Ranking Institutions by Clustered Hospitalization Records

Yan Wang * David Zingmond † Honghu Liu ‡

Abstract

Ranking institutions according to different measures clustered within certain domains is very common in many fields, e.g. Universities are ranked each year based on majors according to faculties' reputation, students' performance, source of funding, etc.; Facilities are ranked based on diseases according to doctors' specialty, patients' complications, outcome of treatments, etc. Often these measures are not independent and are always clustered within certain domains. This problem can be more complicated because the collection of the data and the validity of each measure. The situation can be simplified as that how the weight for each measure is being assigned statistically and how the final score for ranking is calculated by taking the data complications and the institution characteristics into account. We propose the ranking methods based on statistical models (Survey Logistic model and Proportional Hazard model) to rank institutions. Finally, the methods will be applied to the analytic administrative data collected by California licensed hospitals to rank their performance in stroke acute care.

Key Words: Clustered Observation, Survey Logistic Model, Cox Model, Variable Selection, Rank, Calibration

1. Introduction

Ranking institutions is very common in many fields. The measures collected from the institutions are highly correlated within the institutions. For example, Universities are ranked each year based on majors according to faculties' reputation, students' performance, source of funding and many other factors. The faculties in specific University are not completely independent because of their intensive collaboration. The students are clustered within certain department and they share the resources and take the same classes. The measures related to the students and faculties are highly correlated. In this paper, we talk about ranking institutions based on statistical models by taking the nature of clustered observations into account. We rank the California licensed hospitals according to the stroke related outcomes of the acute care episode, based on the analytic administrative data set of California hospitalization records.

Acute ischemic (a blocked artery in the brain) and hemorrhagic stroke (bleeding into the brain caused by a ruptured blood vessel) are responsible for approximately 50,000 hospitalizations annually. The outcomes of stroke are severe. For example, stroke was the third-leading cause of death for Americans and the leading cause of disability in 2005. Office of Statewide Health Planning and Development (OSHPD) contracted with University of California at Los Angeles (UCLA) to develop validated outcome measures for monitoring ischemic stroke care in California hospitals. This is a statewide effort to develop meaningful, valid outcome measures of stroke for patients treated in California general acute care hospitals.

This paper is part of the results from the Stroke Outcomes and Validation Study (SOVS). We create analytic models to predict the stroke mortality and calculate the scores that can

*Fielding School of Public Health, UCLA

†David Geffen School of Medicine, UCLA

‡School of Dentistry, David Geffen School of Medicine, Fielding School of Public Health, UCLA

Table 1: Create Episode of Stroke Care

History	Acute Episode of Stroke Care			Outcomes
Two-year Look back window	Acute ED Encounter for stroke not resulting in admission at this hospital	Index Hospitalization for stroke 1st stroke admit with 90 day look-back	Acute inpatient transfers (any diagnosis) Continuous acute level inpatient care	Post-discharge Outcomes readmission, ED visits, and deaths

be applied for ranking hospitals using Patient Discharge Database (PDD) records. We also identify whether the key outcome measures may affect the hospital ranking.

Ranking hospitals based the treatment of certain outcome, e.g. Stroke, is important, especially for assessment of quality of care and making decision in health policies. Usually, the hospitals are ranked simply based on all-cause mortality, which may be affected by lots of reasons. For example, the complications of the patients, the size of the hospital, and sometimes the distance for the patients have to travel to the hospital may all affect the outcome of the treatment. Another factor affect the result of hospitalization in large medical centers can be that many complicated/severe cases that have higher mortality rate are transferred to these centers. The aim of the paper is to discuss the statistical methods on how to assign an analytic score to each hospital based on the identified statewide cohort of all incident ischemic stroke patients admitted from December 2006 (after the date December 1, 2006) to November 2009 (before the date November 30, 2009) and therefore, we can rank the California licensed hospitals based on stroke related outcomes (In patient mortality, 30-day all cause mortality, 30-day all cause readmission of the stroke records). We will compare the ranking results based on different outcomes and different scenarios.

The administrative data are rich and collected cumulatively over years. There is no missing data because each hospital is required by the State to report all the hospitalization records. The data were collected annually and over the years by OSHPD with the same data structure, although the diagnosis code changed in the PDD data since the third quarter of 2007. We have taken the change into account when we define the complications and comorbidities by the diagnosis codes. As in Table1, we define a two year look back window to check the history of stroke outcomes and are able to compare the annual ranks, the performance of the hospitals. We define the annual window as from last November to December. The data can be treated as both cross-sectional and longitudinal. We have different population of the same data structure over years. Within the data structure, it is possible for the same person to have multiple hospitalization records.

The final analytic data consist of PDD data, which is the main database for the index records, Emergency Department and Ambulatory Surgery Data (EDD), Death Statistical Master Files (DSMF) and Census Data of year 2000. We identify the stroke hospitalization records with acute level of care based on the primary diagnosis code. The primary diagnosis with ICD9-CM codes of 434.xx, 436 and 437.x is defined as the stroke record. These codes are the highest reported accuracy for ischemic stroke. We exclude those who have evidence of prior ischemic or hemorrhagic stroke within 180 days of the stroke admission. We exclude those patients who are younger than 18 years old and who are transferred from within the hospital or from another acute care hospital. The acute episode of care, outcomes and prior stroke history is illustrated in Table1.

The index hospitalization is the acute hospitalization for ischemic stroke that meets the previously defined inclusion and exclusion criteria. The acute episode of care is defined as the index hospitalization plus transfers from an outside emergency department and transfers

to outside hospitals for acute care. The look back window for stroke history and prior admissions include all hospitalizations and ED encounters two years prior to the index hospitalization. The detail description of acute episode of the care is in the published report from OSHPD for the SOVS study (Zingmond, 2012). In this paper, we focus on illustration the ranking method used in the report.

2. Methods

For each ischemic stroke index admission, a longitudinal record is created for the acute care episode, including the index admission, prior admissions, and subsequent outcomes (death and readmission). Because reliable cause-specific outcomes were uncommon, we focus on all-cause outcomes (mortality and readmission) in the modeling of patient outcomes. We model the three binary outcomes inpatient death, 30-day death (from admission), and 30-day readmission (with censoring for death) and one time to event outcome for 30-day death (from admission). Because each PDD record is collected from the hospital and the records are highly correlated in terms of the performance of that hospital, we use survey logistic model for the binary outcomes variable and use proportional hazard (Cox) model for the time to event outcome for clustered observations to predict the probability of the event.

The patient stroke related outcome is a function of patient demographics, stroke severity (neurologic signs, symptoms, and disability due to the acute stroke), co-morbid illness (the measure that predict mortality in the general hospitalized patient population), prior cerebrovascular disease history, patient health habits (smoking and alcohol use), ED transfer and hospital to hospital transfer.

Prior strokes are defined as: (1) ischemic, (2) hemorrhagic, (3) unknown, and (4) total occurrence of temporary ischemic attacks (TIAs), measured during the index hospitalization and prior hospitalizations (within two years of the index hospitalization) and any prior emergency department visits. Patient demographics includes not only the direct measure, such as age, gender, ethnicity and so on, but also the indirect measures of socioeconomic status based upon Census measures and travel distance to closest high volume stroke hospitals (a measure of access to stroke specialty care). These indirect measures are based upon U.S. Census results by zip code that can be linked to the data. Because arrival time to a hospital is critical towards initiating treatment and because treatment in a hospital where there is experience in acute (administering thrombolytic) and post-acute (rehab) stroke care, distance to the closest high volume (or otherwise determined specialty) stroke center is a good proxy for access to specialty stroke care.

In the following, we will explain in detail how we assign scores to hospitals by different models in order to rank these hospitals and classify the hospitals into different categories based on their relative performance of acute stroke care in California. The mortality and readmission have been identified to be useful outcomes related to the performance. The formulas for calculating the scores for each hospital is based on the ratio of observed counts of the outcome events and expected (predicted) counts of the outcome events based on the models, and then multiply the State mean (overall from late 2006 to 2009) as given by (1),

$$\text{Score for Hospital}_i = \frac{\text{Observed Counts}_i}{\text{Expected Counts}_i} \times \text{State Mean} \tag{1}$$

We do not limit the number of digits of the scores to avoid ties. We define the total number of stroke records within each hospital as N_i. The ratio of the two counts of the outcome

events can be further expressed as,

$$\begin{aligned} \frac{\text{Observed Counts}_i}{\text{Expected Counts}_i} &= \frac{\text{Observed Counts}_i/N_i}{\text{Expected Counts}_i/N} \\ &= \frac{\text{Observed Rate}_i}{\widehat{\text{Expect Probability of Event}_i}} \end{aligned} \tag{2}$$

The expected probability of the event (death or readmission) for the hospital is,

$$\widehat{\text{Expect Probability of Event}_i} = \bar{\hat{p}}_{i1} \tag{3}$$

Here $\bar{\hat{p}}_{i1}$ is the mean of the predict probabilities for each individual in hospital$_i$ for the outcome event. The 98% confidence interval (which is usually used in the government report) for the scores is calculated. We assume the observed count of the events to be Poisson distribution, and therefore, the mean and the variance are the same. Given the predicted rate, we can calculate interval range for the rate of the score. Then the classification of the hospitals based on this interval range is as following algorithms, for the hospitals with more than 30 observations,

1. If the lower limits $>$ State mean, we classify the performance of the hospital$_i$ as poor performance.
2. If lower limits $\leq$ State Mean $\leq$ upper limits, we classify the performance of the hospital$_i$ as normal (No better than State Average performance).
3. If the upper limits $<$ State Mean, we classify the performance of the hospital$_i$ as good performance.
4. If the predict count is significantly different from the observed count (based on Paired T test), we classify the hospital$_i$ as outliers.

In Step 4, we use multiple tests for the sample model. We applied the Bonferroni correction (multiple comparison correction adjustment) for the p-values, using the cut-off p-value as α/m to define the outliers, where m is the number of tests. In the data, $m = 375$ is the number of hospitals in the final model. We use the cut-off p-value as 0.00013. In this case, we can check whether the prediction of a hospital is an outlier of the model when we classify the hospital as the poor/good performance.

Based on the score assigned to each hospital, we can rank the hospitals by their performance. The lower score indicates better performance, and therefore, be given a higher rank. Results of these models are compared across a number of measures. Global measures of goodness of fit (R-squared and log-likelihood) and local measures (z-score and p-values for included predictors) are examined across the models. Finally, we examine goodness of fit within models by comparing model performance in predicting actual events using C-Statistic and Calibration risk, allowing for dynamic comparison depending upon cutoff thresholds for classifying predicted probabilities.

The goodness of fit is checked by the Calibration of the risk, which refers to the ability of a risk model to match predicted mortality with observed mortality. We extend the idea to survey logistic model for clustered observations. The data are partitioned into 10 roughly equal size groups by their sorted predicted risks of events (mortality or readmission). Group10 is the highest risk group and group1 is the lowest risk group. If none of the 10 risk groups have either significantly fewer or more death events than the predicted number of events by the model, we conclude the model has good predictive power. The output for the main model and model for the most recent year (2009) is displayed in the

result section to compare the predicted mortality and observed mortality. Good calibration means these two numbers are very close to each other and the observed counts are with the 95% confidence interval of the predicted counts.

The agreement of the ranks is assessed by the scatter plots, correlation between the ranks, the Kappa statistics and weighted Kappa statistics, quantile categorize of the ranks. At data level, we compare the ranks of different set of predictors (whether the symptoms are present at admission, whether the complication is present at any duration of the care episode, whether the symptoms are only present at the index record). Detailed comparison between different set of predictors is in the report (Zingmond, 2012). We compare the ranks by using hospitalization records after December 2007, by using hospitalization records during individual year from 2007 to 2009. We have included a large number of predictors to achieve the great predictive power. We use the stepwise variables selection method for Survey Logistic model (Wang, 2011) to check the agreement of ranking.

All the analysis in the paper is performed by SAS 9.2.

3. Results

After exclusions for age and inpatient transfers preceding patient stroke, the overall state wide stroke cohort between December 2006 and November 2009 totaled 104,918 individual cases. Of these, 2,884 are likely transfers (based upon source of admission, but not evidence of preceding hospitalization in the PDD records) and are excluded based upon criteria used in other inpatient outcomes reporting. The observations used to calculation the score is 102,034 cases. Among eligible patients, 6,098 were seen and transferred from an emergency department at a different hospital, 4,346 were transferred to another acute care hospital after the initial stroke admission, and 305 patients were both transferred from an ED to the stroke hospital and transferred to another hospital after the stroke admission. There are 365 observations that are excluded from the model because of missing census information. As a result, in the model we are left with 101,669 individual stroke cases.

The main outcome is all cause mortality in 30 days. The output of survey logistic model is displayed in Table2. We use 0.05 as the cut-off for NS (not significant). We report the estimate of the parameter, the odds ratio, the confidence interval of the odds ratio and p-value in the table for the mode.

Table 2: Main Model for 30 day all-cause Mortality

Predictors	Estimate	Odds	95% CI		P-value
Intercept	-6.3503				<.0001
Age in Years at Admission	0.0518	1.053	1.050	1.053	<.0001
Male	-0.0205	0.980	0.928	0.980	NS*
African American	-0.7201	0.487	0.435	0.487	<.0001
Latino	-0.4289	0.651	0.600	0.651	<.0001
Asian	-0.5175	0.596	0.524	0.596	<.0001
Other Ethnicity	-0.2837	0.753	0.662	0.753	<.0001
0/1 ED transfer	0.1490	1.161	1.029	1.161	0.0154
0/1 HH transfer	0.1396	1.150	0.971	1.150	NS
0/1 ED + HH transfer	0.2062	1.229	0.795	1.229	NS
Nursing source of Admit	0.5501	1.734	1.590	1.734	<.0001
Other source of Admit	0.1212	1.129	0.920	1.129	NS
Distance to High Volume Hospitals	0.0002	1.000	0.999	1.000	NS

Continued on next page

Table 2 – continued from previous page

Predictors	Estimate	Odds	95% CI		P-value
Distance to Middle Volume Hospitals	0.0011	1.001	0.999	1.001	NS
Distance to Low Volume Hospitals	-0.0015	0.999	0.996	0.999	NS
$\geq$40% Populations are rural dwellers	0.0262	1.027	0.915	1.027	NS
%Adults with $\geq$4years college	-0.0047	0.995	0.993	0.995	0.0006
%Adults with Income<200% Poverty	-0.0032	0.997	0.994	0.997	0.0359
Facial palsy	-0.0997	0.905	0.839	0.905	0.0096
Dysarthria (inability to articulate)	-0.3972	0.672	0.627	0.672	<.0001
Any post-stroke disability	-0.0668	0.935	0.810	0.935	NS
Aphasia	0.2795	1.322	1.246	1.322	<.0001
Hemiplegia/Hemiparesis	0.5611	1.753	1.640	1.753	<.0001
Other paralysis	0.3134	1.368	1.109	1.368	0.0035
Hemineglect	0.2902	1.337	1.038	1.337	0.0243
Vision loss	-0.1898	0.827	0.713	0.827	0.0125
Apraxia	-0.8625	0.422	0.256	0.422	0.0007
Ataxia	-0.8910	0.410	0.356	0.410	<.0001
Decreased consciousness, altered mental status, coma	1.6092	4.999	4.429	4.999	<.0001
Seizure or seizure disorder	0.3147	1.370	1.261	1.370	<.0001
Conjugate deviation of eyes	1.2246	3.403	1.521	3.403	0.0029
Other cerebral ischemic signs or symptoms	-0.1781	0.837	0.698	0.837	NS
Perenteral nutrition	0.5023	1.653	1.172	1.653	0.0042
Dysphagia	0.1644	1.179	1.086	1.179	<.0001
Admission elevated glucose	0.1411	1.151	0.975	1.151	NS
Acute myocardial infarction	0.8083	2.244	1.973	2.244	<.0001
Left sided valvular heart disease	-0.0868	0.917	0.726	0.917	NS
Right sided valvular heart disease	-0.0827	0.921	0.750	0.921	NS
Atrial fibrillation	0.4771	1.611	1.537	1.611	<.0001
Cardiopulmonary arrest	1.7183	5.575	4.948	5.575	<.0001
Systolic heart failure	-0.0733	0.929	0.450	0.929	NS
History of CHF (L heart failure, cardiomyopathy)	0.4358	1.546	1.458	1.546	<.0001
Any Ischemic Heart Disease: CAD, angina, AMI, prior MI	0.0650	1.067	1.010	1.067	0.0216
Hyperlipidemia	-0.4642	0.629	0.600	0.629	<.0001
Dementia or Alzheimers Disease	0.2554	1.291	1.178	1.291	<.0001
Low platelet count	0.0226	1.023	0.521	1.023	NS
Bleeding Disorders (no platelet disorders)	0.5160	1.675	1.231	1.675	0.0010
Anticoagulation	-0.1963	0.822	0.755	0.822	<.0001
Hypercoagulable state	0.4177	1.519	1.080	1.519	0.0164
Falls	-0.0135	0.987	0.882	0.987	NS
Current Smoker	-0.0359	0.965	0.893	0.965	NS
Recurrent Strokes	0.0452	1.046	0.976	1.046	NS
Former TIA	-0.0247	0.976	0.897	0.976	NS
TIA resolved	0.1595	1.173	1.068	1.173	0.0008
Fever 48 h	0.6203	1.860	1.343	1.860	0.0002

Continued on next page

Table 2 – continued from previous page

Predictors	Estimate	Odds	95% CI		P-value
Valvular disease	-0.0945	0.910	0.840	0.910	0.0203
Pulmonary circulation disease	-0.0353	0.965	0.836	0.965	NS
Peripheral vascular disease	0.1959	1.216	1.124	1.216	<.0001
Hypertension	-0.0896	0.914	0.866	0.914	0.0013
Paralysis	0.2537	1.289	1.200	1.289	<.0001
Other neurological disorders	0.0441	1.045	0.952	1.045	NS
Chronic pulmonary disease	0.0432	1.044	0.974	1.044	NS
Diabetes w/o chronic complications	0.1073	1.113	1.051	1.113	0.0003
Diabetes w/ chronic complications	0.0542	1.056	0.960	1.056	NS
Hypothyroidism	-0.1602	0.852	0.798	0.852	<.0001
Renal failure	0.2626	1.300	1.206	1.300	<.0001
Liver disease	0.4081	1.504	1.226	1.504	<.0001
Peptic ulcer Disease x bleeding	0.4557	1.577	0.625	1.577	NS
Acquired immune deficiency syndrome	0.2475	1.281	0.239	1.281	NS
Lymphoma	0.3519	1.422	1.019	1.422	0.0386
Metastatic cancer	1.9593	7.095	6.061	7.095	<.0001
Solid tumor w/out metastasis	0.6631	1.941	1.679	1.941	<.0001
Rheumatoid arthritis/collagen vas	0.0082	1.008	0.854	1.008	NS
Obesity	-0.2142	0.807	0.720	0.807	0.0002
Weight loss	0.2259	1.253	1.093	1.253	0.0012
Fluid and electrolyte disorders	0.2478	1.281	1.206	1.281	<.0001
Chronic blood loss anemia	-0.0088	0.991	0.747	0.991	NS
Deficiency Anemias	-0.0545	0.947	0.884	0.947	NS
Alcohol abuse	-0.0409	0.960	0.831	0.960	NS
Drug abuse	0.2485	1.282	1.019	1.282	0.0338
Psychoses	-0.2461	0.782	0.675	0.782	0.0010
Depression	-0.1059	0.900	0.828	0.900	0.0128

We have a large number of predictors in order to achieve large predictive power. The detail outputs of other models and the reason for selection of these predictors can be found in the final report (Zingmond, 2012). According to the classification by 98% confidence interval method, the scores calculated by the all cause 30-day mortality classified the 309 California licensed hospital ($\geq$ 30 stroke index records) into three different classes. There are 10 hospitals (3.24%) being classified as poor performance and none of these 10 hospitals are outliers for the model prediction. There are 28 hospitals (9.06%) being classified as good performance and 15 of these 28 good performance hospitals are classified as the outliers for the model. The rest of the hospitals are classified as normal performance (87.70%) with no outliers found.

Based on the scores assigned by the survey logistics model, we rank the hospitals with at least thirty or more observations. The rank will be higher if the score is lower, i.e. the lower score indicate the better performance. We assess the variation of the ranks by different stroke related outcomes by the correlation among the ranks. The correlations of the ranks by other stroke outcomes (In patient 30-day mortality, all cause 30 day readmission, 30 day mortality exclude the inpatient death and using the Cox model for 30 day mortality) and main outcome (the all cause 30-day mortality) are shown in Table3.

Table 3: Correlation of ranks between 30-day mortality and other Outcomes

30-day all cause mortality	Inpatient mortality	30-day readmission	30-day mortality (COX)
	0.6840 <0.0001 309	-0.1276 0.0251 308	0.4075 <0.0001 308

Table 4: 30-Day Mortality versus Inpatient Mortality

30D Mortality	Q 1	Q 2	Q 3	Q 4	Q 5	Total
Q 1	**33**	16	9	2	1	61
Q 2	13	22	16	8	3	62
Q 3	8	16	15	19	4	62
Q 4	6	7	12	21	16	62
Q 5	1	1	10	12	**38**	62
Total	61	62	62	62	62	309

Except for the rank by 30-day any reason readmission, the other two outcomes are highly correlated with the rank by 30-day all cause mortality and are statistically significant (p-values < 0.0001). The 30-day any cause readmission provide an alternative way to assess the performance of the acute care. The three categorical classification method based on 98% confidence interval method classifies 10 of the 308 hospitals with 2 outliers as good performance and 1 hospital with no outlier as poor performance. All these 11 hospitals are classified as normal performance when using 30-day mortality as the outcome.

The high correlation between the ranks for inpatient mortality and all-cause mortality shows strong agreement between these two outcome measures. The five equal quantile classification between the two ranks in Table4 explores more detail about the two measures. The clustering around the diagonal with the greatest agreement occurring at the highest quantile and the lowest quantile ($\kappa = 0.27$ and $\kappa_w = 0.48$). The weighted κ statistics indicates strong agreement between the two measures.

The scatter plot of the two rank measures in Figure3 illustrates the trend of agreement. The trend along the positive diagonal is preferred. In this situation, the two measures classify certain hospitals poor or good performance together at the same time.

The model diagnosis is performed among the model level and among the data level. Among the different models, we compared the C-statistics (area under the curve) between

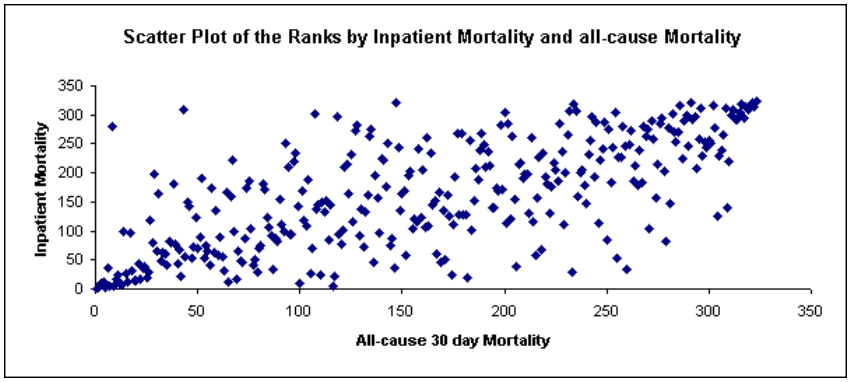

Figure 1: Scatter plot of 30-Day mortality versus inpatient death

Table 5: Calibration of risk model for 30-day mortality

group	N	Observed	Predicted	Difference	95% CI of predicted
1	10202	108	78.11	-29.89	(63.63,96.50)
2	10203	128	164.80	36.80	(136.00,200.77)
3	10203	198	256.17	58.17	(212.59,310.11)
4	10203	276	371.42	95.42	(309.37,447.45)
5	10203	449	521.19	72.19	(435.76,625.15)
6	10203	663	723.35	60.35	(606.78,864.49)
7	10203	949	1003.49	54.49	(844.19,1193.93)
8	10203	1591	1434.77	-156.23	(1210.26,1699.17)
9	10203	2537	2204.57	-332.43	(1872.25,2583.69)
10	10208	4521	4624.68	103.68	(4096.55,5162.66)
	102034	11420	11382.6	-37.44	

the models. We also compare the calibration risk and Pseudo R-square for the survey logistic models. The main outcome is the all cause 30-day mortality (11,382 of 101,669 observations has a death outcome within 30 days of admission) with Pseudo R-square as 0.2791 and C-statistics as 0.831. The Calibration of risk model is in Table5.

There are 5,736 inpatient death events. The model diagnosis statistics are Pseudo R-square as 0.2434 and C-statistics as 0.834. There are 12,881 records with a result of all cause readmission. The model diagnosis statistics for modeling readmission are Pseudo R-square as 0.0433 and C-statistics as 0.629. Compare with the model using outcome as all cause mortality but exclude those inpatient death events, the Pseudo R-square as 0.2692 and C-statistics as 0.825 for the Cox model. The detailed outputs of these models are in the report (Zingmond, 2012).

We build up the proportional hazard Cox model to further comparing the ranks, if using more detailed measures. Instead using the binary outcome for the event, we use the time to event outcome variable with censored observations at 30 days after admission date to compare with the all cause 30 day mortality binary outcome. The C-statistics is being calculated by simple random sample 3% of the data (4000 records), using the method as (Liu, 2009) mentioned in the paper. The C-statistics of the Cox model is 0.040 with confidence interval $(0.028, 0.053)$. It is much lower than the one from Survey Logistic Model.

As noticed in Table2, we have a large number of predictors, many of which are not statistically significant. We use the stepwise selection of the survey logistic model (with option of select in and select out value 0.2) to verify whether the reduced model can attain the similar diagnosis statistics as the full model. The C-statistics for the reduced model is 0.827 and the Pseudo R-square is 0.2706.

Besides compare the model diagnosis statistics, we check the model generalizability by limiting years range to the most recent years (Using records with admission date after December 2007) to compare the difference between the models, the C-statistics is 0.829 overall (C-statistics is 0.826 for each individual year). For the most recent one year (from December 2008 to November 2009), the Calibration of risk model is in Table6.

In addition, we consider giving credit to the transferred hospitals. If there are two PDD records (the index hospitalization records), the length of stay in the first record is less than 2 (< 2) days, and the stay in the second record is greater than one (≥ 1) day, we give the credit (episode of care) to the second record. If there are three or more records, we right now only consider the first two PDD records. The C-statistics for this type of scenario is 0.825, very close to the model in Table2.

Table 6: Calibration of risk model for 30-day mortality in 2009

group	N	Observed	Predicted	Difference	95% CI of predicted
1	3373	40	26.96	-13.04	(19.43,38.00)
2	3374	43	56.95	13.95	(42.02,78.29)
3	3374	77	89.02	12.02	(66.36,120.80)
4	3374	96	128.61	32.61	(96.62,172.97)
5	3374	170	179.51	9.51	(135.47,239.79)
6	3374	211	248.20	37.20	(188.42,329.03)
7	3374	335	341.15	6.15	(259.44,449.50)
8	3374	546	483.37	-62.63	(370.06,629.30)
9	3374	820	739.07	-80.93	(568.45,949.59)
10	3375	1506	1536.16	30.16	(1263.63,1816.21)
	33740	3844	3829.00	-15.00	

4. Discussion

Ischemic stroke is a common, severe acute illness with high short-term mortality and prolonged recovery. Thus reporting on stroke outcomes in California hospitals will reflect care delivered in most California hospitals. Although not a random sample of hospitals, it may prove to be a cost effective approach and could serve as a template for a new approach to outcomes reporting by OSHPD. In the future, we may expand the scope of the outcomes reporting program to include measurement of clinical care - processes of care and their indications - and the ability to audit such measures.

The ranking methods used in this paper to rank the California hospitals are based on the hospitalization records from OSHPD administrative data, clustered within hospitals. The non-missing rich data source enables the great prediction power of the model. The main stroke related outcome is the all cause 30 day mortality of the stroke cases from late 2006 to November 2009. The measures in the data have been valid in another part of the study in Chart review. The overall all cause 30 day mortality outcome has overcome the potential bias in the inpatient death. An alternative measure we considered as to measure the performance of the hospitals is the 30-day readmission. The rank based on the model for the readmission measure is negatively associated with the rank based on the model for the mortality in 30 days with Pearson Correlation Coefficient -0.1276 and it is statistically significant (p-value is 0.0251) in Table3. The two outcome measures classify the performance of care on different perspectives.

The statistical methods used in this paper can be easily extended to much broader usage. For example, we can sample the records of all the hospitals within the United States, with the primary sampling unit (PSU) as the States and the Hospitals as the stratum to rank the national hospitals. Based on a specific disease related outcomes, for example, stroke in the paper, we use all cause 30-day mortality of stroke records as the outcome. We can assign the scores based on the predicted mortality rate of survey logistic model to each hospital. Based on the scores, we rank these hospitals in terms of their performance on acute stroke care.

The 365 records excluded from the model, is because when matching the zip code information of the patients to the previous census data of 2000 for socioeconomic status of each patient. The missing information of socioeconomic status can be treated as missing at random when build up the models. As a result, this does not affect the ranking. However, a type of missing problem ignored in the paper is the records being dropped because the missing of social security number (SSN), which is used to create the acute episode of care

and to link the records between data sets, for example when we link the EDD data to PDD records and link the PDD records to DSMF data. This type of missing is not missing at random, especially in California. We need a better resource in health policy to track those patients without SSN.

When we calculate the scores, we assume the number of events in the hospitals follow Poisson distribution with different parameters. This is an extremely strong assumption, based on which we calculate the confidence intervals and classify the performance of the hospitals into three different categories, poor, normal and good. In the future, we may develop more complex methods to find the interval range and weaken the distribution assumptions.

5. Acknowledgment and Reference

We would like to thank Rachel Louie for part of the data management work.

REFERENCES

Zingmond, D. etc.(2012), "Ischemic Stroke Outcomes Validation Study in California, 2006–2009," *California Office of Statewide Health Planning and Development*.

Robert, B., etc.(2012), "The California Report on Coronary Artery Bypass Graft Surgery 2009 Hospital Data California CABG Outcomes Reporting Program," *California Office of Statewide Health Planning and Development*.

Liu, L., etc.(2009), Fitting Cox Model Using PROC PHREG and Beyond in SAS.*SAS Global Forum 2009 Statistics and Data Analysis*

Wang, F. etc. (2011), Model Selection Macros for Complex Survey Data Using PROC SURVEYLOGISTIC/SURVEYREG. *Midwest SAS? Users Group (MWSUG) 2011 Proceedings*.