12 Hamilton W, Peters TJ, Round A, Sharp D. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax* 2005;60b:1059-65.
13 Crosland A, Jones R. Rectal bleeding: prevalence and consultation behaviour. *BMJ* 1995;311:486-8.
14 Chapuis PH, Goulston KJ, Dent OF, Tait AD. Predictive value of rectal bleeding in screening for rectal and sigmoid polyps. *BMJ* 1985;290:1546-8.
15 Nakama H, Kayano T, Katsuura T, Kamigaito T, Shimada S, Nishikawa N, et al. Comparison of predictive value for colorectal cancer in subjects with and without rectal bleeding. *Hepato-Gastroenterology* 1999;46:1730-2.
16 Ferraris R, Senore C, Fracchia M, Sciallero S, Bonelli L, Atkin WS, et al. Predictive value of rectal bleeding for distal colonic neoplastic lesions in a screened population. *Eur J Cancer* 2004;40:245-52.
17 Hamilton W, Sharp D. Diagnosis of colorectal cancer in primary care: the evidence base for guidelines. *Fam Pract* 2004;21:99-106.
18 Hamilton W, Round A, Sharp D, Peters TJ. Clinical features of colorectal cancer before diagnosis: a population-based case-control study. *Br J Cancer* 2005;93:399-405.
19 Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991;302:766-8.
20 Bruyninckx R, Buntinx F, Aertgeerts B, Van Casteren V. The diagnostic value of macroscopic haematuria for the diagnosis of urological cancer in general practice. *Br J Gen Pract* 2003;53:31-5.
21 Summerton N, Mann S, Rigby AS, Ashley J, Palmer S, Hetherington JW. Patients with new onset haematuria: assessing the discriminant value of clinical information in relation to urological malignancies. *Br J Gen Pract* 2002;52:284-9.
22 Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R, Global Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol* 2006;101:1900-20.

Accepted: 20 March 2007

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

Paul Aylin,[1] Alex Bottle,[1] Azeem Majeed[2]

EDITORIAL by Mohammed

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

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

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

BMJ 2007;334:1044-7
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 AAA, 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 (such as www.healthcarecommission.org.uk,www.drfoster.co.uk, www.ihi.org/IHI/Programs/Campaign/).[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.

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 hospital episode statistics (HES) administrative data, with published models based on data from the clinical databases and assessed the ability of each model to predict death. Details of the clinical datasets and HES data are on bmj.com.

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

This article is an abridged version of a paper that was posted on bmj.com on 23 April 2007. Cite this version as: *BMJ* 23 April 2007, doi: 10.1136/bmj.39168.496366.55 (abridged text, in print: *BMJ* 2007;334:1044-7)

(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). 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. We used secondary diagnosis fields to create comorbidity variables used to make up the Charlson index.[2] Further factors considered specific to each index procedure group were also considered (box). The two variables we used that were not adjusted for in the models from the clinical databases were financial year and fifth of 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 the box.

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.[3 4] 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.[5] 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. 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.[6] We also checked for influential data points via their Cook's statistic, which have values greater than 1[7] and standardised residuals.

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

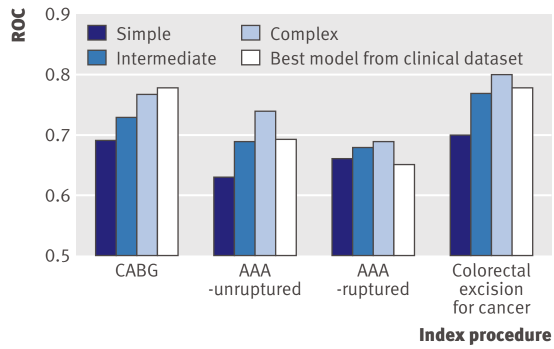
Variables used in complex models for predicting risk of death in hospital for isolated coronary artery bypass graft (CABG), repair of abdominal aortic aneurysm (AAA), and colorectal excision procedures for cancer

Variables common to all four index procedures
- Age in years (divided into five year bands to ≥85, but with those aged <45 combined)
- Sex
- Method of admission (emergency, elective)
- Year of admission (treated as continuous variable)
- Fifth of deprivation
- Charlson comorbidity score (capped at 6, treated as continuous variable)
- Number of previous emergency admissions (treated as continuous variable)

Additional variables for AAA
- Part of aorta repaired (unspecified, endovascular, thoracic, infrarenal, suprarenal)
- Previous abdominal surgery (yes, no)

Additional variables for CABG
- Previous heart operation (one or more, none)
- Previous coronary artery bypass graft (yes (that is, revision procedure), no (that is, first time CABG)
- Number of previous admissions for ischaemic heart disease (treated as continuous variable)
- Recent admission for myocardial infarction (≥1, 0)
- Number of arteries replaced (4, 3, 2, 1, unknown or other specified, other bypass of coronary artery, connection of thoracic artery to coronary artery)

Additional variables for colorectal cancer excision
- Part of colon/rectum removed (left, right, transverse, sigmoid, segment of colon not stated, anteroposterior resection, anterior resection, whole colon and/or rectum, Hartmann's, unspecified)
- Previous abdominal surgery (yes, no)



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)

2004-5), and 7.4% for colorectal resections for cancer (1999-2000).

Tables showing the odds ratios for all the variables for each index procedure are available on bmj.com. 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 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). The figure 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).

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, are on bmj.com.

DISCUSSION

We used HES data to build statistical models with good discrimination for predicting postoperative death as an outcome and 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).[8] 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).[9]

Goodness of fit

When we used the method developed by Hosmer and Lemeshow[6] 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.[10] 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, although quality has improved in recent years.[11] 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.[3] Within the Association of Coloproctology database, 39% of patients had missing

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

data for the risk factors included in their final model.[5] A comparison of numbers of vascular procedures recorded within HES and the national vascular database found four times as many cases recorded within HES.[12]

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, which will allow longer term outcomes to be monitored.

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.[13]

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. 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.[14] 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.

Contributors: See bmj.com.

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 Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, GhaliWA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004;57:1288-94.
3 Society of Cardiothoracic Surgeons of Great Britain and Ireland. *Fifth national adult cardiac surgical database report 2003*. Henley-upon-Thames: Dendrite Clinical Systems, 2004.
4 Vascular Society of Great Britain and Ireland. *Fourth national vascular database report 2004*. Henley-upon-Thames: Dendrite Clinical Systems, 2005.
5 Tekkis PP, Poloniecki JD, Thompson MR, Stamatakis JD. Operative mortality in colorectal cancer: prospective national study. *BMJ* 2003;327:1196-201.
6 Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley, 1989.
7 Cook RD. Detection of influential observations in linear regression. *Technometrics* 1977;19:15-8.
8 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.
9 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.
10 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.
11 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=&ProdID=4D598AF6-3894-401d-AA48-1076125DA38D&SectionID=toc#
12 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.
13 Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. *Health Technol Assess* 2005;9:1-106.
14 Keogh B. Surgery for congenital heart conditions in Oxford. *BMJ* 2005;330:319-32.

Accepted: 23 February 2007