Hemoglobin A1c Measurement Impact on Hospital Readmission Rates.

Matthieu Marinangeli[1]

[1]École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

This analysis shows that the measurement of the hemoglobin A1c, a measure of glucose control, for patient being diagnosed with diabetes reduce the probability that it will be readmitted within 30 days. Data is taken from a large clinical database available at the UCI Machine Learning Repository and around 70,000 encounters were selected for this analysis. Multivariate logistic regression, random decision forest and gradient boosting classifier were used to fit the relationship between the measurement of HbA1c and early readmission. This work was greatly inspired from the original analysis [1] performed on these data and consisted to reproduce results with slightly different techniques using scikit-learn[2].

1 Introduction

The interest of management of hyperglycemia for hospitalized patient is increasing as it might have a significant bearing on outcome, in terms of both morbidity and mortality. This has led to the development of new intensive care protocols with rigorous glucose targets in many institutions in the United States. This analysis is meant to show if measurement of the hemoglobin

A1c for patient diagnosed with diabetes is associated with a reduction in readmission rates in individuals admitted to the hospital using 10 years of data from US hospitals.

2 Data and Method

2.1 Initial Dataset Description

This study is using the Health Facts database (Cerner Corporation, Kansas City, MO), a data warehouse in the United States collecting comprehensive clinical records accros hospitals throughout the country, available from the UCI Machine Learning Repository [3]. The data represents 10 years (1999-2008) of clinical care at 130 hospitals and integrated delivery networks in the United States and consist, after preprocessing, of a database of 101,766 encounters fulfilling the following criteria:

1. It is an inpatient encounter (a hospital admission):
2. it is a "diabetic" encounter, that is, one during which any kind of diabetes was entered to the system as diagnosis;
3. the length of stay was at least 1 day and at most 14 days;
4. medications were administered during the encounter.

This dataset contains a set of 55 features, selected by clinical experts and described in Table 1, potentially associated with the diabetic condition or management.
The readmission attribute we are interested in has been redefined into two values: "readmitted" if the patient was readmitted within 30 days of discharge or "otherwise", which covers both readmission after 30 days and no readmission at all.

The measurement of the Hemoglobin A1c (HbA1c) gives important indication of the glucose control and it is widely applied to measure performance of diabetes care [4, 5]. The measurement of HbA1c at the time of hospital admission offers a unique opportunity to assess the efficacy of current therapy and to make changes in that therapy if indicated (e.g., HbA1c $> 8.0\%$ on current regimen). The attribute "A1c test result" indicates if the test was done or not and if yes gives the range of the result. Both the frequency of HbA1c test ordering and the response to its result, which is defined as a change in diabetic medications (any dosage change (increase or reduction) as well as change to a drug with a different generic name), were examined.

We consider four groups of encounters: (1) no HbA1c test performed, (2) HbA1c performed and in normal range, (3) HbA1c performed and the result is greater than 8% with no change in diabetic medications, and (4) HbA1c performed, result is greater than 8%, and diabetic medication was changed.

2.2 Data Selection

The initial dataset described before contains incomplete, redundant and noisy information as expected in any real-world data. Several features has been discarded due to high percentage of missing values such as the weight (97%) and the payer code (40%), the medical specialty of the admitting physician was maintained even with 47% of missing values. In the dataset there are multiple inpatient visits for some patient and the observations could not be considered as statistically independent, therefore we kept only encounter per patient. In particular we maintained only the first encounter for each patient as the primary admission and then determined whether or not they were readmitted within 30 days. In addition all encounters that resulted in either discharge to a hospice or patient death were removed in order to avoid biasing the analysis.
After the above-described requirements 69774 encounters are left in the dataset. To summarize, the dataset consists of hospital admissions of length between one and 14 days that did not result in a patient death or discharge to a hospice. Each encounter corresponds to a unique patient diagnosed with diabetes, although the primary diagnosis may be different. During each of the

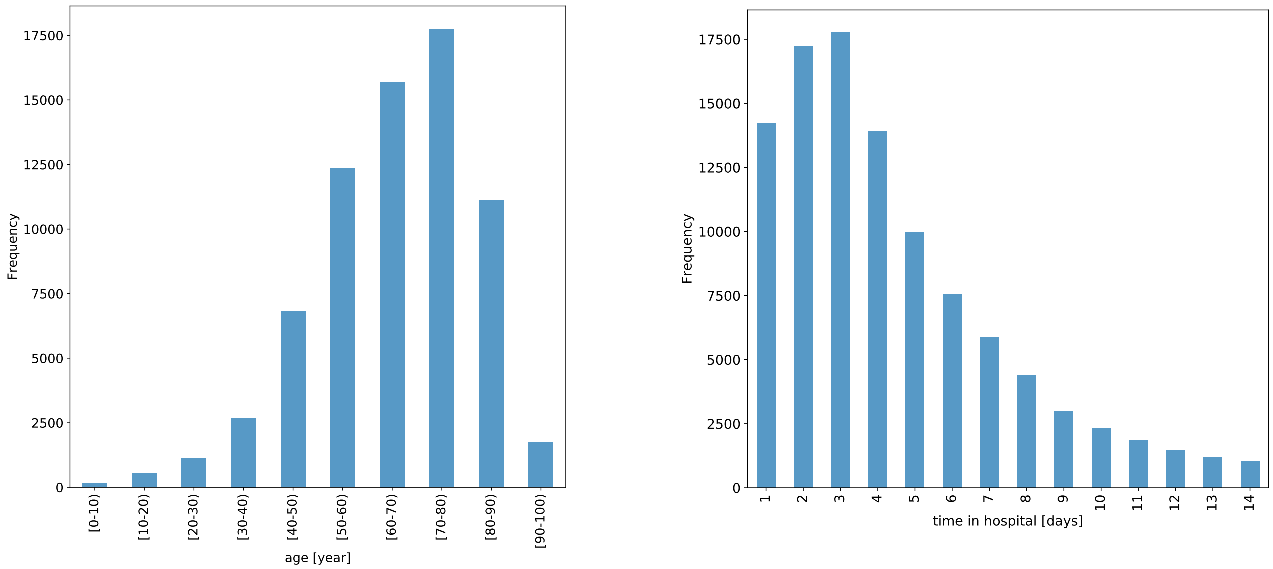

Figure 1: Age distribution and distribution of the time spent in the hospital of the patients.

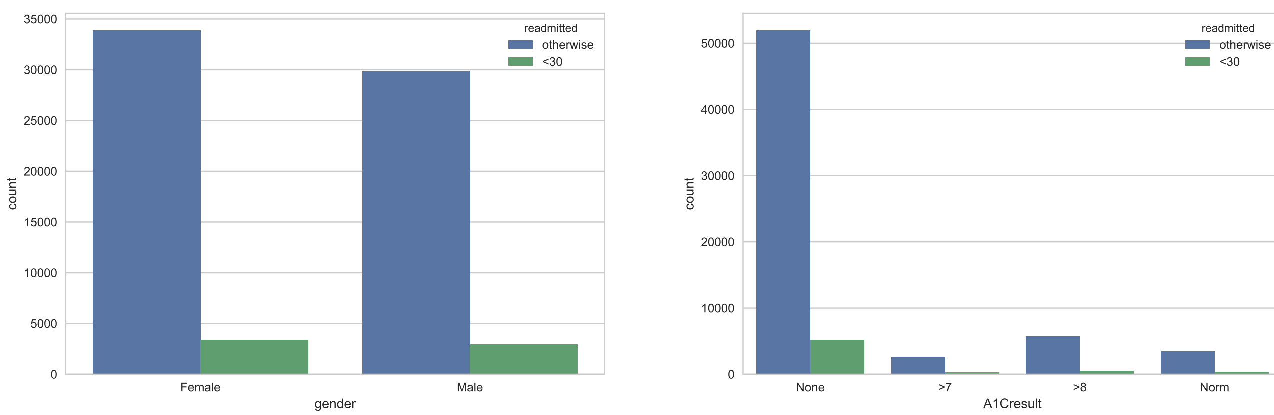

Figure 2: Distributions of patients races and A1C tests in the case the patients are readmitted within 30 (green) days or not (blue).

analyzed encounters, lab tests were ordered and medication was administered.

Distributions of some the demographic and illness severity variables in the dataset are shown in Figures 1, 2 an 3.

All values of the diagnosis features have been grouped into categories, the Table 2 shows

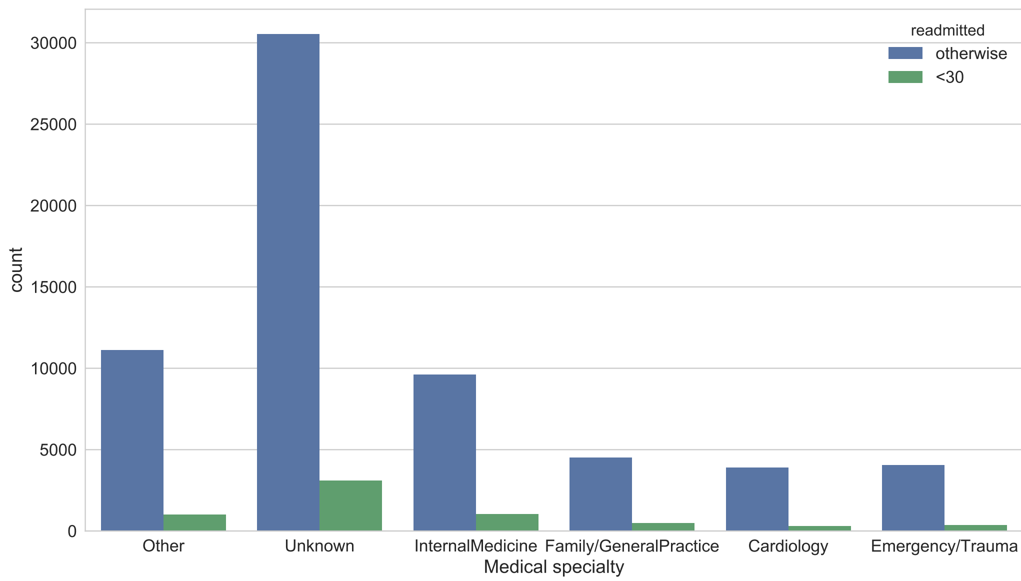

Figure 3: Specialty of the admitting physician distribution in case the case the patients are readmitted within 30 days (green) or not (blue).

these categories for the first diagnosis and and their distributions in the dataset.

2.3 Strategy

Multivariate logistic regression, random decision forests and gradient boosting classification from scikit-learn are used to fit the relationship between the measurement of HbA1c and early readmission. As we are dealing with a lot of categorical (some with high cardinality) and numerical features several assumptions and requirements have been done in order to reduce the number of features. The number of medical specialties is reduced in order to keep only the most significant ones (the rest is grouped under an ”Other” category), a threshold on the number of encounters is set for the types of admissions and discharge dispositions (encounters under this threshold are put in an ”Other” category) and finally the 24 diabetes medications features are hashed into a lower number of outputs; all of these requirements will be optimized.

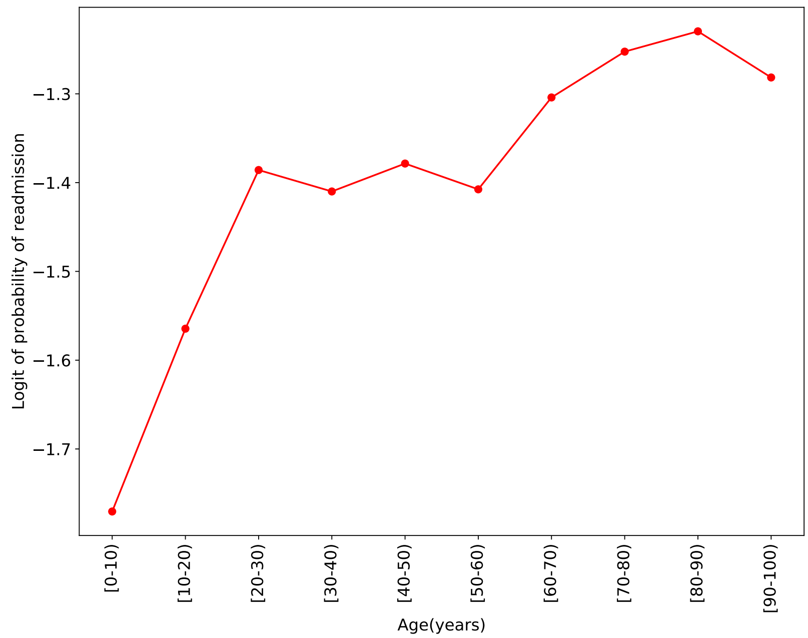

Figure 4: Relationship of age and the logistic function of the readmission rate.

3 Results and Discussion

A primary analysis has been done using logistic regression. The requirements to reduce the number of features have been optimized using this classifier and results in keeping only 7 medical specialties, a threshold of 100 encounters for types of admission and 500 for types of discharge disposition. Finally the diabetes medications have been hashed into 8 outputs.

Fig. 4 shows that we can divide the age feature into three groups, indeed there are three regimes for the logit of probability of readmission rate. For the next steps we can defined three ages categories from 0 to 20 year, from 20 to 60 years and from 60 to 100 years.

Logistic regression, random forest decision and gradient boosting classifier have been trained on this newly defined dataset and the gender feature has been discarded from the dataset due to a poor significance for those classifiers. All classifiers are able to predict if a patient was readmitted within 30 days with an accuracy of 91%.

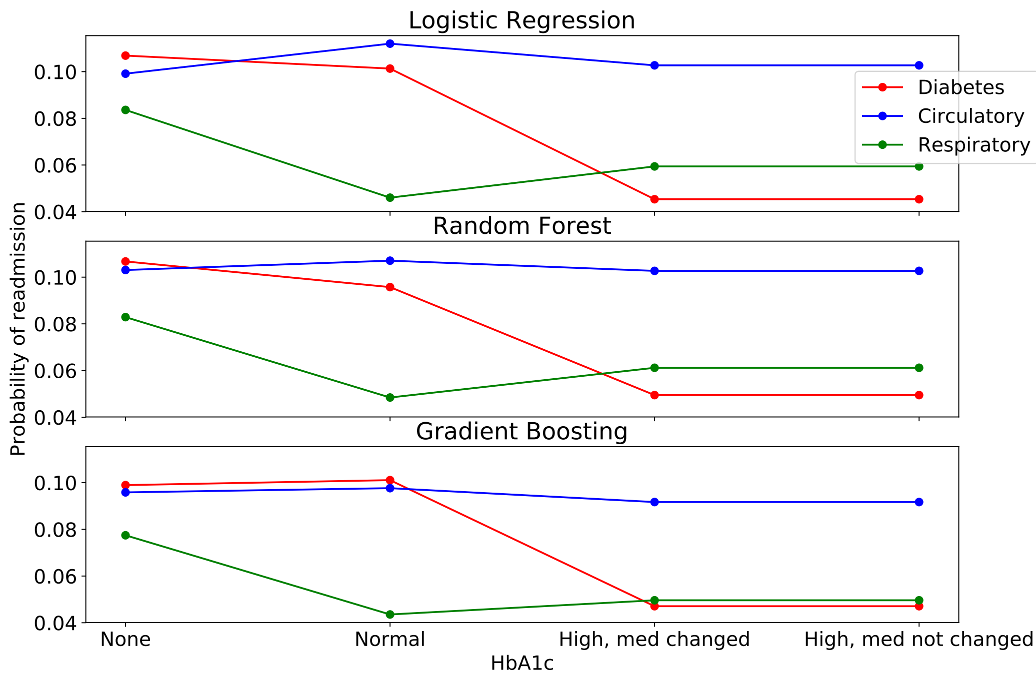

Figure 5: Readmission rates for the four groups of encounters described in Section 2.1 and for different primary diagnosis (Diabetes, Respiratory and Circulatory).

Fig. 5 shows the probability of readmission estimated from the three classifiers for the four groups of encounters defined in Section 2.1 for encounters being primary diagnosed to have a diabetes, circulatory or respiratory issue. We see observe a decrease of the probability of readmission for patient diagnosed with diabetes when a HbA1c test is done will all classifiers.

4 Conclusion

The hypothesis that the measurement of the hemoglobin A1c is associated with a reduction in readmission rates has been shown to be true. HbA1c is therefore a useful predictor of readmission rates which might be valuable in the development of strategies to reduce readmission rates and costs for the care of individuals with diabetes mellitus. The analysis also shown that the profile of readmission differed significantly for patients where HbA1c was checked in the setting of a primary diabetes diagnosis compared to patients with a primary circulatory disorder (for which the readmission rates are the highest). It seems the readmission rates for patient with diabetes is associated with the decision to test for HbA1c but not if diabetes medications were changed or not..

References

1. Beata Strack, Jonathan P. DeShazo, Chris Gennings, *et al*., ”Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records,” *BioMed Research International*, vol. 2014, Article ID 781670, 11 pages, 2014. doi:10.1155/2014/781670

2. Scikit-learn: Machine Learning in Python, Pedregosa *et al*., JMLR 12, pp. 2825-2830, 2011

3. A. Frank and A. Asuncion, *UCI Machine Learning Repository*, University of California, School of Information and Computer Science, 2010.

4. D. Baldwin, G. Villanueva, R. McNutt, and S. Bhatnagar, ”Eliminating inpatient sliding-scale insulin: a reeducation project with medical house staff,” *Diabetes Care*, vol. 28, no. 5, pp. 1008?1011, 2005.

5. H. Anwar, C. M. Fischbacher, G. P. Leese *et al*., ”Assessment of the under-reporting of diabetes in hospital admission data: a study from the Scottish diabetes research network epidemiology group,” *Diabetic Medicine*, vol. 28, no. 12, pp. 1514?1519, 2011.

Table 1: List of features and their descriptions in the initial dataset (the dataset is also available at the website of Data Mining and Biomedical Informatics Lab at VCU (`http://www.cioslab.vcu.edu/`)).

Feature name	Type	Description and values	% missing
Encounter ID	Numeric	Unique identifier of an encounter	0%
Patient number	Numeric	Unique identifier of a patient	0%
Race	Nominal	Values: Caucasian, Asian, African American, Hispanic, and other	2%
Gender	Nominal	Values: male, female, and unknown/invalid	0%
Age	Nominal	Grouped in 10-year intervals: [0, 10), [10, 20), . . ., [90, 100)	0%
Weight	Numeric	Weight in pounds.	97%
Admission type	Nominal	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available	0%
Discharge disposition	Nominal	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available	0%
Admission source	Nominal	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital	0%
Time in hospital	Numeric	Integer number of days between admission and discharge	0%
Payer code	Nominal	Integer identifier corresponding to 23 distinct values, for example, Blue Cross\Blue Shield, Medicare, and self-pay	52%
Medical specialty	Nominal	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family\general practice, and surgeon	53%
Number of lab procedures	Numeric	Number of lab tests performed during the encounter	0%
Number of procedures	Numeric	Number of procedures (other than lab tests) performed during the encounter	0%
Number of medications	Numeric	Number of distinct generic names administered during the encounter	0%
Number of outpatient visits	Numeric	Number of outpatient visits of the patient in the year preceding the encounter	0%
Number of emergency visits	Numeric	Number of emergency visits of the patient in the year preceding the encounter	0%
Number of inpatient visits	Numeric	Number of inpatient visits of the patient in the year preceding the encounter	0%
Diagnosis 1	Nominal	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values	0%
Diagnosis 2	Nominal	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values	0%
Diagnosis 3	Nominal	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values	1%
Number of diagnoses	Numeric	Number of diagnoses entered to the system	0%
Glucose serum test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured	0%
A1c test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>8” if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal” if the result was less than 7%, and “none” if not measured.	0%
Change of medications	Nominal	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: “change” and “no change”	0%
Diabetes medications	Nominal	Indicates if there was any diabetic medication prescribed. Values: “yes” and “no”	0%
24 features for medications	Nominal	For the generic names: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: “up” if the dosage was increased during the encounter, “down” if the dosage was decreased, “steady” if the dosage did not change, and “no” if the drug was not prescribed	0%
Readmitted	Nominal	Days to inpatient readmission. Values: “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was readmitted in more than 30 days, and “No” for no record of readmission.	0%

Table 2: Values of the primary diagnosis in the final dataset. In the analysis, groups that covered less than 3.5% of encounters were grouped into "other" category.

Group name	icd9 codes	Number of encounters	% of encounter	Description
Circulatory	390–459, 785	21,411	30.6%	Diseases of the circulatory system
Respiratory	460–519, 786	9,490	13.6%	Diseases of the respiratory system
Digestive	520–579, 787	6,485	9.3%	Diseases of the digestive system
Diabetes	250.xx	5,747	8.2%	Diabetes mellitus
Injury	800–999	4,697	6.7%	Injury and poisoning
Musculoskeletal	710–739	4,076	5.8%	Diseases of the musculoskeletal system and connective tissue
Genitourinary	580–629, 788	3,435	4.9%	Diseases of the genitourinary system
Neoplasms	140–239	2,536	3.6%	Neoplasms
Other (17.3%)	780, 781, 784, 790–799	2,136	3.1%	Other symptoms, signs, and ill-defined conditions
	240–279, without 250	1,851	2.6%	Endocrine, nutritional, and metabolic diseases and immunity disorders, without diabetes
	680–709, 782	1,846	2.6%	Diseases of the skin and subcutaneous tissue
	001–139	1,683	2.4%	Infectious and parasitic diseases
	290–319	1,544	2.2%	Mental disorders
	E–V	918	1.3%	External causes of injury and supplemental classification
	280–289	652	0.9%	Diseases of the blood and blood-forming organs
	320–359	634	0.9%	Diseases of the nervous system
	630–679	586	0.8%	Complications of pregnancy, childbirth, and the puerperium
	360–389	216	0.3%	Diseases of the sense organs
	740–759	41	0.1%	Congenital anomalies