

Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models

Paul Aylin, clinical senior lecturer,¹ Alex Bottle, lecturer,¹ Azeem Majeed professor of primary care and social medicine²

¹Dr Foster Unit, Imperial College London, London EC1A 9LA

²Department of Primary Care and Social Medicine, Imperial College London

Correspondence to: P Aylin
p.aylin@imperial.ac.uk

doi: 10.1136/bmj.39168.496366.55

ABSTRACT

Objective To compare risk prediction models for death in hospital based on an administrative database with published results based on data derived from three national clinical databases: the national cardiac surgical database, the national vascular database and the colorectal cancer study.

Design Analysis of inpatient hospital episode statistics. Predictive model developed using multiple logistic regression.

Setting NHS hospital trusts in England.

Patients All patients admitted to an NHS hospital within England for isolated coronary artery bypass graft (CABG), repair of abdominal aortic aneurysm, and colorectal excision for cancer from 1996-7 to 2003-4.

Main outcome measures Deaths in hospital. Performance of models assessed with receiver operating characteristic (ROC) curve scores measuring discrimination (<0.7 =poor, $0.7-0.8$ =reasonable, >0.8 =good) and both Hosmer-Lemeshow statistics and standardised residuals measuring goodness of fit.

Results During the study period 152 523 cases of isolated CABG with 3247 deaths in hospital (2.1%), 12 781 repairs of ruptured abdominal aortic aneurysm (5987 deaths, 46.8%), 31 705 repairs of unruptured abdominal aortic aneurysm (3246 deaths, 10.2%), and 144 370 colorectal resections for cancer (10 424 deaths, 7.2%) were recorded. The power of the complex predictive model was comparable with that of models based on clinical datasets with ROC curve scores of 0.77 (v 0.78 from clinical database) for isolated CABG, 0.66 (v 0.65) and 0.74 (v 0.70) for repairs of ruptured and unruptured abdominal aortic aneurysm, respectively, and 0.80 (v 0.78) for colorectal excision for cancer. Calibration plots generally showed good agreement between observed and predicted mortality.

Conclusions Routinely collected administrative data can be used to predict risk with similar discrimination to clinical databases. The creative use of such data to adjust for case mix would be useful for monitoring healthcare performance and could usefully complement clinical databases. Further work on other procedures and diagnoses could result in a suite of models for performance adjusted for case mix for a range of specialties and procedures.

INTRODUCTION

Routine administrative databases are increasingly being used for performance monitoring in healthcare in the United Kingdom (such as www.healthcarecommission.org.uk and www.drfooster.co.uk), United States (such as www.ihl.org/IHI/Programs/Campaign/), and elsewhere.¹ In comparisons of performance between clinicians or organisations it is essential to adjust for several parameters including comorbidity and severity of disease (case mix). Routine data, however, might contain insufficient information for adequate adjustment. Clinical databases, run by various bodies including professional societies, could potentially record more detailed clinical information and might permit better adjustment for case mix. A survey of 105 multi-centre clinical databases (which included hospital episode statistics, the administrative database available within England) found that their distribution was uneven and that their scope and the quality of the data was variable.² The report from the public inquiry into deaths at a paediatric cardiac unit at Bristol criticised this “dual” system as “wasteful and anachronistic.”³ It also suggested that hospital episode statistics should be supported as a major national resource and used to undertake monitoring of a range of healthcare outcomes.

We examined mortality for three index procedures (coronary artery bypass graft, abdominal aortic aneurysm repair, and colectomy for bowel cancer) used in three large clinical datasets (the national adult cardiac surgical database, the national vascular database, and a colorectal cancer database collected by the Association of Coloproctology of Great Britain and Ireland). We compared risk adjustment models for mortality, based on administrative data, with published models based on data from the clinical databases and assessed the ability of each model to predict death.

Background

The Society of Cardiothoracic Surgeons has collected voluntary data from its members for over 25 years and individual patient level data since 1996 and in 2003 introduced the national cardiac surgical database (NCSD). Some 40 units contribute to the database, which contains information on over 210 000 individual records. The central cardiac audit database (CCAD) is

now used for all cardiac procedures and will incorporate the national cardiac surgical database. The society has published outcomes using several different risk prediction scores including that of Parsonnet et al,⁴ the EuroSCORE,⁵ and scores from both simple and complex models.⁶ The score accepted by most UK clinicians is the EuroSCORE, which is based on age, sex, and factors related to the patient (such as the presence of chronic pulmonary disease, cardiac factors such as the presence of unstable angina, and other factors related to the operation such as whether or not the admission was an emergency). Adult cardiac surgery was one of the key performance indicators for the Healthcare Commission.⁷

The Vascular Surgical Society of Great Britain and Ireland (VSSGBI) runs the national vascular database (NVD), which collects data voluntarily from surgeons on three procedures: repair of abdominal aortic aneurysm, carotid endarterectomy, and infra-inguinal bypass. At the time of the 2004 report, 259 surgeons in 99 hospitals were contributing data and there were 12 389 records on the database. Information collected

includes details of the operation performed, the surgical and anaesthetic staff involved, the patient's history and risk factors, biochemical and haematological parameters, and 30 day postoperative morbidity and mortality.⁸

The Association of Coloproctology of Great Britain and Ireland (ACPGBI) bowel cancer audit collects clinical data on patients with a diagnosis of bowel cancer, recorded either by consultant surgeons or dedicated audit staff. The database for April 2001 to March 2002 contained information from 93 healthcare trusts or hospitals with details of 10 613 cases of bowel cancer. Data from this audit have been used to create a model for predicting outcomes from colorectal cancer surgery.⁹ Models for predicting mortality include age, sex, the American Society of Anaesthesiology grade,¹⁰ Dukes's stage, urgency of the operation, and cancer excision.

Data on hospital activity have been collected since 1949 from all NHS hospitals in the UK.¹¹ Hospital episode statistics (HES) were introduced in 1986 and measure all hospital inpatient and day surgery activity for

Table 1 | Odds ratios (95% confidence intervals) for mortality in hospital for isolated coronary artery bypass (CABG), repair of ruptured abdominal aortic aneurysm (AAA), repair of unruptured AAA, and colorectal excision procedures for cancer for variables applying to all four index procedures

Variable and value	Isolated CABG	AAA with rupture	AAA without rupture	Colorectal excision
No of cases in training and validation set	152 523	12 781	31 705	144 370
No (%) of deaths in training and validation set	3247 (2.1)	5987 (46.8)	3246 (10.2)	10 424 (7.2)
Age (years):				
≥85	20.1 (12.5 to 32.42)	8.18 (5.17 to 12.96)	8.49 (5.55 to 13)	19.00 (14.43 to 25.02)
80-84	9.97 (6.90 to 14.41)	5.87 (3.79 to 9.08)	6.40 (4.3 to 9.51)	12.23 (9.3 to 16.1)
75-79	5.40 (3.89 to 7.77)	4.43 (2.88 to 6.80)	4.69 (3.18 to 6.94)	8.23 (6.26 to 10.83)
70-74	3.43 (2.44 to 4.84)	3.33 (2.17 to 5.12)	3.59 (2.43 to 5.31)	5.65 (4.29 to 7.44)
65-69	2.42 (1.72 to 3.42)	2.34 (1.52 to 3.61)	2.48 (1.67 to 3.69)	3.67 (2.78 to 4.86)
60-64	1.72 (1.22 to 2.44)	1.66 (1.07 to 2.58)	1.97 (1.31 to 2.97)	2.83 (2.13 to 3.76)
55-59	1.22 (0.85 to 1.75)	1.89 (1.19 to 2.99)	1.76 (1.13 to 2.74)	1.74 (1.29 to 2.34)
50-54	1.00 (0.68 to 1.47)	1.71 (1.02 to 2.86)	1.05 (0.6 to 1.84)	1.69 (1.23 to 2.32)
45-49	1.11 (0.73 to 1.68)	0.76 (0.37 to 1.60)	1.79 (1 to 3.22)	1.36 (0.94 to 1.97)
≤44	1	1	1	1
Sex:				
Female	1.39 (1.29 to 1.51)	1.06 (0.96 to 1.16)	1.23 (1.12 to 1.35)	0.77 (0.74 to 0.80)
Male	1	1	1	1
Method of admission:				
Emergency	1.54 (1.40 to 1.69)	1.38 (1.23 to 1.54)	2.76 (2.53 to 3.00)	3.46 (3.31 to 3.63)
Elective	1	1	1	1
Per year since 1996	0.93 (0.91 to 0.94)	0.96 (0.94 to 0.97)	0.95 (0.93 to 0.96)	0.99 (0.98 to 1.00)
Fifth of deprivation:				
5 (most deprived)	1.17 (1.04 to 1.31)	1.31 (1.16 to 1.47)	1.37 (1.2 to 1.55)	1.21 (1.13 to 1.29)
4	1.04 (0.92 to 1.16)	1.14 (1.01 to 1.28)	1.30 (1.15 to 1.48)	1.14 (1.07 to 1.22)
3	1.02 (0.91 to 1.14)	1.04 (0.93 to 1.17)	1.26 (1.11 to 1.42)	1.01 (0.95 to 1.08)
2	0.96 (0.85 to 1.08)	1.07 (0.95 to 1.2)	1.11 (0.97 to 1.26)	0.99 (0.93 to 1.06)
Unknown	0.79 (0.62 to 0.99)	0.92 (0.69 to 1.23)	1.00 (0.71 to 1.41)	0.77 (0.59 to 0.99)
1 (least deprived)	1	1	1	1
Per unit increase in Charlson comorbidity score (capped at 6)	1.72 (1.66 to 1.78)	1.32 (1.26 to 1.38)	1.70 (1.64 to 1.77)	1.23 (1.22 to 1.25)
Per previous emergency admission	1.20 (1.15 to 1.24)	1.10 (1.03 to 1.17)	1.08 (1.03 to 1.13)	1.17 (1.14 to 1.20)

England. The basic unit of activity is the finished consultant episode, covering the period a patient is under the care of one consultant. Every NHS hospital in England must submit data items of HES electronically for each episode in every patient's stay in that hospital. The data items are entered from the patient's notes onto the hospital's patient administration systems by trained clinical coders. The items include date of birth, sex, home postcode, and clinical data such as primary and secondary diagnoses and dates and details of any operations performed within the patient's stay. Diagnoses are coded with ICD-10 (international statistical classification of diseases, tenth revision); procedures use the UK Office of Population Censuses and Surveys classification (OPCS4). Since 1991, HES has been used for contracting in the internal market and now contain some fourteen million records per financial year.

HES data are often regarded as unreliable by clinicians because of considerable problems in the early years after their inception in 1986. McKee et al summed up the poor reputation of routine data in 1994: "Many clinicians have concluded that, despite a massive investment in technology, routinely collected data still fail . . . and that separate systems are still required."¹² Data quality has since improved considerably,^{13,14} and, if suitable predictive models could be developed using this routinely collected information source, they would be a valuable tool for generating measures of performance adjusted for case mix.

METHODS

We extracted data on all admissions in England for isolated coronary artery bypass graft (CABG, OPCS4 codes K40-K46), repair of abdominal aortic aneurysm (OPCS4 codes L18-L21), and colorectal excision (OPCS4 H06-H11, H33) for cancer (ICD10 C18-C20) for the period 1996-7 to 2003-4. After we linked episodes belonging to the same admission, we excluded records with invalid date of birth, sex, length of stay, or method of admission and duplicated records. We also excluded records for CABG if the procedure was preceded in the same admission by an angioplasty because we then considered it to be a "rescue" rather than the primary intended procedure. We divided repairs of abdominal aortic aneurysm into ruptured and non-ruptured (according to whether the primary diagnosis was I710, I711, I713, I715, or I718) to enable comparison with published results. We divided colorectal excisions into procedure subgroups by OPCS code. Data extracts were split randomly and equally into training sets and validation sets. Within HES, death in hospital in the same admission or after transfer to another unit was taken as the outcome.

Operations were classified as elective (admission method (ADMIMETH) 11 to 13) or non-elective (all other ADMIMETH values) as HES does not have an "urgent" category, unlike US admissions data or those from the Society of Cardiothoracic Surgeons. Age was divided into five year bands to ≥ 85 , but with those aged < 45 combined. We used secondary diagnosis fields to

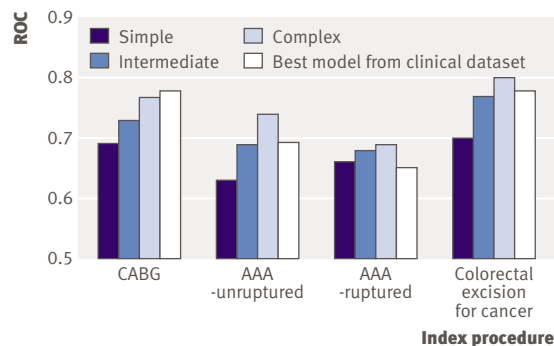


Fig 1 | ROC curve areas comparing simple (year, age, sex), intermediate (including method of admission), and complex models derived from HES with best model derived from clinical databases for index procedures (CABG=coronary artery bypass graft; AAA=abdominal aortic aneurysm)

create comorbidity variables used to make up the Charlson index.¹⁵ Further factors considered specific to each index procedure group were also considered (tables 1 and 2). The two variables we used that were not adjusted for in the models from the clinical databases were financial year and socioeconomic deprivation. Our measure of deprivation was the index of multiple deprivation for 2004 at super output area, linked through the patient's postcode.

We plotted each variable against the death rate to determine whether the relation, if any, was linear or if the variable should be categorised (age group and all dichotomous variables were automatically fitted as factors—that is, as categorical variables rather than as continuous covariates). We then used logistic regression to fit three models for each index procedure: a simple model—year, age, and sex only; an intermediate model—year, age, sex, method of admission, diagnostic, or operation subgroup; and a complex model—all appropriate variables in tables 1 and 2.

We compared these HES based models with the best published predictive risk model based on data from the clinical databases. For CABG and abdominal aortic aneurysms we used the most recent society reports available.^{6,8} For colorectal resection we used the published model in the report on risk adjusted outcomes from the Association of Coloproctology of Great Britain and Ireland.^{9,16} We compared models using receiver operating characteristic (ROC) curve scores (c statistics). The c statistic is the probability of assigning a greater risk of death to a randomly selected patient who died compared with a randomly selected patient who survived. A value of 0.5 suggests that the model is no better than random chance in predicting death. A value of 1.0 suggests perfect discrimination. In general, values less than 0.7 are considered to show poor discrimination, values of 0.7-0.8 can be described as reasonable, and values above 0.8 suggest good discrimination. The models were calibrated by plotting observed versus predicted numbers of deaths by tenth based on risk. A model that closely fits the observed outcome is desirable, and this can be tested

using a χ^2 type statistic developed by Hosmer and Lemeshow measuring goodness of fit.¹⁷ This test compares the number of observed cases with the number of predicted cases for each tenth of risk. As the performance of this test depends on sample size, we also inspected the proportion of residuals whose absolute values were greater than 1.96 (5% are expected to be greater than this value). We also checked for influential data points via their Cook's statistic, which have values greater than 1.¹⁸

RESULTS

Overall 3.4% of HES admissions in 1996-7 and 2.4% in 2003-4 had missing or invalid age, sex, admission method, or length of stay. After we excluded these records, in the eight year period there were 152 523 isolated CABGs with 3247 deaths in hospital (2.1%),

12 781 repairs of ruptured abdominal aortic aneurysm (5987 deaths, 46.8%), 31 705 repairs of unruptured abdominal aortic aneurysm (3246 deaths, 10.2%), and 144 370 colorectal resections for cancer (10 424 deaths, 7.2%). In the clinical databases mortality was 2.0% for isolated CABG (2003), 41.0% for ruptured abdominal aortic aneurysm (2001-2 to 2004-5), 6.8% for unruptured abdominal aortic aneurysm (2001-2 to 2004-5), and 7.4% for colorectal resections for cancer (1999-2000).

Tables 1 and 2 show the odds ratios for all the variables for each index procedure. As expected, patient's age was a strong predictor of mortality but many of the other variables in HES were also significant predictors of mortality (for example, deprivation and comorbidity). Models derived from the training and validation datasets gave similar odds ratios and c statistics. We

Table 2 | Odds ratios (95% confidence intervals) for mortality in hospital for isolated coronary artery bypass (CABG), repair of ruptured abdominal aortic aneurysm (AAA), repair of unruptured AAA, and colorectal excision procedures for cancer for variables specific to each index procedure

Variable and value	Isolated CABG	AAA with rupture	AAA without rupture	Colorectal excision
Coronary artery bypass graft (CABG):				
Revision	1.70 (1.33 to 2.18)	NA	NA	NA
First time	1	NA	NA	NA
Per previous admission for IHD	0.85 (0.80 to 0.91)	NA	NA	NA
Recent admission for myocardial infarction:				
One or more	1.34 (1.14 to 1.57)	NA	NA	NA
None	1	NA	NA	NA
Previous heart operation:				
One or more	0.70 (0.31 to 1.57)	NA	NA	NA
None	1	NA	NA	NA
No of arteries replaced:				
Four	1.39 (1.15 to 1.68)	NA	NA	NA
Three	1.30 (1.12 to 1.49)	NA	NA	NA
Two	1.09 (0.95 to 1.26)	NA	NA	NA
One	1	NA	NA	NA
Unknown or other specified	3.07 (2.42 to 3.90)	NA	NA	NA
Other bypass of coronary artery	1.30 (0.57 to 2.97)	NA	NA	NA
Connection of thoracic artery to coronary artery	0.96 (0.84 to 1.11)	NA	NA	NA
Part of aorta repaired:				
Unspecified	NA	0.67 (0.5 to 0.89)	0.4 (0.31 to 0.51)	NA
Endovascular	NA	0.83 (0.61 to 1.12)	1.48 (1.11 to 1.96)	NA
Thoracic	NA	0.54 (0.4 to 0.72)	0.68 (0.52 to 0.9)	NA
Infrarenal	NA	0.58 (0.45 to 0.76)	0.53 (0.4 to 0.7)	NA
Suprarenal	NA	1	1	NA
Part of colon/rectum removed:				
Left, right, transverse, sigmoid	NA	NA	NA	0.89 (0.72 to 1.09)
Segment of colon not stated	NA	NA	NA	1.65 (1.31 to 2.07)
APR	NA	NA	NA	0.96 (0.77 to 1.2)
Anterior resection	NA	NA	NA	0.97 (0.79 to 1.2)
Whole colon and/or rectum	NA	NA	NA	1.73 (1.35 to 2.21)
Hartmann's	NA	NA	NA	1.48 (1.19 to 1.83)
Unspecified	NA	NA	NA	1
Previous abdominal surgery:				
Yes	NA	0.82 (0.60 to 1.11)	0.97 (0.79 to 1.18)	0.93 (0.87 to 0.99)
No	NA	1	1	1

IHD=ischaemic heart disease; NA=not applicable.

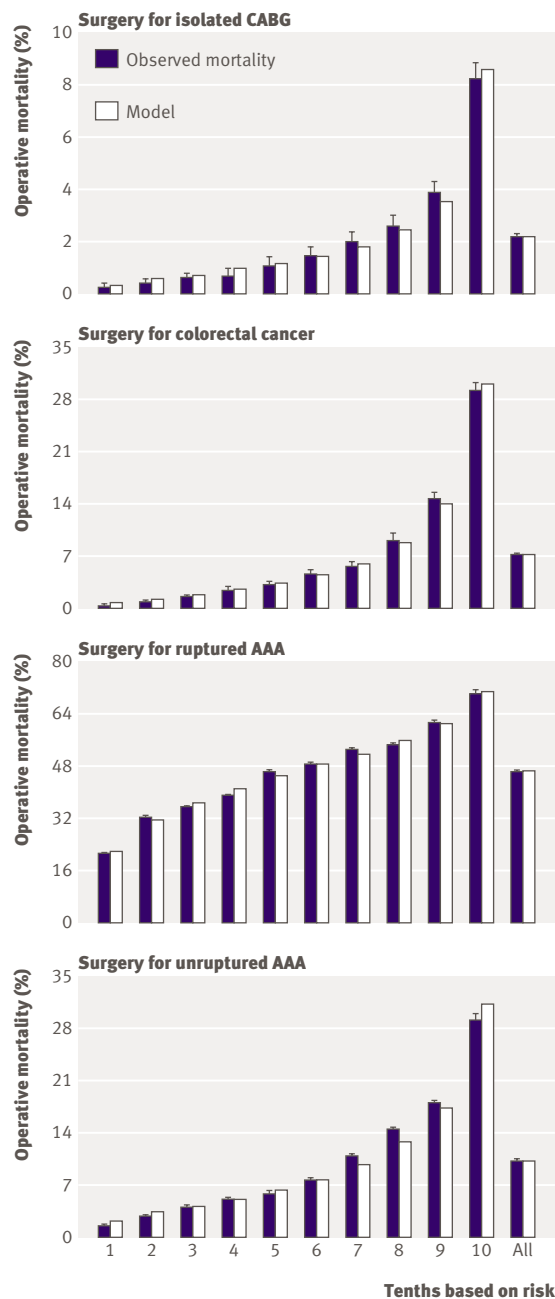


Fig 2 | Calibration plots for complex risk prediction models based on HES data for index procedures showing observed number of deaths against predicted based on validation set (CABG=coronary artery bypass graft; AAA=abdominal aortic aneurysm)

also trained the models on operations from 1996-7 to 2001-2, testing them on 2002-3 to 2003-4 so that the latter two years represent a “future” dataset to the training set. The c statistics differed by at most 0.02 (with the test set having the higher values for each procedure). Figure 1 shows the ROC c statistics for the three HES based models and published models based on clinical databases. For repairs of abdominal aortic aneurysm and colorectal excision for cancer, the model based on HES had better discrimination than that based on

the clinical database. For isolated CABG, the c statistic was similar (0.768 in HES *v* 0.783 from the national cardiac surgical database).

Figure 2 shows calibration plots based on the most complex models for the index procedures giving observed versus predicted deaths for tenths of risk from the Hosmer-Lemeshow test. Although there is generally some slight overestimation of risk for patients at low risk of death and an underestimation of patients at high risk, which is perhaps to be expected with a linear model, there seems to be close agreement between the observed and predicted numbers and therefore a high goodness of fit. The Hosmer-Lemeshow statistics in table 3, however, suggest a highly significant difference between the deaths predicted from our models and those observed for colorectal excision of cancer ($P=0.001$ for complex model). Repair of unruptured abdominal aortic aneurysm also shows borderline significance ($P=0.077$). The proportion of standardised residuals outside the range -1.96 to 1.96 , however, was 2.1% for CABG, 0.5% for repair of ruptured abdominal aortic aneurysm, and 5.1% for colorectal procedures. No influential points were discovered for any model.

DISCUSSION

We used HES data to build statistical models with good discrimination for predicting postoperative death as an outcome that were comparable with those derived from clinical databases in their predictive power. We now assess two key aspects of the models—discrimination and goodness of fit—and consider data quality and other issues relating to HES and clinical datasets.

Discrimination

HES data lack many clinical variables and have been criticised for being inadequate for monitoring performance, but for the index procedures examined in our study, the ROC curves were comparable with those from clinical datasets. Other than lacking clinical variables, the HES models differed in several ways from the clinical datasets: they included the year, area level fifth of deprivation, narrower age bands, and information derived from previous emergency admissions. The degree to which the non-HES models would improve with the use of five year age bands is unknown. We could not apply our HES age groups to the clinical datasets but the clinical models are validated and considered by the relevant surgical bodies to be the best currently available. A US study developed a model based on an administrative dataset (Veterans Affairs patient treatment file) for mortality after cardiac bypass surgery with a c statistic of 0.70 compared with a value of 0.76 from a clinical dataset model (clinical improvement in cardiac surgery programme).¹⁹ In a similar study looking at predicting mortality after non-cardiac surgery, the performance of the model ranged from good to fair (0.83 for orthopaedic surgery to 0.65 for thoracic surgery).²⁰

Simplified models of risk prediction might be as effective in predicting outcome as some complex models currently in use.^{21 22} The authors of the US study derived their own clinical groups to adjust for comorbidity after excluding conditions that might have arisen as a complication after surgery, and recognised that in so doing they may also have excluded comorbid diseases that were important for some patients.¹⁹ The Charlson index, which we used, also tries to exclude potential complications.¹⁵ For example, it excludes acute renal failure (ICD10 N17) and includes chronic renal failure (N18); however, it also includes unspecified renal failure (N18), which in practice will include some acute cases, of which some will be complications. We also fitted the components of the Charlson index as dummy variables instead of one continuous variable and obtained the same c statistic. Exclusion of the renal disease variable from our complex model reduced the c statistic for CABG marginally from 0.76 to 0.75.

Goodness of fit

When we used the method developed by Hosmer and Lemeshow¹⁷ the goodness of fit of at least one of the complex models seemed to be poor. For small samples, the test is known to have poor power to detect badly fitting models and the resulting P value may differ between software packages.²³ For large samples, which we clearly have for national data, even small (clinically unimportant) differences between observed and predicted numbers will seem significant. The calibration plots show good agreement between observed and predicted numbers of deaths and examinations of the residuals suggests that the small P value from the Hosmer-Lemeshow statistic is because of the large sample size. A better method for testing the goodness of fit in such cases might be to examine the residuals and check for influential points. With these criteria, all the complex models exhibit good fit.

Data quality

Concerns remain about the quality of HES data.¹³ The overall percentage of admissions with missing or

invalid data on age, sex, admission method, or dates of admission or discharge was 2.4% in 2003. For the remaining admissions, 47.9% in 1996 and 41.6% in 2003 had no secondary diagnosis recorded (41.9% and 37.1%, respectively, if day cases are excluded). In contrast to some of the clinical databases, if no information on comorbidity is recorded, we cannot tell whether there is no comorbidity present or if comorbidity has not been recorded. Despite these deficiencies, our predictive models are still good. In the most recent report of the Society of Cardiothoracic Surgeons, 30% of records had missing EuroSCORE variables.⁶ Within the Association of Coloproctology database, 39% of patients had missing data for the risk factors included in their final model.⁹ A comparison of numbers of vascular procedures recorded within HES and the national vascular database found four times as many cases recorded within HES.²⁴

A comparison of analyses based on HES and the Society of Cardiothoracic Surgeons' own data concluded that statistical correlation was good, although counts of operations were consistently lower within HES.⁶ This was probably because of our stricter definition of what constitutes an isolated CABG in HES. For complex specialist procedures, the OPCS4 coding system may not be suitable for monitoring outcomes,²⁵ but a revised version (v4.3) of the system is now available that should improve the recording of newer types of procedures. With the introduction of a system based on reimbursement for providers of health care for each individual case treated ("payment by results"²⁶), there are financial incentives to record diagnoses more thoroughly,^{27 28} which may help to improve completeness and accuracy in data abstraction and coding within the NHS even further.

Like the clinical databases in this study, HES does not capture deaths out of hospital, which will reduce mortality in hospital in trusts that discharge patients early. We were able to capture most deaths occurring after transfers to other NHS hospitals and so were missing only deaths after discharge home or to residential homes. National mortality data are now linked to HES,

Table 3 | Area under ROC curve (SE) and Hosmer-Lemeshow (H-L) statistic (P value*) for hospital episode statistics (HES) models and best models based on clinical databases for isolated coronary artery bypass (CABG), repair of ruptured abdominal aortic aneurysm (AAA), repair of unruptured AAA, and colorectal excision procedures for cancer

	Isolated CABG		AAA without rupture		AAA with rupture		Colorectal excision	
	ROC	H-L	ROC	H-L	ROC	H-L	ROC	H-L
HES based on validation set:								
Simple model (year, age, sex)	0.687 (0.007)	8.8 (0.551)	0.621 (0.007)	9.0 (0.532)	0.629 (0.007)	5.8 (0.832)	0.707 (0.004)	8.2 (0.609)
Intermediate model (including method of admission)	0.724 (0.006)	22.2 (0.014)	0.687 (0.007)	5.4 (0.863)	0.638 (0.007)	3.6 (0.964)	0.772 (0.003)	28.1 (0.002)
Complex model	0.768 (0.006)	15.1 (0.128)	0.740 (0.006)	16.9 (0.077)	0.660 (0.007)	4.0 (0.947)	0.803 (0.003)	30.3 (0.001)
Clinical datasets	0.783 (0.018)†	NA	0.695 (0.039)‡	NA (0.975)	0.651 (0.025)§	NA (0.665)	0.775 (0.016)	6.34 (0.610)

NA=not applicable.

*Based on 10 df.

†National adult cardiac surgical database—logistic EuroSCORE.⁶

‡National vascular database—Bayesian model validation dataset.⁸

§National vascular database—V-BHOM validation dataset.⁸

¶Association of Coloproctology of Great Britain and Ireland—validation data.⁹

WHAT IS ALREADY KNOWN ON THIS TOPIC

Routine administrative databases are increasingly being used to monitor performance. Clinical databases can potentially record more detailed clinical information and might permit better adjustment for case mix, but their scope and the quality of data are variable.

WHAT THIS STUDY ADDS

Good risk prediction with discrimination comparable with that obtained from clinical databases is possible with routinely collected administrative data. Creative use of administrative data to adjustment for case mix could be useful for monitoring provider performance and could usefully complement clinical databases.

which will allow longer term outcomes to be monitored.

We compared the performance of different models on different databases and it is important to remember that the performance of any model is also a reflection of the quality of the database and the type of patients it covers. HES and the other databases are not strictly comparable because of the high proportion of missing data in the Association of Coloproctology database and because the national vascular database and the Society of Cardiothoracic Surgeons' database include some hospitals outside England, unlike HES. If our comparisons of model prediction do not strictly compare like with like, they are still important because they reflect reality of the two sets of databases in their present form. The question we have asked is: which one is currently the best predictor of death?

Implications for practice

Clinical databases are expensive to compile and maintain. An exercise to look at the utility of electronic health data to assess new health technologies estimated costs per record ranging from around £10 (UK cardiac surgical register) to £60 (Scottish hip fracture audit) compared with £1 per record for HES.²⁹ Despite these costs, mortality adjusted for case mix by unit or surgeon is still not in the public domain from any of the three databases covered in our report, with the recent exception of unit level mortality adjusted for case mix for heart surgery published by the Healthcare Commission.³⁰

We selected our three index procedures a priori because they were common and because the models for risk prediction derived from clinical databases were published and easily accessible. Although work needs to be carried out on other procedures and diagnoses, we have shown the potential utility of administrative data for performance monitoring with adequate adjustment for case mix. There may, of course, be other clinical specialties where it is not possible to generate comparable risk models from routinely collected data (see www.icnarc.org/). Clinical databases also exist for reasons other than performance monitoring, including audit, case finding, and research. Our findings suggest that for monitoring outcomes, administrative databases may be as good as clinical databases. Administrative databases also have the advantage that they are available for the entire NHS and do not

depend on voluntary participation by individual clinicians and providers. Hence, they can be used to generate performance measures on all relevant provider units, adjusted for case mix and other relevant variables. These adjusted measures of performance are likely to be fairer and more accurate measures of the performance of clinicians and providers than the cruder measures generally available now. Furthermore, as the content of administrative databases in different countries is often broadly similar, methods of using these databases to generate outcome measures may be applicable in healthcare systems in many developed countries.

CONCLUSIONS

We have shown that for three common procedures, it is possible to use routinely collected administrative data to predict risk of death with discrimination comparable with that obtained from clinical databases. Ideally, clinical and administrative datasets should function as one and clinicians should take a role in institutional data collection.³¹ The creative use of administrative data for risk prediction and adjustment for case mix for monitoring mortality might be useful for performance monitoring and could usefully complement the outputs from clinical databases. Further work on other procedures and diagnoses could result in a suite of models for adjustment for case mix in several specialties. This would then allow the publication of better measures of performance of clinicians and providers and allow patients, primary care physicians, and healthcare purchasers to make more informed choices when selecting specialist services. Administrative databases would then start to justify the investments made in them by providing information for use in clinical audit and to help improve the quality of health care received by the population.

Contributors: PA and AB were involved in the original research question and extracted and analysed data. All authors drafted the paper and contributed comments on drafts. PA is the guarantor.

Funding: Dr Foster Intelligence.

Competing interests: The Dr Foster Unit at Imperial College is funded by a grant from Dr Foster Intelligence (an independent health service research organisation).

Ethical approval: We have approval under Section 60 granted by the Patient Information Advisory Group (PIAG) to hold patient identifiable data and analyse them for research purposes. We also have approval from St Mary's local research ethics committee.

- 1 Quan H. ICD-10-CA/CCI coding algorithms for defining clinical variables to assess outcome after aortic and mitral valve replacement surgery. *Can J Cardiol* 2006;22:153-4.
- 2 Black N, Barker M, Payne M. Cross sectional survey of multicentre clinical databases in the United Kingdom. *BMJ* 2004;328:1478.
- 3 *Learning from Bristol: the report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995*. Bristol: Bristol Royal Infirmary Inquiry, 2001 (CM5207). www.bristol-inquiry.org.uk/final_report/index.htm
- 4 Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;701(suppl):13-112.
- 5 Nashef SA, Roques F, Michel P, Gauducheau E, Lemeschou S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999 Jul;16(1):9-13.
- 6 Society of Cardiothoracic Surgeons of Great Britain and Ireland. *Fifth national adult cardiac surgical database report 2003*. Henley-upon-Thames: Dendrite Clinical Systems, 2004.

- 7 Healthcare Commission. *Performance ratings. Adult cardiac surgery*. London: Healthcare Commission, 2005. <http://ratings2005.healthcarecommission.org.uk/Trust/Indicator/indicatorDescriptionShort.asp?indicatorId=1400>
- 8 Vascular Society of Great Britain and Ireland. *Fourth national vascular database report 2004*. Henley-upon-Thames: Dendrite Clinical Systems, 2005.
- 9 Tekkis PP, Poloniecki JD, Thompson MR, Stamatakis JD. Operative mortality in colorectal cancer: prospective national study. *BMJ* 2003;327:1196-201.
- 10 Dripps RD, Lamont A, Eckenhoff JE. The role of anesthesia in surgical mortality. *JAMA* 1963;178:261-6.
- 11 Rowe RG, Bremner W. *Hospital activity analysis. Computers in medicine series*. London: Butterworths, 1973.
- 12 McKee M, Dixon J, Chenet L. Making routine data adequate to support clinical audit. *BMJ* 1994;309:1246-7.
- 13 Audit Commission. *Information and data quality in the NHS. Key messages from three years of independent review*. 2004. www.audit-commission.gov.uk/reports/NATIONAL-REPORT.asp?CategoryID=8&ProdID=4D598AF6-3894-401d-AA48-1076125DA38D&SectionID=toctoc#
- 14 Hansell A, Bottle A, Shurlock L, Aylin P. Accessing and using hospital activity data. *J Pub Health Med* 2001;21:51-6.
- 15 Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004;57:1288-94.
- 16 Tekkis PP, Poloniecki JD, Thompson MR, Stamatakis JD. *ACPGBI colorectal cancer study 2002. Part B: risk adjusted outcomes*. London: Association of Coloproctology of Great Britain and Ireland, 2003. www.acpgbi.org.uk/pdfs/ACP_report_part_B.pdf
- 17 Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley, 1989.
- 18 Cook RD. Detection of influential observations in linear regression. *Technometrics* 1977;19:15-8.
- 19 Geraci JM, Johnson ML, Gordon HS, Petersen NL, Shroyer AL, Grover FL, et al. Mortality after cardiac bypass surgery: prediction from administrative versus clinical data. *Med Care* 2005;43:149-58.
- 20 Gordon HS, Johnson ML, Wray NP, Petersen NJ, Henderson WG, Khuri SF, et al. Mortality after noncardiac surgery. Prediction from administrative versus clinical data. *Med Care* 2005;43:159-67.
- 21 Sutton R, Bann S, Brooks M, Sarin S. The surgical risk scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *Br J Surg* 2002;89:763-8.
- 22 Jenkins J, Gauvrau K, Newburger J, Spray T, Moller J, Lezzoni L. Consensus-based method for risk adjustment for surgery for congenital heart disease. *J Thorac Cardiovasc Surg* 2002;123:110-8.
- 23 Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997;16:965-80.
- 24 Aylin P, Lees T, Baker S, Prytherch D, Ashley S. Descriptive study comparing routine hospital administrative data with the Vascular Society of Great Britain and Ireland's national vascular database. *Eur J Vasc Endovasc Surg* 2007;33:461-5.
- 25 Paediatric Cardiac Surgery Steering Group. *Report of the paediatric cardiac surgery steering group*. Oxford: Thames Valley Strategic Health Authority, Oxford Radcliffe Hospitals NHS Trust, 2005. www.tvsha.nhs.uk/papers/PaedCardiacReport.pdf.
- 26 Dixon J. Payment by results—new financial flows in the NHS. *BMJ* 2004;328:969-70.
- 27 McKee M. Routine data: a resource for clinical audit? *Qual Health Care* 1993;2:104-11.
- 28 Carter GM, Newhouse JP, Relles DA. How much change in the case mix index is DRG creep? *J Health Econ* 1990;9:411-28.
- 29 Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. *Health Technol Assess* 2005;9:1-106.
- 30 Healthcare Commission. *Heart surgery in Great Britain*. London: Healthcare Commission and Society for Cardiothoracic Surgery in Great Britain and Ireland, 2006. <http://heartsurgery.healthcarecommission.org.uk/>
- 31 Keogh B. Surgery for congenital heart conditions in Oxford. *BMJ* 2005;330:319-32.

Accepted: 23 February 2007