institute for Health and Clinical Excellence (NICE) released a clinical guideline for recognition of and response to acute illness in adults in hospital (Armitage et al., 2007) following reports that deteriorating patients frequently experienced delays in receiving adequate care leading to significant numbers of avoidable deaths (National Patient Safety Agency, 2007). Care of the deteriorating patient is currently a focus of the NPSA / NHSI 'Patient Safety First' campaign (Feinmann, 2009). In accord with the theoretical proposition supporting failure to rescue as a quality indicator there is evidence that failure to rescue measures are more closely associated with hospital characteristics and less influenced by patient characteristics than mortality (Silber et al., 2007). Thus failure to rescue offers a partial solution to the problems of case mix/risk adjustment that remain so controversial for mortality measures such as HSMR (Mohammed et al., 2009). In addition to being associated with registered nurse staffing, failure to rescue has been associated with a range of potentially quality-related hospital characteristics including teaching hospital status, hospital size and skill mix of nursing staff (Silber et al., 2007).

Our previous report recommended further work to assess the validity and utility of failure to rescue as a quality indicator for nursing in England (Griffiths et al., 2008). The Next Stage Review of the NHS defined quality of care in terms of three dimensions: patient experience, clinical effectiveness, and patient safety. Failure to rescue is intended to measure the patient safety dimension of quality.

Failure to rescue indicators

There remain competing specifications for failure to rescue indicators. These include those supported by the Association for Healthcare Research and Quality (AHRQ) in the US ("death among surgical in-patients with serious treatable complications"), the original definition developed by Silber and colleagues (Silber et al., 1992) and a more recently developed alternative which uses extended hospital stays as a proxy measure of failure to rescue (Rafferty et al., 2007). Most commonly, the indicators are mortality rates amongst a sub-set of patients who experience a complication, which, although life threatening, is amenable to successful treatment if there is timely and appropriate intervention. Generally, the indicators are based on information collected from hospital administrative data. Because the indicators rely on the identification of a group of patients who experience particular complications, the validity of such indicators is potentially compromised if coding of secondary diagnoses (such as complications) is poor.

The under recording of secondary diagnoses in such databases is a known issue and previously McKee et al (1999) reported that English Hospital Episode Statistics (HES) from 1996/7 and 1997/8 were unsuitable for deriving failure to rescue measures. McKee's study raised a number of objections to deriving failure to rescue from these data in England at the time. Levels of secondary diagnostic coding for some conditions in the failure to rescue definition were markedly lower than in the US where the indicator was validated. Although this could reflect real differences in case mix, this is unlikely to explain fully the differences. For some conditions with clear and objective parameters (e.g. atrial fibrillation) the level of secondary coding was nearly identical between the US and England whereas others (e.g. renal dysfunction) were coded over 50 times more often in US data.

However, while problematic, low levels of secondary coding do not necessarily invalidate the indicator and are not the only criterion against which it should be judged. McKee also assessed whether failure to rescue was associated with age and complication rates. He argued that the rationale behind failure to rescue suggested that it should show low correlations with age and complications compared with mortality if it truly reflects quality of care more strongly than the patient's underlying condition. McKee also assessed the stability of rates of failure to rescue and complications over time and the association between features of hospitals that might be related to quality, such as hospitals with high staff turnover and teaching status. McKee's findings cast doubt on the validity of the indicator at that time, although the main problem identified was the lack of accurate secondary diagnostic coding.

In response to such limitations, two alternative approaches to deriving the indicators have been proposed, both of which attempt to get around the problem of low coding of complications. The first, proposed by the originator of failure to rescue indicators, is to assume that all surgical deaths are in fact examples of failure to rescue since, in most cases, death results from a complication whether or not such a complication is recorded in the administrative data set (Silber et al., 2007). The resulting indicator means that providers with poor coding are not 'rewarded' by excluding some deaths from the 'basket' of cases considered for failure to rescue. The second approach is predicated on the recognition that not all failures to rescue will result in a death but that complicated recoveries are longer than uncomplicated ones. If failure to rescue results in serious deterioration, which in turn leads to extended hospital stay, then stays which fall well outside the norm can also be used as an (indirect) indicator of failure to rescue (Rafferty et al., 2007).

Given these competing approaches and the fact that several policy initiatives in following years, including Payment by Results (PbR), have strongly incentivised improvement in the quality of patient data and so, the previous objections to FTR may have been resolved in this report we consider three definitions of failure to rescue for among surgical patients:

- The AHRQ definition which counts deaths among a subset of surgical patients experiencing certain complications (FTR-A)
- A revised indicator based on Silber et al (2007) which includes all surgical deaths as cases of failure to rescue (FTR-S)
- A 'long stay' based indicator which counts all patients with abnormally long (longer than the 75^{th} percentile) hospital stays for the diagnosis / procedure as failure to rescue (FTR-L).

While there is considerable, interest in failure to rescue indicators for both medical and surgical patients there has been significantly more validation among surgical patients. While early definitions of failure to rescue patient safety indicators from the US Agency for Healthcare Research and Quality (AHRQ) included medical patients (Agency for Healthcare Research and Quality, 2003), they were dropped from subsequent versions due to the problem of adequately identifying failure to recue cases using administrative data sets. Furthermore, while researchers have consistently found associations between quality inputs, such as high staffing levels, and failure to rescue in surgery, evidence of such association for medical patients is generally lacking (Kane et al 2007).

Thus, the primary purpose of this project was to assess if there was evidence of a substantive change in coding practice in England that might significantly alter conclusions about the potential to derive a valid failure to rescue indicator for surgical patients and to assess the relative methods of alternative approaches to deriving FTR indicators derived from English data. Assessment of indicator validity is based on the approaches undertaken by McKee et al (1999) and Silber et al (2007) and considers associations between failure to rescue indicators and other variables which are generally associated with quality or which have a predictable relationship with a valid failure to rescue indicator. Further, we sought evidence that failure to rescue was specifically linked to nurse staffing as opposed to other staff groups in order to determine its possible role as a nurse sensitive indicator. We consider:

- Whether coding of secondary diagnoses has increased since the previous assessment – indicating improved potential for deriving mortality based indicators
- How rates derive from English data compare to the US – in order to determine if findings from UK data are plausible
- Whether FTR rates are associated with age, rates of complications and coding practices – in order to determine sources of bias and the need for further risk adjustment
- Whether FTR rates are stable over time – indicating that it is a relatively stable property as would be expected if it indicates quality
- Whether FTR is associated with Hospital Standardised Mortality (HSMR) – to assess whether there are plausible associations with other recognised mortality indicators and whether FTR indicators give additional information
- Whether FTR rates are associated with hospital factors that are generally supported as linked to quality (e.g. nurse staffing levels, teaching status) – to assess the validity of the claim that the indicator reflects some dimension of underlying ‘quality’.

Methods

Data sources

We used hospital discharge data from the Commissioning Data Sets (CDS) which includes a summary record of all NHS hospital care to identify all surgical admissions (including day cases) to general acute hospitals. Data were available from 147 acute NHS trusts from April 1997 to March 2009. The data were initially processed by Imperial College then passed in an anonymised form to Dr Foster Intelligence. Where trusts had merged, all historical data had been mapped to the new trust by Imperial prior to our analysis. Less than 1% of records were rejected due to data quality reasons.

The data included all episodes in an NHS general acute hospital under a consultant from a surgical speciality that had at least one surgical procedure. In the CDS, a single stay in hospital can be made up from several "consultant episodes" as a patient is transferred between services and specialties in the hospital. To identify a surgical admission, consultant episodes were first grouped into spells (a complete admission) using the NHS Information Centre's HRG grouper. We searched for a surgical procedure in the first two consultant episodes and checked that the procedure took place within two days of the admission date.

From these data, we derived three failure to rescue indicators based on (see box 1 for a summary).

- The AHRQ definition which counts deaths among a subset of surgical patients experiencing certain complications (FTR-A)
- A revised version of the AHRQ indicator based on Silber et al (2007), which includes all surgical deaths as cases of failure to rescue (FTR-S)
- A 'long stay' based indicator which counts all patients with abnormally long (longer than the 75th percentile) hospital stays for the diagnosis / procedure as failure to rescue (FTR-L).

We based our definition of complications that might lead to failure to rescue on the Agency for Health Research and Quality patient safety indicator "Death among Surgical Inpatients with Serious Treatable Complications" (previously called failure to rescue) (Agency for Healthcare Research and Quality, 2007, Agency for Healthcare Research and Quality, 2009).

> *".. surgical discharges age 18 years and older ... defined by specific DRGs or MS-DRGs and an ICD-9-CM code for an operating room procedure, principal procedure within 2 days of admission OR admission type of elective (ATYPE=3) with potential complications of care (e.g., pneumonia, DVT/PE, sepsis, shock/cardiac arrest, or GI hemorrhage/acute ulcer) (Agency for Healthcare Research and Quality, 2009).*

Because there is a different coding system in use in England, (ICD 10 / OPCS), we had to 'translate' the AHRQ indicator specification. We mapped the AHRQ specification onto the NHS data dictionary (http://www.datadictionary.nhs.uk/) and the ICD 9 codes for complications by inspection of the ICD 10 codebook. We supplemented our 1CD-9 to ICD-10 mapping by an inspection of codes generated by an automated translation from ICD 9 to ICD 10. We did not rely on the automated translation exclusively because of known issues with such mapping (Schulz et al., 1998). See appendix 1 for detail of the mappings and final specification.

Designation	Denominator	Numerator
FTR-A	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission *who experienced a complication* listed Appendix 1	*Number of deaths* amongst *surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission *who experienced a complication* listed Appendix 1
FTR-S	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission *who experienced a complication* listed Appendix 1	*Number of deaths* amongst surgical patients (elective or emergency) aged between 18 and 90 who were operated on within two days of admission (*whether or not they were recorded as experiencing a complication*)
FTR-L	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission.	*Number of surgical patients* (elective or emergency) aged between 18 and 90 who were operated on within two days of admission. *who stay longer than the 75th percentile* for their HRG

Box 1: Summary of FTR Implementations used in this paper

FTR-A

Our first definition is based as closely as possible on the then current AHRQ specification *(Agency for Healthcare Research and Quality, 2009).* This definition is an expanded version of what Silber refers to as FTR-N (Silber et al., 2007), which was originally developed to better focus on nursing quality of care and was deployed in studies which identified a link between nurse staffing and patient outcome (Needleman and Buerhaus, 2007). This definition identifies the proportion of deaths among surgical patients aged between 18 and 90 who, whether elective or emergency admissions who were operated on within two days of admission and who experienced one of the potential complications of care (appendix 1).

FTR-S

We also derived an alternative FTR indicator using the same denominator (i.e. surgical patients who experienced one of the potential complications of care), but including *all* surgical deaths in the numerator. Silber argues that this approach has advantages over FTR-A as almost all surgical deaths are the result of a complication (Silber et al., 2007). If a complication is not coded in a patient who died, it is assumed to have occurred but remained unrecorded. For our assessment purposes this indicator has the additional advantage of reducing bias that might arise if coding practices are related to the outcome, for example if complications were more likely to be coded where a patient died.

FTR-L

Subsequent to McKee's work, which rejected FTR indicators derived from routine data in the UK, an alternative or proxy indicator of failure was proposed. Rafferty et al argued that in presence of poor levels of clinical coding, the number of patients experiencing abnormally long hospital stay could be used as a proxy for failure to rescue (Rafferty et al., 2007) because many delayed stays are caused by complications that are not successfully / swiftly resolved. Because it is based on primary procedure coding, it is not affected at all by secondary coding.

For this indicator, FTR-L, the numerator is the number of surgical patients with stays longer than the 75th percentile for that patient's Healthcare Resource Group (HRG). HRGs are standard groupings of clinically similar treatments, which use common levels of healthcare resource. HRGs were assigned using the NHS HRG classifications (v3.5).

Validation and control variables

We identified a number of organisational characteristics to use as validation variables. These factors had been identified or suggested by previous studies (e.g. McKee et al 1999, Jarman et al 1999, Silber et al 2007) as being associated with safety related quality of care. We also identified organisational factors outside the control of the acute hospital that need to be controlled for in order to ensure proper comparison. Control variables were primarily factors used in previous assessments of the link between organisational characteristics and HSMR (Jarman et al., 1999) which need to be considered in examining any link between quality and organisational characteristics. We also considered percentage of patients with a failure to rescue complication as a control for variation in coding practices, since this number will be influenced by the extent to which secondary diagnoses are coded in a trust.

These data were obtained from Dr Foster and the NHS Information Centre (see appendix 2 for detailed list of data sources). The validation variables used were teaching status (university hospital flag), hospital size (number of inpatient beds), bed occupancy, nursing staff stability (proportion of staff at year-end still employed a year later) and numbers of professionally qualified clinical staff and clinical support staff. We also obtained HSMR data for each of the included trusts from Dr Foster Intelligence. Variables used as control variables were London hospitals, number of people who died in hospital as a percentage of all deaths and GPs per 100,000 population. Because levels of different staff groups were highly correlated with each other, we could not include staff to bed ratios for all groups of staff in simultaneously in regression models. Therefore, our regression models considered total numbers of professionally qualified staff per bed and the relative numbers (ratios) of each staff group to nurses to give an indication of the effect of both staffing levels and skill mix.

Analysis

All analysis was undertaken using R 2.10.1 software. We used bivariate correlations and multiple Poisson regression models to determine associations between FTR rates and quality related organisational factors derived from routine NHS data. Cameron &. Trivedi (1998) recommend using Poisson regression when modelling problems such the count of number with people with failure to rescue. In such models, the count is offset against a measure of the number of people at risk. For all Poisson regression models, the log transfer function was used. For FTR-A and FTR-S, the log of the number of people with FTR complications was used as the offset. For the FTR-L model, the log of the total number of surgical cases was used. So that the coefficients within each model could be compared, the independent variables were standardised as z scores prior to being used in the model so that model coefficients give an accurate indication of the relative influence of each variable on the dependant variable (FTR). Backwards stepwise regression, based on minimising the Bayesian Information Criterion (Cameron and Trivedi, 1998), was used to remove variables that contributed little to the model.

A key assumption of multiple regression is that the factors included in the model are not highly correlated with each other. If this assumption is not met, the model becomes invalid. The Belsey collinearity diagnostic (Belsey, 1991) indicated that doctors per bed, nurses per bed and clinical support staff per bed were highly correlated (i.e. the model suffered from collinearity). Therefore, we calculated a new variable, total number of clinically qualified staff per bed. In order to examine the contribution of each individual staff group we created variables giving the relative numbers (ratios) of each staff group to nurses to give an indication of the effect of team composition. The Belsey diagnostic confirmed that this revised approach did not suffer from collinearity.

Results

Volume of cases, depth of cases and change over time

Between 1997/8 and 2008/9, we found 66,100,672 surgical admissions. The total number of secondary diagnoses that are recorded in this group has increased over time. The median percentage of surgical admissions to a trust with at least one secondary diagnosis recorded was 26% in year 1997/8 and had increased to 40% in 2008/9. The 25th percentile for 2008/9 (33.3%) is higher than the median in 1997/8 (Figure 1).

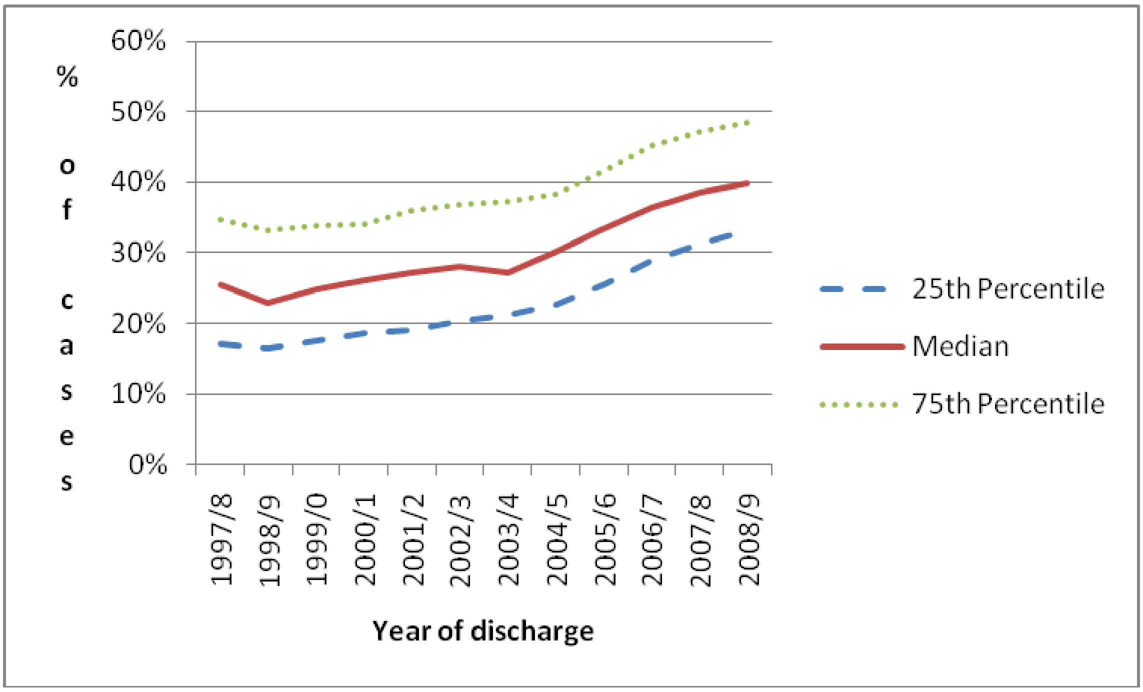


Figure 1: Percentage of surgical patients with a secondary diagnosis by year, 1997/8 to 2008/9

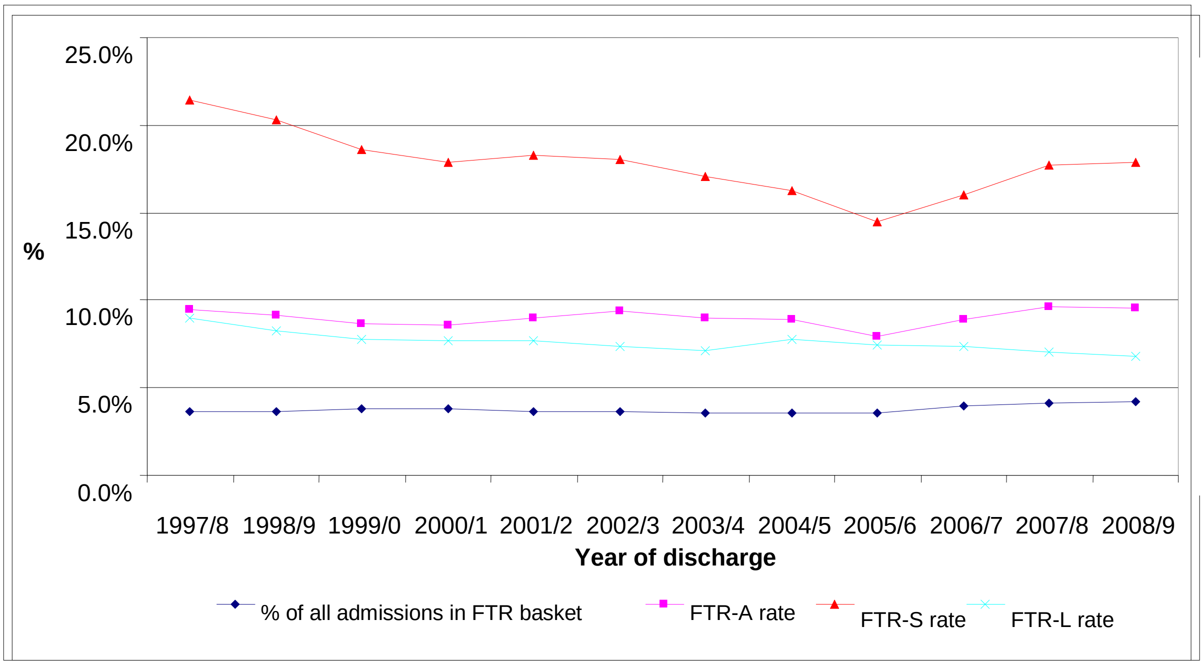
Of these, 2,496,356 admissions (3.8%) were included in the FTR-A / FTR L denominator by virtue of having an eligible complication. We refer to these cases as being in the FTR 'basket'. Of these 226,237 died (0.3%). 4,993,863 (7.6%) were classified as long stay (table 1). As secondary coding has increased, so the number of people counted in the FTR 'basket' has also increased over time. In 2008/9, 4.2% of all surgical admissions had a recorded FTR complication compared with 3.6% in 1997/8. Over the same time, the number of deaths fell from 0.8% to 0.7% of all surgical admissions. In 2008/9 the median number of patients in the FTR basket was 1,801 per trust (IQR 1,328-2,574) and the median number of deaths in that group was 186 (IQR 126-257).

Table 1 Surgical admissions, FTR complications, and FTR rates 1997-2009

	All years		April 1997 to March 1998		April 2008 to March 2009	
	Count	% of number of eligible surgical admissions	Count	% of number of eligible surgical admissions	Count	% of number of eligible surgical admissions
Total number of eligible surgical admissions	66,100,672	100.0%	4,465,970	100.0%	7,015,024	100.0%
All eligible surgical deaths	442,462	0.7%	34,852	0.8%	52,531	0.7%
Long stay patients (FTR-L)	4,993,863	7.6%	401,628	9.0%	478,588	6.8%
Patients with FTR complications	2,496,356	3.8%	162,400	3.6%	294,235	4.2%
Patients with FTR complications who died	226,237	0.3%	15,436	0.3%	28,128	0.4%
FTR-A rate	9.1%		9.5%		9.6%	
FTR-S rate	17.7%		21.5%		17.9%	
FTR-L rate	7.6%		9.0%		6.8%	

Figure 2 shows how the various FTR indicators have changed over time. If all FTR indicators steadily increased in line with coding depth, it might raise a concern that this is all they measure. FTR-A, which is most susceptible to coding depth, has reminded broadly static being 9.5% in 1996/7 and 9.6% in 2008/9, whereas the other indicators have reduced over this time. FTR-S has reduced from 21.5% to 17.98%. FTR- L has reduced from 9.0% in 1997/8 to 6.8% in 2008/9.

Figure 2: Change in FTR indicators over time, 1997/8 to 2008/9



FTR indicators derived for recent years are stable year on year. The inter-year correlation between 2007/8 and 2008/9 for FTR-A was 0.917 (p<0.001) and for FTR-L the inter-year correlation was 0.940 (p<0.001). For FTR-S the inter-year correlation was somewhat lower, although still high at 0.723 (p<0.001).

Comparison across countries and by age groups

Table 2: English and US FTR rates by age and gender

Demographic	English data (2008-09)			US rate 2007
	Numerator	Denominator	Rate (Adjusted rate)*	
FTR-A	*FTR-A deaths*	*FTR complications*		
Male	14,969	154,748	9.7%	-
Female	13,159	139,487	9.4%	-
Age 18 to 39	624	40,238	1.6%	6.9%
Age 40 to 64	5,126	112,823	4.5%	12.1%
Age 65 to 74	5,727	60,287	9.5%	14.6%
Age 75+	16,651	80,887	20.6%	-
FTR-S	*deaths*	*FTR complications*		
Male	27,040	154,748	17.5% (16.2%)	
Female	25,491	139,487	18.3 %(16.8%)	
Age 18 to 39	1,289	40,238	3.2 %(3.2%)	6.9%
Age 40 to 64	9,436	112,823	8.4 %(8.1%)	12.1%
Age 65 to 74	10,672	60,287	17.7 %(16.4%)	14.6%
Age 75+	31,134	80,887	38.5 %(32.6%)	
FTR-L	*Long stayers*	*admissions*		
Male	177,353	3,010,543	5.9%	
Female	301,213	4,004,035	7.5%	
Age 18 to 39	138,146	1,795,844	7.7%	
Age 40 to 64	130,623	2,629,355	5.0%	
Age 65 to 74	82,528	1,293,160	6.4%	
Age 75+	127,291	1,296,665	9.8%	

* Unlike Silber's definition (FTR-A), the adjusted rate included all deaths in the denominator as well as the numerator

Table 2 compares the age and gender breakdown of the various FTR indicators for England and the USA... If the English rates are similar to the previously validated US rates, it might strengthen their credibility. There are marked differences in levels of FTR in the UK compared with the most recently available US data. FTR-A rates for England were lower in all age groups than in the most recently available US data (Agency for Healthcare Research and Quality, 2007).

Silber's approach to deriving FTR (FTR-S) adds all surgical deaths to the numerator (as they are presumed to have suffered an unrecorded complication) but does not add them to the denominator of all those experiencing complications. This will have the tendency to exaggerate FTR-S rates in the face of poor coding of complications. We derived an adjusted FTR-S rate by adding the deaths without complications to the denominator as well. The adjusted FTR-S is nearly the same as the FTR-S rate for younger age groups but the difference increased steadily with age, reaching a maximum difference of 5.9% in the 75+ age group (table 2). In all but the oldest age group rates of FTR-S were still lower than US rates.

Both US and English data show a trend of increasing with age for the mortality based FTR indicators but this is more marked for the English data where there is a 12-fold increase in FTR-A rates from the youngest (18-39) to oldest (75+) age groups (table 2). However, there was no clear trend for age with FTR-L with rates in the middle age groups (40-74) lower than in the youngest (18-39) and oldest age groups (75+).

Correlates of FTR

Table 3 is the correlation matrix between a trust's various failure to rescue indicators, complication rate, Hospital Standardised Morality Ratio (HMSR) and the percentage of patients' records with at least one secondary diagnosis recorded. FTR-A and FTR-S show a significant strong positive correlation (r=0.854). Neither is correlated with FTR-L (r<0.1). FTR-S shows a significant but weak positive correlation with HSMR (r=0.187) but FTR-A and FTR-L show no correlation with HSMR (r<0.1).

Table 3: Correlation matrix for trust-level indicators for 2008/9

		FTR-A	FTR-S	FTR-L	HSMR	Complication rate	Coding depth
FTR-A	Pearson Correlation	1	0.854(**)	0.012	0.037	-0.174(*)	0.085
	P		<0.001	0.882	0.661	0.035	0.305
FTR-S	Pearson Correlation	0.854(**)	1	0.049	0.187(*)	-0.323(**)	0.107
	p	<0.001		0.554	0.024	<0.001	0.198
FTR-L	Pearson Correlation	0.012	0.049	1	-0.036	0.305(**)	0.006
	p	0.882	0.554		0.665	<0.001	0.942
HSMR	Pearson Correlation	0.037	0.187(*)	-0.036	1	0.003	-0.148
	p	0.661	0.024	0.665	<0.001	0.976	0.075
Complication rate	Pearson Correlation	-0.174(*)	-0.323(**)	0.305(**)	0.003	1	0.150
	p	0.035	<0.001	<0.001	0.976		0.070
Coding depth	Pearson Correlation	0.085	0.107	0.006	-0.148	0.150	1
	p	0.305	0.198	0.942	0.075	0.070	

*significant at the 0.05 level, **0.01 level (2-tailed). N=147 except for HSMR where data are available from 145 trusts

None of the indicators is significantly correlated with overall coding depth for secondary diagnoses. There are weak to moderate associations between FTR rates and the proportion of patients with FTR complications recorded (the FTR "basket"). FTR-A shows a significant weak negative correlation with the proportion of cases in the FTR "basket" (r=-0.174) while FTR-S has a significant moderate negative correlation with the proportion of cases in the FTR basket (r=-0.323). FTR-L shows a moderate significant positive correlation with the complication rate (r=0.305) (table 3).

Tables 4-5 give the bivariate and regression associations and coefficients for the variables retained in the regression model. Bivariate correlations (uncontrolled for the effects of other variables) are given in columns 2-4. The regression model, showing relationships when controlled for the influence of other variables, is given in columns 5-7. There were significant bivariate associations, in the expected direction, between FTR-A and most of the validation variables, including levels of nurse staffing. A higher number of nurses per bed was associated with lower failure to rescue. However, the relationship between the nurse stability index (a measure of staff turnover) and FTR-A was significant but in the opposite direction to that expected – i.e. the more stable the workforce the higher the rate of FTR-A.

In the regression model, larger hospitals (more beds) and more professionally qualified clinical staff per bed were associated with lower rates of failure to rescue. Higher bed occupancy was associated with higher rates of failure to rescue. Although not significant in the bivariate correlations a higher proportion of nurses in the clinical workforce (more nurses per doctor) was associated with higher rates of failure to rescue in the regression model and a more stable nursing workforce (higher nurse stability index). The relationship between the nurse stability index (a measure of staff turnover) and FTR-A remained significant and in the opposite direction to that expected with higher stability associated with higher failure to rescue.

Table 4: Poisson regression model for FTR-A for 2008/9 data

Variable	Bivariate Correlation			Regression Correlation		
Validation variables	Relative risk	95% CI	p	Relative risk	95% CI	p
Professionally qualified clinical staff per bed	0.94	(0.93-0.96)	<0.001	0.96	(0.95-0.98)	<0.001
Doctors per bed	0.95	(0.94-0.97)	<0.001			
Nurse per bed	0.94	(0.93-0.95)	<0.001			
Support to clinical staff per bed	0.99	(0.98-1.01)	0.224			
Nurse per doctor	1.00	(0.99-1.01)	0.777	1.04	(1.03-1.06)	<0.001
Support to clinical staff per nurse	1.05	(1.04-1.06)	<0.001			
Hospital is Teaching Hospital	0.92	(0.90-0.95)	<0.001			
Number of beds	0.96	(0.95-0.97)	<0.001	0.95	(0.94-0.96)	<0.001
Nurse stability index	1.05	(1.04-1.07)	<0.001	1.04	(1.03-1.06)	<0.001
Average % occupancy	1.05	(1.03-1.06)	<0.001	2.17	(1.70-2.77)	<0.001
Control variables						
% cases with a co morbidity	1.01	(1.00-1.02)	0.057			
% of patients with an FTR Complication	0.003	(0.001-0.01)	<0.001	0.95	(0.94-0.97)	<0.001
GPs per 100,000 population	0.95	(0.94-0.96)	<0.001	0.98	(0.97-1.00)	0.005
Hospital is in London	0.88	(0.85-0.91)	<0.001	0.88	(0.84-0.92)	<0.001
Number of people who died in hospital as a % of all deaths	1.01	(1.00-1.02)	0.147	1.04	(1.03-1.06)	<0.001
Number of discharges from hospital	0.96	(0.95-0.97)	<0.001			

Similar results were obtained for FTR-S except that teaching hospital status was also included in the multi-variable model; contrary to expectation, these hospitals were associated with higher failure to rescue rates (table 5).

Table 5: Poisson regression model for FTR-S for 2008/9 data

Variable	Bivariate Correlation			Regression Correlation		
Validation variables	Relative risk	95% CI	p	Relative risk	95% CI	p
Professionally qualified clinical staff per bed	0.95	(0.94-0.96)	<0.001	0.96	(0.95-0.97)	<0.001
Doctors per bed	0.96	(0.95-0.96)	<0.001			
Nurse per bed	0.95	(0.94-0.96)	<0.001			
Support to clinical staff per bed	1.00	(0.99-1.01)	0.452			
Nurse per doctor	1.01	(1.00-1.01)	0.166	1.03	(1.02-1.04)	<0.001
Support to clinical staff per nurse	1.06	(1.05-1.07)	<0.001			
Hospital is Teaching Hospital	0.98	(0.97-1.00)	0.018	1.11	(1.09-1.14)	<0.001
Number of beds	0.98	(0.98-0.99)	<0.001	0.96	(0.95-0.97)	<0.001
Nurse stability index	1.06	(1.05-1.07)	<0.001	1.03	(1.02-1.05)	<0.001
Average % occupancy	1.03	(1.02-1.04)	<0.001	1.79	(1.53-2.10)	<0.001
Control variables						
% cases with a co morbidity	1.01	(1.00-1.02)	0.022			
% of patients with an FTR Complication	0.00	(0.00-0.00)	<0.001	0.92	(0.91-0.93)	<0.001
GPs per 100,000 population	0.96	(0.95-0.97)	<0.001	0.98	(0.97-0.99)	<0.001
Hospital is in London	0.84	(0.82-0.86)	<0.001	0.86	(0.84-0.89)	<0.001
Number of people who died in hospital as a % of all deaths	1.00	(0.99-1.01)	0.966	1.03	(1.02-1.04)	<0.001
Number of discharges from hospital	0.98	(0.98-0.99)	<0.001			

FTR-L showed a somewhat different set of relationships with the validation variables (table 6). The bivariate tests showed a significant positive association between all levels of professionally qualified staff and FTR-L (i.e. the more professionally qualified staff the more extended stays). In the multivariable model, there was no relationship between the number of nurses per doctor and failure to rescue. Nurse stability was associated with lower failure to rescue while higher bed occupancy, teaching hospital status, larger hospitals, and more professionally qualified staff per bed were all associated with higher levels of failure to rescue (table 6).

Table 6: Poisson regression model for FTR-L for 2008/9 data

Variable	Bivariate Correlation			Regression Correlation		
Validation variables	Relative risk	95% CI	p	Relative risk	95% CI	p
Professionally qualified clinical staff per bed	1.08	(1.08-1.08)	<0.001	1.04	(1.02-1.05)	<0.001
Doctors per bed	1.09	(1.09-1.09)	<0.001			
Nurse per bed	1.08	(1.08-1.09)	<0.001	1.00		
Support to clinical staff per bed	0.99	(0.98-0.99)	<0.001			
Nurse per doctor	1.06	(1.05-1.06)	<0.001			
Support to clinical staff per nurse	0.90	(0.90-0.90)	<0.001			
Hospital is Teaching Hospital	1.20	(1.29-1.20)	<0.001	1.11	(1.09-1.14)	<0.001
Number of beds	0.98	(0.97-0.99)	<0.001	1.02	(1.01-1.03)	<0.001
Nurse stability index	0.93	(0.93-0.933)	<0.001	0.97	(0.96-0.99)	<0.001
Control variables						
% cases with a co morbidity	1.00	(1.00-1.00)	0.030			
% of patients with an FTR Complication	5.77	(4.33-7.70)	<0.001	1.06	(1.05-1.07)	<0.001
Average % occupancy	1.04	(1.04-1.04)	<0.001	1.33	(1.11-1.60)	<0.001
GPs per 100,000 population	1.05	(1.04-1.05)	<0.001	1.02	(1.01-1.03)	<0.001
Hospital is in London	1.25	(1.24-1.26)	<0.001			
Number of people who died in hospital as a % of all deaths	1.09	(1.08-1.09)	<0.001	1.08	(1.07-1.09)	<0.001
Number of discharges from hospital	1.02	(1.02-1.02)	<0.001	1		

Distribution of FTR rates across trusts

We made funnel plots of the rates of failure to rescue against the size of the trust (Figure 3). Funnel plots are a technique from statistical process control to identify trusts that have an FTR rate outside what is expected through random variation after allowing for sample size. Each dot represents a trust. The ‘funnel’ is made by plotting a line that represents a number of standard deviations from the mean against the sample size (size of the denominator for the trust). Standard deviations are higher for small samples and so a ‘funnel’ is formed as the standard deviation becomes smaller for larger samples. Most observations would be expected to lie within two standard deviations of the mean and very few would be expected to lay more than three standard deviations from the mean. These lines are referred to as control limits. Observations lying outside them are unlikely to be a result of ‘random’ variation.

Considerable over-dispersion (more variation than would be expected by sampling error alone) is clear in all funnel plots of the failure to rescue rate against the failure to rescue denominator (Figure 3 gives the funnel plot of number of FTR-A). FTR-A has the least amount of over-dispersion with 37% of cases outside the 3SD control limits (Table 7). FTR-L has the most with over 78% of trusts outside these control limits. The number of trusts classified as high and low is broadly symmetrical in FTR-A and FTR-S. However, this is not the case in FTR-L where many more trusts are classified as low than high (table 8).

Table 8: Summary of the number of outliers identified by the funnels plots

	FTR-A	FTR-S	FTR-L
Trusts outside High 3SD limit	27 (18%)	41 (28%)	46 (31%)
Trusts inside control limits	93 (63%)	64 (44%)	33 (22%)
Trusts outside Low 3SD limit	27 (18%)	42 (29%)	68 (46%)

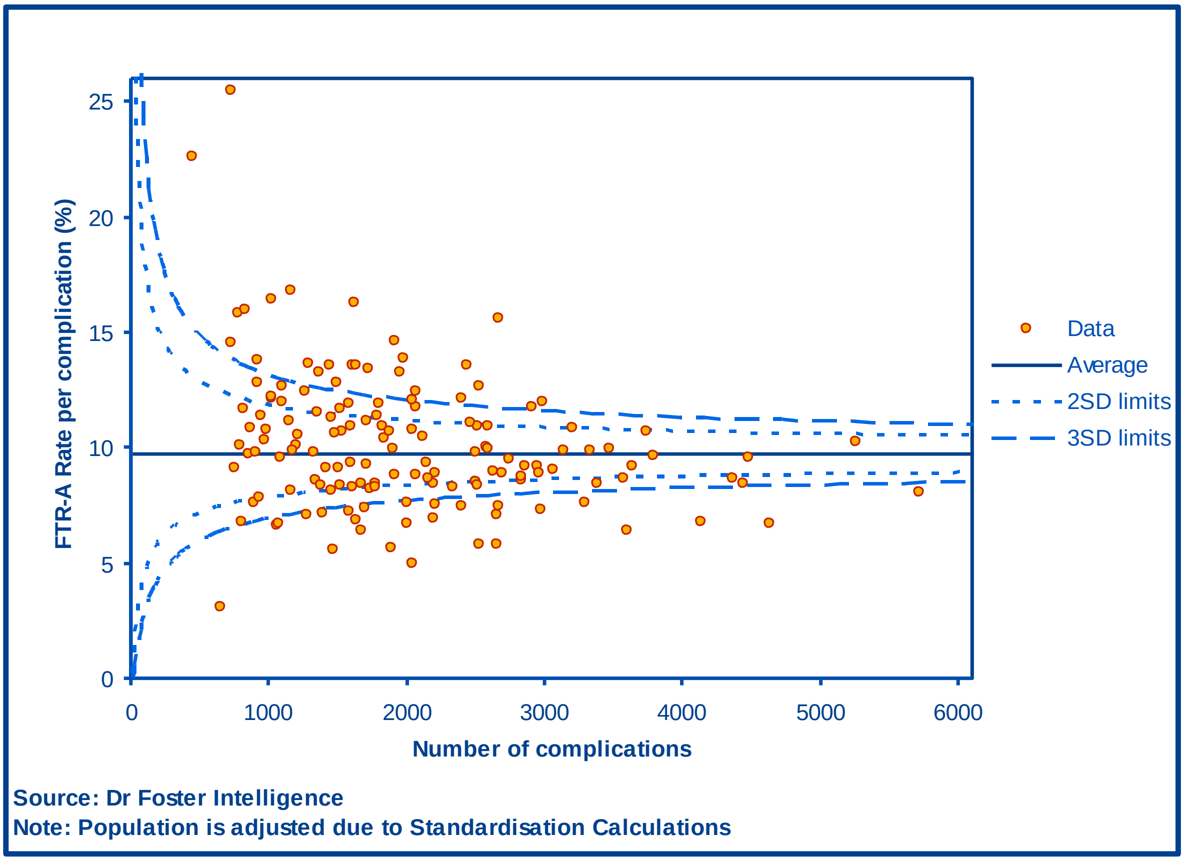


Figure 3: Funnel plot of FTR-A rate, 2008/9 data

Discussion & Conclusions

Summary of results

Rates of secondary coding have increased substantially over the years studied. FTR rates seem to be stable over time suggesting that they are measuring a relatively stable property, as would be expected of a quality measure. None of the FTR indicators is significantly correlated with coding depth (average number of secondary diagnoses coded), which suggests an absence of systematic bias caused by coding practices, although there are some associations between FTR rates and the proportion of surgical patients with a complication from the FTR 'basket' recorded.

Mortality based FTR rates (FTR-A, FTR-S) were associated with a number of hospital characteristics that are plausibly associated with quality and have been shown to be associated with safety in previous studies. Bivariate correlations supported most of the hypothesised associations with organisational characteristics and supported the hypothesised association between FTR and numbers of registered nurses per bed. However, some results ran counter to expectation. For example, a more stable nursing workforce was associated with *higher* rates of failure to rescue. Regression models suggested that while a higher number of qualified clinical staff was associated with lower rates of failure to rescue (as predicted) a greater number of nurses relative to doctors was associated with higher rates of FTR-A and FTR-S. While this finding is plausibly explained by a richer skill mix in the clinical workforce, it does not give strong support to FTR being specifically nurse sensitive. FTR-L was also plausibly associated with quality in some respects but many of the hypothesised relationships were not found, for example in regression models and hospitals with more professionally qualified clinical staff per bed were associated with *higher* rates of failure to rescue.

Although FTR was predicted to be relatively independent of age, we found that FTR-A and FTR-S rates are strongly age related. This age gradient is not reflected in US FTR rates. FTR-A rates for England are substantially lower in all age groups than in the US. Rates of FTR-S came closer to US rates but were still lower, except in the 65-74 age groups. All FTR indicators showed more variation than would be expected by sampling error alone.

Discussion

McKee (1999) argued that both the general lack of secondary coding and its variability meant that the routine data should not be used to derive failure to rescue indicators. Rates of secondary coding have increased substantially since McKee's study. Whereas McKee *et al.* (1999) found that failure to rescue rates showed a relatively low year on year correlation we found strong year on year correlations for all our failure to rescue rates over recent years. Whilst the quality of data has clearly improved, there is still scope for improvement amongst some trusts, although FTR rates are not significantly correlated with a trust's coding depth (average number of secondary diagnoses coded), which suggests an absence of systematic bias caused by coding practices.

However, the increase in secondary coding is far greater than the increase in cases coded with a FTR complication. In contrast to a 54% overall increase, the proportion of patients with a complication in the FTR ‘basket’ increased by only 17% from 1997/8 to 2008/9. While tariffs are designed to recognise additional expense in providing care in the face of comorbidities they are not designed to ‘reward’ extra costs incurred due to complications that arise as a result of treatment (Jameson and Reed, 2007). While Payment by Results will assign payment based on the most expensive HRG associated with the patient, surgical procedures tend to attract the highest tariffs and so complications in the FTR ‘basket’ are unlikely to add to the tariff. It may well be that trusts are prioritizing clinical coding that results in an increase in tariff at the expense of other types of secondary coding. Of course, it might be that secondary coding of complications in the FTR basket was already better than that of other comorbidities due to their significance and seriousness and thus the scope for improvement is less.

Rates of failure to rescue using the mortality-based indicators (our FTR-A and FTR-S) are considerably lower in England than observed in the US in all age groups other than for FTR-S in the 65 to 74 age group, where rates are higher than the US. The most recently available US figures use an older version of the AHRQ definition than that used here but these definitions are broadly comparable. Given the inclusion of *all* surgical deaths in the numerator for our FTR-S definition, general under coding of complications cannot be a complete explanation for the discrepancy, since the estimate is potentially inflated by including miscoded patients from outside the denominator. The increasing discrepancy between FTR-S and our adjusted FTR-S is suggestive of more under coding in older people, as the upward biasing of the estimate is greater as age increases. However, it is only in the older age group that the English FTR-S rate matches the US rate. Further work is required to examine to what extent this difference can be explained by differences in quality of care, clinical practices, case mix variation between England and US samples and difference in definitions (e.g. the ICD-9-CM to ICD-10 mapping).

Although not correlated with overall depth of coding FTR-A and FTR-S were negatively correlated with rates of complication in the FTR basket. McKee suggested that one possible explanation for this is that trusts with low complication rates might tend to code complications only when a patient died. This is certainly possible but if this effect was at its most extreme no additional cases would be added to the numerator for FTR-L in hospitals with low coding (since all deaths would be subject to ‘selective’ secondary coding leading to equality between FTR-A and FTR-S). By contrast, in hospitals with less selective coding, (and by McKee’s hypothesis overall higher coding), FTR-L would be higher than FTR-A. If this was the case, FTR-L should show a lower correlation with complications than FTR-A However, the correlation between FTR-L and complications was stronger than for FTR-A. An alternative hypothesis might be that hospitals that see more complications are better at managing them.

Selective coding of complications among patients who died does provide a potential explanation in the discrepancy between UK and English rates. Such selective coding will inflate the estimate of failure to rescue rates because the size of the basket is reduced and this might explain why US rates are so much higher. If this is the case, the overall problem of under coding, and in particular selective coding, may be less in the UK than in the US although we were unable to find any US data that could allow us to explore this further.

The fact that while FTR-A and FTR-S are highly correlated with each other, but are uncorrelated with FTR-L, suggests that these mortality based indicators measure very different things from FTR-L. We also found little correlation between FTR rates and a hospital's HSMR. Mortality has been shown to be an insensitive indicator of quality of care (Hofer and Hayward, 1996, Mant and Hicks, 1995). Whereas HSMR is based on the premise that residual variation after statistical adjustment for co-morbidity and other patient level factors can be ascribed to variation in quality of care (Jarman et al., 1999) failure to rescue is proposed as a superior approach which separates the development of complications, which is seen as largely determined by the patient's underlying condition, from the success in treating the complication, which is more strongly associated with the quality of care (Silber et al., 1992). Thus, it is thought to be both more sensitive to quality and less vulnerable to patient level variation. HSMR is also based on a diagnostic 'basket' containing both medical and surgical admissions. The absence of correlation is thus not entirely unsurprising.

In contrast to McKee, who found little association between age and failure to rescue, we found a clear pattern of increasing rates of both FTR-A and FTR-S with increasing age. FTR-L showed considerably less association with age. McKee argued that invariance with age was evidence for the utility of the indicator, establishing its independence from pre-admission severity. Although it is impossible to eliminate the possibility that the relationship we observed reflects the widely acknowledged variation in quality of care with age (Delamothe, 2008) this seems unlikely to be a full explanation. We conclude that any failure to rescue indicator would need to be age adjusted, despite the claims that the indicator is insensitive to patient level variation.

The regression models showed a number of associations between the various failure to rescue indicators and presumed markers of quality. These associations tend to lend support to the mortality based FTR indicators as valid measures of safety. Bivariate relationships with FTR-A and FTR-S were generally plausible and supported associations demonstrated in other research. However, some of the relationships in the regression models were perplexing in the light of expected relationships. While lower FTR-A and FTR-S rates were associated with higher levels of professional staffing, the opposite was true for FTR-L. While there is a plausible causal link between high levels of professional staffing and quality there is no obvious explanation for the opposite finding. Similarly, while there is a relationship between a more stable nursing workforce and lower rates of FTR-L, the opposite relationship applies for FTR-A and FTR-S. Again, no plausible causal explanation suggests itself. It may be that there is residual confounding at the organisational level or at the level of patient case mix that warrants further exploration. This is particularly the case for FTR-L. We would conclude that non-risk adjusted FTR rates cannot be used to make comparisons between providers.

Given the very different relationships observed between these validation variables and FTR-L, coupled with the lack of correlation with the other FTR measures, it is hard to conclude that FTR-L is a true measure of patient safety. FTR-L is defined quite differently from Silber's definition of failure to rescue. FTR-L will be associated with a number of factors beyond the hospital's control, such as local availability of long-term care, community health, and social care services. Although FTR-L performed well in some respects, the absence of specific problems relating to secondary coding in the other measures (which this was designed to avoid) means that there was no clear advantage for this over the mortality based measures (FTR-A and FTR-S). The advantage of this indicator is further undermined by the need for additional risk adjustment, which will likely rely on secondary coding. There is no logical reason why pre-existing severity should not contribute to extended stays, and the positive association between rates of the FTR complications and FTR-L suggests that the underlying patient condition may indeed contribute significantly to rates of FTR-L and that further risk adjustment will be required. However, while 'true' failure to rescue could theoretically contribute to extended stays in a risk-adjusted indicator (even though we found no evidence that it did) it seems inappropriate to consider it a valid proxy for failure to rescue. If it is a quality indicator, it represents an aspect of whole system performance and not specifically the safety of hospital care. If it is to be further considered as an indicator for hospital performance FTR-L should be examined in patients discharged home only, where external drivers of stay may be fewer.

It is notable that previous US work establishing a link between nurse staffing and failure to rescue has not generally considered the number of doctors (Griffiths, 2009). Our bivariate correlations showed an association between higher levels of nurse staffing per bed and lower rates of failure to rescue, as seen previously. However, while we could not directly include both nurse and medical staffing in our regression models there was an association between a high nurse to doctor ratio and high rates of FTR-A and FTR-S (i.e. higher proportion of nurses more failure to rescue). When combined with the negative association with levels of professionally qualified staff (more qualified staff was associated with lower FTR) this suggests that medical staffing levels may make a more important contribution than nurse staffing levels. The absence of an association between FTR-L and the nurse to doctor ratio suggests that this indicator is not sensitive to the nurse-doctor skill mix. While this could support a degree of nurse-doctor substitution, it does little to establish this indicator as particularly nurse sensitive. However, given our conclusion that the measures need to be risk adjusted, a final conclusion on this matter may be premature.

Over-dispersion in the funnel plots gives cause for concern. The picture is of a process that is not in control (in a statistical sense) and an indicator which is not sufficiently sensitive to quality to be useful (Spiegelhalter, 2005). There are several possible explanations. First, and most pessimistically, it might be that clinical coding practices in trusts vary so much that it is not meaningful to compare the different trust rates. However, since we find no other evidence of a significant problem related to clinical coding, this explanation seems unlikely. A second explanation is that over-dispersion is the result of inadequate patient level risk adjustment or adjustment for institutional factors independent of quality (Spiegelhalter, 2005). This explanation is partially supported by a clear pattern of increasing rates of both FTR-A and FTR-S with increasing age. A third explanation might be that considerable variation in actual quality of care exists. At present it is unclear which of the reasons contributes most to the over-dispersion. However, developing an appropriate risk adjustment model and case note audit of failure to rescue cases are important first steps in understating the underlying causes for this over dispersion. The existence of these variations in tandem with otherwise positive findings for the FTR-A and FTR-S indicator supports the case for further research.

Because FTR-S counts all surgical deaths as part of the numerator but not the denominator it provides an active disincentive to trusts to 'game' by under coding secondary diagnoses and thus omitting them from the indicator. However, because it fails to add the additional cases to the denominator it is inherently inaccurate and will tend to exaggerate FTR rates in hospitals with poor coding. FTR-A showed more year on year stability and less over-dispersion than FTR-L and we conclude that FTR-A is the most promising indicator, although further work is required. While we have found no serious obstacles to deriving it from English data the validation of FTR measures using approaches other than associations with variables they are expected to correlate with has been limited so far.

Conclusion and Implications

A number of our findings suggest that low levels of recording of secondary diagnoses in hospital administrative data, previously a significant reason for concluding that mortality based failure to rescue indicators could not be derived from English data, is no longer a significant obstacle. Our FTR-A indicator is a plausible patient safety related quality indicator although it must be risk adjusted and cannot be used to make comparisons between providers without such adjustment. Further assessment of a risk-adjusted measure is required and further validation through case note audit is warranted to give further confidence that it is a true measure of safety. Although a final conclusion is premature, we found no evidence to support failure to rescue, as measured by any of our indicators, to be specifically nurse sensitive.

Implications for Policy

- **There is potential to derive mortality based failure to rescue indicators for surgical patients from routine administrative data in England**
- **Failure to rescue indicators offer some advantages over standardised mortality measures such as HSMR for surgical patients**
- **Our FTR-A indicator, based on the AHRQ definition, is worthy of further exploration**
- **FTR-L may have a role as a system level indicator although its precise meaning remains unclear**
- **Unadjusted failure to rescue cannot be used to compare the quality of care between NHS trusts**
- **Failure to rescue does not appear to be a specifically nurse sensitive indicator although this conclusion should be regarded as tentative**

Implications for research

- **Further assessment is required to develop and test risk adjustment models**
- **A case note audit needs to be performed to verify the extent that cases identified by the indicator are indeed cases of 'failure to rescue'**
- **The indicator could be further validated against the successful implementation of quality improvement initiatives which focus on interventions to reduce failure to rescue (e.g. global track and trigger tools)**
- **FTR-L should be examined in patients discharged home only, where external drivers of stay may be fewer**

References

Agency for Healthcare Research and Quality, 2007. Guide to Patient Safety Indicators.

Agency for Healthcare Research and Quality, 2007. Patient Safety Indicators (PSI) Version 3.1 Comparative Data Agency for Healthcare Research and Quality,. http://www.qualityindicators.ahrq.gov/downloads/psi/psi_provider_comparative_v31.pdf Accessed December 2009

Agency for Healthcare Research and Quality, 2009. PSI Technical Specifications, Version 4.0. http://www.qualityindicators.ahrq.gov/psi_download.htm Accessed 16 December 2009

Armitage, M., Eddleston, J., Stokes, T., Guideline Development Group, 2007. Recognising and responding to acute illness in adults in hospital: summary of NICE guidance. BMJ 335 (7613), 258-259.

Belsey, D., 1991. A guide to using the collinearity diagnostics. Computer Science in Economics and Management 4 (1), 33-50.

Cameron, A., Trivedi, P., 1998. Regression analysis of count data. Cambridge University Press.

Delamothe, T., 2008. Universality, equity, and quality of care. BMJ 336 (7656), 1278-1281.

Department of Health, 2008. High Quality Care for All: NHS Next Stage Review Final Report. Stationary Office, Norwich.

Feinmann, J., 2009. Safety first. BMJ 338 (feb06_1), b420-.

Griffiths, P., 2009. RN+RN=better care? What do we know about the association between the number of nurses and patient outcomes? International Journal of Nursing Studies 46 (10), 1289-1290.

Griffiths, P., Jones, S., Maben, J., Murrells, T., 2008. State of the Art Metrics for Nursing: a rapid appraisal. London: King's College London.

Hofer, T., Hayward, R., 1996. Identifying poor-quality hospitals: Can hospital mortality rates detect quality problems for medical diagnoses? Medical Care 34 (8), 737-753.

Jameson, S., Reed, M.R., 2007. Payment by results and coding practice in the National Health Service: THE IMPORTANCE FOR ORTHOPAEDIC SURGEONS. Journal of Bone and Joint Surgery. British Volume 89-B (11), 1427-1430.

Jarman, B., Gault, S., Alves, B., Hider, A., Dolan, S., Cook, A., Hurwitz, B., Iezzoni, L.I., 1999. Explaining differences in English hospital death rates using routinely collected data. BMJ 318 (7197), 1515-1520.

Kane, R., Shamliyan, T., Mueller, C., Duval, S., Wilt, T., 2007. The Association of Registered Nurse Staffing Levels and Patient Outcomes: Systematic Review and Meta-Analysis. Medical Care 45 (12), 1195-1204.

Luettel, D., Beaumont, K., Healey, F., 2007. Recognising and responding appropriately to early signs of deterioration in hospitalised patients. National Patient Safety Agency.http://www.nrls.npsa.nhs.uk/resources/?entryid45=59834

Mant, J., Hicks, N., 1995. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. BMJ 311 (7008), 793-796.

Marshall, M.N., Shekelle, P.G., Davies, H.T.O., Smith, P.C., 2003. Public Reporting On Quality In The United States And The United Kingdom. Health Aff 22 (3), 134-148.

McKee, M., Coles, J., James, P., 1999. 'Failure to rescue' as a measure of quality of hospital care: the limitations of secondary diagnosis coding in English hospital data. Journal of Public Health 21 (4), 453-458.

Mohammed, M.A., Deeks, J.J., Girling, A., Rudge, G., Carmalt, M., Stevens, A.J., Lilford, R.J., 2009. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. BMJ 338 (mar18_2), b780-.

National Patient Safety Agency, 2007. Safer care for the acutely ill patient: learning from serious incidents. Patient Safety Observatory 5.

Needleman, J.P., Buerhaus, P.I.P.R.N.F., 2007. Failure-to-Rescue: Comparing Definitions to Measure Quality of Care. Medical Care 45 (10), 913-915.

Rafferty, A.M., Clarke, S.P., Coles, J., Ball, J., James, P., McKee, M., Aiken, L.H., 2007. Outcomes of variation in hospital nurse staffing in English hospitals: cross-sectional analysis of survey data and discharge records. International Journal of Nursing Studies 44 (2), 175-182.

Schulz, S., Zaiss, A., Brunner, R., Spinner, D., Klar, R., 1998. Conversion problems concerning automated mapping from ICD-10 to ICD-9. Methods of Information in Medicine 37 (3), 254-259.

Silber, J.H., Romano, P.S., Rosen, A.K., Wang, Y., Even-Shoshan, O., Volpp, K.G., 2007. Failure-to-Rescue: Comparing Definitions to Measure Quality of Care. Medical Care 45 (10), 918-925.

Silber, J.H., Williams, S.V., Krakauer, H., Schwartz, J.S., 1992. Hospital and patient characteristics associated with death after surgery: a study of adverse occurrence and failure to rescue. Medical Care 30 (7), 615-629.

Spiegelhalter, D.J., 2005. Handling over-dispersion of performance indicators. Quality and Safety in Health Care 14 (5), 347-351.

Appendix 1: FTR specification mapping

AHRQ definition

Numerator: All discharges with a disposition of “deceased” (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.

Denominator: All surgical discharges age 18 years and older ....defined by specific DRGs or MS-DRGs and an ICD-9-CM code for an operating room procedure, principal procedure within 2 days of admission OR admission type of elective (ATYPE=3) with potential complications of care listed in Death among Surgical definition (e.g., pneumonia, DVT/PE, sepsis, shock/cardiac arrest, or GI haemorrhage/acute ulcer).

Mappings

AHRQ requirement	Our interpretation
Patients aged between 18 and 90 at time of discharge	**Age at start of episode[HES]** for discharging episode between **18** and **90**
Have a code for a specific DRGs and an ICD-9-CM code for an operating room procedure	**Treatment Function either** begins with 1 or is equal to 502. AND **Primary procedure** is present and corresponds to a valid OPCS procedure. (This approach does not quite match the AHRQ deflation and may need refining)
Principal procedure within 2 days of admission OR admission type of elective	Have an elective admission (**admission method** equal to **11, 12** or **13**) or an emergency admission (**admission method** equal to **21,22,23,24** or **28**) with principal procedure within 2 days of admission (**Primary Procedure Date - Admission Date (Hospital Provider Spell) <3)**
Not transferred to an acute care facility	**Discharge Destination (Hospital Provider Spell** not equal to **48, 49,50,51,52,** and **53)**. We have extended our definition to any healthcare provider due to the quality of coding for this field.
potential complications of care listed in Death among Surgical definition	See below for our mapping from ICD 9-CM to ICD-10 diagnosis codes

Unlike the data used to calculate the US AHRQ indicators, the English data are based around consultant episode and not hospital stay. A consultant episode is defended as:

“...the time a patient spends in the continuous care of one consultant using Hospital Site or Care Home bed(s) of one Health Care Provider or, in the case of shared care, in the care of two or more consultants.” Connecting for Health (2009)

For elective care, spells and consultant episodes are almost equivalent. However, this is not the case with emergency admissions. We grouped episodes into spells using the NHS Information Centre’s

HRG grouper. We have only searched for a surgical procedure in the first two consultant episodes and checked that the procedure took place within two days of the admission date (and not just the start of the episode).

Complication codes for analysis

An initial set of codes were derived using a converter of ICD9 to ICD10 from the web. The second set ('de novo') were derived using the AHRQ specifications as a starting point for an inspection of the ICD10 book, with appropriate codes from the first set added if necessary – it is these codes that were used for this study. This process revealed a number of inaccuracies with the converter, particularly involving complications of prostheses and devices and GI haemorrhage.

Comparison of ICD codes derived from conversion from ICD9 ("from ICD9") and those derived by inspection of ICD10 book ("de novo")

ICD10	Description	From ICD9	De novo	Used
FTR1				
N17		Y	Y	N
O084	Renal failure following abortion and ectopic and molar pregnancy	Y	Y	N
O904	Postpartum acute renal failure	Y	Y	N
O756	Delay delivery after spontaneous or unspec. rupture of membranes	Y	N	N
FTR2				
I26	Pulmonary embolism	Y	Y	Y
I80	Phlebitis and thrombophlebitis	Y	Y	Y
I828	Embolism and thrombosis of other specified veins	Y	N	N
I829	Embolism and thrombosis of unspecified vein	Y	N	N
FTR3				
J13	Pneumonia due to Streptococcus pneumoniae	Y	Y	Y
J14	Pneumonia due to Haemophilus influenzae	Y	Y	Y
J15	Bacterial pneumonia, not elsewhere classified	Y	Y	Y
J16	Pneumonia due to other infectious organisms NEC	Y	Y	Y
J18	Pneumonia, organism unspecified	Y	Y	Y
A481	Legionnaires' disease	Y	Y	Y
J69	Pneumonitis due to solids and liquids	Y	Y	Y
J81	Pulmonary oedema	Y	Y	Y
FTR4				
A40	Streptococcal septicaemia	Y	Y	Y
A41	Other septicaemia	Y	Y	Y
A483	Toxic shock syndrome	Y	Y	Y
A499		Y	Y	Y
R651	Not sure – not in standard ICD10	Y	N	Y
T802	Infections following infusion transfusion & therapeutic injection	N	Y	Y
T814	Infection following a procedure, not elsewhere classified	N	Y	Y
T811	Shock during or resulting from a procedure, not elsewhere classified	Y	Y	N
T826	Infection and inflammatory reaction due to cardiac valve prosthesis	Y	Y	Y
T827	Infection and inflammatory reaction due to other cardiac and vascular devices, implants and grafts	Y	Y	Y

T835	Infection and inflammatory reaction due to prosthetic device, implant and graft in urinary system	Y	Y	Y
T836	Infection and inflammatory reaction due to prosthetic device, implant and graft in genital tract	Y	Y	Y
T845	Infection and inflammatory reaction due to internal joint prosthesis	Y	Y	Y
T846	Infection and inflammatory reaction due to internal fixation device [any site]	Y	Y	Y
T847	Infection and inflammatory reaction due to other internal orthopaedic prosthetic devices, implants and grafts	Y	Y	Y
T857	Infection and inflammatory reaction due to other internal prosthetic devices, implants and grafts	Y	Y	Y
FTR5				
A419	Septicaemia, unspecified (septic shock)	Y	Y	Y
A483	Toxic shock syndrome	N	Y	Y
I21	Acute myocardial infarction	N	N	N
I22	Subsequent myocardial infarction	N	N	N
I46	Cardiac arrest	I469	Y	Y
O083	Shock following abortion and ectopic and molar pregnancy	Y	Y	Y
O150	Eclampsia in pregnancy	Y	N	Y
O751	Shock during or following labour and delivery	Y	Y	Y
O754	Other complications of obstetric surgery and procedures	N	Y	Y
O908	Other complications of the puerperium	Y	N	N
R092	Respiratory arrest	Y	Y	Y
R57	Shock, not elsewhere classified	Y	Y	Y
R96	Other sudden death, cause unknown	N	Y	Y
T780	Anaphylactic shock due to adverse food reaction	N	Y	Y
T782	Anaphylactic shock, unspecified	Y	Y	Y
T805	Anaphylactic shock due to serum	Y	Y	Y
T811	Shock during or resulting from a procedure NEC	Y	Y	Y
T821, 2, 5	Mechanical complications of cardiac devices etc	Y	N	N
T830,1 , 3, 4	Mechanical complications of GU prosthetic devices etc	Y	N	N
T882	Shock due to anaesthesia	Y	Y	Y
T886	Anaphylactic shock due to adv effect of correct drug or med prop admin	N	Y	Y
FTR6				
I60	Subarachnoid haemorrhage	N	N	N
I61	Intracerebral haemorrhage	N	N	N
I62	Other nontraumatic intracranial haemorrhage	N	N	N
I850	Oesophageal varices with bleeding	Y	Y	y
I982A	Oesophageal varices in diseases classified elsewhere	Y	Y	Y
K226	Gastro-oesophageal laceration-haemorrhage syndrome	Y	Y	Y
K228	Other specified diseases of oesophagus (haemorrhage)	Y	Y	Y
K250	Gastric ulcer, acute with haemorrhage	Y	Y	Y
K251	Gastric ulcer, acute with perforation	Y	Y	Y
K252	Gastric ulcer, acute with both haemorrhage and perforation	Y	Y	Y
K253	Gastric ulcer, acute without haemorrhage or perforation	Y	Y	Y
K259	Unspec. as acute or chronic w'out	Y	Y	Y

	haemorrhage or perforation			
K260	Duodenal ulcer, acute with haemorrhage	Y	Y	Y
K261	Duodenal ulcer, acute with perforation	Y	Y	Y
K262	Duodenal ulcer, acute with both haemorrhage and perforation	Y	Y	Y
K263	Duodenal ulcer, acute without haemorrhage or perforation	Y	N	Y
K269	Unspec. as acute or chronic w'out haemorrhage or perforation	Y	N	Y
K270	Peptic ulcer, acute with haemorrhage	Y	Y	Y
K271	Peptic ulcer, acute with perforation	Y	Y	Y
K272	Peptic ulcer, acute with both haemorrhage and perforation	Y	Y	Y
K273	Peptic ulcer, acute without haemorrhage or perforation	Y	N	Y
K279	Unspec. as acute or chronic w'out haemorrhage or perforation	Y	N	Y
K280	Gastrojejunal ulcer, acute with haemorrhage	Y	Y	Y
K281	Gastrojejunal ulcer, acute with perforation	Y	Y	Y
K282	Acute with both haemorrhage and perforation	Y	Y	Y
K283	Acute without haemorrhage or perforation	Y	N	Y
K289	Unspec. as acute or chronic w'out haemorrhage or perforation	Y	N	Y
K290	Acute haemorrhagic gastritis	Y	Y	Y
K3182	Other specified diseases of stomach and duodenum (ICD9 asked for Angiodysplasia of stomach and duodenum with hemorrhage)	Y	N	Y
K5522	Angiodysplasia of colon	Y	Y	Y
K570	Diverticular dis of small intestine with perf and abscess	N	Y	Y
K5711	Diverticular dis of small intestine without perf or abscess	Y	Y	Y
K5713	Diverticular dis of small intestine without perf or abscess	Y	Y	Y
K5721	Diverticular dis of large intestine with perf and abscess	Y	Y	Y
K5723	Diverticular dis of large intestine with perf and abscess	Y	Y	Y
K574	Diverticular dis of both small and large intest with perf + abscess	N	Y	Y
K578	Diverticular dis of intest part unspec. with perf and abscess	N	Y	Y
K625	Haemorrhage of anus and rectum	Y	Y	Y
K920	Haematemesis	Y	Y	Y
K921	Melaena	Y	Y	Y
K922	Gastrointestinal haemorrhage, unspecified	Y	Y	Y

Appendix 2: independent variables used in the model

Variable	Motivation	Source	Data period
Control Variables			
% cases with a co morbidity	Mohammed et al (2009) reported links with mortlity rates	Dr Foster	2008/9
% cases with a complication from the FTR basket	Alternative measure of co morbidity	Dr Foster	2008/9

GPs per 100,000 population	Jarman et al (1999) reported links with mortlity rates	NHS IC and ONS	Sep-2008
Number of people who died in hospital as a % of all deaths	Based on the assumption that end of life provision in surrounding area may effect a hospitals mortality rate	ONS	2006
Number of discharges from hospital	Jarman et al (1999) reported links with mortlity rates	**Dr Foster**	**2008/9**
Validation Variables			
Dr per Bed *	Full Time Equivalent (FTE) Number of hospital doctors (all grades) per bed	**NHS IC and DH HES Website**	**2008/9**
Nurse per Bed	FTE Qualified nurses per hospital bed	**NHS IC and DH HES Website**	**2008/9**
(Prof Qualified scientific, therapeutic & technical staff) per bed *	FTE Allied health professions + Qualified healthcare scientists + Other qualified scientific §, therapeutic and technical per bed	NHS IC and DH HES Website	Sep-2008
Support to clinical staff per bed *	FTE Support to doctors & nursing + Support to ST&T +Support to ambulance service § per bed	NHS IC and DH HES Website	Sep-2008
NHS infrastructure support per bed *	FTE administrative and management staff per bed	NHS IC and DH HES Website	Sep-2008
Non medical stability index	Number of non-medical staff showing in 2008 and remaining in 2009. Effectively the percentage of staff that have been in post for over a year. Presumption was that staff with more stable groups might find it easier to provide a high quality service.	**NHS IC**	**2008/9**
Nurse stability index	Number of nursing staff showing in 2008 and remaining in 2009. Effectively the percentage of staff that have been in post for over a year.	**NHS IC**	**2008/9**
Number of beds	Jarman et al (1999) reported links with mortlity rates	**DH HES Website**	**2008/9**
Hospital is in London (y/n)	Jarman et al (1999) reported links with mortlity rates	Dr Foster	2008/9
Hospital is Teaching Hospital (y/n)	Jarman et al (1999) reported links with mortlity rates	Dr Foster	2008/9
Length of stay	Jarman et al (1999) reported links with mortlity rates	Dr Foster	2008/9
Spells per Bed	Measure of hospital throughput	Dr Foster and DH HES Website	2008/9
% occupancy	Jarman et al (1999) reported links with mortlity rates	DH HES Website	2008/9

*to avoid collinearity problems this variable was implemented as number FTE staff per nurse
§ for more detailed composition see www.ic.nhs.uk/workforce

Further copies of this report can be obtained from:

The National Nursing Research Unit at King's College London

Waterloo Road

London

SE1 8WA

Telephone 020 7848 3057

Fax 020 7848 3069

http://www.kcl.ac.uk/schools/nursing/nnru