Can valid and practical risk-prediction or casemix adjustment models, including adjustment for comorbidity, be generated from English hospital administrative data (Hospital Episode Statistics)? A national observational study

Alex Bottle,[1]* Rene Gaudoin,[1] Rosalind Goudie,[1] Simon Jones[2] and Paul Aylin[1]

[1]Dr Foster Unit at Imperial, Department of Primary Care and Public Health, Imperial College London, London, UK
[2]Department of Health Care Management and Policy, University of Surrey, Surrey, UK

*Corresponding author

Declared competing interests of authors: The Dr Foster Unit at Imperial College London is principally funded by Dr Foster Intelligence, an independent health-care information company. The Dr Foster Unit is affiliated with the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre. Dr Alex Bottle and Professor Paul Aylin are part funded by a grant to the unit from Dr Foster Intelligence outside the submitted work.

Published November 2014
DOI: 10.3310/hsdr02400

Scientific summary

Risk-prediction or casemix adjustment models from HES data

Health Services and Delivery Research 2014; Vol. 2: No. 40
DOI: 10.3310/hsdr02400

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

England's NHS has a wealth of administrative hospital data covering inpatient and outpatient activity that are increasingly being used to measure and monitor hospital performance and explore variations in outcomes. These databases are complex and imperfect, but their low cost, national coverage and richness make them appealing to researchers and regulators alike – if their utility and limitations can be assessed and appreciated. Some of the value of Hospital Episode Statistics (HES) has been demonstrated in two public inquiries, covering Bristol Royal Infirmary and Mid Staffordshire NHS Trust, by revealing those hospitals' high mortality, but a more thorough assessment of their usefulness in the area of risk adjustment and risk prediction, including for outcomes other than mortality, is warranted.

Objectives

Our main objectives were:

1. to derive robust casemix adjustment models for these outcomes adjusting for available covariates using HES
2. to update the weights and codes for the widely used Charlson index of comorbidity, recalibrate it for the NHS and assess its use for mortality and also non-mortality outcomes
3. to assess if more sophisticated statistical methods based on machine learning such as artificial neural networks (ANNs) outperform traditional logistic regression (LR) for risk prediction
4. to assess the usefulness of outpatient data for these models.

The first of these may be considered an overarching aim, with the other three as elements of that aim.

Methods

We assessed the quality of HES records first by considering published evidence from Audit Commission coding inspection reports and a previously published systematic review. Secondly, we determined the coverage by hospital for accident and emergency (A&E) records, comparing against the online Accident and Emergency Quarterly Monitoring Data Set counts, and the completeness of key fields for all records.

We defined a number of outcome measures: in-hospital and total 30-day mortality, unplanned readmission within 30 days of live discharge, unplanned return to theatre (RTT) within 90 days of the index operation, outpatient department (OPD) non-attendance for first appointment and for first post-hospitalisation appointment, and other unplanned readmission measures such as subsequent bed-days. For OPD non-attendance, we used HES to track later hospital activity in our patient cohorts. RTT was defined by taking the set of procedures dated between 1 and 29 (or 89) days of the index procedure and consulting with senior clinicians and a coding expert to identify those procedures that were not considered a planned second phase to the index procedure. We included a number of patient groups, including those admitted for heart failure (HF), acute myocardial infarction (AMI), and colorectal or orthopaedic surgery.

To determine the best way to adjust for comorbidity, we undertook a systematic review of studies comparing two or more comorbidity indices or approaches. Other variables we included depended on the outcome and the purpose, but age, sex, method of admission, year of discharge and areal deprivation quintile were always included.

We built risk-adjustment models using LR for a range of procedure and diagnosis groups. For AMI and colorectal surgery, we then compared several methods for deriving risk-adjustment models: LR (with and without adjusting for the clustering of patients within hospital), ANNs, support vector machines and random forests. Patient groups were split into training and testing portions, and standard measures of discrimination and calibration were calculated for each method, patient group and data portion. Hospital-level relative risks were derived for mortality and readmission by summing patient-specific predicted probabilities and actual outcomes and calculating the ratio of the latter sum to the former sum. The numbers of funnel plot outliers at 95% and 99.8% control limits were counted.

For a cohort of patients admitted for HF we undertook additional analyses. For unplanned readmission within 30 days, we divided these by primary diagnosis (HF vs. non-HF) and correlated the resulting hospital-level rates with published performance measures from the 2011 national HF audit. We tried several ways of incorporating their previous hospital contacts (OPD and inpatient activity) and considered several measures of future contacts beyond just the first unplanned readmission within 30 days. We aimed to predict the number of bed-days and the number of unplanned readmissions within a year of discharge, split by those for HF and those for any other primary diagnosis. We also allocated patients with non-zero activity to one of five 'buckets' of equal total activity and aimed to predict membership of the highest-activity bucket.

Results

Our assessment of HES data quality led us to include inpatient, day case and OPD, but not A&E records, in the analysis. Most OPD records still lack diagnostic information.

Our systematic review of studies comparing comorbidity indices included 54 studies. The commonest outcome was mortality, which we divided into short term (30 days or fewer) and long term (more than 30 days). This led us to choose a combination of the Elixhauser set and dementia from Charlson, with weights to be determined using our own data rather than any published set. HES-based weights for the Charlson index revealed that human immunodeficiency virus (HIV) status is no longer a significant predictor but that dementia merits a much higher weight now than in Charlson's original formulation.

Logistic regression models for mortality and readmission were often poorly calibrated, with overprediction of low risk and underprediction of high risk commonly responsible, but discrimination for many of the mortality models was high. Overfitting was common with random forests, and results from the machine learning methods were little better than from LR. Discrimination (*c*-statistic) was often good for mortality but moderate or modest for other outcomes and lowest for readmission.

The 90-day RTT rate was 2.1% for hip replacement and 1.8% for knee replacement. These are comparable to 3-year revision rates but require only 90 days' follow-up and offer a useful additional measure. Patient factors explained little of the variation by surgeon or hospital for either index procedure. Hierarchical modelling showed that the majority of a surgeon's RTT rate is explained by factors other than patient factors or the hospital at which they operate.

The literature review identified many reasons for patients missing their OPD appointment. While many of the foregoing factors such as personal circumstances are not available in HES, several key ones are. We found young and very old age, male gender, area-level deprivation and prior non-attendance to be key predictors. The time interval between inpatient discharge and the first subsequent appointment showed a weak relation for our set of acute diagnosis groups combined, with a small reduction in the non-attendance rate for between 3 and 6 weeks compared with 12 or more weeks. This effect was not seen for most of the individual diagnosis groups. Discrimination was moderate ($c = 0.67$), and there was significant overprediction of low risk. Patients who did not attend their first post-discharge appointment had more emergency admissions, total inpatient bed-days and further non-attendances in the subsequent year than

those who did; however, they had fewer elective admissions and total OPD appointments. The rates were different but the patterns the same following a first general medical or general surgical appointment after GP referral. Given that patients who did not attend differed in various ways such as age and gender from those who did, we ran LR with death in the year after the index appointment as the outcome, adjusting for the factors in the tables plus the fact of non-attendance as an extra predictor. These models suggested that non-attendance was associated with about a 50% higher odds of death, only slightly reduced from the unadjusted figure.

Finally, we focused on readmissions to patients admitted for HF. Predictors were similar across all follow-up periods for all-cause readmissions with the exception of same-day discharge, which was important for 7-day readmission, but they sometimes differed by cause of readmission. Thirty-day rates for HF ranged from 1.5% to 9.0% (median 5.4%), while 30-day rates for non-HF ranged from 7.6% to 17.6% (median 13.6%). Rates showed negative but modest correlations with publicly available quality of care measures from the National HF Audit. It was notable that rates for HF did not appreciably correlate with rates for non-HF and that the associations between readmission rates and process measure performance existed only for readmissions for HF.

Of our various count models that we employed to predict inpatient activity in the year following index discharge, the negative binomial hurdle and zero-inflated models performed best, though convergence was sometimes hard to achieve. These models showed that a number of comorbidities were associated with higher odds of readmission but were much less important in predicting future bed-days.

Conclusions

Robust casemix adjustment models for a range of outcomes adjusting for available covariates can be derived using LR with HES data, though recalibration will often be necessary, for instance by using an extra step in the regression. Mortality models had much higher discrimination than readmission-type outcomes, with OPD non-attendance between the two. The OPD records add useful information when predicting readmission-type outcomes.

The Charlson comorbidities needed new weights for use with HES, with HIV less important and dementia more important than in the original study. For general-purpose comorbidity adjustment, we recommend using the Elixhauser set plus dementia with extra *International Classification of Diseases* codes besides those originally given.

The machine learning methods that we tried offered fairly little above LR with these outcomes and data, and are much less straightforward to implement.

Recommendations for future research

The A&E portion of HES improved in coverage from its first two years to 2009/10 onwards, but gaps and frequent lack of diagnostic information still limit its utility. Analyses that exclude hospitals with poor data are suggested. The A&E records could be used both in risk-adjustment models and also for outcome measures.

We have outlined the process for producing RTT metrics, which involves empirical analysis and expert clinical and coding input, and this could be replicated for other index procedures. Elucidation of the relations between RTT and outcomes such as mortality and quality of life would help determine its value as an indicator of quality of care.

For readmissions, we considered the first and also the total number, but more sophisticated approaches such as multistate analysis or cluster analysis to look for patterns of activity could be usefully employed.

Funding

The National Institute for Health Research Health Services and Delivery Research programme.

Health Services and Delivery Research

ISSN 2050-4349 (Print)

ISSN 2050-4357 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HS&DR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hsdr. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Services and Delivery Research* journal

Reports are published in *Health Services and Delivery Research* (HS&DR) if (1) they have resulted from work for the HS&DR programme or programmes which preceded the HS&DR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

HS&DR programme

The Health Services and Delivery Research (HS&DR) programme, part of the National Institute for Health Research (NIHR), was established to fund a broad range of research. It combines the strengths and contributions of two previous NIHR research programmes: the Health Services Research (HSR) programme and the Service Delivery and Organisation (SDO) programme, which were merged in January 2012.

The HS&DR programme aims to produce rigorous and relevant evidence on the quality, access and organisation of health services including costs and outcomes, as well as research on implementation. The programme will enhance the strategic focus on research that matters to the NHS and is keen to support ambitious evaluative research to improve health services.

For more information about the HS&DR programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hsdr

This report

The research reported in this issue of the journal was funded by the HS&DR programme or one of its proceeding programmes as project number 09/2001/32. The contractual start date was in September 2010. The final report began editorial review in March 2014 and was accepted for publication in August 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).