National administrative data produces an accurate risk prediction model for short term and 1-year mortality following cardiac surgery

[1,2]Dincer Aktuerk, [1]David Mcnulty, [1]Daniel Ray, [1]Irena Begaj, [2]Neil Howell,

[1,3]Nick Freemantle, and [1,2]Domenico Pagano.

[1]Quality and Outcomes Research Unit, and [2] Department Cardiothoracic Surgery, University Hospital Birmingham UK, [3]Department of Primary Care and Population Health, University College London, UK

Word count abstract:

Word count article:

Aknowledgemnts: We would like to thank Professor B. Bridgewater and Dr G. Hickey for providing the methodology to calculate the Recalibrated Logistic Euroscore and Mrs V. Barnett for data collection and data management.

Correspondence

Professor D. Pagano
Consultant Cardiothoracic Surgeon
Clinical Director Quality and Outcomes Reseach Unit
University Hospital Birmingham
B15 2TH, UK
E. domenico.pagano@uhb.nhs.uk

Abstract

Objective: Various models exist to predict short-term risk-adjusted outcomes after cardiac surgery. Many models are calibrated using administrative data and are subject to controversy. This study describes a procedure specific risk prediction model based on administrative NHS Health Episode Statistics (HES) data for England and compares its performance with the EuroSCORE.

Methods: Procedure specific models were built on administrative data and tested using an independent dataset sampled at a later period. Models are applicable to patients having isolated CABG, isolated valve or combined CABG and valve surgery. Outcomes at hospital discharge and 1 year are linked with national death registrations. Comparisons between the models are conducted on a local cohort of patients between 2010 and 2013 using c-statistic for performance and expected vs observed mortality plots to examine calibration across different risk strata.

Results: A total of 84,791 patients in England and 1,174 locally were identified. HES models have higher or comparable predictive performance compared to the re-calibrated logistic EuroSCORE. Model calibration demonstrates good performance with a median difference between observed and predicted mortality rates of 0.15% for CABG, 0.39% for valve and 0.63% for CABG+Valve surgery.

C-statistics test for risk prediction models							
		CABG		CABG and valve		Valve	
		In-hospital (%)	1-year (%)	In-hospital (%)	1-year (%)	In-hospital (%)	1-year (%)
HES	England data	81.6	78.4	76.4	72.0	78.6	77.8
	Local data	80.5	76.3	73.9	72.0	78.0	76.1
Recalibrated logistic EuroSCORE		78.8	75.7	71.6	68.8	71.2	70.5

Conclusion: The national administrative dataset has produced and accurate short and one year mortality prediction after cardiac surgery.

Introduction

Risk stratification is an important part of modern cardiac surgery and there are multiple validated risk predicting statistical models used for numerous purposes (Euroscore, $\log_e$ euroscore, euroscore 2 and sts 3 references valve bacbg and cabg and valves). These are not without limitations. Firstly, there are issues regarding their construction based upon variably sized datasets collected in clinical speciality specific registries.

Although it is recognised that this approach may provide more detailed information for individual patients, there are a number of problem areas. Firstly national clinical registries that underpin clinical audits are expensive (ref HQUP 14 million). Secondly, once a registry is established, by its very nature it is unable to adapt to include emerging prognostically important variables. Finally, particularly when used for governance purposes (both institutional and individual level), given their information is usually entered by clinicians, it may lead to inaccuracies and bias ((J Am Coll Cardiol 2012;59:2309–16)).

Furthermore, most risk predicting scores are modelled to predict an early end point of mortality. There is debate as to exactly what constitutes such an endpoint: death in hospital, death within 30 days a combination of the two, and this lack of standardisation makes benchmarking comparisons difficult amongst different healthcare systems.

Finally, given that most cardiac surgical procedures are performed for their perceived late prognostic value, there are no tools to predict such outcomes to date.

Administrative datasets are known to have several pitfalls and have not been recommended as a principal datasets to construct risk-scoring systems (REF Siregar et al.). The Hospital Episode Statistics (HES) dataset linked to the Office of National Statistics (ONS) provides an opportunity to use routinely collected patient level national data with provision to identify in

hospital and post-discharge long term outcome, and we have demonstrated its accuracy when used to create mortality based risk algorithms for a national health care system (QUORUM paper NBJ Open). The aim of this study was therefore to develop this methodology for cardiac surgery by constructing procedure specific prognostic models both for early (in-hospital and 30 days) and up to one year and assess their performance.

Materials and methods

Data Source

Data were extracted from the National Hospital Episode Statistics (HES) dataset (http://www.hscic.gov.uk/hes). Mortality data were obtained from the Office of National Statistics and linked to HES data by the Health and Social Care Information Centre, England. For the purpose of this study patients were selected on the basis of having undergone isolated coronary artery bypass grafting (CABG), isolated valve surgery or a combined valve and CABG procedures (CABG+Valve). Patients having undergone aortic surgery, procedures for adult congenital heart disease, thoracic organ transplantation and implantation of a primary ventricular assist device or postinfarction VSD repair were excluded. The sample and independent variables i.e. the additive and logistic EuroSCOREs were extracted from the local cardiac surgical database.

Statistical Analysis and Model development

This study was conducted in two stages. In stage 1 HES data was used to construct models predicting short (in-hospital and 30 days from operation) and midterm mortality (90 days, 180 days and 365 days from operation). The sample and independent variables for the HES models were derived from clinical, demographic and administrative data contained within NHS Health Episode Statistics (http://www.hscic.gov.uk/hesdatadictionary). In stage 2 the HES based predictions were compared with the additive EuroSCORE, recalibrated EuroSCORE and logistic EuroSCORE.

Development of HES model

National patient data for the financial years 1 April 2008/ 31 March 2009 to 1 April 2010/ 31 March 2011 were extracted from HES and split 70:30 as development and test samples. The same procedure was repeated for the 3 clinical groupings (CABG, valve and CABG+Valve surgery) and for the different mortality end points (in-hospital, 30, 90, 180 and 360 postoperative days). A second independent dataset was extracted from HES in financial years 2011/2012 and 2012/13 to assess the time stability of models.

All procedures were identified using Office of Population, Censuses and Surveys taxonomy (OPCS codes) selected by a team of cardiac surgeons. In admissions with multiple episodes the earliest episode having cardiac surgery was chosen.

Explanatory variables were selected in the year preceding the admission considering existing knowledge on items likely to be associated with mortality in admissions to the NHS, and based upon previous experience (11). The following explanatory variables were included: age, gender, ethnicity, deprivation score (Index of Multiple Deprivation, IMD) (Ref: Noble M, Wright G, Dibben C, et al. Indices of deprivation 2004. Report to the Office of the Deputy Prime Minister. London: Neighbourhood Renewal Unit, 2007.) hospital trust, individual components of the Charlson comorbidity score12, type and method of admission (determined on the basis of admissions ending within the specified time period), total number of previous emergency admissions, worst Charlson score on previous admissions, hospital length of stay and trust – year (identifying the patients admitted to each trust over each of the three included years) . The index of multiple derivation comprises of seven indicators on a scale of 0 (least deprivation) to 100 (highest deprivation). The individual domains and their weights are: Income deprivation 22.5%; Employment deprivation 22.5%; Health deprivation and disability 13.5%; Education, skills and training deprivation 13.5%; Barriers to housing and services 9.3%; Crime 9.3%; Living Environment deprivation 9.3%. Charlson components relating to acute events i.e. stroke and heart attack were excluded, as it is uncertain in the administrative

record whether they occur before or after the surgical intervention. Year and Month were used to investigate long term and seasonal time trends respectively. Models were constructed to predict mortality using stepwise logistic regression on a 70% development sample from April 2008 until March 2011. After validation on the 30% test sample models were recalibrated on the full dataset. Individual logistic mortality models were constructed for CABG, valve and combined valve and CABG procedure at hospital discharge and at 30, 90 and 365 days post operation.

Performance of HES model

The risk models were evaluated in terms of discrimination and calibration. Discrimination was assessed by determining the C-Statistics, also known as the area under the receiver operating characteristic (ROC) curve. A C-Statistic of 100% indicates perfect discriminative power and 50% no discriminative ability. The Calibration was not assessed using the Hosmer-Lemeshow test since is not informative in large sample sizes, but we constructed calibration plots comparing observed vs predicted mortality across different risk strata. We assesed the performance of the HES model within the development dataset (years 2008-2011) and on a yearly basis (2008-2013) for all the mortality endpoints.

The second part of this study assessed the performance of the HES model against the EuroSCORE and its variants. Access to EuroSCOREs was limited to patients admitted locally. Additive and logistic EuroSCOREs were extracted for patients with the same medical profile as in stage 1 i.e. patients who received CABG, valve surgery and CABG + valve surgery. A recalibrated EuroSCORE (Rec-ES) (compensating for drift in the mortality calibration was calculated using the methodology provided by the National Institute for Cardiovascular Outcomes Research (Eur J Cardiothorac Surg (2014) 45 (2): 225-

233).. HES risk scores were calculated using HES model, but may also be calculated from local episode tables providing prior emergency and non-emergency admissions. The final data set contained:

- Additive EuroScore, Logistic EuroScore, Re-calibrated Euroscore
- HES scores predicting In hospital, 30, 90, 180 and 365 day mortality post operation
- ONS mortality recorded In hospital, 30, 90, 180 and 365 day mortality post operation

from 2008/09 to 2013/14.

C-Statistics were compared using bootstrapping (Reference: Efron b. and Tibshirani R. 1993 and introduction to the Bootstrap. (Chapman & Hall/ CRC monographs on statistics & applied probability.). One thousand c-statistics were generated by sampling the original data with replacement and calculating the c-statistic on each sample for each score. The resultant dataset contains paired data and facilitates direct comparison of c-statistics

Analyses were performed using SAS 9.2. This study has been approved by the University Hospitals Birmingham Clinical Governance Board (code: CARMS-00050).

Results

The HES model was developed using data of patients undergoing CABG (35115), valve surgery (18353) and CABG+valve (8392) between 1/4/2008 and 31/3/2011 in England. In hospital mortality for CABG was 1.82% (650/35115), for valve surgery 3.6% (666/18353) and for CABG+valve 6.0% (505/8392). Patients' summary statistics including type of cardiac surgery and in-hospital outcomes are described in Table 1. The model variables to predict in hospital mortality are illustrated in Table 2. the remaining models predicting mortality at 30, 90, 180 and 365 days post surgery are available in supplemental file 1. The predictive model for in-hospital mortality was tested also on patients undegoing surgery frorm 1-/4/-2011 to 31/-03/-2013.

Performance statistics for the HES models for mortality in –hospital at 30, 90, 180 and 360 days post surgery are described in Table 3 (discrimination) and Figure 1 (calibration). Calibration of the HES model demonstrates a small difference between observed and expected mortality and delivers a good estimate of risk of mortality for patients undergoing CABG, valve or combined CABG and valve surgery. Across the 10 risk strata and all time periods the median absolute difference between observed and expected mortality rates is 0.15% for CABG, 0.39% for Valve and 0.63% for combined CABG and valve; the upper quartile for differences are 0.21%, 0.65% and 1.1% respectively. Analysis of In-hospital deaths during the developmental (2008/09, 2009/10, 2010/11) and validation periods (2011/12, 2012/13) shows under prediction on 2008/09 and good prediction within the bounds of error for remaining years on isolated CABG and valve procedures and concomitant CABG and valve. Kaplan Meier survival curves post surgery are illustrated in Figure 2 (for the entire patient population- a, and for the 3 groups of surgery- b.) and counts of deaths post surgery in Table 4 (a and b). Visual analysis indicates that that survival has improved over time. After

an initial sharp decline survival becomes linear after 120 days from surgery for the entire population and at different times for different procedures.

Comparison of the HES models against the EuroSCOREs is on a much smaller sample i.e. the local hospital dataset with around 500 patients per year, and to compensate performance statistics are accumulated over five years from April 2008 until March 2013. Across all five years and all three sets of procedures there were 2580 patients and 88 deaths. HES surgical classifications as CABG, valve and combined procedures decoded from OPCS codes agree with surgical classification input directly into the local cardiac database; 99.8%, 99.2%, 98.0% respectively. Comparison data of the HES model with EuroSCORE are described in Table 5 (discrimination) and Table 6 (calibration).

Additional bootstrap results for the in-hospital deaths indicate the discrimination of the EuroSCORE variants is the same within each type of operation; CABG 81%, combined CABG and Valve 74%, and Valve 69%. The HES based model has the same dicrimination as the EuroSCORE for CABG and combined CABG and Valve procedures. HES discrimination of the Valve only procedure is 9.4% higher than EuroSCORE alternates. There was no significant difference between the HES and the Rec-ES in the local dataset. Nonetheless an increase in the c-stat of the Valve model from 65% to 75% lifts a predictor from being relatively poor to good.

Table 1

Demographic characteristics of the study population for the HES model development .

		CABG		Valve		CABG + Valve		
Characteristic	**Value**	**Survive**	**Die**	**Survive**	**Die**	**Survive**	**Die**	**Value**
Gender	Male	28,037	454	9,930	335	5,574	310	Male
	Female	6,428	196	7,757	331	2,313	195	Female
Ethnicity	White	27,883	509	15,184	567	6,855	420	White
	Asian	2,310	54	468	22	199	24	Asian
	Other	941	21	555	27	124	11	Other
	Unknown	3,331	66	1,480	50	709	50	Unknown
		CABG		**Valve**		**CABG + Valve**		
	Value	**Survive**	**Die**	**Survive**	**Die**	**Survive**	**Die**	**Value**
Age	20	8		223	5			20
	25			279		7		25
	30	14	11	310	8		8	30
	35	70		337	8	8		35
	40	346		489	13	19		40
	45	958		642	9	60		45
	50	2,025	12	813	21	108	7	50
	55	3,188	17	1,050	27	256	9	55
	60	4,992	44	1,686	31	593	24	60
	65	6,226	77	2,325	50	992	45	65
	70	6,399	116	2,660	101	1,481	82	70
	75	6,080	173	2,946	149	1,990	131	75
	80	3,244	128	2,273	142	1,593	117	80
	85	852	66	1,238	72	673	71	85
	90	60	6	366	24	107	11	90
	95			50	6			95
		CABG		**Valve**		**CABG + Valve**		
	Value	**Survive**	**Die**	**Survive**	**Die**	**Survive**	**Die**	**Value**
IMD	0 - 10	12,187	212	7,106	250	3,186	190	0 - 10
	11 - 20	9,569	179	5,056	178	2,208	152	11 - 20
	21 - 40	8,634	176	3,820	145	1,746	109	21 - 40
	41 - 60	3,382	74	1,405	74	610	49	41 - 60
	61 +	693	9	300	19	137	5	61 +
		CABG		**Valve**		**CABG + Valve**		
	Value	**Survive**	**Die**	**Survive**	**Die**	**Survive**	**Die**	**Value**
Current Admission	Elective	22,772	278	14,507	392	6,014	298	Elective
	Emergency	4,246	185	1,450	147	689	96	Emergency
	Other	7,447	187	1,730	127	1,184	111	Other
Emergency Admissions ending within 365 days	0	22147	371	11717	353	5221	285	0
	1	8836	171	3811	152	1808	127	1
	2	2355	64	1339	74	551	51	2
	3+	1127	44	820	87	307	42	3+
Other Admissions ending within 365 days	0	11300	305	3849	197	1655	161	0
	1	15433	181	7729	211	3686	178	1
	2	5220	103	3981	130	1671	92	2
	3	1621	38	1301	79	573	41	3
	4+	891	23	827	49	302	33	4+

Table 2 Comparison of models to predict in-hospital mortality

		CABG		Valve		CABG + Valve	
Effect	**Level**	**Estimate**	**Probt**	**Estimate**	**Probt**	**Estimate**	**Probt**
Intercept		-4.928	<.0001	-4.471	<.0001	-3.570	<.0001
Gender	Male	0.000		-0.308	0.000	-0.398	0.000
	Female	0.257	0.005	0.000		0.000	
Age (B-Spline)	Age_0	2.766	0.008	-0.127	0.665	0.935	0.596
	Age_1	-2.632	<.0001	0.000		-0.513	0.494
	Age_2	0.000		1.334	<.0001	0.000	
	Age_3	0.924	0.001			0.483	0.068
	Age_4	1.816	<.0001			0.641	0.034
	Age_5					2.108	0.020
Ethnicity	Unknown	0.176	0.203				
	Other	0.381	0.080				
	Asian	0.410	0.007				
	White	0.000					
CHF	Yes	1.413	<.0001	1.232	<.0001	1.150	<.0001
	No	0.000		0.000		0.000	
LvDis	Yes	3.256	<.0001	2.595	<.0001	3.608	<.0001
	No	0.000		0.000		0.000	
RenDis	Yes	0.805	<.0001	1.132	<.0001	0.792	<.0001
	No	0.000		0.000		0.000	
PVD	Yes	0.631	<.0001			0.329	0.023
	No	0.000				0.000	
PeUl	Yes					1.026	0.002
	No					0.000	
CHF*RenDis	CHF = Y ; RenDis = Y			-0.473	0.018		
	CHF = Y ; RenDis = N			0.000			
	CHF = N ; RenDis = Y			0.000			
	CHF = N ; RenDis = N			0.000			
Current Admission	Other	0.314	0.003	0.512	<.0001	0.224	0.096
	Emergency	0.769	<.0001	0.694	<.0001	0.635	<.0001
	Elective	0.000		0.000		0.000	
Emergency Admissions in last Year	3+	0.744	<.0001				
	2	0.312	0.038				
	1	0.217	0.022				
	0	0.000					
Other Admissions in Last year	4+	0.200	0.370				
	3	-0.007	0.973				
	2	0.046	0.717				
	1	-0.251	0.015				
	0	0.000					
Other Admissions in the last 3 Months (91 days)	2+					0.331	0.057
	1					-0.195	0.095
	0					0.000	
Days in hospital from admissions ending in the Last Year				0.017	<.0001	0.014	0.004
No_History				0.429	0.001		
Age B-Spline Details		Degree = 2 ; Min = 19 ; Max = 94 ; Knots = 62, 67		Degree = 1 ; Min = 18 ; Max = 99 ; Knots = 60		Degree = 2 ; Min = 18 ; Max = 99 ; Knots = 60, 74, 81	

Unk= Unknown; CHF = congestive heart failure; PeUl= Peptic ulcer; PVD=Peripheral Vascular Disease;

Table providing details of the In Hospital mortality models on a $\log_e$ odds scale. Parameters for each model have been aligned to compare which terms are included and excluded within each model. All classification levels are included for each factor. Reference levels are identified in the table as having a zero estimate and no associated p-value. Continuous factors e.g. Age are incorporated using B-Splines and have similar form/shape to the mortality rates in Table 1. Interaction terms are indicated by the * between the two candidate variables.

Table 3

Discrimination of the HES models (C- Statistics, %)

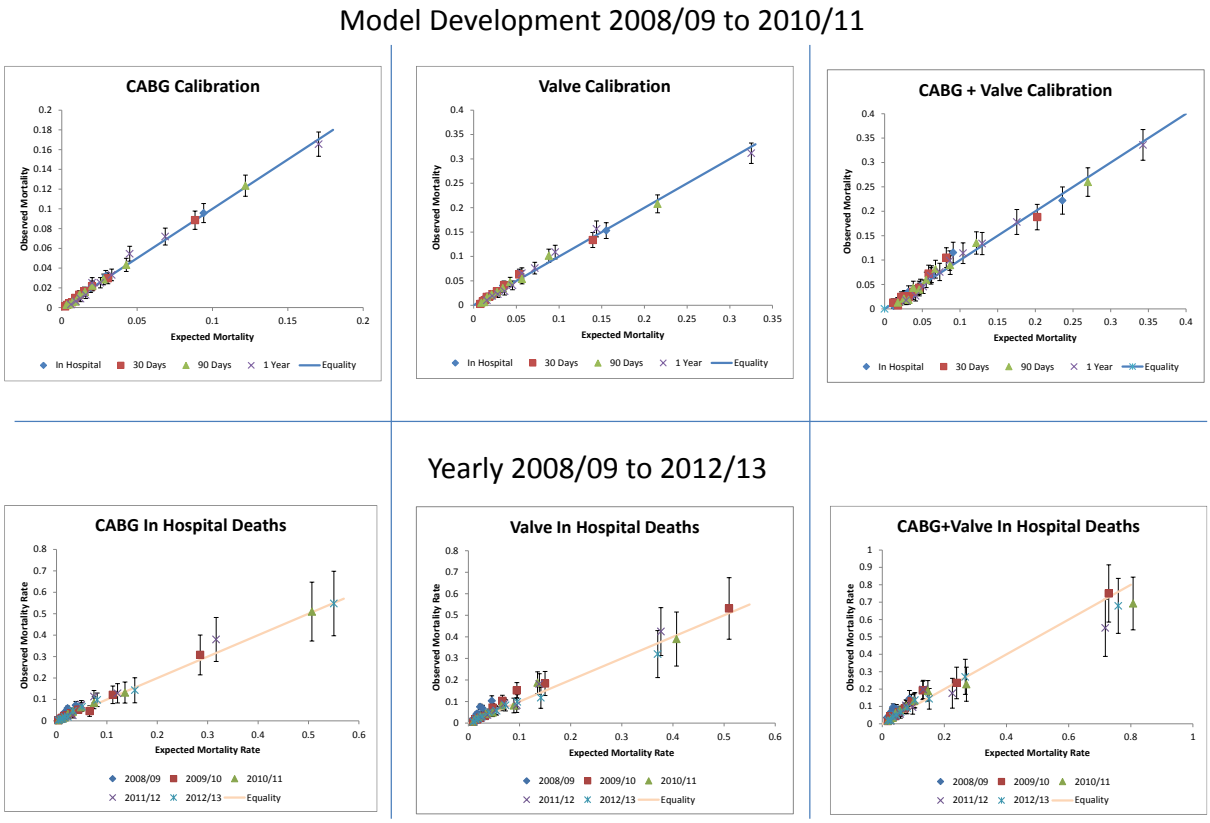
Model	Mortality	Development statistics			Yearly Performance				
		Development (%)	Testing (%)	Combined (%)	2008/09 (%)	2009/10 (%)	2010/11 (%)	2011/12 (%)	2012/13 (%)
CABG	In-hospital	81.8	81.4	81.6	72.4	78.6	80.8	83.4	82.4
	30 days	80.0	78.6	79.6	68.9	77.6	79.4	79.5	77.7
	90 days	80.6	79.9	80.4	70.5	78.2	80.1	79.9	77.3
	180 days	79.7	79.1	79.5	70.5	76.9	79.8	78.6	76.7
	365 days	78.9	77.3	78.4	70.5	76.7	78.4	77.5	76.9
Valve	In-hospital	79.0	77.6	78.6	68.0	79.6	75.0	78.5	76.0
	30 days	76.5	77.6	76.8	66.5	77.2	71.7	77.1	73.3
	90 days	77.9	77.2	77.7	68.8	78.0	74.2	75.6	74.4
	180 days	78.3	77.1	78.0	69.3	78.0	74.1	75.6	74.0
	365 days	78.1	77.0	77.8	69.0	77.1	74.5	75.6	74.1
CABG + Valve	In-hospital	77.8	73.1	76.4	61.8	74.0	75.8	74.9	74.7
	30 days	75.7	69.7	73.9	63.3	71.1	72.3	72.0	73.8
	90 days	75.6	69.9	73.9	66.0	71.8	72.8	72.8	73.1
	180 days	74.7	69.6	73.1	66.2	71.6	71.9	71.9	71.9
	365 days	73.4	69.0	72.0	66.2	71.1	71.0	71.1	71.8

CABG= Coronary Artery Bypass Grafting; % indicates the C-Stats area under the ROC curve;

.

The C-statistics of the HES model applied to three procedure groups (CABG, valve, CABG and valve) demonstrates a good performance up to one year post surgery. Extended testing to patients undergoing surgery in subsequent financial years (2011/2012 and 2012/2013) shows that the models maintained performance outside of the development period.

Figure 1.


Model Development 2008/09 to 2010/11
CABG Calibration
Observed Mortality
Expected Mortality
In Hospital
30 Days
90 Days
1 Year
Equality
Valve Calibration
Observed Mortality
Expected Mortality
In Hospital
30 Days
90 Days
1 Year
Equality
CABG + Valve Calibration
Observed Mortality
Expected Mortality
In Hospital
30 Days
90 Days
1 Year
Equality
Yearly 2008/09 to 2012/13
CABG In Hospital Deaths
Observed Mortality Rate
Expected Mortality Rate
2008/09
2009/10
2010/11
2011/12
2012/13
Equality
Valve In Hospital Deaths
Observed Mortality Rate
Expected Mortality Rate
2008/09
2009/10
2010/11
2011/12
2012/13
Equality
CABG+Valve In Hospital Deaths
Observed Mortality Rate
Expected Mortality Rate
2008/09
2009/10
2010/11
2011/12
2012/13
Equality

Legend to Figure 1.

Calibration plot of the HES risk prediction models. Panel A: observed vs. expected risk of mortality (±95% confidence interval) for risk strata of the HES models. Panel B: year by year calibration for In-hospital deaths.

Figure 2 a. Survival from operation for all groups

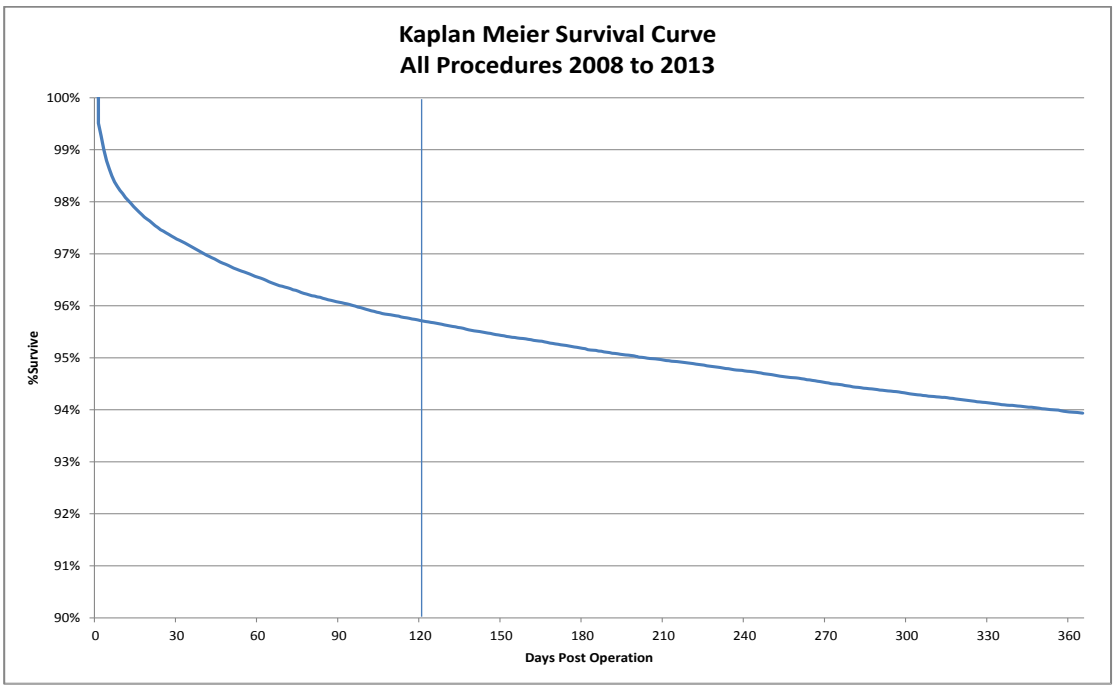
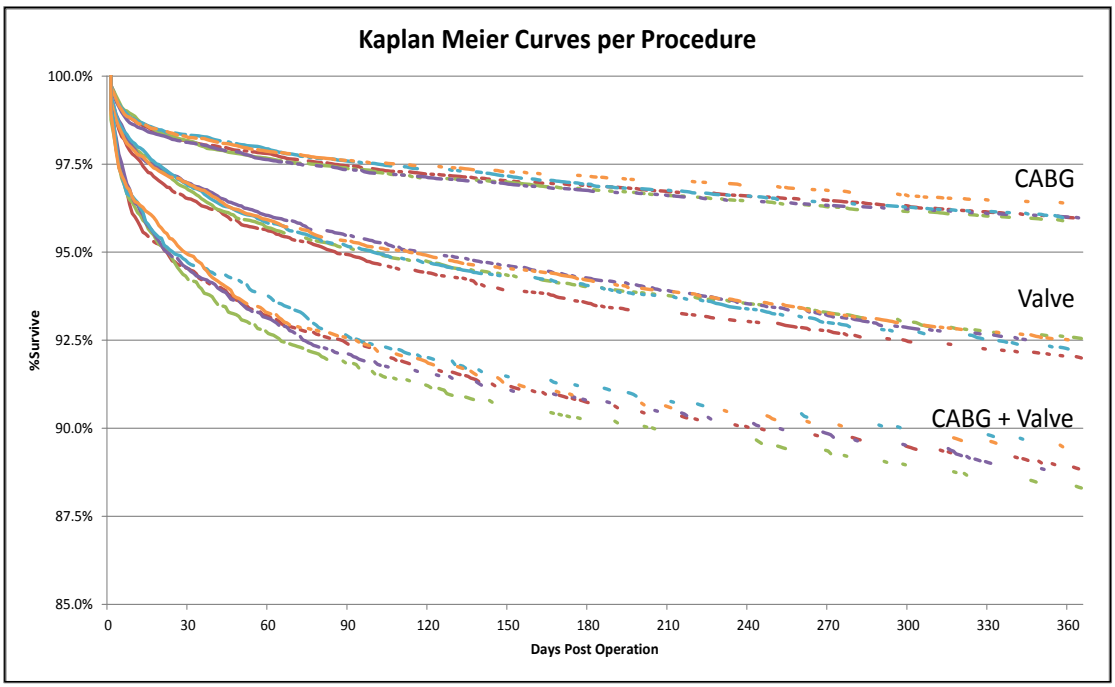


Figure 2 b. Survival by procedure



Legend to Figure 2 (a and b).

a. Kaplan Meier survival plot for all patients undergone surgery between finacial years 2008-2013. Note the accelerated survival becomes linear at 120 days post surgery.

b. Kaplan Meiar survival plot for all patients undergone surgery between finacial years 2008-2013 grouped by the 3 procedure .

Table 4 a: Count of deaths over time following different operations

Deaths Post operation aggregated from 2008 to 2013									
	CABG		Valve		CABG + Valve		Total		
Days post Operation	In Hospital	Discharged	In Hospital	Discharged	In Hospital	Discharged	In Hospital	Discharge	Grand total
0 - 30	1296	171	1233	172	997	98	3526	441	3967
31 - 60	119	232	161	283	112	166	392	681	1073
61 - 90	32	218	44	227	42	139	118	584	702
91 - 120	16	164	23	180	23	107	62	451	513
121 - 150	4	138	9	166	5	95	18	399	417
151 - 180	2	143		145	4	72	6	360	366
181 - 210	3	124	2	128	2	67	7	319	326
211 - 240	2	113	3	132	2	52	7	297	304
241 - 270		126	2	139	1	62	3	327	330
271 - 300		112	1	124		64	1	300	301
301 - 330	1	109		101		57	1	267	268
331 - 360		107		99		50		256	256
Total for Year	1475	1757	1478	1896	1188	1029	4141	4682	8823

Table 4 b: Percentages of deaths over time following different operations

Deaths Post operation aggregated from 2008 to 2013									
	CABG		Valve		CABG + Valve		Total		
Days post Operation	In Hospital	Discharged	In Hospital	Discharged	In Hospital	Discharged	In Hospital	Discharge	Grand total
0 - 30	87.9%	9.7%	83.4%	9.1%	83.9%	9.5%	85.1%	9.4%	45.0%
31 - 60	8.1%	13.2%	10.9%	14.9%	9.4%	16.1%	9.5%	14.5%	12.2%
61 - 90	2.2%	12.4%	3.0%	12.0%	3.5%	13.5%	2.8%	12.5%	8.0%
91 - 120	1.1%	9.3%	1.6%	9.5%	1.9%	10.4%	1.5%	9.6%	5.8%
121 - 150	0.3%	7.9%	0.6%	8.8%	0.4%	9.2%	0.4%	8.5%	4.7%
151 - 180	0.1%	8.1%	0.0%	7.6%	0.3%	7.0%	0.1%	7.7%	4.1%
181 - 210	0.2%	7.1%	0.1%	6.8%	0.2%	6.5%	0.2%	6.8%	3.7%
211 - 240	0.1%	6.4%	0.2%	7.0%	0.2%	5.1%	0.2%	6.3%	3.4%
241 - 270	0.0%	7.2%	0.1%	7.3%	0.1%	6.0%	0.1%	7.0%	3.7%
271 - 300	0.0%	6.4%	0.1%	6.5%	0.0%	6.2%	0.0%	6.4%	3.4%
301 - 330	0.1%	6.2%	0.0%	5.3%	0.0%	5.5%	0.0%	5.7%	3.0%
331 - 360	0.0%	6.1%	0.0%	5.2%	0.0%	4.9%	0.0%	5.5%	2.9%
Total for Year	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0 %

Table 5. Discrimination of the HES model and the EuroSCORE models

Period	Model	CABG	Valve	CABG + Valve
In hospital	HES	80.5%	78.0%	73.9%
	Rec-ES	78.8%	71.2%	71.6%
	ES	79.0%	70.9%	71.6%
	Add-ES	79.2%	70.5%	71.1%
30 Days	HES	78.4%	74.4%	72.2%
	Rec-ES	78.6%	72.3%	73.2%
	ES	78.8%	72.2%	73.4%
	Add-ES	78.9%	71.7%	73.0%
90 Days	HES	82.3%	76.2%	74.0%
	Rec-ES	78.8%	71.5%	73.7%
	ES	78.8%	71.3%	73.9%
	Add-ES	79.0%	71.0%	73.3%
180 Days	HES	80.7%	72.4%	74.6%
	Rec-ES	78.3%	71.9%	72.8%
	ES	78.3%	71.9%	72.9%
	Add-ES	78.0%	71.7%	72.7%
1 Year	HES	76.3%	76.1%	72.0%
	Rec-ES	75.7%	70.5%	68.8%
	ES	75.8%	70.5%	68.7%
	Add-ES	75.2%	70.5%	68.9%

HES: Hospital Episode Statistics, refers to the administrative dataset from which the risks models have been constructed. Rec-ES: Recalibrated EuroSCORE; logES =logistic EuroSCORE; add-ES: additive EuroSCORE.

The discrimination for the HES model for In-hospital deaths for CABG and CABG+valve is similar to the EuroSCORE models (Rec-ES 0.79, logES 0.79, addES 0.79, for CABG) and (Rec-ES 0.72, logES 0.72, addES 0.71 for CABG+valve). For patients undergoing valve surgery however, discrimination of the HES model is superior (0.78 vs. 0.71). The c-statistics of the EuroSCORE models for longer periods are numerical lower than that of the HES model.

Table 6

Observed versus expected deaths by 10 different risk strata.

In hospital Risk stratification						
Model	Strata	Expected	Observed	Patients	Exp.Mortality	Observed Mortality
Recalibrated Logistic EuroSCORE	1	6.4	16	1180	0.5%	1.4%
	2	6.4	12	424	1.5%	2.8%
	3	6.4	6	275	2.3%	2.2%
	4	6.5	5	206	3.2%	2.4%
	5	6.4	5	153	4.2%	3.3%
	6	6.3	8	115	5.5%	7.0%
	7	6.6	12	91	7.2%	13.2%
	8	6.4	9	62	10.3%	14.5%
	9	6.5	8	46	14.2%	17.4%
	10	6.5	7	29	22.3%	24.1%
HES	1	5.2	6	961	0.5%	0.6%
	2	5.3	11	460	1.1%	2.4%
	3	5.2	7	331	1.6%	2.1%
	4	5.2	8	257	2.0%	3.1%
	5	5.3	9	195	2.7%	4.6%
	6	5.2	8	138	3.8%	5.8%
	7	5.3	8	102	5.2%	7.8%
	8	5.3	11	72	7.3%	15.3%
	9	5.2	11	46	11.4%	23.9%
	10	5.3	7	18	29.6%	38.9%
Risk is stratified by expected deaths to mitigate low counts within each strata						

Table illustrates the observed versus expected number of deaths for the Recalibrated Logistic EuroSCORE model compared to the HES model by risk strata. To avoid reporting restrictions due to low numbers the table is partitioned with respect to the expected number of deaths in each risk strata. The Recalibrated Logistic EuroSCORE shows over prediction in the lower risk strata, and the HES model under-prediction in the higher risk strata, in this limited dataset.

Discussion

Analysing HES data for NHS England linked to ONS, a large, complete and objective data source we have obtained time to death following adult cardiac surgical procedures and have produced specific models that have both good discrimination and calibration, and ~~-~~are stable over time. ~~ and we believe capture the majority of post procedural deaths.~~ We have also demonstrated that administrative data sets, when used correctly including complete patient specific data may be produce risk prediction ~~algorhytm comparable~~algorithm comparable to those based on large clinically specific registries without some of their potential disadvantages.

The HES data is comprehensive of each hospital admission and its level of completeness has significantly increased in recent years. Specialised clinical coders who have to have a nationally recognised qualification enter patient level data. HES data depth of coding, a measure of completeness of patient level comorbidity often used as a measure of data quality has also improved. This dataset is used for high profile institutional level national governance, by health care quality regulators, forms the basis of reimbursement and is regularly audited by independent national auditing bodies.

Clinical registry based data provide the most robust information from which to construct disease/clinical specialty specific risk predicting models. The Society of Thoracic Surgery (STS) have suggested that data derived from clinical registries are superior than those derived from administrative sources, especially when its models are based upon large and robust data sources such as the STS NCD (REF) . European centres have also demonstrated the superiority of clinical registries over administrative data. A recent Dutch study demonstrated significant inferiority of a risk model based on administrative data compared to a clinical

model on both model discrimination and calibration thus recommending caution when interpreting benchmarking information from administrative data in the Netherlands (REF Siregar et al).

It could be hypothesised that one of the reason for inferiority of previously developed administrative registries is that they often tried to replicate the disease specific dataset variables that clinical registries collect and then, because of a lack of specific prognostic data fields such as: active endocarditis or presence of pulmonary hypertension in the Dutch administrative data, the derived risk algorithm may produce and inferior performance. The advantage of large administrative data sets such as the National Health Service Hospital Episode Statistics for England, is that it collects information on powerful prognostic covariates such as the presence of liver disease, the level of social deprivation, ethnicity, and previous hospital admission history which are not usually included in disease specific registries (REF our BMJ paper, QURUM model). It is possible that our approach of incorporating this patient specific powerful prognostic information in our models offsets the loss of disease specific data such as left ventricular function, degree of coronary artery disease which are known to be powerful disease specific prognostic variables normally collected in registries. This is in accord with Lilford's 2004 study (REF get from heart editorial) that suggested that variation in hospital outcomes was a composite of data quality/definitions, case-mix healthcare quality and chance i.e. that outcome can be influenced by unmeasured or unrecorded factors that carry significant prognostic importance.

Siregar et al. (REF) reported the difficulty in matching the administrative hospital episodes with those recorded in the registries, a well-recognised limitation of such an approach. In our study we found a high level of agreement between registry and HES data (>95%) for the three procedure groupings, although this finding is based only on a local registry. Shold this level of agreement be demonstrated on a larger dataset, it would suggest that at least at institutional

level in England, the HES model could be used for outcome and benchmarking analysis in adult cardiac surgery, and a similar approach could be implemented where good quality administrative data is available but clinical registries are not well developed. Some national governance programme, such as the one in UK (REF website SCTS.ORG page on patient Outcomes) are using data from clinical registries for surgeon specific outcome monitoring and this approach relies on the exact attribution of eligible procedures to a individual practitioner. This remains a recognised area for improvement for the HES administrative dataset.

Both the EuroSCORE and the STS scores are affected by a phenomenon called "calibration drift" which commonly leads to significant overestimation of the expected mortality end point with the time. This phenomenon has important repercussions, including the over-estimation of risk for patients with higher risk profile, thus potentially denying them from surgical treatment. Furthermore a model that over predicts risk when used as a benchmarking tool may falsely ~~reassure us~~reassure us about clinical performance. Calibration drift is due to several reasons, which include changes in clinical practice, lack of clarity on variable definitions and inaccurate data entry. One example is that in the original EuroSCORE definition of recent myocardial infarction had a significant weight in predicting increased risk of in-hospital death following cardiac surgery, in keeping with the significant increase risk of early surgery shown in studies such as the VANQUISH trial (REF). The introduction in routine clinical practice of high sensitivity Troponin analysis has now changed the threshold for the diagnosis of myocardial infarction, and now it includes large numbers of patients with less myocardial necrosis and a less risk following surgery. When applying the EuroSCORE definition and its original statistical weight to this group it will lead to overestimation of their risk.

The commonest method of dealing with calibration drift is to recalibrate the risk model to produce an adjustment factor in the equation. This may need to be done as frequently as every year to maintain its accuracy raising concerns over model fit over several adjustments. In our study the HES model proved to be stable over two subsequent years after its development and in some cases its performance (C-Statistics) improved which could reflect the continuous improvement in the HES dataset. The stability of our risk scoring system, which needs to be proven over a longer period of time, could be due to the fact that is based on powerful patient level covariates that are well defined and unlikely to change their impact on prognosis over a short time as described above.

The STS ~~have~~ has improved their risk model prediction by developing procedure specific models. They have validated models for CABG, isolated valve and valve + CABG procedures, however they models for more specific valve procedures such as mitral valve surgery proved difficult to construct (REF STS Valve Model Paper Annals T S). We have followed a similar approach. In our study, the model fit and its calibration are good for isolated CABG surgery (based on 35115 patients' data) and is good for valve only procedures (based on 18353 patient data). The number of patients undergoing combined valve and CABG procedures was adequate to construct a further well-calibrated model but a reduction in fit is noted for this group, as demonstrated by the difference in c-statistics between construction and validation set (ref that STS used). Whilst "one for all" procedures risk prediction algorhytm have been proven to be adequate for overall institutional risk adjusted outcome reporting, our findings suggest that patients undergoing more sub-specialised and less common procedures such as aortic surgery may need a different approach to both risk estimation and quality and outcome benchmarking programs.

The EuroSCORE and the STS risk models have been developed to predict early mortality (in hospital and / or 30 days). Whilst this could have been an acceptable approach in the pioneering era of cardiac surgery and it remains a common approach for governance and institutional benchmarking initiatives, we feel that in modern multi-modality practice, we also need to be able to predict the longer-term benefit that should be associated with these complex surgical procedures. In our study we developed tools to predict early and up to one-year outcomes, and each model was specifically developed for the different mortality endpoints. In doing so we also identified variables that seem to be important for long-term prognosis such as the level of social deprivation, as we have previously demonstrated (BMJ Paper). We also found, at least using data from our local institution, that the EuroSCORE variables produce a longer-term prediction of similar accuracy, confirming the long-term values of its variables. Longer-term prediction could be improved by linking national administrative and clinical registry data, thus potentially identifying a "mixed" set of patient variables that underpin a more robust and accurate production on longer term. This approach could help to identify patients that are more likely to benefit from cardiac surgery.

Recently, it has been recognised that there is a need to identify and standardise measurable clinically relevant endpoints for each major procedure (VARC-2 Reference EJCTS). This approach is useful for clinical reporting of outcomes, to benchmark like-with-like, for designing clinical trials and for patients level outcome estimation. Historically in conventional cardiac surgery early post-procedural mortality (30 days or in-hospital) has been used as a standard endpoint. We have previously discussed that this pragmatic approach, at least when used for governance may have limitations (Heart Editorial), particularly as it does not take into account different practices of discharge among several health care systems thus missing some early deaths from the counts. In addition, whilst there might be value in using an early end point for institutional based outcome measure, longer-term outcomes may

depend more on health care secondary care. Our time-to-death analysis (Figure 2 and Table 4) provides information that would help inform a consensus on mortality end-points to be used and their potential pitfalls.

In summary we have produced a risk prediction model based on English administrative data, which comprises a small list of pronostic variables (max 10) and allows accurate prediction for early and up to one year mortality following cardiac surgery. Our risk model's performance is comparable to that of commonest ~~most~~ risk scoring systems used. Investing in improving the accuracy of national administrative data ~~could~~ may provide a more cost effective platform to underpin quality improvement programmes in ~~several~~ many clinical specialties. Future research is necessary to develop tools to predict more accurately serious complications and morbidity, which may include the use of new predictors including genetic profiling (REF Duke paper), in order to aid correct interventional decisions in an era of multimodality approach.

.

Funding: none

Conflict of Interest: none declared.

References

1. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg.* Jul 1999;16(1):9-13.
2. Head SJ, Holmes DR, Jr., Mack MJ, et al. Risk profile and 3-year outcomes from the SYNTAX percutaneous coronary intervention and coronary artery bypass grafting nested registries. *JACC Cardiovasc Interv.* Jun 2012;5(6):618-625.
3. Bohensky MA, Jolley D, Pilcher DV, Sundararajan V, Evans S, Brand CA. Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. *J Crit Care.* Aug 2012;27(4):422 e411-421.
4. Brinkman S, Abu-Hanna A, van der Veen A, de Jonge E, de Keizer NF. A comparison of the performance of a model based on administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ratios of intensive care units. *Crit Care Med.* Feb 2012;40(2):373-378.
5. Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SL. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation.* Mar 27 2007;115(12):1518-1527.
6. Glance LG, Dick AW, Osler TM, Mukamel DB. Accuracy of hospital report cards based on administrative data. *Health Serv Res.* Aug 2006;41(4 Pt 1):1413-1437.
7. Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *European heart journal.* May 2003;24(9):881-882.
8. Wendt D, Osswald BR, Kayser K, et al. Society of Thoracic Surgeons score is superior to the EuroSCORE determining mortality in high risk patients undergoing isolated aortic valve replacement. *The Annals of thoracic surgery.* Aug 2009;88(2):468-474; discussion 474-465.
9. Dewey TM, Brown D, Ryan WH, Herbert MA, Prince SL, Mack MJ. Reliability of risk algorithms in predicting early and late operative outcomes in high-risk patients undergoing aortic valve replacement. *The Journal of thoracic and cardiovascular surgery.* Jan 2008;135(1):180-187.
10. Dunning J, Gao H, Chambers J, et al. Aortic valve surgery: marked increases in volume and significant decreases in mechanical valve use--an analysis of 41,227 patients over 5 years from the Society for Cardiothoracic Surgery in Great Britain and Ireland National database. *J Thorac Cardiovasc Surg.* Oct 2011;142(4):776-782 e773.
11. Freemantle N, Richardson M, Wood J, et al. Weekend hospitalization and additional risk of death: an analysis of inpatient data. *Journal of the Royal Society of Medicine.* Feb 2012;105(2):74-84.
12. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases.* 1987;40(5):373-383.