Impact of Diabetics' Demographic Characteristic and Hospital Treatment on Readmission Rates

Chen, Shupeng

University of Toronto

1. Introduction

For the past few years, Canadian government agencies and medical system have paid more and more attention to the 30-day readmission rate to determine the complexity of their patient population and improve treatment quality. Diabetes, like other chronic diseases, is associated with an increased risk of readmission. Research shows that a large proportion of hospitalization costs are incurred by a small number of patients, particularly those with chronic medical conditions.[1] However, diabetics are not treated as they should be, investigations have shown that inpatient management is arbitrary, or results in no treatment at all or widespread fluctuations in glucose.[2] Therefore, the purposes of this study is to examine the association among diabetics' demographic characteristic, hospital treatment and readmission rates. This is important because by analyzing the characteristic of diabetic patients and their hospitalization, it can help to predict the probability of readmission and thus improve patient outcomes and lower cost of inpatient care.

Chen, Shupeng
1004465321

2. Methods

<u>Choice of Methods :</u> Since we are primarily interested in factors that lead to early readmission, multivariable logistic regression was used to predict the probability of a patient being readmitted. Notice that the number of observations is greater than the number of patients, so to ensure the independence of observation, I removed the duplicated data by choosing the first encounter's record. In order to assess whether the candidate covariates were significantly associated with readmission, I used the logistic model to fit the relationship between readmission and one covariate at a time while controlling for other covariates such as gender, severity of the disease, and type of admission. Lastly, I dichotomized the response variable in to two categories "No Readmission" and " <30" for the convenience of logistic regression model

<u>Variable Selection :</u> From the given database, there are 48 features describing the diabetic encounters. However, some information is incomplete and redundant. These include"weight" ," payer code","medical specialty" and "age".(see table 1). So, weight and payer code were removed because missing too many values , medical specialty and age were kept because they are significant factors in predicting readmission possibility. Variables like "encounter_id" "Examide" and "Citogliption" were removed because their outcomes are not helpful towards predicting readmission. Variables "Glucose serum test" and "A1C test" result are usually highly correlated, so we just pick one of them. In our dataset, A1C test has more taken. So, we keep A1Ctest and remove glucose serum test. On the other hand , it is necessary to combine several similar variables into one variable. They are the following: "number_outpatient_" , "number_inpatient" and "number_emergency" were merged into one variable that is more representative. "number_of_visit", "num_lab_procedures" , "num_procedures" and "num_medications" were merged into variable "num_of_service" which is more concise. Variables such as “Metformin” and the other twenty-two similar medication variable were merged into “ HbA1C”since we only care about the impact of medication change on A1C result[3]. Additionally, medical specialty categories are grouped based on Wikipedia. After regrouping variable age[4] and race, model shows patients who are Caucasian age between 30 – 60 and has higher possibility of being readmitted. So, age and race have significant influence on readmission. In addition , I regrouped cases cover less than 10% of encounters were grouped into “other” category. For example, “discharge_disposition_id” was regrouped into ‘home’ and ‘other’ 2 categories because over half of the patient were discharged to home and their possibility of readmission are lower than those who were discharged to other places. Length of stay was also kept because patients who stay in hospital longer have higher possibility of being readmitted. “Admision_type_id” is also important as the encounters who were admitted Emergency have higher possibility of being readmitted. At last , values like ‘NA’ ‘unknown’and any result that is not helpful to the result were removed. For example , in variable “discharge_disposition_id”categories ‘11’ ‘13’ ‘14’, ‘19’, ‘20’, ‘21’ were removed because patients are related to death or hospice and they cannot be readmitted[5]. Therefore , covariates : race, age ,discharge disposition, length of stay, admission type id , number of diagnoses , number of visits , number of services, HbA1C[6] and medical specialty were selected.

Variable	Type	Description	%missing
Race	Nominal	Caucasian, Asian, African American ,Hispanic, and other	2%
Weight	Numeric	Weight in pounds	97%
Payer code	Nominal	Integer identifier	40%
Medical Specialty	Nominal	Values	49%

Table 1 : List of variables that are missing a large percent of data from given data set

Chen, Shupeng
1004465321

Model Violations / Diagnostics : Since we assumed independence of observations, we removed duplicated data which could result in deleting helpful data value. Also, response variable originally has 3 categories, but I divided into 2 categories which could lead to inaccurate result. In order to check violations, we can first use residual check to test if the model follows normal distribution or if they are independent. We can also use AUC plot to determine the difference between predict value and observed value. Lastly, in order to test prediction accuracy , we divided dataset into 2 parts train and test set. We then create a test set which contains a random selection of 20000 patients and use this training set to fit the model and compare with the true value in the test set.

3. Result

Description of Data : As shown in table 2, ,measurement of HbA1C was infrequent. However, 9.1 percent of the patients who did not get HBA1C test were readmitted to the hospital, whereas the rate of readmission was lower for those who did, and only 7.3 percent for those who had higher HBA1C test results and no change in diabetic medication .That is, with respect to readmission and taken as a whole without adjusting for covariates, measurement of HbA1C was associated with a significantly reduced rate of readmission. Another significant difference is discharge disposition, only 7 % patients who were discharged to home were readmitted whereas 12.5 % of patients who were discharged to other places were readmitted.

			Readmitted	
Variable	#of encounters	%of population	# of encounters	%of population
HbA1C				
Result high, diabetic medication changed	4058	5.8%	348	8.5%
Result high, diabetic medication not changed	2181	3.1%	161	7.3%
Normal result	6606	9.4%	569	8.6%
Not measured	57125	81.6%	5199	9.1%
Gender				
Female	37229	5.3%	3361	9%

Variable	#of encounters	%of population	Readmitted # of encounters	%of population
Male	31741	4.6%	2916	8.9%
Discharge Disposition id				
Home	44315	6.3%	3078	6.9%
Other	25655	3.6%	3199	12.5%
Admission type id	13784	0.197	1141	0.0828
Elective	35464	0.507	3252	0.0917
Emergency	7920	0.113	736	0.0929
Other	12802	0.183	1148	0.0897
Urgent				
Medical Specialty				
General	4978	0.0711	485	0.0974
Internal Medicine	17691	0.253	1642	0.0928
Missing	33637	0.481	3104	0.0923
Other	7597	0.109	563	0.0741
Surgery	6067	0.0867	483	0.0796
Race				
African American	12625	0.180	1093	0.0866
Caucasian	52292	0.747	4801	0.0918
Missing	1916	0.0274	141	0.0736
Other	3137	0.0448	242	0.0771
Age				
30-60	21869	0.313	1573	0.0719
60-100	46293	0.662	4592	0.0992
< 30	1808	0.0258	242	0.0619

Following are numerical data

Variable	Min	1^{st} Qu	Median	Mean	3^{rd} Qu	Max
Length of Stay	1	2	3	4.27	6	14
Num of Diagnose	1	6	8	7.22	9	16
Num of Services	2	45	60	59.97	76	170
Num of Visits	0	0	0	0.56	1	49

Table 2： Distribution of selected variables values and readmissions

Process of obtaining Final Model : we first remove all the duplicated data and then randomly choose 20,000 patients and divide them into train and test 2 sets in which training set contains the patients who are in the randomly chosen set. After that, we fit a model with 10 covariates.(see table 3)

Intercept	Estimate	P value
Race		
African American	Reference	
Caucasian	-0.0433454	0.306
Missing	-0.2642516	0.018
Other	-0.1577238	0.082
Discharge		
Home	Reference	
Other	0.5509398	2e-16
Age		
30-60	Reference	
60-100	0.2026782	9.83e-08
<30	0.0538414	0.6637
Length of Stay	0.0153653	0.0107
Medical Specialty		
Internal Medicine	0.0456415	0.4932
Missing	-0.0444054	0.4817
Other	-0.1408539	0.0748
Surgery	-0.159469	0.0624
General	Reference	
Num of Diagnoses	0.0283474	0.0019
Num of Services	0.0026384	0.0006
Num of Visits	0.0902807	2e-16
HbA1C		
Result high, changed	Reference	
Result high, not changed	-0.0861975	0.4714
Normal result	-0.0588107	0.4964
Not measured	0.0701116	0.3268

Table 3 : Coefficients of covariates estimated from the logistic regression model.

<u>Goodness of Final Model</u> : Regression diagnostics are useful in checking model assumptions and in identifying any unusual points. We can compute residuals as a difference between observed and fitted values. Unfortunately, the plot is not particularly useful. We will now construct a binned residual plot. First, we add the residuals and linear predictor to the data frame. Then create the bins, compute the mean of the residuals and linear predictors in each bin. The deviance residuals are not constrained to have mean zero, so the mean level of the plot is not of interest. We observe an even variation as the linear predictor and fitted values vary — thus the plots do not detect any inadequacies in the model, except some outliers at the right tail. We can also plot the binned residuals against the predictors. For example, we can group by Length of stay. we conclude that if length of stay is not too large, the longer stay, the higher readmission probability. However, this is not always the case. Especially, when the length of stay is close to 2 weeks. we assume those patients are well treated in their current encounter, then they are less likely to have an early readmission in the next 30 days. We can also detect unusual observations by examining the leverages. we can use a half-normal plot for this. We can identify those two outlying points[7]. These two individuals have relatively higher number of services and number of visits, and both don't have HbA1c test measured given the relatively large size of the dataset and the fact that these points are not particularly extreme, there is no need to be concerned.

In order to check goodness of fit , we first divide the observations up into bins based on the linear predictor. We then take the mean response and mean predicted probability within each bin and plot the observed proportions against the predicted probabilities. For a well-calibrated prediction model, the observed proportions and predicted probabilities should be close. Although there is some variation, there is no consistent deviation from what is expected. We have computed approximate 95% confidence intervals using the binomial variance. The line passes through most of these intervals confirming that the variation from the expected is not excessive. We will now compute the test statistic and p–value for the Hosmer-Lemeshow test(see table 4). This test formalizes the procedure above. Since the p–value is greater than 0.05,we detect no lack of fit. For relatively small (but non-significant) p–values, it might be worth experimenting with different numbers of bins in order to see if the test ever becomes significant. Now let us move on to sensitivity and specificity. The ROC curve perhaps the best way to assess goodness of fit for binary response models[8].

		P-value
127.3577	100.0000	0.0247

Table4 : test statistic and p–value for the Hosmer-Lemeshow test

Note: the p–value is above a chosen threshold — for example, well above 0.05 — then the test does not detect any lack of fit This statistic has an approximate $\chi 2$ distribution with J−2 degrees of freedom.

4. Discussion

Final Model Interpretation and Importance : In conclusion, the decision to obtain a measurement of HbA1C for patients is a useful predictor of readmission rates. For instance, our analysis showed that the profile of readmission differed significantly in patients where HbA1C was checked in the setting of a primary diabetes diagnosis, when compared to those who don't . Additionally , patient ages from 30 to 60 years old have higher possibility of being readmitted. These two results perfectly accomplish the goal of this article. That said, in order to improve treatment outcomes and reduce rehospitalization costs for diabetics, hospitals should do more HbA1C tests and pay more attention to diabetics between the ages of 30 and 60

Limitation of Analysis : I recognize that the results of the current analysis are preliminary observations of the limitations inherent in such a large health record. In addition to the limitations of using large clinical data sets, this study was limited by the design of non-randomized studies. Moreover, in this GLM model, a large amount of data was removed to maintain independence which results in a less predictable model. However, the results of this study should be taken seriously in the detection of glycated hemoglobin in diabetic patients aged 30 ~ 60 years.

Chen, Shupeng
1004465321

Reference

1. Johansen H, Nair C, Bond J. Who goes to the hospital? An investigation of high users of hospital days. Health Rep. 1994;6(2):253–277d

2. S.E.Siegelaar ,Hoekstra, and J. H. Devries, “Special considerations for the diabetic patient in the ICU; targets for treatment and risks of hypoglycaemia,” Best Practice and Research: Clinical Endocrinology and Metabolism, vol. 25, no. 5, pp. 825–834, 2011

3. Beata Strack. Impact of HbA1c Measurement on Hospital Readmission Rates.
Volume 2014 |Article ID 781670

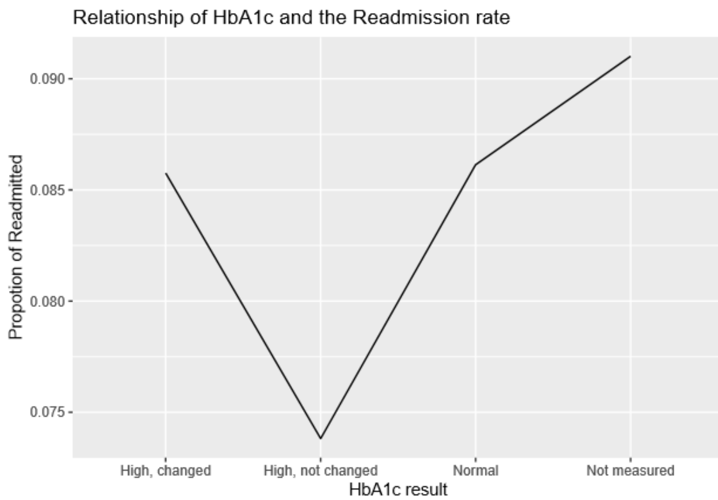
4. This preliminary plot was the motivation to divide the age variable into three categories, <30,[30,60) and 60+.

.

5. Andrew Long. Using Machine Learning to Predict Hospital Readmission for Patients with Diabetes. Oct 21, 2018

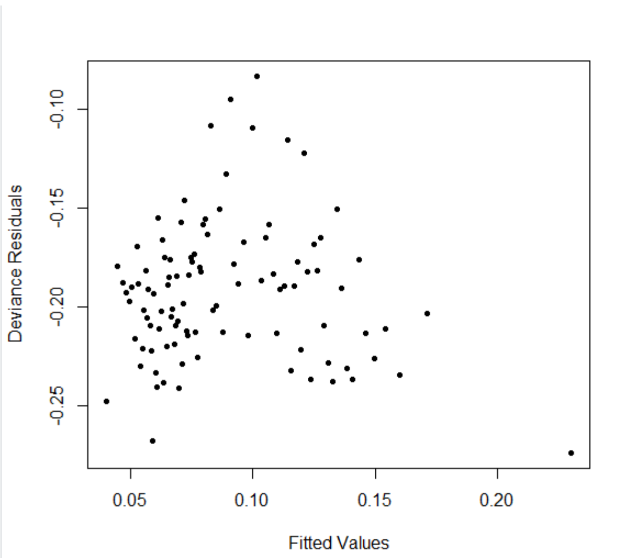
6 This plot was the motivation to divide HbA1C into 4 categories

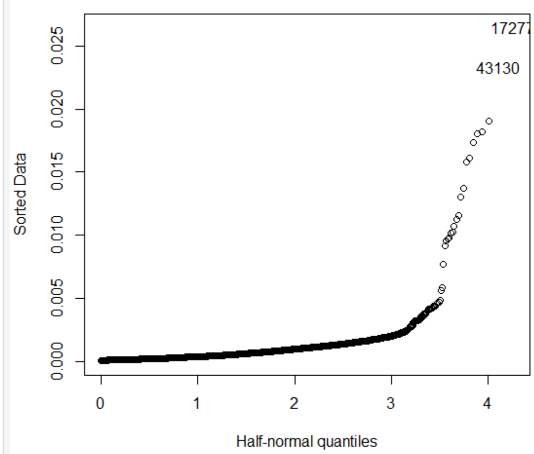
Chen, Shupeng
1004465321



.

7. outliers at the right tail and unusual observations





.8. correspoonding AUC curve

Chen, Shupeng
1004465321