NIHR Health Services Research programme

NHS National Institute for Health Research

HSR Protocol – project ref: 09/2001/32
Version: 01
Date: 11/10/10

Can valid and practical risk-prediction or casemix adjustment models, including adjustment for co-morbidity, be generated from English hospital administrative data (Hospital Episode Statistics)?

Chief investigator: Dr Robert Alex Bottle

Sponsor: Imperial College London

Funder: HSR Programme

NIHR Portfolio number:

ISRCTN registration (if applicable):

Can valid and practical risk-prediction or casemix adjustment models, including adjustment for co-morbidity, be generated from English hospital administrative data (Hospital Episode Statistics)?

090/2001/32 – Bottle protocol version: 01
25 November 2010

Aims and objectives

Comparison of healthcare performance is an essential component of driving quality improvement. It is imperative to appropriately adjust for casemix when making such comparisons. Statistical models to predict risk form the bases of such adjustments. The aim of this project is to use a core NHS dataset, Hospital Episode Statistics (HES), to develop risk prediction and casemix adjustment models to predict and compare between different healthcare units mortality, unplanned readmission, unplanned returns to theatre for selected specialties and non-attendance in outpatients departments.

Using HES data, the project has these objectives:

1. To derive robust casemix adjustment models for mortality (in and out of hospital, the latter via existing linkage with ONS death registrations) and the other outcomes listed above adjusting for available covariates
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 whether more sophisticated statistical methods based on machine learning such as artificial neural networks outperform traditional multilevel logistic regression for risk prediction
4. To assess the usefulness of outpatient data (part of HES since 2003/4) for the above models

Background

<u>The use of hospital patient outcomes and relevance to NHS policy</u>

The NHS collects a wealth of data that are currently underutilised to support improvements in health care provision. Potentially these data can be used from an individual level to provide information through to organisational level to drive improvements. At an individual patient level, predicting the risk of an adverse event, such as post-operative mortality, is important. Knowledge of a patient's pre-operative risk aids clinical decision making and informs clinical discussion with the patient, allowing for a more informed dialogue. At an organisational level, adjusting for differences in patient risk (also known as case-mix) between different units such as surgeons or hospitals is crucial in making valid and fair comparisons of a unit's performance, for example in 'provider profiling' and pay-for-performance schemes. These data are also vital in quality improvement efforts, placing "quality at the heart of everything we do" with a focus on measurement as per the 2008 Darzi report, High Quality Care for All, the final report of the NHS Next Stage Review. At present, few patient outcomes are available by hospital on the NHS Choices website (the site's risk adjustment is based on our methodology). This could be improved and extended to cover more outcomes to drive quality improvement. Such public reporting is increasing internationally and administrative data such as those held by the NHS will be in greater demand for this purpose – provided that they can be shown to be of adequate quality and timeliness and that inter-unit differences in case-mix can be accounted for sufficiently (given that no dataset can capture all case-mix factors).

<u>Use of routine data for modelling patient outcomes</u>

Indicators of quality of care need data for their calculation and case-mix adjustment. The NHS has often been described as "data rich, information poor" and benefits from a number of core datasets. Although maintained for primarily administrative purposes, particularly with HES, they are well

established as a source of information on patient outcomes. Administrative data have some key attractions including national coverage, availability and low cost and, if their limitations are taken into account, can provide useful knowledge [1]. Their use particularly in mortality models is common in many countries such as the USA, Canada, Australia, Japan as well as the UK [2] [3] [4] [5]. We have shown with three common procedure groups that models based on HES can predict mortality at least as well as models based on clinical databases [6]. As well as mortality, these data can be used for other outcome measures of interest to patients, clinicians and managers, and this project will focus on several key ones listed under the objectives. Specification of these other outcomes is less straightforward than for death and requires the input from people with different backgrounds.

Measuring comorbidity in administrative databases

Patient comorbidity is a common potential confounder in health services research and can be derived from administrative datasets. The two most commonly used indices are the Charlson and Elixhauser, originally described using ICD9 with US data. In each index, points are given for the presence of a set of codes representing diseases associated with higher or sometimes lower than average risk. The points are then summed to give a score for the admission. The Charlson index is now over twenty years old and both indices need calibrating on the dataset of interest; to our knowledge, very little has been published from the UK on this. The weights (or scores) for these two indices may be inappropriate for the UK due to differing populations and coding practices. We have to date been using an Australian version of Charlson [7] in our risk-adjustment models, but discussions with clinical coders raised questions over the suitability of some codes when used in the UK. Recent work in Canada [8] found that a modified Elixhauser [9] outperformed Charlson, but this appears to be because it includes some acute conditions that could represent complications rather than comorbidities. The authors also rightly concluded that it needed external validation, such as in another country. Some comorbidities were associated with lower risk of mortality and were given negative scores. One reason for this was given by Elixhauser: low-risk patients may be given more codes for less-acute problems compared with acutely ill people for whom coders will focus on problems relevant to the acute situation. The presence of codes for non-life-threatening disease may therefore be a marker for relatively healthy patients and this warrant an assessment.

These comorbidity scores ignore any interaction between components. For example, a patient with both congestive heart failure and diabetes without long-term complications will have a total Charlson of 2, 1 from each condition, but may have a higher risk of death than a score of 2 would indicate. In contrast, the additional presence of codes for non-threatening disease mentioned above may reduce the risk but by different degrees depending on what codes they accompany. These ‘interactions’ between diagnosis codes may show a non-linear relation with the outcomes; these are easier to spot using machine learning methods. In short, the effect on risk of combinations of codes is unpredictable and we hope to describe and quantify this problem.

Other groups have tried linking together admissions belonging to the same patient to help identify comorbidities [10]. This can overcome the problem of variations in coding between hospitals if, for example, a patient with diabetes is admitted to two hospitals but only one records this condition. A related approach would be to inspect the primary diagnosis field in previous admissions for acute events such as stroke and AMI that form part of the Charlson index. These facts can be included in the risk-prediction model and their effects evaluated.

Incorporating previous health service contacts into the risk model

This idea of tracking back into the full electronic record can be extended to incorporate operations that the patient has had in the past. This will be particularly important when predicting the outcome of a

revision procedure such as a second CABG, as revisions are known to be more complicated than first-time procedures, but may also be useful for risk models for medical patients. For example, the most predictive factor for future unplanned admissions is the number of previous unplanned admissions [11] [12].

We currently adjust our mortality and readmission models [13] for the number of previous unplanned admissions that a patient had in the 12 months prior to their index admission. This is a relatively simple but crude way of trying to account for the influence of factors relating to disease severity and admission thresholds. It could be extended to include outpatient appointments and A&E attendances, which are now part of HES. It is not clear, however, how this should best be done, how important it is and to which outcomes and patient groups it relates, and we aim to resolve this in this project.

Definition of outcome measures

Mortality has long been used as an indicator of quality of care and is of special interest to patients. HES includes in-hospital deaths, and in recent years includes (with a time lag) linkage with ONS death registrations so that total mortality rates, such as 30-day post-operative rates, may be calculated. In England, around 40% of deaths occur in hospital. As the time since admission increases, the proportion of deaths occurring in the community increases and the more the in-hospital and total mortality rates diverge. 30-day post-operative death rates are commonly used in surgery and are typically based on only the in-hospital portion; the linkage with ONS will enable an assessment of what type of patients die in the community but within 30 days of the procedure and help with the validity the use of in-hospital death rates. This linkage also enables longer-term mortality to be modelled such as within a year as we have done in studies of laparoscopy in GI surgery [14].

Unplanned readmission to hospital is a commonly used indicator of quality of care and is also of interest to hospital managers. Readmission rates have been rising for some years in the UK. It is often due to the disease processes and a relation with comorbidity has been shown [15]. We have found that confounder control (as measured by discrimination using the c statistic) using HES is generally poorer for readmission than for mortality. It is hoped that better control for comorbidity and for the patient's health status, by analysing the combination of secondary diagnosis codes in the index and also in previous admissions, may improve the prediction. When used as a measure of hospital performance, readmission rates often use a 28-day timeframe, though some use "early readmissions" such as within 7 or 14 days instead as being more specific to quality of care issues [16]. However, when used for case management or resource need as in frequently admitted patients sometimes known as high-impact users, a longer time frame such as a year is used [11] [12]. A year may be too long, particularly in high-mortality groups, and we will therefore look at the reasons for subsequent unplanned admissions (from the primary diagnosis) within 3 and 6 months of discharge. We may thereby be able to provide a more realistic approach for health and social care interventions to reduce readmissions. As subsequent admissions may be longer than the initial one, we will also record the number of bed days utilised in each and in total.

There is considerable interest amongst surgeons in unplanned returns to theatre (RTT). HES captures the dates of all procedures carried out (we have found these to have high levels of completeness) and it is therefore possible to define RTT within a given time frame by inspecting the dates and the OPCS procedure codes within HES, providing that codes that represent planned stages of the operation are excluded.

Non-attendance for appointments in hospital outpatients departments (OPD) is a major burden on healthcare systems and costs the NHS an estimated £790 million per year [17]. It also increases waiting

times with poorer outcomes for non-attenders and a loss of continuity of care [18]. Non-attendance in the OPD is readily identified in the dataset. As diagnosis and procedure codes are as yet sparse in these records, the analysis will be stratified by specialty. A number of patient and hospital factors related to non-attendance have been identified, such as age, sex, deprivation, specialty and also communication and administration problems [19] [20]. Using HES we will be able not just to build models predicting non-attendance but also follow up non-attenders over time to study some of their outcomes, including if they have higher rates of unplanned admissions or other hospital contacts. For follow-up appointments, the timing may be important. When a patient has been hospitalised, they may be given an OPD appointment for a few weeks after discharge; the length of time since discharge may influence their likelihood of attendance. HES data may be used to derive an empirically optimal length of time between discharge and appointment, thereby reducing non-attendances.

Choice of statistical methods

There is a huge literature on risk models in a range of specialties. Prediction of binary outcomes such as mortality or readmission is usually done via logistic regression with records from single institutions, and is therefore prone to problems such as poor reproducibility due to small sample sizes and variations in patient characteristics between study centres [21]. Combining a larger number of institutions would increase the sample size, but the modeller then in principle needs to account for the "clustering" of patients within institutions (using multilevel models for example). As an alternative to regression, researchers have recently been applying machine learning methods (especially artificial neural networks, ANNs), and the initial results have been promising [22] [23] [24]. A Bayesian approach circumvents the traditional problem with ANNs of over-complex models that arise from over-fitting the dataset. An exciting recent development in machine learning is the discovery and elaboration of Support Vector Machines (SVMs) [25] [26] for classification problems such as predicting mortality. SVMs differ from neural networks in a number of ways, and offer some advantages over them. Song et al [27] used Canadian administrative data to predict death, comparing logistic regression, various neural network models and least squares SVMs (although only linear kernels were used). They found that both ANNs and SVMs performed slightly better in terms of discrimination than logistic regression. We believe that these more advanced methods warrant further consideration. They are not well served by existing software if very large datasets are used, such as with HES, and some time will be required to write code in the free downloadable package R (http://www.r-project.org/) that is potentially ideal for this purpose. The essential question here is: do these sophisticated methods offer substantial benefits over the humble logistic regression?

Need

The need for this project falls into several categories. By constructing models that account for case-mix, hospitals can be compared more fairly in terms of their patient outcomes. This assists the NHS in decision-making to help bring about improvement. By using a combination of advanced statistical methods, coding expertise and clinical input, we will build on existing knowledge in the area of risk prediction and case-mix adjustment. Stratification of the output, such as by ethnicity or small-area deprivation score, enables the assessment of health inequalities. The specification of metrics beyond mortality, in particular emergency readmission (here not just defined using the common 28-day time lag) and unplanned returns to the operation theatre, will make our findings more generalisable to the many specialties and patient groups for which mortality is not the most useful indicator.

In our models of OPD non-attendance, we aim to determine the time interval between discharge as an inpatient and OPD follow-up to minimise non-attendance, which is relevant to the organisation and

delivery of healthcare. Comparing subsequent outcomes of non-attenders with patients who do attend will add to the body of knowledge in this area.

The whole project will make advanced use of routinely collected NHS data and increase our understanding of their strengths and limitations, which will be of value to researchers and the NHS management community alike.

Methods

In this section we will first describe the databases that we hold and plan to use, the general modelling framework and the groups of patients (defined by diagnosis groups, procedure groups and specialty). Then we consider the definition of each outcome measure in more detail together with the research questions and analysis plan for that measure. Lastly we will give the outputs by each year of the project.

Databases

We hold (with approval under Section 251 (formerly Section 60) granted by the Patient Information Advisory Group) HES data from 1996/7 to 2005/6. Applications for 2006/7 and 2007/8 are pending, but we already hold equivalent data from the Secondary Uses Service (SUS) from 2006/7 to October 2009 and obtain a full extract from SUS each month. We therefore hold some 13 years of inpatient and day case records with pseudonymised patient identifiers for linkage. We also hold OPD records since 2003/4, again updated monthly from SUS. HES records for 2000/1 to 2005/6 have been linked with ONS death registrations via the HES patient ID ("HESID"), giving us the date of death (up to 2006) for all hospitalised patients during that period. We have applied for the cause of death field to be added to the file containing the HESID and date of registered death, and if we receive this we will be able to include this information in some of the mortality analyses.

We anticipate receiving the pending HES data in 2010 which, for 2007/8, will include for the first time attendances in A&E department; a HES report from their website indicates that coverage is incomplete, but for trusts with good data (such as those whose HES-based counts match their Quarterly Monitoring of Accident and Emergency (QMAE) returns) this represents potentially useful extra information.

We have linked via the patient's full postcode the small-area Carstairs deprivation score and calculated its population-weighted quintile, which we have attached to every record.

Modelling framework

The terms risk prediction and risk adjustment are closely related despite their differing aims but a model for predicting mortality, for example, might not include the same set of variables as a risk-adjustment models used to compare hospitals' mortality rates. The former needs to be interpretable, whereas the latter may include a number of two- or even three-way interaction terms in the interests of confounder control. Risk-prediction models in health services research could encompass factors such as staffing and bed numbers or other factors that are (at least partly) under the hospital's control, whereas this would be wrong for risk-adjustment models for comparing providers. A notable example of the difference concerns elective paediatric cardiac surgery at the Bristol Royal Infirmary that gave rise to a public inquiry [28]. The surgeons at Bristol delayed the procedure in these infants until nearly one year of age because they had had poor outcomes at younger ages, when all the other specialist units were operating on their patients in accordance with the guidelines. Adjustment for age – with elective surgery such as this, the age at operation was under the control of the hospital – would have obscured at least some of Bristol's excess risk. This is an unusual case of age being excluded from a risk-adjustment model and demonstrates the care that needs to be taken. In this proposal, we will aim to

derive practical models that are suitable for both purposes, allowing extra complexity, for example with interactions between variables, only when it is clearly superior for risk-adjustment models.

In the USA the healthcare market is more used to competition than in the UK. Nationwide public reporting of hospital outcomes is generally done using administrative data and requires case-mix adjustment to minimise the risk of misclassifying "well-" and "poor-" performing hospitals. Krumholz et al [29] set out to define the preferred attributes of case-mix-adjustment models and to develop criteria against which different models can be compared. We will take their framework as a basis and adapt it to suit HES and non-cardiovascular conditions, extending it where necessary by incorporating standard measures of model performance [30] [31]. We will also be influenced by the National Quality Board and the NHS Information Centre's development of an indicator library and also hope to contribute to it.

We will construct each model in the following steps:

1. For each outcome, choose the set of patients (see "patient groups" section below)
2. Build logistic regression model, ignoring the clustering of patients within hospitals
3. Adjust for clustering using multilevel modelling
4. Apply machine learning methods (ANN and SVM)
5. Compare the sets of models, predicted risks and hospital-level adjusted outcome rates
6. Arrive at final recommended models

Logistic regression, including multilevel modelling, will be carried out using the package SAS (v9.2). Machine learning methods will be implemented in R. We will develop the models on earlier data years (the "training" dataset) and test them using later data years (the "validation" dataset). This internal validation is common practice and is particularly important for machine learning methods, in which "over-training" is a well-recognised problem.

An issue with large datasets such as administrative ones is the retention of unimportant variables whose apparent relation with the outcome measure is in fact spurious. For each model we will assess this using bootstrapping to determine how often each variable is retained [32].

Our application of ANNs in this project falls within the framework of Bayesian learning techniques. The overall aims of classical (backpropagation) learning in ANNs, which are to provide an optimal solution to the data fitting problem that simultaneously generalises well, are properly supported by the application of Bayesian statistics. Such ideas are well established in the literature [33] [34] [35] and have been shown to yield powerful methods for controlling, comparing and using adaptive neural networks. The search in model space can then be treated as an inference problem, in which we infer the relative probability of alternative models given the data. The development of ANNs along these lines can yield simple yet powerful models [36] but the plethora of competing approaches that currently exist in the research literature warrant comparison. What distinguishes ANN models developed in the Bayesian framework from the traditional backpropagation-trained ANNs is that the Bayesian approach does not seek an output classifying function utilising a minimum of the sum of squares cost function. Instead, the output function of a Bayesian ANN is an average over all models weighted by the evidence for the model. In this way we will obtain error bars and confidence intervals for the predictions of the network.
We will assume Gaussian priors for given network models and compute the posterior distribution of weights. The usual Bayesian integrals in this setting are best estimated by the hybrid Monte Carlo approach [37] allied with simulated annealing [38]. These techniques will provide us with our Bayesian ANN classifier for the risk of mortality and other binary outcomes. The evidence used to weight the

Bayesian integrals will be obtained from the project datasets used in the development of the competing models.

Apart from ANNs, we will be using state-of-art classifiers such as Support Vector Machines (SVM). There are three practical issues that will be considered in our application: kernel selection, efficient parameter selection and scalability to large data sets. The primary challenges in applying SVM methods to a given domain lies in the selection of the kernel and its parameters as well as the magnitude of the soft margin, which allows some level of training data misclassification. Once a performance measure has been selected, the parameter selection problem can be formulated as an optimization problem and we plan on using genetic algorithms to determine the optimal tuning parameters. We will use standard model performance indices such as the area under the receiver operating characteristic (ROC) curve. There are also computational issues involved with training SVMs on very large data sets such as ours. SVMs are generally solved by quadratic programming algorithms where the number of variables equals the number of training data, which can result in very slow training. In order to accelerate training in very large data sets we will decompose the global optimization problem defined by the SVM into more manageable sub-problems. Finally, an extension of the single-SVM approach consists in fitting several SVMs and then combining the individual models by optimally weighting their importance. This approach, referred to as "learning with experts" or "classifier combination" [39 ref Kuncheva], has often been proved to increase the out-of-sample model performance. The underlying idea is to build local models that specialize in given sub-domains of the entire data space, and then combine the models to determine the optimal prediction.

Sets of models will be compared in two principal ways. The first will use standard measures of model performance such as the area under the ROC curve, or c statistic, which measures discrimination (the ability of a model to give a higher, though not necessarily accurate, risk of death to a person who died than to a person who survived), and calibration plots from the Hosmer-Lemeshow test.
The second is not a test of which model is "better" but of its effect when used for benchmarking and comparing providers. In the feedback loop of outcome monitoring, such as by regulatory bodies but also by the hospitals themselves for internal quality improvement efforts, hospitals identified as outliers in this way would be investigated to determine the cause(s) of their high outcome rate. An important aim of the modelling is to reduce the number of "false positives" – hospitals who only appear to be outliers because of insufficient case-mix adjustment. Trust-level risk-adjusted outcome rates will be put onto funnel plots. The proportion of trusts whose rates lie outside the 95% and 99.8% control limits in each model will be noted, together with how many trusts change "outlier" status from one model to the next. This has been used in comparisons of models based on administrative and clinical databases [40] [41], with the latter being used as the gold standard despite neither database being perfect. We have no gold standard available because clinical databases are often beset by problems of coverage and completeness [42] [43], but we can compare some of our new models against those that performed well against published models derived from UK clinical databases [6].

Patient groups

For mortality and unplanned readmission, this will be by primary diagnosis (we have grouped together all the ICD10 codes into the Agency for Healthcare Research and Quality's 259 Clinical Classification System groups) or by set of OPCS procedure codes, whereas for OPD non-attendance we will choose a small number of specialties. The set of patients for each outcome is given in Table 1.

Table 1. Set of patient groups by outcome measure

Outcome	Diagnoses	Procedures	Specialties
Mortality	Top 20 causes of in-hospital death, including acute myocardial infarction, stroke, COPD	Commonest procedures with non-negligible risk of death, including CABG, AAA repair, colorectal resection, oesophagectomy	n/a
Unplanned readmission i) within 7 and 28 days ii) for readmissions within 3 and 6 months	As per mortality COPD and heart failure	As per mortality n/a	n/a n/a
Unplanned returns to theatre	n/a	As per mortality plus primary hip replacement, primary knee replacement	
OPD non-attendance i) referrals from GP ii) following discharge after an emergency admission	n/a n/a	n/a n/a	general medicine, general surgery general medicine

Outcome measures and analysis plan for each measure

Mortality

Mortality objectives:

1. Update weights and ICD codes for the Charlson index of comorbidity, calibrated for the NHS
2. Assess and quantify potential coding biases in which patients with a lower risk of in-hospital death have various "minor" ICD codes recorded
3. Improve risk prediction models for 30-day mortality for a range of common diagnosis and procedure groups
4. Compare trust-level risk-adjusted rates for in-hospital and total 30-day mortality

The first task is to update the Charlson index specification for use in the NHS. This has two parts – choice of ICD10 codes and derivation of up-to-date weights – but needs to be considered in the wider context of assessing the relation between groups of codes (and groups of groups) and risk of death. Having updated Charlson in mortality models, we will try the new definition in models of the other outcomes below.

We will begin by asking expert clinical coders to review the set of ICD10 codes used in the Sundararajan version of the Charlson index and suggest modifications based on UK guidelines and practice. We will also ask them about the large number of "minor" codes (in particular the Z chapter) to tell us in what circumstances they would perhaps not be recorded, for example if the maximum number of diagnosis fields had been reached (some HES records have entries for all 14 fields, and SUS now allows 25 fields), or they would be more likely to be recorded, such as in patients with less acute conditions attending hospitals with keen coders.

Next will come the empirical analysis, beginning with descriptive statistics for the proposed Charlson index. This will include comparison with the existing version and the use of backtracking into patients' previous admissions to look at consistency of recording. We will construct logistic regression models using dummy variables for the amended Charlson index components to derive their weights for 2007/8 and then 2008/9 in the standard way [44]. We will then do subanalyses to look at the effect of the presence of the "minor" codes on these weights for each component; this will need to be done for sets of patients with the same primary diagnosis such as AMI and COPD and/or for the same main procedure such as colectomy, before trying to do this for all patients combined. This effect of the presence of the "minor" codes will be evaluated in terms of difference in the crude mortality rates and in the empirical weights of the Charlson index. It may be possible to find one or more groups of these codes that are associated with a lower mortality rate and could therefore be added to the Charlson index definition to indicate low-risk patients. Machine learning methods are particularly adept at looking for such hidden relations between variables.

We will conduct sensitivity analyses on the effect of coding depth on the Charlson weights and on the hospital-level risk-adjusted mortality rates. One simple way to do this is to use just the first N secondary diagnosis codes to calculate Charlson, where N ranges from 1 to 13 (or 12 if, for example, an external cause ICD code accompanies the primary diagnosis) [45]. It may be possible to account for variations in coding depth at, say, an institutional level by incorporating an "inflation factor" for low-coding hospitals, akin to imputing missing values. This would need information on the amount of missing comorbidity coding at each institution, for example from annual coding audits. There are two sources for this. The Audit Commission has carried out two reports on coding quality (available on their website). Trusts are also required to undertake internal clinical coding audits and make them available under Freedom of Information legislation: some do this by putting them on their website. They also have to comply with the Audit Commission's Data Assurance framework.

It is common to consider all deaths in hospital for diagnosis groups and deaths within 30 days of the operation for procedure groups, though 30 days may be used as the cut-off for diagnoses too (e.g. on the public Hospital Compare website run by Medicare). We routinely link admissions ending in transfer to another hospital to capture post-transfer (but still in-hospital) deaths, but it is not always possible to identify the transfer part. Even small differences in transfer rates may affect the adjust mortality rates [46]. Many hospital administrative databases do not capture deaths following discharge, but, with the file containing HES records linked to ONS death registrations that we have, we will be able to model total mortality. We will therefore compare at a trust level case-mix adjusted death rates calculated with and without these out-of-hospital deaths, noting which kind of patients die out of hospital but within 30 days of admission or the operation and which trusts have particularly high proportions of such patients. As well as risk prediction using the updated Charlson index derived above, the main aim here is to determine under what circumstances in-hospital death rates may be used as proxies for total mortality rates. There are relatively few such comparisons in the literature [47] [48] [49] with studies tending to focus simply on the two sets of standardised mortality ratios rather than patient characteristics.

Unplanned readmissions

Readmissions objectives:

1. Improve risk prediction models for 7-day and 28-day unplanned readmissions for a range of common diagnosis and procedure groups
2. Compare trust-level risk-adjusted rates for 7-day and 28-day unplanned readmission
3. Compare patients with COPD or heart failure (as indicator conditions) readmitted within 3 months with those readmitted within 6 months; construct risk prediction models for each

We will firstly model those within 28 days of discharge, which is the most common definition when being used as an indicator of quality of care, but will compare the results with "early" readmissions (those within 7 days) in terms of the types of patients readmitted "early" (within 7 days) and "late" (between 7 and 28 days) and trust-level adjusted rates. To obtain sufficient numbers at trust level for a given patient group, it will probably be necessary to combine several years of admissions. We will repeat the process for redefining the Charlson index as described under "mortality" above. This will inform whether the weights are still appropriate or whether they are markedly different for non-mortality outcomes. The 7-day cut-off is considered to be more specific to potential quality of care issues than the 28-day cut-off. One might therefore expect the relations between comorbidity and the two sets of readmission rates to differ, with comorbidity being more strongly related to the 28-day version.

We will secondly compare the characteristics of patients readmitted within 3 months with those readmitted between 3 and 6 months. For this analysis we will include primary diagnoses of firstly COPD and secondly heart failure and also possibly related symptom codes such as dyspnoea and chest pain, providing that at least one of these admissions including the index is for COPD or heart failure. We have found that such symptom codes are common in general but also in the admission histories of patients with chronic conditions. We will take advice on these from expert clinical coders. Regarding comorbidity, we will use the findings from the above analysis of 28-day readmissions in the predictive models in this section.

Unplanned returns to theatre
Return-to-theatre objectives:

1. Define return-to-theatre metrics for some common procedures
2. Construct risk prediction models for return-to-theatre metrics for some common procedures

These will be calculated within 28 days of the index procedure. With HES, there is unfortunately no flag to indicate that the return was unplanned, so an inspection of the OPCS codes is required. To this end we will build on the literature and our own experience such as in orthopaedics [50] and colorectal surgery. Specifically, in colorectal surgery we have already classified re-interventions occurring in the perioperative period following major elective colonic and rectal resections for cancer as well as restorative proctocolectomy for ulcerative colitis. For each procedure, we will obtain from the data a list of OPCS codes with dates of between 1 and 27 days following the index procedure. It will not generally be possible to include same-day returns because of the shortage of codes specific to reinterventions, although we will investigate this. Using a combination of clinical and coding knowledge, including that from the relevant literature, we will use those codes that represent unplanned returns in our metrics prior to deriving risk models for them using the usual approach. Regarding comorbidity, we will use the findings from the mortality analyses to help fit the models. It may be necessary to define a new set of Charlson weights for this set of outcomes.

OPD non-attendance
Non-attendance objectives:

1. Construct risk-prediction models for non-attendance at medical and surgical OPD departments following GP referral
2. Assess the non-attendance rate by lag between discharge following emergency hospitalisation and follow-up OPD appointment to derive an "optimal" lag time in terms of maximal attendance

3. Quantify future hospital contacts in patients who miss OPD appointments compared with those who keep them

We will begin by assessing the data quality of OPD records in terms of completeness of key fields such as age, sex and specialty and compare them with inpatient records for the same period. Linkage between inpatient and outpatient records will be tested for some common elective operations that would be expected to be followed up in the OPD, matching records using all available identifiers (NHS number, HESID for HES years, date of birth, sex and postcode and hospital number).

Non-attendance is defined using the "ATTENDED" field. Anyone can forget an appointment once (a common reason for non-attendance), but some people fail to attend more often and a number of factors are associated with this and are included in HES. We will examine two groups of patients. The first group will be those referred from their GP; we will divide these into medical and surgical by specialty. The second group will be those admitted as an emergency and given an OPD appointment for within two months of discharge; we will limit this to medical patients. Of interest in the second group will be the time between discharge and the appointment. The hypothesis is that patients asked to come to clinic six or eight weeks after discharge may be less likely to attend than those asked to come sooner because their illness may have resolved and they may therefore not see the need for the appointment. We will take 2007/8 as the index year for each analysis and track back to count attended and missed appointments in any of the previous four years.
We will then follow-up attenders and non-attenders in HES for a year to capture and compare their future inpatient, day case and outpatient contacts. We will estimate the cost associated with this subsequent activity using HRG tariffs. We will look at overall activity and also try to identify activity related to their index OPD appointment where possible; for example, for a patient with an appointment following an admission for COPD, we will identify future contacts with general medicine and respiratory medicine and/or later admissions for COPD and related symptom codes such as dyspnoea.

Constructing risk models for OPD non-attendance including linkage with past and previous episodes in HES will provide valuable practical experience on the quality and potential of this relatively new set of electronic records. The introduction of OPD tariffs has not yet led to a sizeable increase in the use of diagnosis and procedure codes in OPD HES, but if this improves during the course of this project, then we will take advantage of the extra information.

Contribution to collective research effort and research utilisation

The project will deliver the following outputs. See the "plan of investigation and timetable" for more detailed timings.

Year 1
Updated weights and ICD codes for the Charlson index of comorbidity calibrated for the NHS; this may need to be specific to patient group (output O1)
Assessment and quantification of potential coding biases in which patients who are less ill have various "minor" ICD codes recorded (output O2)
Improved risk prediction models for 30-day mortality for a range of common diagnosis and procedure groups (output O3)
Definitions of return-to-theatre metrics for some common procedures (output O4)

Year 2
Assessment of whether current machine learning methods can identify any hidden relationships between sets of comorbidities in mortality models (output O5)
Comparison of trust-level risk-adjusted rates for in-hospital and total 30-day mortality (output O6)
Risk prediction models for return-to-theatre metrics for some common procedures (output O7)
Improved risk prediction models for 7-day and 28-day unplanned readmissions for a range of common diagnosis and procedure groups (output O8)
Comparison of trust-level risk-adjusted rates for 7-day and 28-day unplanned readmission (output O9)
Assessment of whether current machine learning methods can improve the models for unplanned readmission (output O10)
Comparison of patients with COPD or heart failure readmitted within 3 months with those readmitted within 6 months; risk prediction models for each (output O11)

Year 3
Risk prediction models for non-attendance at medical and surgical OPD departments following GP referral (output O12)
Assessment of the non-attendance rate by lag between discharge following emergency hospitalisation and follow-up OPD appointment (output O13)
Quantification of future hospital contacts in patients who miss OPD appointments compared with those who keep them (output O14)
Assessment of how much current machine learning methods offer in terms of risk prediction over traditional logistic regression (output O15)
Final report with recommendations on outcome measure specifications and level of case-mix adjustment required (output O16)

We will aim to disseminate these findings initially via several peer-reviewed publications in general medical and surgical journals, with presentations at appropriate national and international conferences. Secondly, we will present our methods and findings to the Information Centre to determine whether they can incorporate the metrics on the NHS Choices website and include them in their indicator library for use by NHS management. Locally, we would envisage the metrics being used by Imperial College NHS Trust.

Plan of investigation and timetable

Table 2. Output timetable

Output/Task	Brief description	Year	Month
O0	Recruit RA	1	1-2
O1	Update Charlson index	1	1-6
O2	Assess coding biases	1	7-12
O3	Fit mortality models	1	7-12
O4	Define RTT metrics	1	1-2
O5	Apply machine learning methods to mortality	2	1-6
O6	Compare trust-level mortality	2	7
O7	Fit RTT models	2	1-3
O8	Fit readmission models	2	4-6
O9	Compare trust-level readmission	2	8-9
O10	Apply machine learning methods to readmission	2	10-12
O11	Fit COPD, HF readmission models	2	10-12

O12	Fit OPD DNA models for medical and surgical appts	3	1-3
O13	Fit OPD DNA model for post-discharge appts	3	4-5
O14	Compare hospital contacts for OPD DNAs and attenders	3	6-7
O15	Summarise value of machine learning methods	3	1-9
O16	Write final report	3	10-12

Approval by Ethics Committees

We hold sufficient HES data for this project and have approval under Section 251 (formerly Section 60) granted by the Patient Information Advisory Group (now the NIGB). We have had approval for using these data for research from St Mary's local ethics committee since 2002 and are in the process of updating this approval for our Unit's work (which would cover this project and our other activities). We expect to obtain this local approval in early 2010.

Project management

The project manager will be Dr Paul Aylin, clinical reader and assistant director of the Dr Foster Unit (DFU). The Project Management Group will consist of the co-applicants and the researcher and will discuss the analysis plan, the results, the draft papers and the study progress principally by e-mail. All the applicants are based in London and can meet in person or by teleconference as necessary in addition. The researcher will be based at the DFU and will meet weekly with Drs Aylin and Bottle; the latter will lead the analysis and write-up.

Service users/public involvement

This project consists of the exact (re)specification of some existing outcome measures and the development of a number of statistical models. The measures are known to be of interest to patients, clinicians and managers. We therefore do not plan any involvement with patients or the public during the project. However, if we discover during the project that the data quality and the models are robust enough, we envision the inclusion of the measures and models on the NHS Choices website, whose intention is to help with the engagement of patients in decision-making regarding their care.

Expertise and justification of support required

This project requires a combination of statistical, clinical and coding expertise as well as experience in processing and analysing these complex data. The applicants listed below cover all these criteria in addition to project management. For detailed coding expertise we will engage a senior coding consultant. If additional clinical expertise is required, we will be able to draw upon our natural and well-established links with Imperial College NHS Trust for advice.

Alex Bottle (statistician) has used the principal dataset for this project (HES data) for over ten years to construct logistic regression models and for other health services research. His work on risk prediction and statistical process control underpins a monitoring tool (for mortality and other outcomes) provided by Dr Foster Intelligence (who fund the Dr Foster Unit on a research grant basis), currently used by around 75% of NHS hospital trusts in England. He will lead the analysis and assist with the processing of the data and writing SAS programs. He will also lead the recruitment of the full-time project researcher and the report and paper writing.

Paul Aylin (clinical reader in public health and epidemiology) has great experience in the analysis and presentation of results from routine data sources including HES and in leading research groups such as the Dr Foster Unit. He will manage the project and assist with the recruitment of the full-time project researcher. He and the other investigators will also contribute to the methodological approach of the research.

Giovanni Montana (mathematician/statistician) has been developing and applying machine learning methods to problems arising in bioinformatics and other domains for several years. He has developed software packages in R and other higher level languages. His main contribution to the project will be in training and applying support vector machines.

Simon Jones (mathematician/statistician) has tremendous programming expertise, particularly in R and including writing code to implement neural networks. He is experienced in using the HES dataset. His main contribution to the project will be in training and applying artificial neural networks.

Andy Grieve (professor of statistics) has a wealth of research experience and, as a well-known Bayesian expert, will advise on the Bayesian perspective on artificial neural network implementation.

Omar Faiz (consultant colorectal surgeon) has experience in using HES data to look at patient outcomes and inter-hospital variations in surgical approach. As a practising surgeon used to data analysis, his main contribution to the project will be in providing clinical advice, particularly for the specification and interpretation of return-to-theatre metrics.

Derek Bell (professor of acute medicine) has academic research interests relating to the quality and organisation of care, particularly acute medical care, and in the methods of delivery of care. As a practising physician with particular expertise in chest medicine, his main contribution to the project will be in providing clinical advice, particularly regarding the analysis and interpretation of unplanned readmissions and non-attendance in OPD.

In addition to resources to cover the applicants´ time, we request funding for a full-time post-doctoral researcher in order to implement the range of statistical techniques on the range of patient outcomes detailed in this document in the three-year lifetime of the project.

We would aim to present at major national and international medical and statistics conferences such as those hosted by the International Society for Quality in health care (ISQua), the International Forum on Quality and Safety in Healthcare, the Operational Research Society and the International Conference on Artificial Neural Networks.

Planned or active related research grants

Drs Aylin and Bottle are co-Is on the NIHR-funded Centre for Patient Safety and Service Quality (CPSSQ) at Imperial. Dr Aylin leads a work programme at the CPSSQ on the use of health information that includes work on firstly defining patient safety indicators using routine data and secondly assessing how the indicators are used within the Trust and the barriers to using them for quality improvement. The proposed project therefore fits very well with this work programme.

Prof Bell is director of the NW London Collaboration for Leadership in Applied Health Research and Care (CLAHRC) funded by NIHR. The aim of CLAHRC is to continuously improve the quality of

patient care by accelerating the implementation of evidence based research and innovations into practice. Subprojects include optimising the discharge processes for COPD and case management for chronic disease. The proposed project's analysis of readmissions (with particular focus on COPD and heart failure) and OPD non-attendance is very relevant to CLAHRC's goals.

History of past or existing NIHR programme research

We do not hold and have never held an NIHR programme contract which has been terminated, or extended in time or funding.

References

[1] Hansell A; Bottle A; Shurlock L; Aylin P. Accessing and using hospital activity data. J Public Health Med 2001; 23: 51-56.
[2] Krumholz HM, Wang Y, Chen J, Drye EE, Spertus JA, Ross JS, Curtis JP, Nallamothu BK, Lichtman JH, Havranek EP, Masoudi FA, Radford MJ, Han LF, Rapp MT, Straube BM, Normand SL. Reduction in acute myocardial infarction mortality in the United States: risk-standardized mortality rates from 1995-2006. JAMA 2009; 302(7): 767-73.
[3] Pilote L, Tu JV, Humphries K, Behouli H, Belisle P, Austin PC, Joseph L. Socioeconomic status, access to health care, and outcomes after acute myocardial infarction in Canada's universal health care system. Med Care 2007 Jul; 45(7): 638-46.
[4] Scott IA, Thomson PL, Narasimhan S. Comparing risk-prediction methods using administrative or clinical data in assessing excess in-hospital mortality in patients with acute myocardial infarction. Med J Aust 2008 Mar 17; 188(6): 332-6
[5] Miyata H, Hashimoto H, Horiguchi H, Matsuda S, Motomura N, Takamoto S. Performance of in-hospital mortality prediction models for acute hospitalization: hospital standardized mortality ratio in Japan. BMC Health Serv Res 2008 Nov 7; 8: 229.
[6] Aylin P; Bottle A; Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. BMJ 2007; 334: 1044
[7] 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(12): 1288-94.
[8] van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009; 47(6): 626-33.
[9] Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005; 43(11): 1130-9.
[10] Zhang JX, Iwashyna TJ, Christakis NA. The performance of different lookback periods and sources of information for Charlson comorbidity adjustment in Medicare claims. Med Care 1999; 37(11): 1128-39.
[11] Bottle A; Aylin P; Majeed A. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. J Roy Soc Med 2006; 99: 406-414.
[12] Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. BMJ 2006; 333: 327.
[13] Bottle A, Aylin P. Intelligent Information: a national system for monitoring clinical performance. Health Services Research 2008; 43:10-31
[14] Faiz O; Clark J; Brown T; Bottle A; Antoniou A; Farrands P; Darzi A; Aylin P. Traditional and laparoscopic appendectomy in adults: outcomes in English NHS hospitals between 1996 and 2006. Ann Surg 2008; 248: 800-806.
[15] Leng GC, Walsh D, Fowkes FG, Swainson CP. Is the emergency readmission rate a valid outcome indicator? Qual Health Care 1999; 8(4): 234-8.
[16] François P, Bertrand D, Beden C, Fauconnier J, Olive F. Early readmission as an indicator of hospital quality of care. Rev Epidemiol Sante Publique 2001; 49(2): 183-92.
[17] Atun AR SSR Mohan A.: Uses and Benefits of SMS in Healthcare Delivery. Centre for Health Management. Tanaka Business School. Imperial College London. 2005.
[18] Karter AJ, Parker MM, Moffet HH, Ahmed AT, Ferrara A, Liu JY, Selby JV: Missed appointments and poor glycemic control: an opportunity to identify high-risk diabetic patients. Med Care 2004; 42(2): 110-115.

[19] Collins J, Santamaria N, Clayton L. Why outpatients fail to attend their scheduled appointments: a prospective comparison of differences between attenders and non-attenders. Aust Health Rev 2003; 26(1): 52-63.
[20] Hamilton W, Round A, Sharp D. Patient, hospital, and general practitioner characteristics associated with non-attendance: a cohort study. Br J Gen Pract 2002; 52(477): 317-9.
[21] Kolh P. Importance of risk stratification models in cardiac surgery. Eur Heart Journal 2006; 27(7): 768-769.
[22] Green M, Björk J, Forberg J, Ekelund U, Edenbrandt L, Ohlsson M. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room, Artificial Intelligence in Medicine 2006; 38: 305-318.
[23] Bottaci L, Drew PJ. Artificial Neural Networks Applied to Outcome Prediction for Colorectal Cancer Patients in Separate Institutions. Lancet 1997; 350(9076): 469-473.
[24] Nilsson J, Ohlsson M, Thulin L, Höglund P, Nashef SA, Brandt J. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks, J Thoracic Cardiovascular Surgery 2006; 132: 12-19.
[25] Schölkopf B, Smola AJ. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2001.
[26] Shawe-Taylor J, Cristianini N. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000.
[27] Song X, Mitniski A, Cox J, Rockwood K. Comparison of machine learning techniques with classical statistical methods in predicting health outcomes. MEDINFO 2004; 736-740.
[28] Aylin P, Alves B, Best N, Cook A, Elliott P, Evans SJ, Lawrence AE, Murray GD, Pollock J, Spiegelhalter D. Comparison of UK paediatric cardiac surgical performance by analysis of routinely collected data 1984-96: was Bristol an outlier? Lancet 2001; 358(9277): 181-7.
[29] Krumholz HM, Brindis RG, Brush JE, Cohen DJ, Epstein AJ, Furie K, Howard G, Peterson ED, Rathore SS, Smith SC, Spertus JA, Wang Y, Normand S-L T. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation 2006; 113: 456-462.
[30] Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ 2009; 338: b604.
[31] Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009; 338: b605.
[32] Austin PC, Tu JV. Bootstrap Methods for Developing Predictive Models. The American Statistician 2004; 58(2): 131-137.
[33] MacKay DJC. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, Cambridge 2003.
[34] Neal, RM. Bayesian Learning for Neural Networks. PhD Thesis, Dept of Computer Science, University of Toronto 1994.
[35] MacKay DJC. A practical Bayesian framework for backpropagation networks. Neural Computation 1992; 4: 448-472.
[36] MacKay DJC. Bayesian non-linear modelling for the prediction competition. ASHRAE Transactions 1994; 100(2): 1053-1062.
[37] Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. Physics Letters B 1987; 195: 216-222.
[38] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimisation by simulated annealing. Science 1983; 20 (4598): 671-680.
[39] Kuncheva L. Combining Pattern Classifiers: Methods and Algorithms. Wiley, 2004.

[40] Geraci JM, Johnson ML, Gordon HS, Petersen NJ, Shroyer AL, Grover FL, Wray NP. Mortality After Cardiac Bypass Surgery: Prediction From Administrative Versus Clinical Data. Med Care 2005; 43(2): 149-158.
[41] Gordon HS, Johnson ML, Wray NP, Petersen NJ, Henderson WG, Khuri SF, Geraci JM. Mortality After Noncardiac Surgery: Prediction From Administrative Versus Clinical Data. Med Care 2005; 43: 159-167.
[42] Garout M; Tilney HS; Tekkis PP; Aylin P. Comparison of administrative data with the Association of Coloproctology of Great Britain and Ireland (ACPGBI) colorectal cancer database. Int J Colorectal Dis 2008; 23(2): 155-163.
[43] 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: 461-465.
[44] Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. Stat Med 2004; 23: 1631–1660.
[45] Martins M, Blais R. Evaluation of comorbidity indices for inpatient mortality prediction models. J Clin Epidemiol 2006; 59: 665–669.
[46] Kahn JM, Kramer AA, Rubenfeld GD. Transferring critically ill patients out of hospital improves the standardized mortality ratio. Chest 2007; 131: 68-75.
[47] Vasilevskis EE, Kuzniewicz MW, Dean ML, Clay T, Vittinghoff E, Rennie DJ, Dudley RA. Relationship Between Discharge Practices and Intensive Care Unit In-Hospital Mortality Performance. Evidence of a Discharge Bias. Med Care 2009; 47: 803–812.
[48] Slobbe LCJ, Arah OA, de Bruin A, Westert GP. Mortality in Dutch hospitals: Trends in time, place and cause of death after admission for myocardial infarction and stroke. An observational study. BMC Health Services Research 2008; 8: 52
[49] Rosenthal GE, Baker DW, Norris DG, Way LE, Harper DL, Snow RJ. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. Health Serv Res 2000; 34(7): 1449-68.
[50] Simon S Jameson, Alex Bottle, Ajay Malviya, Scott D Muller, Mike R Reed. The impact of national venous thromboembolic prophylaxis guidelines on lower limb arthroplasty complications. Br J Bone Joint Surg 2010 (in press).

Disclaimer

This report presents independent research commissioned by the National Institute for Health Research (NIHR). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the SDO programme or the Department of Health.