

PREDICTIVE MODEL BASED ON HEALTH DATA ANALYSIS FOR RISK OF READMISSION IN DISEASE-SPECIFIC COHORTS

Posted on March 15, 2021 by Matthew

Category: [Spring 2021](#)

By Md. Shahid Ansari, M.Sc. Tech. PGDAST; Abhay Kumar Alok, PhD; Dinesh Jain, MBBS, MBA; Santu Rana, PhD; Sunil Gupta, PhD; Roopa Salwan, DM; Svetha Venkatesh, PhD

Abstract

Background: Intervention planning to reduce 30-day readmission post-acute myocardial infarction (AMI) in an environment of resource scarcity can be improved by readmission prediction score. The aim of study is to derive and validate a prediction model based on routinely collected hospital data for identification of risk factors for all-cause readmission within zero to 30 days post discharge from AMI.

Methods: Our study includes 2,849 AMI patient records (January 2005 to December 2014) from a tertiary care facility in India. EMR with ICD-10 diagnosis, admission, pathological, procedural and medication data is used for model building. Model performance is analyzed for different combination of feature groups and diabetes sub-cohort. The derived models are evaluated to identify risk factors for readmissions.

Results: The derived model using all features has the highest discrimination in predicting readmission, with AUC as 0.62; (95 percent confidence interval) in internal validation with 70/30 split for derivation and validation. For the sub-cohort of diabetes patients (1359) the discrimination is slightly better with AUC 0.66; (95 percent CI;). Some of the positively associated predictive variables, include age group 80-90, medicine class administered during index admission (Anti-ischemic drugs, Alpha 1 blocker, Xanthine oxidase inhibitors), additional procedure in index admission (Dialysis). While some of the negatively associated predictive variables, include patient demography (Male gender), medicine class administered during index admission (Betablocker, Anticoagulant, Platelet inhibitors, Anti-arrhythmic).

Conclusions: Routinely collected data in the hospital's clinical and administrative data repository can identify patients at high risk of readmission following AMI, potentially improving AMI readmission rate.

Keywords: Logistic regression, readmissions, acute myocardial infraction.

Introduction

Frequent unplanned readmissions degrade the patient care and institutional performance. It also adds to the cost of managing patients. Recent work reveals that one out of five patients admitted with AMI is readmitted within the first 30 days following discharge.¹⁻³ To reduce readmission rate, it is helpful to recognize the high-risk patients during initial admission and do preventive care accordingly. Finding reliable risk factors are challenging despite of many well-documented risk factors⁴ including severe heart failure⁴, multi-vessel disease⁵, living alone⁶, ethnic background⁷,

psychological comorbidity^{8,9} and socioeconomic factors.¹⁰ Individually these risk factors offer only weak predictability. Our current study investigates this issue from Indian context.

In this current work, we collected and examined the entire patient's information admitted with AMI from Max group of hospitals in India. We derived the model and internally validate it using data from hospital information systems to predict risk of unplanned readmission within zero to 30 days after an index admission with AMI. We explored the model in two ways as a model with and without medication features. We used medication data as a surrogate for different class of co-morbidities. Additionally, we also build models for diabetes sub-cohort and report the model performance.

Most of the recent studies on readmission prediction used administrative, demographic and co-morbidity data extracted from EMR to build the predictive models for readmission prediction. Some of these works have tried to predict readmission for AMI patients, however most do not do well to predict the preventable readmissions within the first 30 days after discharge from hospital. Furthermore, the risk factors are not consistent across studies. However, no prediction model has been proposed for AMI patients in the Indian healthcare domain. In view of the above constraints, we decided to develop a prediction model for identification of patients at risk of readmission within zero to 30 days in the Indian healthcare domain.

Methods

This study is based on retrospective analysis of data extracted from the database system of Max group of hospitals in India for patients admitted at one of their facilities between January 2005 and December 2014, with a primary diagnosis for AMI. An AMI admission was identified by ICD-10 code I21 (acute myocardial infarction). We purposefully limited ourselves to the patient data available on the hospital information systems including the EMR, so as to ensure that the model derived from the study can be implemented on real time patient data available on the EMR system. We included patient related information under the broad heads of demographic, administrative, pathology, clinical procedure, radiology, and medication. Our study includes details for index and readmission of 2849 and 523 patients respectively (18.36 percent). The structure of processed data is shown in [Figure 1\(a\)](#).

These 2849 index patients are included in the study on the basis of identification of patient records, with a confirmed diagnosis of AMI at the time of discharge and admission in the hospital within the period defined above. An emergency initiated unplanned admission following the index admission is considered as a readmission. Patient records for index admission are used to define independent variables. Indication of at least one unplanned readmission within zero to 30 days of discharge from the index admission is used as the dependent variable.

The data related to clinical procedures, radiological procedures and medication are transformed or re-coded to a limited number of variables, in an effort to reduce the large number of unique names. A mapping tool is created for transformation of each of the broad head category. We adopted the

principle of parsimony for the transformation, where each procedure could be broken down to a combination of standardized variables, so as to retain maximum information using a smaller set of variables.

In our work for derivation of the readmission prediction model, the cohort of 2849 patients is randomly divided into two parts: two-third (learning set) for developing a prediction model and the remaining one-third for validating the developed model (validation set). The steps involved in deriving the prediction model are shown in [Figure 1](#). We used the information about the patients from their index admissions to build a set of covariates for prediction model. Depending on the information type, the covariates either take binary or continuous values. The methodology used for generating features related to index admission is described in the following paragraph.

To generate the features for index admission, demographic information (age) is divided into bins of 10-year intervals (0-10, 10-20 etc.) and each patient's age is expressed through a binary representation taking a value one in the appropriate bin and zero at other places. Similarly, geographical region, gender, payment mode, marital status, occupation, and their kin relationship are also expressed through binary representation. Length of stay, critical hours, normal hours, and time in emergency are represented as continuous variables. To create features related to pathology, we extracted several statistics such as count, minimum, maximum, and mean value of six different types of diagnostic tests and represented them as continuous variables. For index procedure, each procedure item name is classified into one of the three classes: body system, procedure type and service. Each class is further divided into a set of categories. Count of categories under a particular class is extracted and further this extracted information is used as feature under the header of class name. Procedure features for patient are organized via body system, procedure type, and service type. The features are modeled as continuous variables. The structure to generate procedure features are shown in [Table 1](#). Similarly, for radiology procedure, each procedure name is classified under four classes: modality, class, body system and contrast. These four features for radiology procedure of single patients are included as continuous variables. The structure to generate radiology procedure features are shown in [Table 1](#).

To incorporate medication information, each medication drug is classified into two classes: drug class and medical condition for which it is prescribed. Medication features are generated based on the presence and absence of medication class in the data. The nature of medication features is assumed as binary variables. The structure to generate medication features are shown in [Table 1](#).

After completion of the feature generation process, we prepared a list of 318 independent variables including index admission features (administrative=14, demographic=43, medication=165, pathology=24, radiology=37 and procedure=35). Finally, dependent variable is a binary variable encoded as presence or absence of a readmission in the 0-30 days following the index admission.

For model fitting using the derivation set, we derived the logistic regression with elastic net¹⁴ which

is a regularized regression technique that linearly combines the L_1 and L_2 penalties of lasso¹⁵ and ridge methods. L_1 regularization helps in sparsifying the weight vectors, while L_2 regularization limits the weight value to protect against outliers. Together Elastic net can find a stable and sparse weight vector for logistic regression. The **elastic net estimator** is linked [here](#).

Where N is the number of observations, y_i is the response at observation i , X_i is data, a vector of d values at observation i , λ_1 and λ_2 are positive regularization parameter which interpolates between L_1 and L_2 norm of β , the parameter β is a coefficient of d -vector.

Using the aforementioned method, we used 10 different random splits to build 10 different models for derivation set and generating prediction of subsequent readmission in validation cohort. Thereafter, we average the predictions coming from each 10 different models to attain the more accurate predictions.

The 30-day readmission probability for AMI hospitalization formulation can be found [here](#).

Where x_i are $d=318$ independent variables, β_0 is constant and P is the probability of readmission following an AMI admission.

Results

The current study includes 2849 patients admitted between January 2005 and December 2014. The rate of readmission for different time periods (0-30, 31-60, and 61-365 days) is measured following first index admission only. Observed readmission rates are as 272 patients (9.55 percent) within zero to 30 days, 60(2.1 percent) and 191(6.7 percent) patients for 31-60 and 61-365 days, respectively. As we can see that most of the readmission happens within the 30 days of discharge, hence here, we build the model for the prediction of readmission falling within zero to 30 days only. The cohort description of 2849 patients and the rate of readmissions over different time horizons are given in [Table 2](#). We derive our model for three different modes: all features, all features without medication and for the sub-cohort of diabetic patients using all the features. These are shown in [Figure 1\(b\)](#).

Model designed without using medication features shows average AUC 0.64(95 percent CI) and 0.61(95 percent CI 0.55-0.67) for derivation and validation cohort, respectively. To enhance the model performance, medication features are also added. After that model performance for derivation and validation sets are improved to average AUC 0.65 (95 percent CI) and 0.62 (95 percent CI 0.56-0.68), respectively. For this case derivation and validation sets, PPV and NPV values are obtained as (0.14, 0.13) and (0.94, 0.93).

Next, we develop a predictive model only for diabetic patients, who are almost half of the whole cohort with 1359 (48 percent) patients. In this case the average model performance for derivation and validation sets are obtained as AUC 0.69 (95 percent CI) and 0.66 (95 percent CI), respectively. The performance of models is reported in [Table 3](#). Based on the model performance, we have

extracted the top risky and protective factors for and against readmissions. Top factors are reported in [Table 4](#), respectively.

Discussion

To predict the risk of readmission for AMI patients within a certain period after discharge from a hospital is a very critical issue in the healthcare domain. In Devan et al.¹³, a systematic review has been performed on the recent works on readmission prediction. In Dharmarajan et al.², authors discussed strategies to reduce the readmission within zero to 30 days for heart failure (HF), AMI and pneumonia patients after discharge. In Donzé et al.¹⁴, authors used multivariate logistic regression to build a model for identification of avoidable readmission of patients within 0-30 days using administrative and clinical data. Model was derived with randomly selected two-third patients and remaining one-third was used for validation. Based on the regression coefficient prediction score was evaluated using seven independent factors including hemoglobin at discharge, sodium level at discharge, procedure during index admission, use of an oncology service, number of preadmissions and length of stay. The model performance for validation set was reported as AUC 0.71. In Wallmann et al.¹⁵, authors derived a prediction model as a screening tool for cardiac related emergency readmission within zero to 30 days. Logistic regression was used to derive model on 70 percent randomly selected records and remaining 30 percent was used for validation. The independent variables included patient demography, hospital utilization, procedure, and clinical co-morbidity. As part of the study, eleven risk factors were identified. They include number of previous emergency admissions within the 6 months preceding the index admission, care type, number of procedures during hospitalization, number of major or minor therapeutic procedure during hospitalization, existence of co-morbidities including anemia, hypertension, heart failure, diabetes, acute coronary syndrome, and renal disease. In Shulan et al.¹⁶, authors developed a prediction model for all cause hospital readmission using administrative data only. Logistic regression was applied on 50 percent selected data to derive the model. Variable selection was not automatic. In each model run, the most statistically significant (p -value <0.05) variables were selected and added with some other variables for next model run. In this way, the variables were selected to derive the model. For the development and validation sets, AUC reported as 0.80 and 0.79, respectively.

The institutional performance and clinical care can be improved if patient's data information is routinely collected and simultaneously updated and fed into a structured electronic medical record (EMR).²⁰⁻²² Nowadays, EMRs are being increasingly adopted in varying degrees, which could be harnessed to derive prediction models, thereby providing a supportive tool to clinicians for handling patient specific risks. However, due to complex structure of EMRs, it is hard to analyze and interpret the data, whereby the benefit of EMR may not be realized.

In the current work, we have derived a regularized linear predictive model to predict the risk of

readmission following AMI using the administrative and clinical hospital data from EMR. We have explored the model with and without medication features and analyzed the impacts on the model performance. We find the model performance on validation set to be satisfactory. Also, we analyzed the model performance for the sub-cohort of diabetic patients.

So far, the clinical practice has been to rely on a combination of biomarkers, clinical risk factors and co-morbidity indices for assessment of a higher risk of deterioration in clinical condition after discharge. Although several biomarkers²³⁻²⁴ are known but none is followed routinely in clinical practice and it is not known which biomarkers will give best prediction. There are some clinical and demographic risk factors available for readmission after AMI^{5-8,25}, but model derived on these factors gives unsatisfactory results. It is widely believed that some co-morbidity such as diabetes, hypertension etc. is responsible for risk of readmission after AMI. Adding these co-morbidities may enhance the model performance.²⁶⁻²⁷ In the current work, we have included many of the important factors to achieve a strong prediction model.

Here predictive risk factors include administrative data, patient demography, radiology procedures, pathology test results, other procedures, and patient medication data.

The selected data set includes a representative set of AMI patients; however, not necessarily all the AMI patients fulfilling the inclusion criteria. Hence the distribution of co-morbidity/demographics may not be representative of the disease incidence in the larger population.

We are aware that the current study has certain limitations, which could be addressed in subsequent works. The entire study data was collected from a single center, and the obtained information was not clinically exhaustive, as the present work has fully relied on administrative and clinical data retrieved from hospital electronic databases. Furthermore, the model performance has not been validated with data external to the study.

Here we also identify some constraints to the study, which could limit the generalization of the findings of this work. The hospital readmission data included in the study is based on the information captured from returning patient population, in an Indian private healthcare environment. Typically, in the Indian healthcare model, the choice of healthcare system is made by the patient, who may choose to return or not, to the same hospital. In case the patient chooses not to follow up at the same hospital, the subsequent developments are missed in the hospital data. Another constraint is because of the incomplete clinical information recorded in the hospital information system fields, which are available for analysis. Since co-morbidities for the patient were not mentioned explicitly in the study data set, we chose to derive them from the medication use data, in a most conservative fashion. This approach is potentially limited to the identification of overtly treated clinical co-morbidity and misses any subclinical co-morbidity not requiring intervention through medication. However, in the context of the study, we believe that any significant co-morbid disease would be

actively managed in the immediate post AMI stage. Lack of drug affordability also does not arise in the context of acute care setting for the post-AMI patients, where the cost of drugs constitutes a small percentage of the cost of overall management.

The study indicates that routinely collected hospital data in the hospital's clinical and administrative data repository can be used to identify the patients at high risk of readmission following AMI. The predictive results of the model are seen to be moderately good in identifying patients at risk of readmission within 30 days of discharge post AMI. We plan to implement the derived model in our information systems for a real-time feedback to the clinician on the risk of readmission at the point of discharge with a 30-day follow-up. This could potentially help the clinician in personalizing the post discharge instructions, thereby leading to improvement in patient outcome. In future a follow-up research study is planned to measure and improve the predictive accuracy of the model, once the on-line system implementation is rolled out.

Declaration of conflicting interests

The authors have no conflicts of interest to declare

Support

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

Author Biographies

Md. Shahid Ansari, M.Sc. Tech., PGDAST (shahid.ansari@maxhealthcare.com) is the deputy manager of clinical data analytics at Max Super Specialty Hospital, New Delhi, India.

Abhay Kumar Alok, PhD (abhayalok@gmail.com) is an associate research fellow of applied artificial intelligence institute at Deakin University, Geelong, VIC, Australia.

Dinesh Jain, MBBS, MBA (dinesh.jain@maxhealthcare.com) is the vice president of clinical data analytics at Max Super Specialty Hospital, New Delhi, India.

Santu Rana, PhD (santu.rana@deakin.edu.au) is an associate professor of applied artificial intelligence institute at Deakin University, Geelong, VIC, Australia.

Sunil Gupta, PhD (sunil.gupta@deakin.edu.au) is an associate professor of applied artificial intelligence institute at Deakin University, Geelong, VIC, Australia.

Roopa Salwan, DM (roopa.salwan@maxhealthcare.com) is the senior consultant of cardiac sciences, cardiology at Max Super Specialty Hospital, New Delhi, India.

Svetha Venkatesh, PhD (svetha.venkatesh@deakin.edu.au) is the professor of applied artificial intelligence institute at Deakin University, Geelong, VIC, Australia.

References

1. Dunlay SM, Weston SA, Killian JM, Bell MR, Jaffe AS, Roger VL. Thirty-day rehospitalizations after acute myocardial infarction: a cohort study. *Annals of internal medicine* 2012; 157: 11–18.
2. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LI, Drye EE. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. *Jama* 2013; 309:355–363.
3. Krumholz HM, Lin Z, Keenan PS, Chen J, Ross JS, Drye EE, Bernheim SM, Wang Y, Bradley EH, Han LF, Normand SLT. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *Jama* 2013; 309: 587–593.
4. Desai MM, Stauffer BD, Feringa HH, Schreiner GC. Statistical models and patient predictors of readmission for acute myocardial infarction: a systematic review. *Circulation Cardiovascular Quality and Outcomes* 2009; 2: 500–507.
5. Kociol RD, Lopes RD, Clare R, Thomas L, Mehta RH, Kaul P, Granger CB. International variation in and factors associated with hospital readmission after myocardial infarction. *Jama* 2012; 307: 66–74
6. Bucholz EM, Rathore SS, Gosch K, Schoenfeld A, Jones PG, Buchanan DM, Krumholz HM. Effect of living alone on patient outcomes after hospitalization for acute myocardial infarction. *The American journal of cardiology* 2011; 108: 943–948.
7. Joynt KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. *Jama* 2011; 305: 675–681.
8. Andres E, Garcia-Campayo J, Magan P, Barredo E, Cordero A, Leon Casasnovas JA. Psychiatric morbidity as a risk factor for hospital readmission for acute myocardial infarction: an 8-year follow-up study in Spain. *The International Journal of Psychiatry in Medicine* 2012; 44: 63–75.
9. Reese RL, Freedland KE, Steinmeyer BC, Rich MW, Rackley JW, Carney RM. Depression and rehospitalization following acute myocardial infarction. *Circulation: Cardiovascular Quality and Outcomes* 2011; 4: 626–633.
10. Lindenauer PK, Lagu T, Rothberg MB, Avrunin J, Pekow PS, Wang Y, Krumholz HM. Income inequality and 30 day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. *Bmj* 2013; 346: f521.
11. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005; 67: 301–320.
12. Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of Royal Statistical Society. Series B (Methodological)* 1996; 58: 267–288.
13. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S. Risk

prediction models for hospital readmission: a systematic review. *Jama* 2011; 306: 1688-1698.

14. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA internal medicine* 2013; 173: 632-638.

15. Wallmann R, Llorca J, Gómez-Acebo I, Ortega AC, Roldan FR, Dierssen-Sotos T. Prediction of 30-day cardiac-related-emergency-readmissions using simple administrative hospital data. *International journal of cardiology* 2013; 164:193-200.

16. Shulan M, Gao K, Moore CD. Predicting 30-day all-cause hospital readmissions. *Health care management science* 2013; 16:167-175.

17. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Speroff T. Automated identification of postoperative complications within an electronic medical record using natural language processing. *Jama* 2011; 306: 848-855.

18. Appari A, Eric Johnson M, Anthony DL. Meaningful use of electronic health record systems and process quality of care: evidence from a panel data analysis of US acute-care hospitals. *Health Service Research* 2013; 48:354-375.

19. FitzHenry F, Murff HJ, Matheny ME, Gentry N, Fielstein EM, Brown SH, Speroff T. Exploring the frontier of electronic health record surveillance: the case of post-operative complications. *Medical care* 2013; 51: 509.

20. Hartford M, Wiklund O, Mattsson-Hulten L, Persson A, Karlsson T, Herlitz J, Caidahl K. C-reactive protein, interleukin6, secretory phospholipase A2 group IIA and intercellular adhesion molecule1 in the prediction of late outcome events after acute coronary syndromes. *Journal of internal medicine* 2007; 262: 526-536.

21. Gao Y, Tong GX, Zhang XW, Leng JH, Jin JF, Wang NF, Yang JM. Interleukin-18 levels on admission are associated with mid-term adverse clinical events in patients with ST-segment elevation acute myocardial infarction undergoing percutaneous coronary intervention. *International heart journal* 2010; 51:75-81.

22. Xin H, Chen ZY, Lv XB, Liu S, Lian ZX, Cai SL. Serum secretory phospholipase A2-IIa (sPLA2-IIA) levels in patients surviving acute myocardial infarction. *Eur Rev Med PharmacolSci* 2013; 17: 999-1004.

23. Ephrem G. Red blood cell distribution width is a predictor of readmission in cardiac patients. *Clinical cardiology* 2013; 36: 293-299.

24. Matsudaira K, Maeda K, Okumura N, Yoshikawa D, Morita Y, Mitsunashi H, Nagoya Acute Myocardial Infarction Study (NAMIS) Group. Impact of low levels of vascular endothelial growth factor after myocardial infarction on 6-month clinical outcome. *Circulation Journal* 2012; 76:

1509-1516.

25. Rodriguez F, Joynt KE, López L, Saldaña F, Jha AK. Readmission rates for Hispanic Medicare beneficiaries with heart failure and acute myocardial infarction. *American heart journal* 2011; 162: 254-261

26. Condon JR, You J, McDonnell J. Performance of comorbidity indices in measuring outcomes after acute myocardial infarction in Australian indigenous and nonindigenous patients. *Internal medicine journal* 2012; 42: e165-e173

27. GandjourA, Ku-Goto MH, Ho V. Comparing the validity of different measures of illness severity: a hospital-level analysis for acute myocardial infarction. *Health services management research* 2012, 25:138-143.

There are no comments yet.