RESEARCH

Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals

Mohammed A Mohammed, senior lecturer,[1] Jonathan J Deeks, professor of health statistics,[1] Alan Girling, senior research fellow,[1] Gavin Rudge, data scientist,[1] Martin Carmalt, consultant physician,[2] Andrew J Stevens, professor of public health and epidemiology,[1] Richard J Lilford, professor of clinical epidemiology[1]

[1]Unit of Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham B15 2TT

[2]Royal Orthopaedic Hospital, Birmingham B31 2AP

Correspondence to:
M A Mohammed
M.A.Mohammed@Bham.ac.uk

ABSTRACT

Objective To assess the validity of case mix adjustment methods used to derive standardised mortality ratios for hospitals, by examining the consistency of relations between risk factors and mortality across hospitals.

Design Retrospective analysis of routinely collected hospital data comparing observed deaths with deaths predicted by the Dr Foster Intelligence case mix method.

Setting Four acute National Health Service hospitals in the West Midlands (England) with case mix adjusted standardised mortality ratios ranging from 88 to 140.

Participants 96 948 (April 2005 to March 2006), 126 695 (April 2006 to March 2007), and 62 639 (April to October 2007) admissions to the four hospitals.

Main outcome measures Presence of large interaction effects between case mix variable and hospital in a logistic regression model indicating non-constant risk relations, and plausible mechanisms that could give rise to these effects.

Results Large significant (P≤0.0001) interaction effects were seen with several case mix adjustment variables. For two of these variables—the Charlson (comorbidity) index and emergency admission—interaction effects could be explained credibly by differences in clinical coding and admission practices across hospitals.

Conclusions The Dr Foster Intelligence hospital standardised mortality ratio is derived from an internationally adopted/adapted method, which uses at least two variables (the Charlson comorbidity index and emergency admission) that are unsafe for case mix adjustment because their inclusion may actually increase the very bias that case mix adjustment is intended to reduce. Claims that variations in hospital standardised mortality ratios from Dr Foster Intelligence reflect differences in quality of care are less than credible.

INTRODUCTION

The longstanding need to measure quality of care in hospitals has led to publication of league tables of standardised mortality ratios for hospitals in several countries, including England, the United States,

Canada, the Netherlands, and Sweden.[1-6] Standardised mortality ratios for hospitals in these countries have been derived with methods heavily influenced by the seminal work of Jarman et al,[1] who first developed standardised mortality ratios for National Health Service (NHS) hospitals in England in 1999, and by the subsequent methodological developments by Dr Foster Intelligence.[7 8] This is a former commercial company that is now a public-private partnership, which annually publishes standardised mortality ratios for English hospitals in the national press.

A consistent, albeit controversial,[9-11] inference drawn from the wide variation in standardised mortality ratios for hospitals is that this reflects differences in quality of care. In the 2007 hospital guide for England,[12] Dr Foster Intelligence portrayed standardised mortality ratios for hospitals as "an effective way to measure and compare clinical performance, safety and quality." Although an increasing international trend exists for standardised mortality ratios for hospitals to be developed and published,[13 14] we must be sure that the underlying case mix adjustment method is fit for purpose before inferences about quality of care are drawn.

Case mix adjustment is widely used to overcome imbalances in patients' risk factors so that fairer comparisons between hospitals can be made. Methods for case mix adjustment are often criticised because they can fail to include all the important case mix variables and do not adequately adjust for a variable because of measurement error.[10 11] Despite these criticisms, case mix adjustment is widely done because the adjusted comparisons, although imperfect, are generally considered to be less biased than unadjusted comparisons.

However, a third, more serious problem exists that can affect the validity of case mix adjustment. In a study that compared unadjusted and case mix adjusted treatment effects from non-randomised studies against treatment effects from randomised trials, Deeks et al observed that on average the unadjusted and not the adjusted non-randomised results agreed best with the randomised

comparisons.[15] In this instance, case mix adjustment had increased bias in the comparisons. Nicholl pointed out that case mix adjustment can create biased comparisons when underlying relations between case mix variables and outcome are not the same in all the comparison groups.[16] This phenomenon has been termed "the constant risk fallacy," because if the risk relations are assumed to be constant, but in fact are not, then case mix adjustment may be more misleading than crude comparisons.[16] Two key mechanisms can give rise to non-constant risk relations. The first mechanism involves differential measurement error, and the second one involves inconsistent proxy measures of risk. Each is illustrated below.

Example of differential measurement error

To illustrate the constant risk fallacy we construct hypothetical hospital mortality data with a single case mix variable—a comorbidity index (CMI) that takes values 0 to 6. The relation between in-hospital mortality and CMI value has been modelled for the population, estimating risks of in-hospital death of 0.02, 0.04, 0.08, 0.14, 0.25, 0.40, and 0.57 in the seven CMI categories (equivalent to an odds ratio of two for each unit increase in the index).

Consider two hospitals, A and B, both of which admit 1000 patients a year in each of the seven CMI categories. Assume that the case mix of the groups of patients and the quality of care in the two hospitals are identical and that 1500 deaths are observed in both hospitals. Consider that hospital A correctly codes the comorbidity index, whereas hospital B tends to under-code, such that in hospital B for each true CMI the following are recorded:

- CMI=0: all are coded as 0
- CMI=1: 50% coded 0, 50% coded 1
- CMI=2: 33% coded 0, 33% coded 1, 33% coded 2
- CMI=3: 25% coded 0, 25% coded 1, 25% coded 2, 25% coded 3
- CMI=4: 20% coded 0, 20% coded 1, 20% coded 2, 20% coded 3, 20% coded 4
- CMI=5: 20% coded 1, 20% coded 2, 20% coded 3, 20% coded 4, 20% coded 5
- CMI=6: 20% coded 2, 20% coded 3, 20% coded 4, 20% coded 5, 20% coded 6.

The consequence of this is that rather than observing 1000 patients in each of the seven CMI categories, in hospital B the numbers instead are 2283, 1483, 1184, 850, 600, 400, and 200. It thus looks as if a difference exists in the distribution of the CMI between the two hospitals, with hospital B having on average a lower CMI. Computation of expected numbers of deaths taking into account the reported (rather than true) CMI is done to calculate standardised mortality ratios on the basis of the modelled values.

The expected number of deaths in hospital A is (1000×0.02)+(1000×0.04)+(1000×0.08)+(1000×0.14)+(1000×0.25)+(1000×0.40)+(1000×0.57)=1500, yielding a standardised mortality ratio (observed/expected deaths) of 1500/1500=100.

The expected number of deaths in hospital B is (2283×0.02)+(1483×0.04)+(1184×0.08)+(850×0.14)+(600×0.25)+(400×0.40)+(200×0.57)=743, yielding a standardised mortality ratio of 1500/743=202.

It thus wrongly seems that the mortality in hospital B is twice that in hospital A. Adjustment has changed a fair comparison (1500 v 1500) into a biased comparison. This is an illustration of the constant risk fallacy. Furthermore, modelling the data by using logistic regression reveals that whereas the relation between CMI and mortality in hospital A is the same as in the population (odds ratio=2.0 per category increase), the relation in hospital B is weaker (odds ratio=1.6 per category increase in CMI) (as would be expected through misclassification introducing attenuation bias) and the interaction between hospital B and CMI is clinically and statistically significant (P<0.001). If CMI was measured with equal measurement error in all hospitals the problem would be one of residual confounding caused by regression dilution or attenuation bias (in which case the standardised mortality ratios would be preferable to crude mortality but will not fully adjust for the risk factor). Because measurement errors differ among hospitals, the constant risk fallacy (where standardised mortality ratios may be more misleading than the crude mortality comparison) is a possibility.

Consider two hospitals that are identical in all respects (case mix, mortality, quality of care) except that one hospital (B) systematically under-records comorbidities (measurement error) in its patients. If mortality is case mix adjusted for comorbidity then the expected (but not the observed) number of deaths in hospital B will be artificially depleted, because its patients seem to be at lower risk than they really are. The effect of case mix adjustment is to erroneously inflate the standardised mortality ratio (observed number of deaths/expected number of deaths × 100) for that hospital. The box presents a numerical example of this scenario.

The second mechanism can occur even in the absence of measurement error. Consider, for example, emergency admissions to hospitals. Patients admitted as emergencies are usually regarded as being seriously ill, but if an individual hospital often admits the "walking wounded" (who are not seriously ill) as emergencies, then the risk associated with being an emergency admission in that hospital will be reduced. Variation in this practice across hospitals leads to a non-constant relation between emergency admission and mortality. The standardised mortality ratio for hospitals that admit more walking wounded will receive an unjustified downward case mix adjustment, because elsewhere emergencies are generally the sickest patients and the case mix adjustment will endeavour to reflect this.

A general feature of these two mechanisms that allows identification of case mix variables prone to the constant risk fallacy is that the value recorded for a given patient would change if he or she presented at a different hospital. Comorbidity would be under-coded in one hospital compared with another, whereas the patient may be admitted (and thus coded) as an emergency in some hospitals and elsewhere treated and discharged without being admitted at all. Case mix variables such as age, sex, and deprivation (on the basis of the patients' home address) are not prone to these two mechanisms because their values do not change with different hospitals.

A simple way to screen case mix variables for their susceptibility to non-constant risk relations, on a scale sufficient to bias the case mix adjustment method, is to do a statistical test for interaction effects between hospital and case mix variables in a logistic regression model that predicts death in hospital.[16] If large interaction effects are not found then no apparent evidence of non-constant risk relations exists and the constant risk fallacy (within the limits of statistical inference) may be discounted (although the other challenges in interpreting standardised mortality ratios, such as omitted covariates, will still remain[9 10]). However, if a large interaction effect is found, then this indicates a non-constant risk relation. If this is due to inconsistent measurement practices across hospitals (as in the comorbidity index example in the box), it will result in a misleading adjustment to standardised mortality ratios. If the interaction occurs because the covariate genuinely has different relations with death across

hospitals (as in the emergency admission example above), this too will result in a misleading adjustment to standardised mortality ratios. Alternatively, the interaction could occur if different levels of the covariate were associated with different standards of care across hospitals, in which case the standardised mortality ratio will appropriately reflect the average of the associated increases in mortality. Unfortunately, no statistical method exists for teasing apart these non-exclusive explanations, but they can be explored and resolved, to some extent, by doing "detective work" seeking a likely cause for the observed interaction effect.

In this paper we screened the Dr Foster Intelligence method,[1] which is used to derive standardised mortality ratios for English hospitals and which has been adopted/adapted internationally,[1-6 12] for its susceptibility to the constant risk fallacy. We first tested for the presence of large interaction effects and then, in respect of two key case mix variables (comorbidity index and emergency admission), we did detective work to seek likely explanations.

METHODS

Dr Foster Intelligence case mix adjustment method

The Dr Foster Intelligence case mix adjustment method uses data derived from routinely collected hospital episode statistics.[12] These data include admission date, discharge date, in-hospital mortality, and primary and secondary diagnoses according to ICD-10 (international classification of disease, 10th revision) codes on every inpatient admission (or spell) in NHS hospitals in England. The Dr Foster Intelligence standardised mortality ratio is derived from logistic regression models, which are based on 56 primary diagnosis groups derived from hospital episode statistics data accounting for 80% of hospital mortality. Covariates for case mix adjustment in the model are sex, age group, method of admission (emergency or elective), socioeconomic deprivation, primary diagnosis, the number of emergency admissions in the previous year, whether the patient was admitted to a palliative care specialty, and the Charlson (comorbidity) index (range 0-6), which is derived from secondary ICD-10 diagnoses codes.[17]

Study hospitals and data sources

This study involves four hospitals, representing a wide range of the published case mix adjusted Dr Foster Intelligence standardised mortality ratios (88-143, for the period April 2005-March 2006), which had purchased the Dr Foster Intelligence Real Time Monitoring computer system and so were able to

Table 1 | Interactions between case mix variables and hospital

Variable and year*	Hospital specific odds ratio† (95% CI)				Likelihood ratio test‡; P value
	GEH	MSH	UHC	UHN	
Age (per 10 year age group)					
Charlson index (per unit increase in Charlson index)					
April 2005 to March 2006	1.07 (1.02 to 1.13)	1.06 (1.00 to 1.12)	0.99 (0.96 to 1.02)	1.00 (0.96 to 1.03)	χ^2=13.10; P=0.01
April 2006 to March 2007	1.02 (0.98 to 1.07)	1.01 (0.95 to 1.07)	0.99 (0.95 to 1.02)	0.92 (0.90 to 0.95)	χ^2=26.63; P<0.0001
April to October 2007	1.02 (0.95 to 1.09)	0.96 (0.89 to 1.03)	1.00 (0.96 to 1.05)	0.99 (0.95 to 1.03)	χ^2=1.64; P=0.80
Emergency admission					
April 2005 to March 2006	1.68 (1.21 to 2.34)	1.76 (1.23 to 2.52)	1.44 (1.22 to 1.71)	1.79 (1.46 to 2.20)	χ^2=77.18; P<0.0001
April 2006 to March 2007	2.14 (1.39 to 3.29)	4.55 (2.79 to 7.42)	1.75 (1.46 to 2.11)	3.09 (2.58 to 3.69)	χ^2=322.66; P<0.0001
April to October 2007	2.68 (1.45 to 4.96)	1.85 (1.16 to 2.95)	1.38 (1.10 to 1.74)	1.45 (1.18 to 1.80)	χ^2=42.48; P<0.0001
April 2005 to March 2006	1.01 (1.00 to 1.01)	1.00 (1.00 to 1.01)	1.00 (1.00 to 1.01)	1.00 (1.00 to 1.00)	χ^2=8.11; P=0.09
April 2006 to March 2007	1.00 (1.00 to 1.01)	1.01 (1.00 to 1.01)	1.00 (1.00 to 1.01)	1.01 (1.00 to 1.01)	χ^2=25.00; P=0.0001
April to October 2007	1.00 (0.99 to 1.01)	1.00 (0.99 to 1.01)	1.00 (1.00 to 1.01)	1.00 (0.99 to 1.00)	χ^2=3.01; P=0.56
Previous emergency admissions (per extra admission)					
April 2005 to March 2006	1.02 (0.96 to 1.10)	1.06 (0.98 to 1.15)	1.01 (0.97 to 1.05)	1.00 (0.95 to 1.05)	χ^2=3.20; P=0.53
April 2006 to March 2007	1.06 (0.99 to 1.14)	1.10 (1.02 to 1.19)	1.07 (1.03 to 1.12)	0.99 (0.95 to 1.03)	χ^2=19.61; P=0.0006
April to October 2007	0.92 (0.84 to 1.02)	1.05 (0.95 to 1.16)	1.03 (0.97 to 1.09)	1.00 (0.94 to 1.06)	χ^2=3.97; P=0.41
Sex					
April 2005 to March 2006	1.10 (0.95 to 1.27)	0.91 (0.78 to 1.06)	1.02 (0.92 to 1.13)	0.97 (0.87 to 1.09)	χ^2=3.23; P=0.52
April 2006 to March 2007	1.02 (0.88 to 1.19)	0.89 (0.76 to 1.05)	1.12 (1.01 to 1.25)	1.03 (0.94 to 1.14)	χ^2=7.20; P=0.13
April to October 2007	0.99 (0.80 to 1.22)	0.96 (0.78 to 1.19)	1.07 (0.93 to 1.23)	0.90 (0.79 to 1.03)	χ^2=3.27; P=0.51
Deprivation (per fifth)					
April 2005 to March 2006	1.00 (0.95 to 1.05)	1.02 (0.96 to 1.01)	1.00 (0.97 to 1.04)	1.01 (0.97 to 1.04)	χ^2=0.38; P=0.98
April 2006 to March 2007	1.00 (0.95 to 1.06)	1.02 (0.96 to 1.09)	1.02 (0.96 to 1.09)	0.98 (0.95 to 1.02)	χ^2=3.01; P=0.56
April to October 2007	0.98 (0.91 to 1.06)	0.94 (0.87 to 1.02)	1.02 (0.97 to 1.07)	1.05 (1.00 to 1.09)	χ^2=6.79; P=0.15

GEH=George Eliot Hospital; MSH=Mid Staffordshire Hospitals; UHC=University Hospitals Coventry and Warwickshire; UHN=University Hospital North Staffordshire.
*Year three, April to October 2007, is a part year because these were the most recent data available at the time of study.
†For relation between each case mix variable and mortality over and above that accounted for in Dr Foster Intelligence case mix adjustment equation.
‡Global test for systematic deviation from odds ratio=1 in any hospital; df=4.

Table 2 | Case mix profiles at four hospitals for baseline year April 2005-March 2006. Values are numbers (percentages) unless stated otherwise

Characteristic	Hospital			
	GEH	MSH	UHC	UHN
Published Dr Foster Intelligence SMR	143	127	123	88
Admissions	10 903	13 767	31 307	40 971
In-hospital deaths (% in-hospital mortality)	1083 (9.9)	869 (6.3)	2149 (6.9)	1489 (3.6)
Mean (SD) Charlson index	0.79 (1.18)	0.57 (1.01)	1.09 (1.46)	1.54 (1.72)
Mean (SD) coding depth	1.62 (1.75)	1.49 (1.70)	1.42 (1.52)	2.12 (1.91)
Mean (SD) age (years)	65 (27)	61 (29)	60 (22)	63 (25)
Sex ratio (female/male)	1.08	1.11	1.02	1.00
Deprivation:				
Most deprived	1456 (13.4)	1546 (11.2)	7128 (22.8)	9405 (23.0)
Above average deprivation	2902 (26.6)	2987 (21.7)	7324 (23.4)	10 256 (25.0)
Average deprivation	2780 (25.5)	2766 (20.1)	7106 (22.7)	8415 (20.5)
Below average	2097 (19.2)	3528 (25.6)	5175 (16.5)	7505 (18.3)
Least deprived	1668 (15.3)	2940 (21.4)	4574 (14.6)	5390 (13.2)
Emergency admissions	7292 (66.9)	8883 (64.5)	18 225 (58.2)	17 828 (43.5)
Re-admissions	1072 (9.8)	1196 (8.7)	3596 (11.5)	3564 (8.7)
Length of stay (days):				
Mean (SD)	6.7 (11.7)	5.9 (12.6)	6.7 (15.6)	3.8 (12.6)
Median (interquartile range)	2 (0-8)	1 (0-6)	2 (0-7)	1 (0-4)

GEH=George Eliot Hospital; MSH=Mid Staffordshire Hospitals; SMR=standardised mortality ratio; UHC=University Hospitals Coventry and Warwickshire; UHN=University Hospital North Staffordshire.

provide anonymised output data (including case mix variables, the Dr Foster Intelligence predicted risk of death, and whether a death occurred) for this study. The hospital with the lowest standardised mortality ratio (88) is a large teaching hospital (University Hospital North Staffordshire, 1034 beds); those with higher ratios were one large teaching hospital (123: University Hospitals Coventry and Warwickshire, 1139 beds) and two medium sized acute hospitals (127: Mid Staffordshire Hospitals, 474 beds; 143: George Eliot Hospital, 330 beds).

Our analyses are based on data and predictions from the Real Time Monitoring system, which were available for the following time periods: April 2005 to March 2006 (year 1), April 2006 to March 2007 (year 2), and April to October 2007 (part of year 3—the most recent data available at the time of the study).

Statistical analyses

We constructed logistic regression models to test for interactions to assess whether the case mix adjustment variables used in the Dr Foster Intelligence method were prone to the constant risk fallacy. The Dr Foster Intelligence dataset includes the predicted risk of death for each patient, generated from the Dr Foster Intelligence case mix adjustment model, which we included (after logit transformation) as an offset term in a logistic regression model of in-hospital deaths. To this model we added terms for each hospital (thus allowing for the differences between hospitals in adjusted mortality) and then interaction terms for each hospital

and case mix variable in turn (which estimate the degree to which the relation between the case mix variable and mortality in each hospital differed from that implemented in the Dr Foster Intelligence case mix adjustment model).

Interaction terms that produced odds ratios close to one indicated that the relation between the case mix variable and mortality was constant and so not prone to the constant risk fallacy. The presence of large significant interactions suggested that the case mix variable was potentially prone to the constant risk fallacy, because its relation to mortality differed from the Dr Foster Intelligence national estimate. We tested the significance of interactions by using likelihood ratio tests; we deemed P values ≤0.01 to be statistically significant. We report the odds ratios, including 95% confidence intervals and P values, for each hospital-variable interaction over the three years.

Selected variables

The following patient level variables included in the Dr Foster Intelligence adjustment were available and tested: Charlson index (0-6, a measure of comorbidity), age (10 year age bands), sex (male/female), deprivation (fifths), primary diagnosis (1 of 56), emergency admission (no/yes), and the number of emergency admissions in the previous year (0, 1, 2, 3, or more). We excluded the palliative care variable from our analyses because no admissions to this specialty occurred in the hospitals. We excluded less than 1.5% of all the data from the Real Time Monitoring system because of missing data (for example, age not known, deprivation not known). The total numbers of admissions for each year were 96 948 (April 2005 to March 2006), 126 695 (April 2006 to March 2007), and 62 639 (April to October 2007, a part year).

For two prominent case mix variables—the Charlson index of comorbidity and emergency admission—we did detective work to seek explanations for the presence of large interaction effects, as described below.

Investigation of interaction effects seen with Charlson index

Patients with a lower Charlson index (less comorbidity) have lower expected mortality in the Dr Foster Intelligence model. Therefore, if the Charlson index was systematically under-coded in some hospitals they would be assigned artificially inflated standardised mortality ratios. We investigated the possibility of such misclassification in the Charlson index in two ways.

Firstly, we investigated changes in the depth of clinical coding (number of ICD-10 codes for secondary diagnoses identified per admission) over time within the hospitals and examined the hypothesis that the increase would be most rapid in those starting with the lowest Charlson indices (as they have the greatest headroom to improve through better coding). We formed the contingent hypothesis that any such change would be accompanied by diminished interactions

between Charlson index and mortality across hospitals.

Secondly, we considered that if clinical coding was similarly accurate in all hospitals, then differences in the Charlson index should reflect genuine differences in case mix profiles. We postulated that hospitals with higher Charlson indices were therefore more likely to admit older patients and to have higher proportions of emergency admissions, longer lengths of stay, and a higher crude mortality. If this was not the case, then this finding would corroborate a hypothesis that differences in the Charlson indices across hospitals were primarily attributable to systematic differences in clinical coding practices.

Investigation of interaction effects seen with emergency admission

In the original analyses by Jarman et al, the emergency admission variable was noted to be the best predictor of hospital mortality.[1] We explored this variable in more depth by investigating the proportion of emergency admissions that were recorded as having zero length of stay (being admitted and discharged on the same day). Although clinically valid reasons may exist to admit patients for zero stay, and some patients may die on admission, the practice of admitting less seriously ill patients has been recognised as a strategy that is increasingly used in the NHS to comply with accident and emergency waiting time targets.[18 19] This potentially leads to a reduction in the mortality risk associated with emergency admissions in hospitals that more often follow this practice. We examined the magnitude of differences in the proportion of emergency admissions with zero length of stay both within hospitals over time and between hospitals, as well as the observed risk associated with zero and non-zero lengths of stay.

RESULTS

We determined the extent to which case mix variables used in the Dr Foster Intelligence method had a non-constant relation with mortality across hospitals by examining the odds ratios of interaction terms for hospital and case mix variables derived from a logistic regression model (with death as the outcome). Table 1 reports the odds ratios of tests of interactions for six case mix variables.

Two variables (sex and deprivation) had no significant interaction with hospitals, indicating that these two variables are safe to use for case mix adjustment because they are not prone to the constant risk fallacy. However, the remaining variables had significant interactions. The number of previous emergency admissions was significant in year 2; the three hospitals with high standardised mortality ratios had 6% to 10% increases in odds of death with every additional previous emergency admission over and above the allowance made in the Dr Foster Intelligence model. Age had a significant interaction in year 2, but the effect was small—a 10 year age change was associated with an additional 1% increase in odds of death across the hospitals. Primary diagnosis also had significant interactions in all three years (results not shown as 56 categories and four hospitals produce 224 interaction terms).

The Charlson index had significant interaction effects in year 1 and year 2 but not in year 3. A unit change in the Charlson index was associated with a wide range of effect sizes—up to a 7% increase in odds of death (George Eliot Hospital, year 1) and an 8% reduction in odds of death (University Hospital North Staffordshire, year 2) over and above that accounted for in the Dr Foster Intelligence model. Across the full range of the Charlson index, these correspond to increases in odds of death of 50% or decreases of 39%.

We found significant interactions with being an emergency admission in all years across all hospitals. The effect sizes ranged from 38% (University Hospital North Staffordshire, year 3) to 355% (Mid Staffordshire Hospitals, year 2) increases in odds of death above those accounted for in the Dr Foster Intelligence equation.

Investigation of interaction effects seen with Charlson index

The 96 948 admissions in the four hospitals for 2005/06 had an overall mean Charlson index of 1.17 (median 1, interquartile range 0-2). Table 2 shows the mean Charlson index for the four study hospitals. The hospital with a low standardised mortality ratio (University Hospital North Staffordshire) had the highest mean Charlson index (1.54), whereas the three hospitals with high standardised mortality ratios had mean Charlson index values near or below the median (1).

An indicator of completeness of coding is depth of coding—the number of ICD-10 codes per admission (table 2). University Hospital North Staffordshire had the highest mean coding depth and Charlson index in all years; more importantly, as coding depth increased over the years in all hospitals (table 3), the interaction between the Charlson index and hospitals became smaller and statistically non-significant (table 1). We also explored the extent to which differences in the

Table 3 | Charlson index and coding depth over three years. Values are mean (SD)

Variable and hospital	Year 1: April 2005 to March 2006	Year 2: April 2006 to March 2007	Year 3*: April to October 2007
Charlson index			
George Eliot Hospital	0.79 (1.18)	0.88 (1.28)	1.02 (1.33)
Mid Staffordshire Hospitals	0.57 (1.01)	0.93 (1.15)	0.96 (1.18)
University Hospitals Coventry and Warwickshire	1.09 (1.46)	1.33 (1.54)	1.19 (1.38)
University Hospital North Staffordshire	1.54 (1.73)	1.70 (1.58)	1.44 (1.56)
Coding depth			
George Eliot Hospital	1.61 (1.75)	2.47 (2.23)	2.54 (2.23)
Mid Staffordshire Hospitals	1.49 (1.70)	1.48 (1.57)	1.83 (1.84)
University Hospitals Coventry and Warwickshire	1.42 (1.52)	1.61 (1.58)	1.74 (1.65)
University Hospital North Staffordshire	2.12 (1.91)	2.63 (2.14)	2.65 (2.34)

*Year 3 is a part year—most recent data available at time of study.

Table 4 | Proportions of non-emergency and emergency admissions with zero length of stay in study hospitals over three years. Values are percentages (numbers*)

Admission and year	Hospital			
	GEH	MSH	UHC	UHN
Non-emergency				
April 2005 to March 2006	71.4 (2580/3611)	67.1 (3277/4884)	60.2 (7881/13 082)	72.9 (16 860/23 143)
April 2006 to March 2007	74.1 (2579/3482)	82.4 (6520/7912)	71.7 (11 870/16 545)	83.4 (36 057/43 229)
April to October 2007†	74.2 (1683/2268)	82.0 (3644/4443)	68.8 (5947/8648)	79.6 (13 257/16 658)
Emergency				
April 2005 to March 2006	10.4 (758/7292)	19.6 (1743/8883)	17.7 (3220/18 225)	15.1 (2699/17 828)
April 2006 to March 2007	11.0 (778/7089)	18.1 (1525/8413)	15.8 (2781/17 606)	20.4 (4568/22 419)
April to October 2007†	12.9 (493/3828)	16.0 (764/4784)	14.6 (1387/9521)	17.7 (2212/12 489)

GEH=George Eliot Hospital; MSH=Mid Staffordshire Hospitals; UHC=University Hospitals Coventry and Warwickshire; UHN=University Hospital North Staffordshire.
*Denominator represents total number of elective or emergency admissions; numerator is number of elective or emergency admissions with zero length of stay (days).
†Part year—most recent data available at time of study.

Charlson index between hospitals reflect genuine differences in case mix profiles (table 2). Although University Hospital North Staffordshire serves a more deprived population with a higher proportion of male patients than the other hospitals, the percentage of emergency admissions, readmissions, length of stay, and crude mortality are at variance with the view that this hospital treats a systematically "sicker" population of patients. The evidence from table 2 is therefore inconsistent with the explanation that differences in the Charlson index reflect genuine differences in case mix profiles.

Investigation of interaction effects seen with emergency admission

We investigated the emergency admission variable in more depth by considering proportions of emergency/non-emergency admissions with a zero length of stay (days). Combining data across hospitals, the crude in-hospital mortality for non-emergency admissions was 1/1000 for zero length of stay and 23/1000 for non-zero length stay; the mortality for emergency admissions was 46/1000 for zero length of stay and 107/1000 for non-zero length of stay. Table 4 shows that the proportion of emergency admissions with zero length of stay varied between 10.4% and 20.4% across hospitals. The hospital with the lower case mix adjusted standardised mortality ratio (University Hospital North Staffordshire) had the highest proportion of zero stay emergency patients in years 2 and 3 (20.4% and 17.7%), whereas the hospital with the highest standardised mortality ratio (George Eliot Hospital) had the lowest proportion of zero stay emergency patients in all three years (10.4%, 11.0%, and 12.9%). The large variations in proportions of emergency/non-emergency patients with zero length of stay indicate that systematically different admission policies were being adopted across hospitals. The net effect of this is that the relation between an emergency admission and risk of death varies sustainably across hospitals (that is, the risk of death is not constant), apparently because of differences in hospital admission policies.

DISCUSSION

The league tables of mortality for NHS hospitals in England from Dr Foster Intelligence,[12] compiled by using case mix adjustment methods that have been internationally adopted or adapted,[2-6] have been published annually since 2001 and continue to raise concerns about the wide variations in standardised mortality ratios for hospitals and quality of care.[20] Unsurprisingly perhaps, similar concerns have been raised in other countries that have developed their own standardised mortality ratios.[5 21] Before such concerns can be legitimately aired, we must ensure that methods used by Dr Foster Intelligence are fit for purpose and not potentially misleading.[8 9]

Our results show that a critical, hitherto often overlooked, methodological concern is that the relation between risk factors used in case mix adjustment and mortality differs across the hospitals, leading to the constant risk fallacy. This phenomenon can increase the very bias that case mix adjustment is intended to reduce.[16] The routine use of locally collected administrative data for case mix variables makes this a real concern.[16] A serious problem is that no statistical fix exists for overcoming the challenges of variables susceptible to this constant risk fallacy.[16] It has to be investigated by a more painstaking inquiry.

As the Dr Foster Intelligence method, like other case mix adjustment methods, does not report screening variables for non-constant risk,[1-12] we investigated seven variables and found that three of them—age, sex, and deprivation—were safe in this respect. However, we found that emergency admission, the Charlson (comorbidity) index, primary diagnosis, and the number of emergency admissions in the previous year had clinically and statistically significant interaction effects. For two variables, the Charlson index and emergency admission, we found credible evidence to suggest that they are prone to the constant risk fallacy caused by systematic differences in clinical coding and emergency admission practices across hospitals.

For the Charlson index variable, we showed how the interaction effects seemed to relate to the number of

ICD-10 codes (for secondary diagnoses) per admission —that is, depth of clinical coding.[22] Overall, we reasoned that as the increased depth of coding (over time) was accompanied by a decrease in the interaction effect and as differences in the Charlson index did not reflect genuine differences in case mix profiles, we could reasonably conclude that the Charlson index is prone to the constant risk fallacy largely as a result of differential measurement error from clinical coding practices. Drawbacks in determining the Charlson index by using administrative datasets have been reported previously.[23] Hospitals with a lower depth of coding were disadvantaged because this was associated with a lower Charlson index, which in turn under-estimated the expected mortality and so inflated the standardised mortality ratio. For the emergency admission variable, we found strong evidence of systematic differences across hospitals in numbers of patients admitted as emergencies who were admitted and discharged on the same day. The higher risk usually associated with emergencies would be diluted by the inclusion of zero length of stay admissions in some hospitals. Thus, we judge these two variables— the Charlson index and emergency admission—to be unsafe to use in case mix adjustment methods because, ironically, their inclusion may have increased the bias that case mix adjustment aims to reduce. Further research to understand the mechanisms behind the other variables with large interactions is clearly warranted.

Given that our analyses are based on a subset of hospitals in the West Midlands, our study urgently needs to be replicated with more hospitals (for example, at the national level) to examine the extent to which our findings are generalisable. Furthermore, given the widespread use of standardised mortality ratios for hospitals in other countries (such as the United States,[2 3] Canada,[4] the Netherlands,[5] and Sweden[6]), with similar methods to those of Dr Foster Intelligence, we are concerned that these comparisons may also be compromised by the possibility of the constant risk fallacy. In addition, given the widespread use of case mix adjusted outcome comparisons in health care (for example, for producing standardised mortality ratios to compare intensive care units[8]), we urge that all case mix adjustment methods should screen (and report) variables for their susceptibility to the constant risk fallacy. A similar analysis could also be done within a single hospital, such that a logistic regression model with an offset term could be used to discover which set of the case mix variables has any systematic relation with mortality over and above the original adjustments. This may be an effective way for a hospital to identify variables that are susceptible to the constant risk fallacy and may give hospitals, especially those with a high standardised mortality ratio, a focal point for their subsequent investigations. Hospitals with low standardised mortality ratios may also find this analysis useful in increasing their understanding of their standardised mortality ratio.

Our findings suggest that the current Dr Foster Intelligence method is prone to bias and that any claims that variations in standardised mortality ratios for hospitals reflect differences in quality of care are less than credible.[8 12] Indeed, our study may provide a partial explanation for understanding why the relation between case mix adjusted outcomes and quality of care has been found to be so dubious.[24] Nevertheless, despite such evidence, assertions that variations in standardised mortality ratios reflect quality of care are widespread,[25] resulting, unsurprisingly, in institutional stigma by creating enormous pressure on hospitals with high standardised mortality ratios and provoking regulators such as the Healthcare Commission to react.[20]

We urge that screening case mix variables for non-constant risk relations needs to become an integral part of validating case mix adjustment methods. However, even with apparently safe case mix adjustment methods, we caution that we cannot reliably conclude that the differences in adjusted mortality reflect quality of care without being susceptible to the case mix adjustment fallacy,[10] because case mix adjustment by itself is devoid of any direct measurement of quality of care.[26]

WHAT IS ALREADY KNOWN ON THIS TOPIC

Case mix adjusted hospital standardised mortality ratios are used around the world in an effort to measure quality of care

However, valid case mix adjustment requires that the relation between each case mix variable and mortality is constant across all hospitals (a constant risk relation)

Where this requirement is not met, case mix adjustment may be misleading, sometimes to the degree that it will actually increase the very bias it is intended to reduce

WHAT THIS STUDY ADDS

Non-constant risk relations exist for several case mix variables used by Dr Foster Intelligence to derive standardised mortality ratios for English hospitals, raising concern about the validity of the ratios

The cause of the non-constant risk relation for two case mix variables—a comorbidity index and emergency admission —is credibly explained by differences in clinical coding and hospitals' admission practices

Case mix adjustment methods should screen case mix variables for non-constant risk relations

---

1 Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999;318:1515-20.

2 Institute for Healthcare Improvement. *Moving your dot: measuring, evaluating, and reducing hospital mortality rates*. Cambridge: Institute for Healthcare Improvement, 2003 (available at www.ihi.org/NR/rdonlyres/CC13C05E-9435-4692-8D7B-771BBAF44C9C/3609/FinalMoveDot1.pdf).

3 Institute for Healthcare Improvement. *Reducing hospital mortality rates (part 2)*. Cambridge: Institute for Healthcare Improvement, 2005 (available at www.ihi.org/NR/rdonlyres/CC13C05E-9435-4692-8D7B-771BBAF44C9C/3605/ReducingHospitalMortalityRates2WhitePaper2005.pdf).

4 Canadian Institute for Health Information. *HSMR: a new approach for measuring hospital mortality trends in Canada*. Ottawa: CIHI, 2007.

5 Heijink R, Koolman X, Pieter D, van der Veen A, Jarman B, Westert G. Measuring and explaining mortality in Dutch hospitals; the hospital standardized mortality rate between 2003 and 2005. *BMC Health Serv Res* 2008;8:73.

6 Koster M, Jurgensen U, Spetz C, Rutberg H. [Standardized hospital mortality as quality measurement in healthcare centres and hospitals.] *Lakar Tidningen* 2008:19:8:1391-6. (In Swedish.)

7 Dr Foster Intelligence. What is Dr Foster Intelligence? www.drfosterintelligence.co.uk/aboutUs/.

8 Imperial College London. Dr Foster Unit at Imperial College. 2008. www1.imperial.ac.uk/medicine/research/researchthemes/publicandint/research/drfosters/.

9 Moran JL, Solomon PJ. Mortality and other event rates: what do they tell us about performance? *Crit Care Resusc* 2003;5:292-304.

10 Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147-54.

11 Iezzoni LI. The risks of risk adjustment. *JAMA* 1997;278:1600-7.

12 Dr Foster Intelligence. Hospital guide. www.drfoster.co.uk/hospitalguide.

13 The Guardian. Hospital surgery death rates to be made public. 2008. www.guardian.co.uk/society/2008/may/29/nhs.health1.

14 Jarman B. Medical Meccas: which hospital is best? *Newsweek International* 2006 Oct 30 (available at www.newsweek.com/id/45117).

15 Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7(27).

16 Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* 2007;61:1010-3.

17 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis* 1987;40:373-83.

18 Locker TE, Mason SM. Are we hitting the target but missing the point? Analysis of the distribution of time patients spend in the emergency department. *BMJ* 2005;330:1188-9.

19 Healthcare Commission. *Acute hospital portfolio review. Management of admissions in acute hospitals: review of the national findings—2006*. London: Healthcare Commission, 2006 (available at www.healthcarecommission.org.uk/_db/_documents/Management_of_admissions_national_report.pdf).

20 Smith R. Hospital deaths are a postcode lottery. *Daily Telegraph* 2008 Mar 26. www.telegraph.co.uk/news/main.jhtml?xml=/news/2008/03/25/nmedi125.xml.

21 Canadian Television. Experts hope report helps cut hospital mortality. 2007. www.ctv.ca/servlet/ArticleNews/story/CTVNews/20071129/hsmr_071129/20071129?hub=Specials.

22 Dixon J, Sanderson C, Elliott P, Walls P, Jones J, Petticrew M. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. *J Public Health Med* 1998;20:63-9.

23 Van Doorn C, Bogardus ST, Williams SC, Concato J, Towle VR, Inouye SK. Risk adjustment for older hospitalized persons: a comparison of two methods of data collection for the Charlson index. *J Clin Epidemiol* 2001;54:694-701.

24 Pitches D, Mohammed MA, Lilford R. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality of care? A systematic review of the literature. *BMC Health Serv Res* 2007;7:91.

25 Marshall M, Shekelle P, Leatherman S, Brook R. Public disclosure of performance data: learning from the US experience. *Qual Health Care* 2000;9:53-7.

26 Lilford R, Brown C, Nicholl J. Use of process measures to monitor quality of care. *BMJ* 2007;335:648-50.