Effects of Covariates on Diabetes Readmission

Terence Chan

August 30, 2020

STA303H1

Introduction

Diabetes is a common disease amongst the human population. Millions of individuals around the world are affected by this disease, in which, the severity of the disease ranges for all affected individuals. This purpose of this study is to identify the factors that causes an affected individual to be readmitted into the hospital due to diabetes. In determining the factors, it will provide health care professionals with the necessary knowledge on which groups of individuals are likely to be readmitted and most importantly, it will allow the professionals to create treatment and prevention methods so that the heavily impacted groups will not be as susceptible to diabetes as before.

Methods

The model chosen for this study was a generalized linear model (GLM). This model was chosen because the error distribution in the data was not guaranteed to be a normal distribution. Moreover, a GLM model was more suitable for a large count dataset such that the response variable was made to be a binary variable, i.e. 0 or 1 for no or yes readmission respectively.

Several assumptions were made in this study. Firstly, the response variable, the readmission variable was to be changed to 'yes' for both <30 days and >30 days of readmission while the 'no' outcome would remain unchanged. This was done to satisfy the purpose of the study which focused on the chance of readmission rather than the length of each readmission. Secondly, since the data included an age group rather than an exact age, the midpoint of each age group was taken, i.e. age 5 for ages 0-10, age 15 for ages 10-20, age 25 for ages 20-30 etc. Furthermore, to focus on whether the medication was prescribed or not, the dosage level of each medication was changed to 'yes' for outcomes 'up', 'down' and 'steady', while the 'no' outcome remained. This was applied to metformin, rosiglitazone and insulin. Finally, the variables with large percentages of NA outcomes were removed entirely, this included weight which had 97% NA outcomes and payer code which had 40%.

Data cleaning was further conducted on some variables for simplicity during the analysis. In the 'change' variables, 'no' became 0 and 'Ch' became 1. In the 'diabetesMed' variable, 'no' became 0 and 'yes' became 1. For the 'gender' variable, 'female' became 1 and 'male' became 2. Finally, for the 'race' variable, 'Caucasian' became 1, 'African American' became 2, 'Asian' became 3, 'Hispanic' became 4, 'other' became 5 and 'NA' became 0.

There were initially eleven covariates chosen for the model. This included 'race' because the susceptibility rate of different ethnicities varies (Ahmed 788). Moreover, there can be different rates within the same racial group, such as South Asians having a lower rate than Chinese individuals (Karter 576). The next covariates chosen were gender and age because different genders at various age groups can have a wide range of susceptibility rates (Ahmed 788). Next, 3 medications were chosen. Metformin and insulin were used as covariates because these medications are common first choice medications prescribed to patients (Landon 471). Furthermore, rosiglitazone was chosen as a covariate as there is evidence that it is a strong medication in maintaining glucose levels (Lipska 317). However, the variable 'metformin. rosiglitazone' was not used because only 2 patients were prescribed it. The covariates

‘diabetesMed’, ‘num_medications’ and ‘change’ were chosen as some patients may be prescribed different medications because some medications may not be effective for all patients, while some patients may not have gotten any prescribed medication. Finally, the covariates ‘number_diagnoses’ and ‘length of stay’ were chosen because patients may need to be monitored for longer periods of time as each patient may have a differing number of diagnoses during their admission.

There were initially six potential models. All six models contained all covariates except for different combinations of the medications because not all patients simultaneously had all 3 medications prescribed, if at all. To select the best variable, stepwise AIC and BIC were used. To determine the best model, cross validation, ROC curves and AUC values were computed. Finally, amongst the models chosen from the previous steps, a final fit against the test data was used to select the best model.

Results

After the 6 models were computed, model 2 and model 3 were chosen as those models produced the lowest AIC value of 138200. Both models initially contained the variables race, gender, age, number medications, number diagnoses, diabetesMed, change and length of stay. However, model 2 contained metformin and insulin whereas model 3 contained metformin only. After stepwise BIC was computed, it was concluded that the all the variables in both models would remain except for ‘num_medications’, therefore, this variable was removed from both models. Next, through a fitted cross validation model, it was shown from the plots that model 2 (Figure 1) was closer to the ideal model compared to model 3 (Figure 2), furthermore, model 2 had a smaller mean absolute error than model 3 which meant that model 2 had a stronger ability to predict the true value.

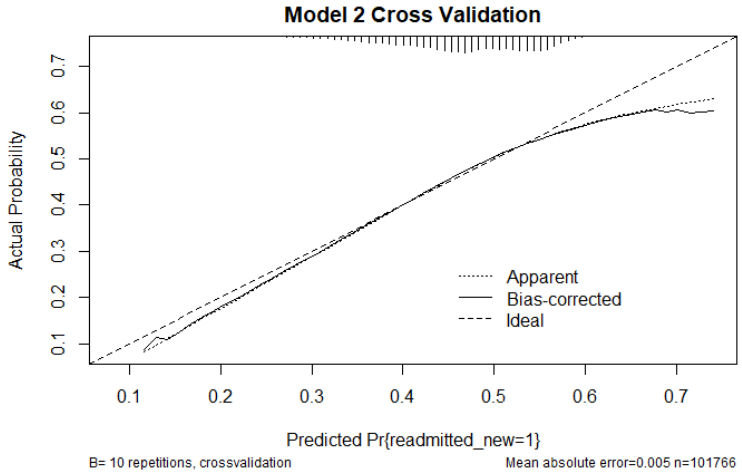

Figure 1

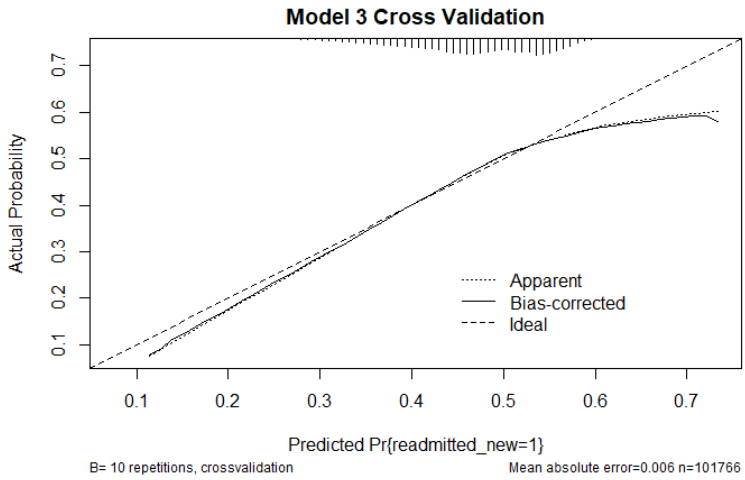

Figure 2

The ROC curves and AUC values were computed to further determine the best model amongst model 2 and model 3 (Appendix Figure 1.1 and Figure 1.2). Since both models produced an AUC value of 0.58, not only did both models produce the same value, it also showed that both models did not have the strongest ability in discriminating between yes and no readmission. Therefore, both models were fitted against the test data. Based on the plots, model 2 (Figure 3) had a model that was closer to the ideal model compared to model 3 (Figure 4).

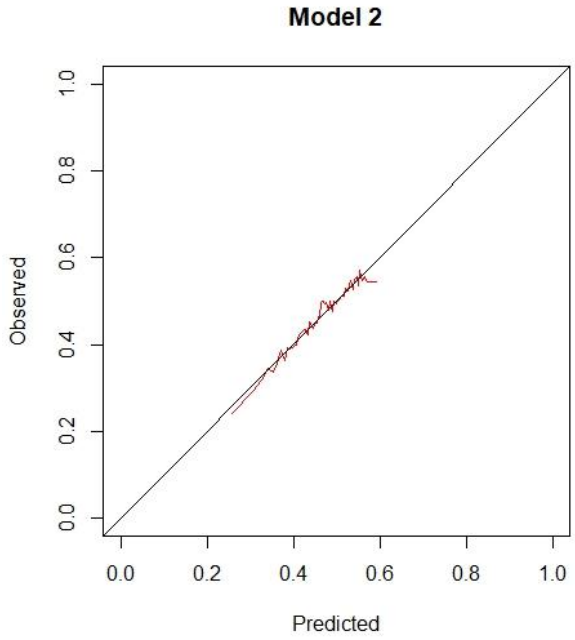

Figure 3

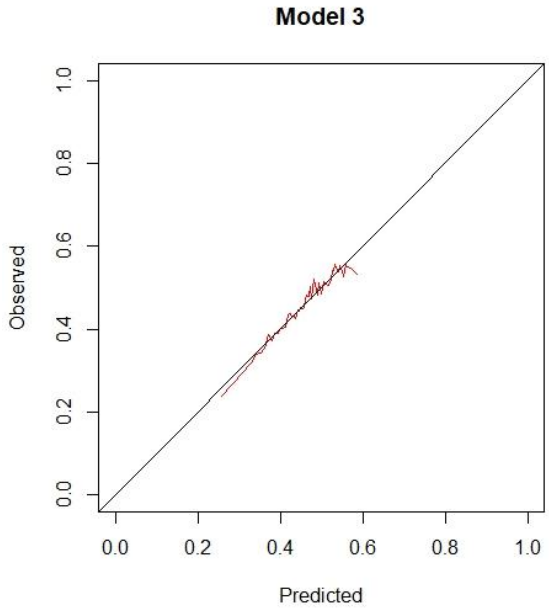

Figure 4

Overall, the final model chosen was new.model 2, which included the covariates: race, gender, age, number diagnoses, diabetesMed, change, length of stay, metformin and insulin (Figure 5). These variables were chosen by stepwise AIC and BIC which produced a value of 138200. Although the model has a weak discrimination ability due to a small AUC value of 0.58, there are no large deviations from the ideal model and it has a small mean absolute error.

```
Call:  glm(formula = readmitted_new ~ race_new + gender_new + age_new +
    number_diagnoses_new + diabetesMed_new + change_new + Length.of.Stay_new +
    metformin_new + insulin_new, family = "binomial", data = train.data)

Coefficients:
              (Intercept)                 race_new1                 race_new2                 race_new3
                -1.7188468                 0.5646571                 0.5577112                 0.1315960
                 race_new4                 race_new5                gender_new2 gender_newUnknown/Invalid
                 0.4363214                 0.2708395                -0.0770258                -8.8693389
                 age_new25                 age_new35                 age_new45                  age_new5
                 0.0911320                -0.0329012                -0.0094462                -0.8795825
                 age_new55                 age_new65                 age_new75                 age_new85
                -0.0676289                -0.0125896                 0.0343423                -0.0002311
                 age_new95       number_diagnoses_new          diabetesMed_new1                change_new1
                -0.3392368                 0.1054569                 0.3399751                 0.1284352
       Length.of.Stay_new          metformin_newYes            insulin_newYes
                 0.0149843                -0.2626963                -0.0934165

Degrees of Freedom: 101765 Total (i.e. Null);  101743 Residual
Null Deviance:      140500
Residual Deviance: 138100       AIC: 138200
```

Figure 5 - Final Model

Discussion

The final model contained the covariates: race, gender, age, number diagnoses, diabetesMed, change, length of stay, metformin and insulin. This model satisfies the purpose of this study because it shows that different racial groups, age, and gender have different chances of being susceptible to readmission. Furthermore, it shows that the use of prescribed medications including the potential use of multiple medications can have an affect on the patient's diabetic health. Moverover, the length of their admission stay has a different affect on each patient as each individual has a chance of having multiple diagnoses per admission, therefore, the varying lengths of admission can affect the period of time that a patient is monitored by health care professionals which can contribute to the stability of the patient's health.

Some limitations that prevented a more accurate prediction included missing data. The data for weight was missing which created inaccuracies because obesity is common amongst patients with diabetes (Maruther 747). Although heavy weight is a key factor for diabetes, weight could not be included due to a lack of evidence. Moreover, there was no exact age given, therefore it was difficult to determine the specific age groups that are most likely to be readmitted. There was also an unequal number of racial groups and gender, i.e. more Caucasian individuals and more females. An equal number would be ideal in determining the specific racial group and gender that is likely for readmission. Finally, there were 72 different medical specialties that initially encountered the patients such as Cardiology, Dentistry and Plastic Surgery. The data would potentially be more accurate if one or a smaller number of specialties admitted the patients as that would allow for a more specialized diagnoses as supposed to a general wide range of diagnoses from various disciplines.

Appendix

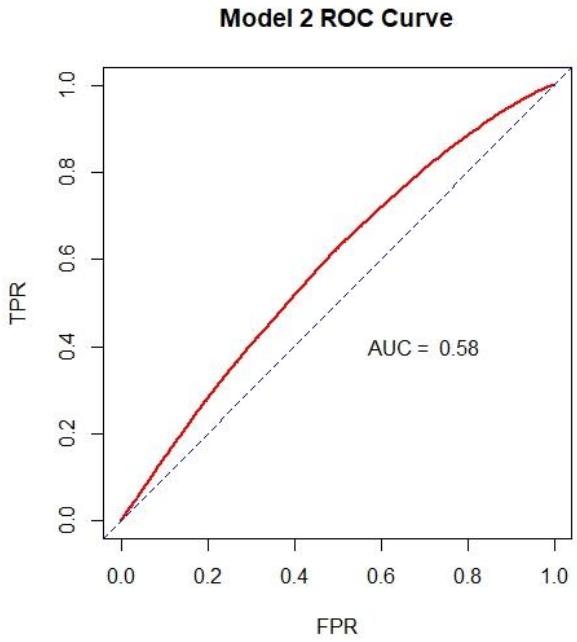

Figure 1. 1

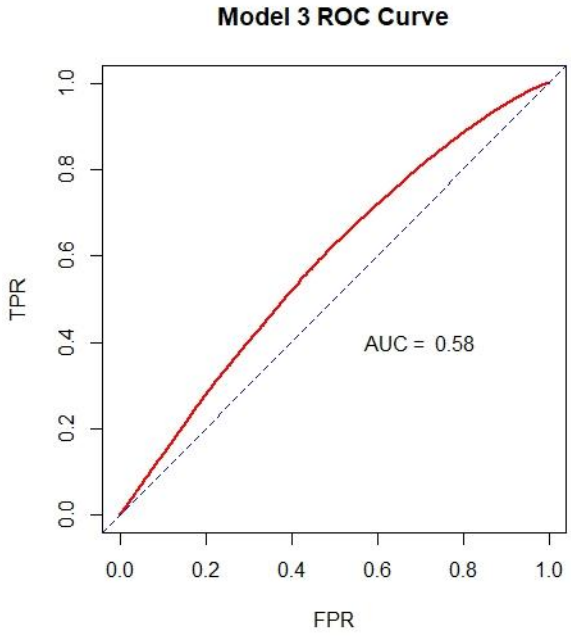

Figure 1. 2

The ROC curves show that the AUC values are equal, therefore, a further model selection test against the test data was done to determine that model 2 is the best model.

References

Ahmed, Shazhan, et al. “Type 2 Diabetes, Medication-Induced Diabetes, and Monogenic Diabetes in Canadian Children.” *Diabetes Care,* American Diabetes Association, Apr. 2010, care-diabetesjournals-org.myaccess.library.utoronto.ca/content/33/4/786.

Karter, Andrew J, et al. “Elevated Rates of Diabetes in Pacific Islanders and Asian Subgroups.” *Diabetes Care,* American Diabetes Association, Mar. 2013, care-diabetesjournals-org.myaccess.library.utoronto.ca/content/36/3/574.

Landon, Bruce E, et al. “Trends in Diabetes Treatment and Monitoring among Medicare Beneficiaries.” *Journal of General Internal Medicine*, Springer US, Feb. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC5880782/?tool=pmcentrez&report=abstract.

Lipska, Kasia J, and Harlan M Krumholz. “Comparing Diabetes Medications- Where do we Set the Bar?” *Jama Internal Medicine,* Jama Network, Mar. 2014, jamanetwork-com.myaccess.library.utoronto.ca/journals/jamainternalmedicine/fullarticle/1790998.

Maruther, Nisa M, et al. “Diabetes Medications as Monotherapy or Metformin-Based Combination Therapy for Type 2 Diabetes.” *Annals of Internal Medicine*, ACP Journals, June 2016, www-acpjournals-org.myaccess.library.utoronto.ca/doi/10.7326/m15-2650.