Global Comparators Project: International Comparison of Hospital Outcomes Using Administrative Data

Alex Bottle, Steven Middleton, Cor J. Kalkman, Edward H. Livingston, and Paul Aylin

Objective. To produce comparable risk-adjusted outcome rates for an international sample of hospitals in a collaborative project to share outcomes and learning.

Data Sources. Administrative data varying in scope, format, and coding systems were pooled from each participating hospital for the years 2005–2010.

Study Design. Following reconciliation of the different coding systems in the various countries, in-hospital mortality, unplanned readmission within 30 days, and "prolonged" hospital stay (>75th percentile) were risk-adjusted via logistic regression. A web-based interface was created to facilitate outcomes analysis for individual medical centers and enable peer comparisons. Small groups of clinicians are now exploring the potential reasons for variations in outcomes in their specialty.

Principal Findings. There were 6,737,211 inpatient records, including 214,622 in-hospital deaths. Although diagnostic coding depth varied appreciably by country, comorbidity weights were broadly comparable. U.S. hospitals generally had the lowest mortality rates, shortest stays, and highest readmission rates.

Conclusions. Intercountry differences in outcomes may result from differences in the quality of care or in practice patterns driven by socio-economic factors. Carefully managed administrative data can be an effective resource for initiating dialog between hospitals within and across countries. Inclusion of important outcomes beyond hospital discharge would increase the value of these analyses.

Key Words. Administrative data, hospitals, quality of care

There has been substantial discussion on differing quality of health care between countries with different approaches to financing that care. However, there are few direct comparisons of clinical outcomes at hospital level between the various health care systems, although many countries have regional or national indicator projects (Groene, Skau, and Frølich 2008). The first step in determining the optimal approach for health care delivery from among these systems is the creation of an international database of clinical outcomes.

This task is not trivial. Administrative data are generally the only feasible resource for this but use a number of coding schemes that require reconciliation. Although the recording of hospital discharge information may be standardized within countries, it is not so across international borders. Differences in diagnostic coding systems create challenges as the relative importance placed on the accuracy of discharge coding will vary between health systems.

We established a mechanism for collecting discharge data from hospitals in the United Kingdom, Europe, and the United States. We then reconciled the differing coding systems and entered the harmonized data into statistical risk-adjustment models. Risk-adjusted outcomes of care were then compared using a purpose-built web interface.

Comparing outcomes across international boundaries has a number of challenges—including logistical, IT-related, and cultural—as the project partners get to know each other, share data, decide on the clinical areas of common interest, and exchange information on processes of care and other relevant issues. In this article, we describe the data and modeling challenges and our approach to tackling them.

METHODS

Our guiding principles were pragmatism, comparability, reproducibility, and transparency. Our main tasks were as follows:

- Compilation of a database and integration of records from five countries in a larger number of differing formats
- Selection of key fields and definitions of variables
- Definition of an inpatient
- Selection of outcome measures
- Classification of diagnoses and procedures into meaningful groups across the different coding systems

Address correspondence to Alex Bottle, Ph.D., Senior Lecturer in Medical Statistics, Dr. Foster Unit at Imperial College London, Department of Primary Care and Public Health, School of Public Health, The Reynolds Building, St Dunstan's Road, London W6 8RP, UK; e-mail: robert.bottle@imperial.ac.uk. Steven Middleton, B.Sc., is with the Dr. Foster Intelligence, London, UK. Cor J. Kalkman, M.D., Ph.D., is with the Division of Anesthesiology, Intensive Care and Emergency Medicine, University Medical Center Utrecht, Utrecht, GA, The Netherlands. Edward H. Livingston, M.D., F.A.C.S., A.G.A.F., is with the JAMA, Chicago, IL. Paul Aylin, M.B.Ch.B., is also with the Dr. Foster Unit at Imperial College London, Department of Primary Care and Public Health, School of Public Health, London, UK.

- Adjustment for comorbidity
- Production of risk-adjustment models
- Extraction of model performance metadata
- Presentation of results

Although some of these issues may appear straightforward, the details can be extremely complex. We now consider each one and describe sensitivity analyses on the effect of the main choices. Alternative approaches are presented in the Discussion.

Data and Key Variables

We obtained electronic inpatient records for 2005–2009 directly from each participating hospital outside of England. For the English hospitals, we used Hospital Episode Statistics (or the equivalent Secondary Uses Service data warehouse), which cover all admissions to NHS (public) hospitals. Late in 2011, we obtained data for 2010 and updated all the models; we report figures for all 6 years in this article. The participants comprised 10 U.S., 10 English, 8 Dutch, 1 Belgian, and 1 Italian hospital.

A number of important variables contained differing levels of detail depending on the country; the lowest level of detail determined that used in the risk modeling. Table 1 shows the mapping of the original to the final version. All the original values were retained and displayed to the user on the web-based tool in their own language.

Selection of Outcome Measures

Three widely used outcome measures were defined: in-hospital mortality, unplanned readmission within 30 days of discharge, and prolonged LOS. Mortality was defined as any in-hospital death for analyses based on diagnoses and in-hospital death within 30 days of the procedure for surgical procedures. Readmissions were defined as unplanned (emergency) admissions to an inpatient unit at the same hospital. Admissions to an observation or assessment ward were not included. A 30-day window was chosen as it seems the most commonly used, although intercountry differences in all windows between 1 and 29 days were found to be consistent.

Prolonged LOS was defined as an LOS greater than the 75th percentile for the diagnosis or procedure group, year, and method of admission based on the entire dataset, as used in our England-wide monitoring system (Bottle and Aylin 2008).

Table 1:   Variables and Mappings of Fields Used in the Risk-Adjustment Models

Variable	Original Values	Value Used in Risk Models	Comments
Method of admission	Elective; planned admission; following day care; voluntary hospitalization; prehospital treatment; nonurgent	Planned	Interaction terms with transfer in also included
	Emergency; urgent; semi-urgent; newborn; other maternity event; unknown; trauma center; obligatory hospitalization	Unplanned	
Source of admission/ transfer in	From usual residence; from temporary residence; within same hospital; outside hospital Emergency Room transfer; born in hospital	Not transferred	Interaction terms with method of admission also included
	Transferred from acute hospital	Transferred from acute hospital	
	Transferred from: skilled nursing facility; intermediate care facility; correctional facility; private health care institution	Transferred from nonacute hospital	
Age group			<1, 1–4 then 5-year bands up to 90+ Excluded if unknown or invalid
Sex			Excluded if unknown or invalid
Urgent admission in previous month			Yes/no. Excludes the current admission
Diagnosis or procedure subgroup			Not present for all groups
Charlson comorbidity score	Derived from set of 0/1 comorbidity flags in secondary diagnoses	Integers from 0 upwards	Interaction terms with age also included. Different weightings for each outcome and group. See text.
Year			Calendar year of discharge

Definition of Inpatient Admission

The project's focus is on inpatient care. We therefore excluded outpatient (ambulatory) care, day-case surgery, and observation ward activity. Some hospitals removed these records prior to sending their data to the data center in London. Others did not, but the greater problem was that in England, there is no flag to identify observation ward or assessment unit patients. These are all unplanned admissions, and patients stay from a few hours to sometimes a few days. As retaining these records would have unduly inflated the denominators, we assessed the impact of overall and hospital-level mortality rates and relative risks from different methods of excluding the observation patients. We reasoned that these patients are low risk, unplanned, and should have a short LOS, often overnight, but probably no longer; we decided upon a maximum LOS of one night (i.e., date of discharge minus date of admission ≤ 1). Some U.S. hospitals transfer high-risk patients within 24 hours to another hospital, so we amended our criteria to retain these admissions; transfers to nonacute settings remained excluded.

For countries other than England, each record represented one admission; transfers out to other hospitals could not be identified. In England, however, each record represents a finished consultant episode, which covers the unbroken period of time during which a patient is under the care of a single consultant or allied health care professional. An admission may therefore comprise more than one episode (around 15 percent of episodes belong to multiepisode admissions). As we wish to avoid multiple counting, we needed to summarize the information across the episodes for the same admission. We allocated the diagnosis group to the first episode unless its primary diagnosis belongs to the ICD chapter on signs and symptoms (ICD10 R chapter), in which case, we use the second episode if present. Admissions ending in transfer may be linked to the posttransfer episode(s) to form what we term a "superspell." For a fairer comparison between countries, we chose to regard any posttransfer portions of English superspells occurring in other participant hospitals as separate admissions. As a result, we "lost" any in-hospital deaths occurring after transfer. This will have particular impact on AMI, for example, in which patients in some areas of the United Kingdom and the Netherlands may be sent from the Accident and Emergency department (emergency room) of one hospital to that of another for a short stay for primary angioplasty before another transfer elsewhere, often on the same day.

Classification of Diagnoses and Procedures

ICD10 was used in England, ICD9-DE in Holland, and ICD9-CM in the other countries. To group our diagnoses, we used the Agency for Healthcare Research and Quality Clinical Classifications Software (HCUP CCS 2012) as it was derived for health services research. This uses 259 clinically meaningful diagnosis groups, some of which have subgroups. All five countries have the same concept of primary diagnosis. For procedures, however, no such recognized grouping exists for any of the three coding systems used (ICD9 in the United States, Italy, and Belgium; CVV in the Netherlands; and OPCS in England). Rather than set up a cross-country working group, we took as a starting point our existing English OPCS groupings used in our national monitoring system (Bottle and Aylin 2008), selecting a sample of 38 groups representative of the major surgical specialties. These groupings have already received considerable clinical and coding input in recent years and acted as a solid base. We then searched for equivalent ICD9 codes in journal articles of studies using, for example, the U.S. Nationwide Inpatient Sample, and other sources such as the National Quality Forum (for CABG and coronary angioplasty) and the Blue Cross Blue Shield Association (for hip and knee replacement). Other ICD9 equivalents were derived from a manual search of the ICD9 online dictionary with input from surgical colleagues. Dutch CVV equivalents were derived from the ICD9 draft set and checked by a Dutch participant. The procedure code equivalencies between coding systems were all reviewed and agreed upon after consultation. These were a starting point, with the expectation that participants could refine the groupings in the next phase.

The three systems support differing levels of detail, and it was sometimes necessary to lose detail from OPCS where it was lacking in ICD9. For example, the ICD9 code for infra-inguinal bypass (39.29, "other peripheral vascular bypass grafting") also refers to upper limbs (Cronenwett et al. 2007). It does not seem possible to filter them out, but they only occur in less than 0.1 percent of cases.

Adjustment for Comorbidity

Many of the datasets had no "present on admission" information for secondary diagnoses, meaning that they cannot distinguish between comorbidities and complications developed in the hospital since admission. We therefore had to exclude this information for consistency with the other records.

Several comorbidity indices suitable for administrative data exist. We initially chose to use the Charlson set of comorbidity variables (Sundararajan et al. 2004) as it is most commonly used in risk-adjustment models. The original Charlson weights were derived from 1-year postdischarge mortality nearly 30 years ago on a relatively small cohort of medical patients (Charlson et al. 1987) and may no longer be appropriate. Our recent systematic review of studies comparing comorbidity indices found the derivation of new weights to be advantageous (Sharabiani, Aylin, and Bottle 2012). We therefore derived new sets of weights for the project dataset, one set per outcome measure. Aggregation of comorbidities into a score facilitates the inclusion of interaction terms between age and comorbidity, particularly for smaller patient groups.

Production of Risk-Adjustment Models

We developed separate logistic regression models for each of the 259 diagnosis and 32 procedure groups for each of the three outcome measures. Automation was therefore highly desirable. We used the software package SAS's inbuilt backward elimination procedure, retaining variables with $p < 0.1$. Two-way interactions, between age and Charlson and between method of admission and transfer, were included as candidates, based on a priori beliefs. To aid model convergence, age groups with fewer than 10 events were iteratively combined with the immediately older group (Jen et al. 2011).

For each model, the area under the ROC curve (c statistic) and McFadden's *r*-squared were obtained. The adequacy (Harrell 2001) of each variable was also calculated, which gives an indication of which variables explain the most variation in the outcome.

Presentation of Outcomes to Users

Dr. Foster Intelligence built a web front end in a similar manner to that used in our England-wide hospital outcomes monitoring system (Bottle and Aylin 2008). Users log in to an initial summary screen that shows a grid with one row for each patient group (diagnosis or procedure) and one column for each of the three outcomes. Cells in the grid are red, blue, or green depending on whether the user's hospital is significantly higher than, neither higher nor lower than, or lower than the benchmark, respectively, using 95 percent confidence intervals without adjustment for multiple testing. The benchmark is all hospitals combined. An example is given in Figure 1.

Figure 1:   Screenshot of Opening Page of Project Tool, Summarizing the Participant Hospital's Outcome Measures by Patient Group. (Red bells indicate significantly poorer than average and green bells significantly better than average performance on the given measure; split-color bells show that performance changed over the time period. White means not significantly different from the average)
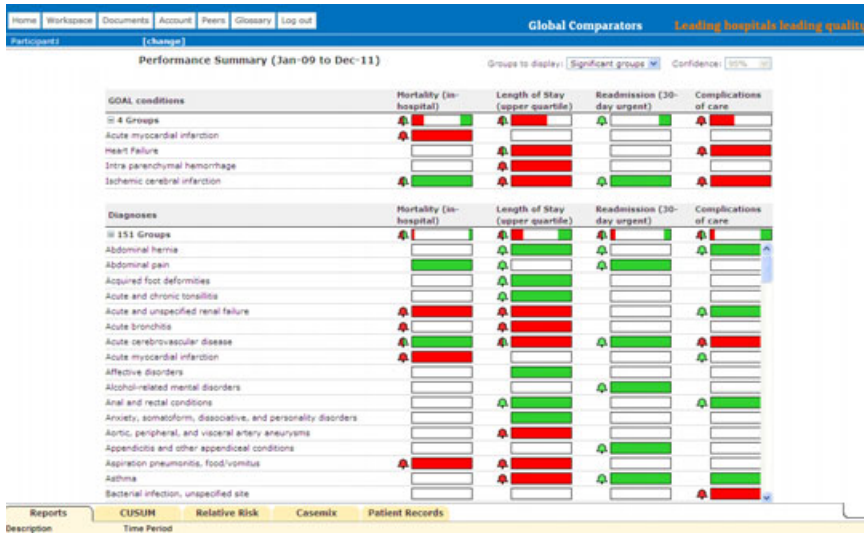

Users can click on a cell to bring up a funnel plot with 95 and 99.8 percent control limits showing the relative risks for each hospital. Rolling the mouse over a point reveals the hospital name. Figure 2 is a screenshot taken from the web tool showing relative risks for mortality as observed-to-expected ratios multiplied by 100 for AMI by hospital. Italy, Belgium, and the Netherlands have been combined to avoid identification of hospitals. Country groups are color-coded.

A large set of drill-down options exist, such as by any of the case mix variables, day of the week, length of stay band, and peer group defined in different ways. As well as prolonged LOS, LOS can be viewed as a histogram with superimposed peer means and medians. The user can also view their individual patient records for the chosen patient group. All participants can see the others' hospital-level figures, but not individual records.

GOAL Groups

The second phase of the first year consisted of the selection of four clinical areas (Global Outcomes—Accelerated Learning or "GOALs") to explore

Figure 2:   Funnel Plot for AMI Mortality for 2008–2010 from Web Tool, with Hospitals in each of the three Country Groups in Different Colors (England in dark purple, United States in blue, the rest in yellow)
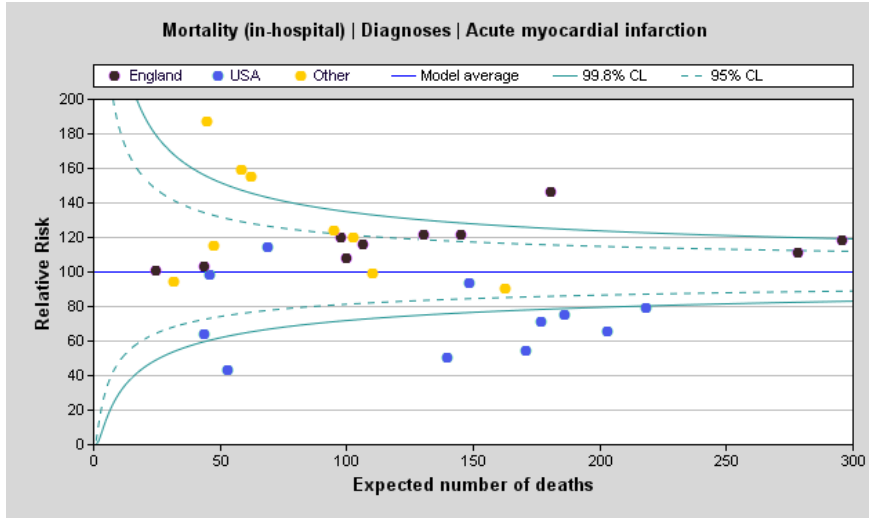
deeper with the help of clinicians from the respective specialties based on either diagnosis or procedure groups. The intention was to use the observed variation to try to identify best practices that might be implemented elsewhere. To maximize collective learning among participants, we produced a shortlist of groups that each showed both inter- and intracountry variation in outcomes. From this shortlist of eight, participants chose four at a mid-year conference in Boston in March 2011: stroke, heart failure, AMI, and colorectal surgery. The aim of the tool was to assist hospitals in understanding potential reasons for the differences and generating some questions that further data collection could answer. Outcome measures cannot tell us how hospitals vary in their practice—they serve only as a starting point for investigation.

Results

Data Quality and Inpatient Definition Analyses

The 30 hospitals submitted a combined total of 9,305,191 records for the 6 years of data 2005–2010. We performed basic data quality checks, such as counting the number of unique patient identifiers for each hospital and

tabulating the key variables, including the secondary diagnosis fields. Upon initial review, four Dutch hospitals were found to have data integrity problems (missing patient identifiers or little procedure or secondary diagnosis coding). They resubmitted their data, after which, only one hospital was excluded entirely from subsequent model building and another had its 2005–2008 records excluded. One further Dutch hospital has been excluded from the comorbidity scoring, but not the risk models, due to very low secondary diagnosis coding. The remaining hospitals had 8,982,436 records after removing any outpatient or day-case records that they submitted.

After checking basic data quality, we considered different inpatient definitions. Table 2 shows the overall death rates by country group with

Table 2:   Overall Inpatient Mortality Calculated by Different Methods as a Sensitivity Analysis to Determine the Impact of Observation Ward Patients

Country Group	All Records	Exclude Unplanned Admissions, LOS = 0	Exclude Unplanned Admissions, LOS < 2 Days	Exclude Unplanned Admissions Ending in Live Discharge, LOS < 2 Days	Exclude Unplanned Admissions Ending in Live Discharge, LOS < 2 Days, Unless Ending in Transfer
Stroke					
England	18.9	19.2	18.1	21.6	21.0
USA	11.3	11.3	9.7	13.1	13.0
NL, It, Bel	13.9	14.2	12.5	16.1	15.8
CHF					
England	14.5	14.0	13.7	15.7	15.7
USA	3.2	3.2	3.2	3.5	3.5
NL, It, Bel	8.7	8.8	8.2	9.5	9.5
AMI					
England	8.5	8.5	7.2	10.7	10.1
USA	4.7	4.7	4.2	5.7	5.7
NL, It, Bel	6.8	6.9	6.3	9.3	8.2
Colorectal excision					
England	4.8	5.2	4.9	5.3	5.3
USA	3.2	3.2	3.0	3.3	3.3
NL, It, Bel	4.7	4.7	4.5	4.8	4.8
All patients					
England	2.8	2.7	3.2	3.9	3.8
USA	1.9	1.9	1.9	2.3	2.3
NL, It, Bel	2.7	2.7	2.6	3.2	3.1

AMI, acute myocardial infarction; CHF, congestive heart failure.

different exclusion criteria. Data are shown for all diagnoses combined and for each of the four GOAL groups.

As expected, removing the short stays had greatest impact on England's mortality rate. U.S. hospitals generally had lower crude in-hospital mortality rates than the other countries, irrespective of the inpatient definition used. Excluding unplanned admissions ending in live discharge but not transfer to another acute center within 2 days led to a final total of 6,737,211 records retained, including 214,622 in-hospital deaths (3.2 percent case fatality rate).

U.S. hospitals typically had shorter LOS (median three nights, interquartile range 2–6, compared with a median of 3 and IQR 2–7 for England and for all countries combined). They had greater use of intermediate care, with 11.4 percent of admissions ending in transfer to an other health care facility (2.5 percent for England, 0.9 percent for the rest, and 5.2 percent for all participants combined) and 1.4 percent ending in transfer to another acute hospital (1.6 percent for England, 6.9 percent for the rest, and 2.7 percent for all participants combined).

Overall, U.S. hospitals had a higher 30-day readmission rate (9.4 percent compared with 6.6 percent in England and 4.9 percent in the other hospitals), a pattern repeated for the four GOAL groups. The pattern was unaffected by the inpatient definition used (figures not shown).

Comorbidity Adjustment and Risk Modeling

Risk models for all three outcomes included adjustment for the Charlson comorbidity score. The empirical Charlson weights are shown for mortality alongside the original published ones in Table 3.

The largest change in weight from the original published set was seen for HIV, the recording of which was now associated with a decreased risk. In contrast, as we have previously found using all English data (Bottle and Aylin 2011), dementia has gained in apparent importance. Weights derived from records from each country group separately were often consistent, although with some large variations, for example, paraplegia and dementia.

Model discrimination was usually best for mortality and poorest for readmission. Age was typically the most important variable, with comorbidity also featuring particularly for mortality. In two thirds of readmission models, the strongest predictor was the number of previous unplanned admissions.

Table 3:   Published and Empirically Derived Weights for Mortality for the Charlson Index

Comorbidity	Original Published	As Used: All Countries Combined	Derived Using England Records	Derived Using U.S. Records	Derived Using Netherlands, Italy, and Belgium Records
Acute myocardial infarction	1	5	4	6	4
Cerebral vascular accident	1	7	10	10	8
Congestive heart failure	1	8	13	11	13
Connective tissue disorder	1	1	3	1	1
Dementia	1	13	16	7	10
Diabetes without long-term complications	1	0	1	−1	1
Mild or moderate liver disease	1	5	8	11	7
Peptic ulcer	1	4	9	4	4
Peripheral vascular disease	1	2	5	2	1
Pulmonary disease	1	2	4	3	3
Cancer	2	7	10	8	5
Diabetes with long-term complications	2	−6	−3	−5	−5
Paraplegia	2	6	2	12	7
Renal disease	2	6	10	6	8
Metastatic cancer	3	9	12	13	10
Severe liver disease	3	11	19	13	8
HIV	6	−3	−4	0	0

DISCUSSION

Using administrative data from 30 mostly academic medical centers in five countries, we produced risk-adjusted outcomes for intercenter comparison on an interactive web-based tool. The resulting analyses, such as we present here showing country-level variation, should not be taken as evidence for differences in quality of care. There are many other potential reasons for the variations in outcome, including case mix, coding, and organization of services. Four patient groups were chosen by the participants for focused learning in the ongoing GOAL phase. In this article, we have outlined the challenges involved in data processing and risk adjustment. We now discuss the issues around each step.

Definition of Inpatient Admission and Handling of Transfers

Only the United States had observation ward status as a data flag, and even this varied between states. The distinction is then blurred as some patients

flagged as being observation patients stayed several days. In England, they are sometimes recorded as emergency department attendances, which are captured in a different dataset that is often incomplete. In any case, their inclusion would lead to denominator inflation that would differ by hospital. We considered taking a diagnosis-specific approach. For example, acute coronary syndrome patients are commonly transferred early for investigation and/or angioplasty. This may be recorded as being from the first hospital's emergency department (which in theory should not generate an admission in the database, but sometimes does) or as a short unplanned inpatient stay. The exclusion criterion of survivors staying <2 days has been used in AMI prediction models (Krumholz et al. 2006) and will reduce the denominator for institutions who transfer quickly. We did this initially but changed to include admissions ending in transfer to an acute unit to capture this important activity.

Deaths are counted only at the center in which they occur, as we do not have the pretransfer center's identity. In contrast, in the Centers for Medicare & Medicaid Services' National Quality Forum-endorsed AMI mortality measure (Joint Commission 2010), patients who are transferred from another acute care or Veterans hospital are excluded because the death is attributed to the hospital where the patient was initially admitted; the patient counts against the second hospital, if the first hospital only sees them in the emergency department, as in this project. The most appropriate attribution of the patient and any death remains unclear with transfers even if they can be linked in the dataset.

Definition of Outcome Variables

There have been some efforts at producing indicators suitable for international comparison. In Europe, the Organization for Economic Co-operation and Development, an intergovernmental economic research institution, launched its Health Care Quality Indicators Project in 2003 (Idänpään-Heikkilä et al. 2006). However, the indicators allow only national and not institutional comparisons. In their selection of cardiovascular indicators, they include, for example, 1-year patient-based mortality following AMI as well as in-hospital mortality for CABG and PTCA. For mortality, we were limited to deaths in hospital, which is clearly affected by discharge policies and the availability and use of intermediate care. We have considered only deaths occurring in the index admission. However, it can be argued that some postoperative complications develop after discharge and result in emergency

readmission, sometimes resulting in death in that readmission, which ought to be attributed to the original surgery. The decision is that what time frame to use, for example, within 90 days of the operation and how to define the complications. The primary diagnosis, LOS, and vital status on discharge of the readmission are made available to users in the tool. Issues with defining hospital mortality indicators are discussed more fully elsewhere (Bottle, Jarman, and Aylin 2011).

Although dichotomizing LOS loses information, "prolonged" LOS is of interest as an indicator of either complications or prolonged processes of care (Faiz et al. 2010). Another advantage is that it is straightforward to model compared with various attempts that have been made to transform it (Vasilakis and Marshall 2005). Normalizing transformations can be complex, as taking logarithms may be inadequate, and compartment models or mixture models (Yau, Leeb, and Ng 2003) are unfeasible, given the large number of patient groups in this project. We also had to use only the portion of an admission covering the time at a participating hospital. Time spent in hospital before or following transfer to or from participating hospitals was not captured; later transfers back were counted as separate admissions, which are inevitable without national coverage by the dataset. However, it may be more appropriate to consider transfers back to the initial hospital as readmissions. If these are for complications of treatment during the index admission, then it may be useful to add the resulting bed days to the index admission's LOS. This approach was taken by the colorectal surgery GOAL group. In general, discussions of LOS within the GOAL groups started with our measure but also covered the mean LOS or time to transfer. Participants also use the web tool to compare the whole LOS distribution at their hospital with that of their peers via histograms.

We were limited to readmissions to the same hospital in this project. In England as a whole in 2008/9, this omitted 11.5 percent of readmissions, but this proportion was unknown for the other participants. Use of the measure for quality improvement at a given hospital will be hampered if this proportion changes over time. A separate issue is choosing the time window for readmissions, with 28 or 30 days perhaps being the commonest. The choice is another trade-off between sensitivity (picking up late complications) and specificity (excluding readmissions due to the patient's underlying disease). Diagnosis-specific windows have been suggested using a mathematical approach (Demir et al. 2008), although quality improvement projects have shown large reductions in readmissions even using 30 days (Hansen et al. 2011).

Adjustment for Comorbidity and Data Validation

We initially allowed the Charlson weights to differ only by outcome measure, but we have since derived separate sets for each diagnosis or procedure group. Other comorbidity indices are available, such as the Elixhauser, which covers 30 conditions (Elixhauser et al. 1998) and whose discrimination was recently found to be generally superior to that of Charlson (Sharabiani, Aylin, and Bottle 2012). All such indices work best if both the levels of secondary diagnosis recording and the distinction between primary and secondary diagnoses are appropriate; the latter distinction may be difficult to make in patients presenting with complex problems.

Including the Charlson score as a linear term in the models invokes several assumptions other than simply a linear relation with the logit of the outcome. First, no interactions between any combinations of two or more comorbidities are modeled—we are currently investigating these using machine learning methods. Second, the effect of each comorbid condition is taken to be the same in all countries. Third, we took the levels of recorded comorbidity at face value for all hospitals, thus assuming that recording is correct or that the degree of underrecording is proportionally equal in every hospital. In England, the diagnosis and procedure coding of a sample of the administrative data is externally audited each year at every hospital. The results give an indirect estimate of the accuracy of the Charlson score for that hospital (Audit Commission 2011). However, equivalents for the other countries were not publically available, although anecdotal evidence points to problems with the Netherlands's LMR database. This is reflected in very low comorbidity levels compared with the other participating countries and sometimes wide gaps between their crude and standardized mortality ratios. Hospitals of each country will have access to other databases for some limited validation of the data that we have used so far. These may include registries or group-specific surveys, for example, MINAP (Healthcare Quality Improvement Partnership 2012) and SINAP (Royal College of Physicians 2012) in the United Kingdom for AMI and stroke, respectively, and NSQIP in the United States for surgery (American College of Surgeons 2012).

Other Issues

As well as no or delayed linkage with postdischarge deaths, administrative data often lack much information on disease severity. This is why the stroke GOAL group is prospectively collecting National Institutes of Health Stroke

Scale information on admission and modified Rankin scores 30 and 90 days later for a subset of patients in participating hospitals. These will be used in further risk models and the results compared with our current ones. This is an example of additional data collection prompted by this project's analysis and limitations of the administrative records. Claims databases have been successfully augmented with physiological variables, for example (Pine et al. 2007; Tabak, Johannes, and Silber 2007). With so many different hospital systems, this will be a challenge for this project, but it is one that we will explore.

Many questions cannot be answered using these data alone. They cannot provide direct evidence for quality of care differences between countries, or say that the shorter LOS in the United States accounts for their lower mortality. For instance, as well as issues with diagnosis and procedure recording, variations in physician diagnostic practice and disease definition can contribute to apparent variations in outcome. Is an AMI in the United States the same as an AMI in Belgium? Do physicians in every country use the same definition for acute coronary syndrome as a whole? We found that the proportion of ACS recorded as AMI rather than unstable angina varies appreciably between countries, which may explain some of the variation in outcomes. The European HCQI project mentioned earlier noted the variation in the diagnostic criteria for AMI and in the diagnosis and coding of heart failure in administrative datasets. Canada and Sweden, for example, base the primary diagnosis on the disease consuming the most resources during the hospitalization, rather than the main problem treated or reason for admission. An analysis of the GUSTO trial data for non-STEMI found that almost all of the intercountry variation in outcomes was explained by patient factors (Chang et al. 2005); in contrast, analysis of AMI patients from another trial did not manage to explain such international differences (Simes et al. 2010). Despite the various data artifacts, quality of care remains one possible explanation of some of the variation in outcomes.

Hospital administrative databases capture patients who were admitted and thereby tell us nothing of those who were not. There are multiple determinants of hospitalization, ranging from psychosocial, supply, demand, and hospital factors. Hospitals that admit sicker patients than other hospitals can be expected to have higher crude outcome rates, and current risk-adjustment modeling may not be able to fully compensate. Patient-based analyses are usually preferable to admission-based ones, but they can be difficult to construct; a suitable starting point from which to start the clock, such as cancer anniversary date or other diagnosis date, will often be unavailable. Nonetheless, such analyses are urgently needed, in particular for diseases such as initial

hospitalization for breast cancer where hospital mortality is not the most logical primary outcome.

CONCLUSIONS

There are considerable challenges in combining administrative databases across countries, with decisions to be made regarding definitions of inpatient admissions and diagnosis and procedure groups. Interpreting the resulting case mix-adjusted in-hospital outcome rates across international boundaries is hampered by differing discharge policies and intermediate care facilities. Standard statistical models cannot take account of this without valid and complete postdischarge information. Nevertheless, the data raise interesting questions and can act as a starting point for more detailed investigation into the reasons for variations in outcomes. By outlining the key limitations of administrative hospital databases for international comparisons of important patient outcomes, this project may help to identify new variables that need to be routinely recorded to explain measured variations in risk-adjusted outcomes of hospital care.

ACKNOWLEDGMENTS

REFERENCES

American College of Surgeons. 2012. National Surgical Quality Improvement Program [accessed on March 15, 2013]. Available at http://www.acsnsqip.org/

Audit Commission. 2011. Payment by Results [accessed on March 15, 2013]. Available at http://www.audit-commission.gov.uk/information-and-analysis/data-assurance-framework/

Bottle, A., and P. Aylin. 2008. "Intelligent Information: A National System for Monitoring Clinical Performance." *Health Services Research* 43: 10–31.

Bottle, A., and P. Aylin. 2011. "Comorbidity Scores for Administrative Data Benefited from Adaptation to Local Coding and Diagnostic Practices." *Journal of Clinical Epidemiology* 64 (12): 1426–33.

Bottle, A., B. Jarman, and P. Aylin. 2011. "Hospital Standardised Mortality Ratios: Strengths and Weaknesses." *British Medical Journal* 342: c7116.

Chang, W.-C., W. K. Midodzi, C. M. Westerhout, E. Boersma, J. Cooper, E. S. Barnathan, M. L. Simoons, L. Wallentin, E. M. Ohman, and P. W. Armstrong. 2005. "Are International Differences in the Outcomes of Acute Coronary Syndromes Apparent or Real? A Multilevel Analysis." *Journal of Epidemiology and Community Health* 59: 427–33.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40 (5): 373–83.

Cronenwett, J. L., D. S. Likosky, M. T. Russell, J. Eldrup-Jorgensen, A. C. Stanley, and B. W. Nolan (VSGNNE). 2007. A Regional Registry for Quality Assurance and Improvement: The Vascular Study Group of Northern New England (VSGNNE). *Journal of Vascular Surgery* 46 (6): 1093–102.

Demir, E., T. J. Chaussalet, H. Xie, and P. H. Millard. 2008. "Emergency Readmission Criterion: A Technique for Determining the Emergency Readmission Time Window." *IEEE Transactions on Information Technology in Biomedicine* 12 (5): 644–9.

Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36 (1): 8–27.

Faiz, O., J. Warusavitarne, A. Bottle, P. P. Tekkis, S. K. Clark, A. W. Darzi, and P. Aylin. 2010. "Nonelective Excisional Colorectal Surgery in English National Health Service Trusts: A Study of Outcomes from Hospital Episode Statistics Data between 1996 and 2007." *Journal of the American College of Surgeons* 210 (4): 390–401.

Groene, O., J. K. H. Skau, and A. Frølich. 2008. "An International Review of Projects on Hospital Performance Assessment." *International Journal of Quality in Health Care* 20 (3): 162–71.

Hansen, L. O., R. S. Young, K. Hinami, A. Leung, and M. V. Williams. 2011. "Interventions to Reduce 30-Day Rehospitalization: A Systematic Review." *Annals of Internal Medicine* 155: 520–8.

Harrell, F. 2001. *Regression Modeling Strategies*. New York: Springer.

HCUP CCS. Healthcare Cost and Utilization Project (HCUP) March 2012. 2012. Rockville, MD: Agency for Healthcare Research and Quality [accessed on March 15, 2013]. Available at www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

Healthcare Quality Improvement Partnership. 2012. Myocardial Ischaemia National Audit Project (MINAP) [accessed on March 15, 2013]. Available at http://www.hqip.org.uk/myocardial-ischaemia-national-audit-project-minap/

Idänpään-Heikkilä, U. M., L. Lambie, S. Mattke, V. McLaughlin, H. Palmer, and J. V. Tu. 2006. "Selecting Indicators for the Quality of Cardiac Care at the Health System Level in Organization for Economic Co-operation and Development Countries." *International Journal of Quality in Health Care* 18 (suppl 1): 39–44.

International Quality Indicator Project. Informational Brochure [accessed on March 15, 2013]. Available at http://www.internationalqip.com/documents/brochure.pdf

Jen, M. H., A. Bottle, G. Kirkwood, R. Johnston, and P. Aylin. 2011. "The Performance of Automated Case-Mix Adjustment Regression Model Building Methods in a Health Outcome Prediction Setting." *Healthcare Management Science* 14 (3): 267–78.

Joint Commission. 2010. Centers for Medicare and Medicaid Services (CMS). Specifications Manual for National Hospital Inpatient Quality Measures, Version 3.1a. Acute Myocardial Infarction (AMI): Hospital 30-Day, All Cause, Risk-Standardized Mortality Rate (RSMR) Following AMI Hospitalization [accessed on March 15, 2013]. Available at http://www.qualitymeasures.ahrq.gov/content.aspx?id=35572

Krumholz, H. M., Y. Wang, J. A. Mattera, Y. Wang, L. F. Han, M. J. Ingber, S. Roman, and S.-L. T. Normand. 2006. "An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates among Patients with an Acute Myocardial Infarction." *Circulation* 113: 1683–92.

Pine, M., H. S. Jordan, A. Elixhauser, D. E. Fry, D. C. Hoaglin, B. Jones, R. Meimban, D. Warner, and J. Gonzales. 2007. "Enhancement of Claims Data to Improve Risk Adjustment of Hospital Mortality." *Journal of the American Medical Association* 297 (1): 71–6.

Royal College of Physicians. 2012. Stroke Improvement National Audit Programme (SINAP) [accessed on March 15, 2013]. Available at http://www.rcplondon.ac.uk/projects/stroke-improvement-national-audit-programme-sinap

Sharabiani, M. T. A., P. Aylin, and A. Bottle. 2012. "Systematic Review of Comorbidity Indices for Administrative Data." *Medical Care* 50 (12): 1109–18.

Simes, R. J., R. L. O'Connell, P. E. Aylward, S. Varshavsky, R. Diaz, R. G. Wilcox, P. W. Armstrong, C. B. Granger, J. K. French, F. Van de Werf, I. C. Marschner, R. Califf, and H. D. White; HERO-2 Investigators. 2010. Unexplained International Differences in Clinical Outcomes after Acute Myocardial Infarction and Fibrinolytic Therapy: Lessons from the Hirulog and Early Reperfusion or Occlusion (HERO)-2 Trial. *American Heart Journal* 159 (6): 988–97.

Sundararajan, V., T. Henderson, C. Perry, A. Muggivan, H. Quan, and W. A. Ghali. 2004. "New ICD-10 Version of the Charlson Comorbidity Index Predicted In-Hospital Mortality." *Journal of Clinical Epidemiology* 57: 1288–94.

Tabak, Y. P., R. S. Johannes, and J. H. Silber. 2007. "Using Automated Clinical Data for Risk Adjustment: Development and Validation of Six Disease-Specific Mortality Predictive Models for Pay-for-Performance." *Medical Care* 45 (8): 789–805.

Vasilakis, C., and A. H. Marshall. 2005. "Modelling Nationwide Hospital Length of Stay: Opening the Black Box." *Journal of the Operational Research Society* 56: 862–9.

Yau, K. W., A. H. Leeb, and A. S. K. Ng. 2003. "Finite Mixture Regression Model with Random Effects: Application to Neonatal Hospital Length of Stay." *Computational Statistics and Data Analysis* 41 (3): 359–66.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.