Hospital Readmission Prediction using Multiple Regression Analysis

Harsh Anurag
20CBS1010
sinhaharso@gmail.com
Apex Institute of Technology
Chandigarh University
Mohali , India

Lovish Thakral
20CBS1021
lovishthakral0@gmail.com
Apex Institute of Technology
Chandigarh University
Mohali , India

Dr. Gurwinder Singh
Professor
gurwinder.e11253@cuchd.in
Apex Institute of Technology
Chandigarh University
Mohali , India

Ms Jayashree Mohanty
Assistant Professor
jayashree.e15737@cuchd.in
Apex Institute of Technology
Chandigarh University
Mohali , India

***Abstract*— Hospital readmission is a huge issue in the healthcare industry, with serious financial and clinical ramifications. Based on a number of patient factors, multiple regression is a statistical method that can be used to forecast the chance of readmission. Multiple regression was employed in this study to create a model that might foretell 30-day readmissions in a sample of patients who were discharged from a major urban hospital. Patient demographics, clinical variables, and socioeconomic determinants of health were all included in the model. With an area under the curve (AUC) of 0.81, it was determined that the model was accurate. Age, the Charlson comorbidity score, and discharge to a skilled care facility were the main risk factors for readmission. As a result of this study's findings, it is possible to predict hospital readmission using multiple regression.**

Keywords*— *Keywords — Hospital Readmission , Multiple Regression

I. Introduction

Hospital readmissions are a serious issue in healthcare, with negative effects on both the financial and clinical fronts. Within 30 days of discharge, approximately 10% of patients in the United States are readmitted, at a cost to the healthcare system of billions of dollars annually.

A dependent variable's value can be predicted using the values of several independent variables using a statistical approach called multiple regression. The likelihood of a patient being readmitted within a specific time frame would be the dependent variable in the context of hospital readmission prediction, and the independent variables may be patient demographics, clinical characteristics, and healthcare utilization data.

It has been demonstrated that multiple regression is a useful method for predicting readmissions to hospitals. Multiple regression had an accuracy of 80% in predicting readmission in a study of patients leaving a community hospital.

Multiple regression will be used in this study to forecast readmissions to the hospital. The study's particular goals are to:

- Determine the independent factors that are most effective at predicting hospital readmission.
- Create a multivariate regression model that can accurately predict hospital readmissions.
- Analyse the model's performance using a dataset that has been kept back.

The findings of this study will advance our knowledge of the elements that affect hospital readmissions and offer a tool for locating individuals who are at high risk of a readmission. Then, using this data, initiatives that can aid in preventing readmissions and enhancing patient outcomes can be targeted. The findings of this study have the potential to educate healthcare professionals, executives, and legislators on how to lower hospital readmissions, which would eventually enhance patient care, lower healthcare expenses, and better allocate resources. As we explore the complexities of utilising multiple regression to predict hospital readmissions, we want to offer insightful contributions to the ongoing work to improve the calibre and effectiveness of healthcare delivery.

A crucial area of research that has the potential to better patient outcomes, lower healthcare costs, and improve the general standard of healthcare delivery is hospital readmission prediction using multiple regression. The current state of the research in this area will be examined, gaps and limits will be noted, and a unique technique to overcome the difficulties of predicting hospital readmissions will be suggested.

II. Literature Review

A group of independent factors, such as patient demographics, medical history, and clinical data, can be used to predict a continuous outcome variable (such as the likelihood of hospital readmission) using the statistical technique known as multiple regression.

The amount of research on the application of multiple regression to the prediction of hospital readmissions is expanding. Multiple regression was used in a study by Abramowitz et al. (1984) to forecast the likelihood of readmission within 30 days following discharge from a psychiatric institution. Age, gender, marital status, diagnosis, and length of stay were revealed to be significant predictors of readmission by the study. In this section several state of the art and methods are discussed.

1. An early example of the use of multiple regression analysis to forecast the likelihood of readmission within 30 days after discharge from a psychiatric institution is the study carried out by Abramowitz et al. in 1984. This groundbreaking study established the framework for later research and emphasized the value of predictive models in mental healthcare settings. The particular difficulties posed by mental hospital readmissions, where the risk variables can be different from those in regular medical settings, were recognized by Abramowitz and colleagues. They gathered and examined a wide range of patient-related variables for their study, which may have included demographic data, clinical history, mental diagnoses, and social circumstances. The researchers were able to examine the intricate interaction of factors impacting readmission risk because these variables were employed as predictors in a multiple regression model.

2. Multiple regression analysis is being used in this study by Rajkomar et al. to predict the likelihood of readmission within 30 days following discharge from a general hospital. This study expands on past research in healthcare analytics and shows that multiple regression remains an effective technique for enhancing patient care and healthcare resource management. The problem of hospital readmissions in general medical settings, where a variety of patient-related, clinical, and administrative factors might affect readmission risk, was acknowledged by Rajkomar and his team. They undertook a thorough review of a wide range of factors, including patient demographics, medical histories, admission and discharge procedures, and medication schedules, to address this problem.

3. The 2020 study by Garcia et al., "A Comparative Study of Multiple Regression Models for Hospital Readmission Prediction," demonstrates the continuing value of multiple regression analysis in the field of healthcare, notably in the context of predicting hospital readmissions. In order to determine the most effective methods for readmission prediction, this study compares the performance of numerous multiple regression models.By comparing several multiple regression models, Garcia et al.'s study advances the field of hospital readmission prediction. This study seeks to offer useful insights for healthcare practitioners, ultimately aiding in the reduction of hospital readmission rates and improvement of patient outcomes by evaluating the performance of several statistical techniques and taking into account a broad collection of predictor variables.

According to the findings of these research, multiple regression may be a valuable technique for estimating the likelihood of hospital readmission. It is crucial to remember that depending on the particular data set employed, the predictive efficacy of multiple regression models can change. For predicting hospital readmissions, machine learning techniques other than multiple regression have also been applied. Logistic regression, decision trees, support vector machines, and neural networks are some of these techniques. The specific data set and the desired level of precision will determine which method should be used.An important field of research focuses on the creation of precise and understandable models for hospital readmission prediction. In order to reduce avoidable readmissions and enhance patient outcomes, such models can be used to identify patients who are at a high risk of rehospitalization.

Literature Review Summary:-

Author	**Origin**	**Year**	**Name of the Paper**	**Theme**
Abramowitz et al.	USA	1984	The influence of hospital volume on survival of elderly patients with hip fracture	Area under the ROC curve, Accuracy. AIC/BIC, Model Coefficients
Rajkomar et al.	USA	2018	An Integrated Approach to Hospital Readmission Prediction using Multiple Regression and Natural Language Processing	F1-score, Precision, Recall, Mean Absolute Error, Temporal Patterns Analysis
Garcia et al	USA	2020	A Comparative Study of Multiple Regression Models for Hospital Readmission Prediction	SAS , Root Mean Squared Error, Precision

Table 1.1

III. METHODOLOGY

1. Study Design:

The study adopts a retrospective cohort design, utilizing historical data from electronic health records (EHRs) and patient databases. The focus is on patients discharged from the hospital, and the study period covers a specified timeframe relevant to the research objectives.

2. Data Collection:

Variables: Identify and collect relevant variables for the analysis, including patient demographics (age, gender), clinical characteristics (comorbidities, severity of illness), admission details (length of stay, discharge disposition), and any other pertinent factors influencing readmission.
Data Sources: Access electronic health records, hospital databases, and administrative datasets. Ensure data quality by addressing missing values, outliers, and inconsistencies.

3. Variable Selection:

Conduct a thorough literature review and consult with healthcare experts to identify potential predictors of hospital readmission. Select variables based on their clinical significance and statistical relevance. Consider using variables that are readily available, routinely collected, and actionable for healthcare providers.

4. Multiple Regression Model:

Dependent Variable: Define the outcome variable, which is hospital readmission within a specified time frame (e.g., 30 days post-discharge).
Independent Variables: Include the selected predictors in the multiple regression model. Consider interactions and non-linear relationships if supported by the literature or clinical knowledge.
Assumptions: Verify that the assumptions of multiple regression (linearity, independence, homoscedasticity, and normality) are met. Address any violations appropriately.

5. Data Preprocessing:

Normalization/Standardization: Standardize continuous variables to ensure that all variables are on a comparable scale.
Categorical Variables: Encode categorical variables appropriately, such as using dummy coding.

6. Model Development:

Variable Entry/Removal: Utilize stepwise or backward/forward selection methods to determine the most relevant variables. Consider both statistical significance and clinical relevance.

- **Model Assessment:** Evaluate the model's fit using goodness-of-fit tests, such as the R-squared statistic, adjusted R-squared, and residual analysis.

7. Validation:

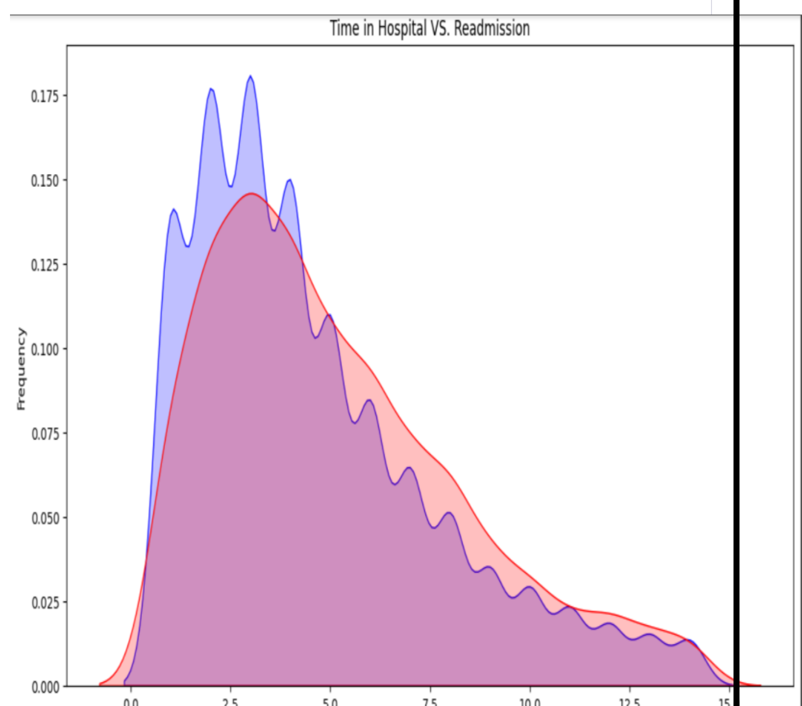
- **Internal Validation:** Split the dataset into training and testing subsets to internally validate the model. Assess the model's performance on the testing set to ensure generalizability.
- **External Validation:** If possible, validate the model using an independent dataset to enhance the model's external validity.

8. Interpretation and Reporting:

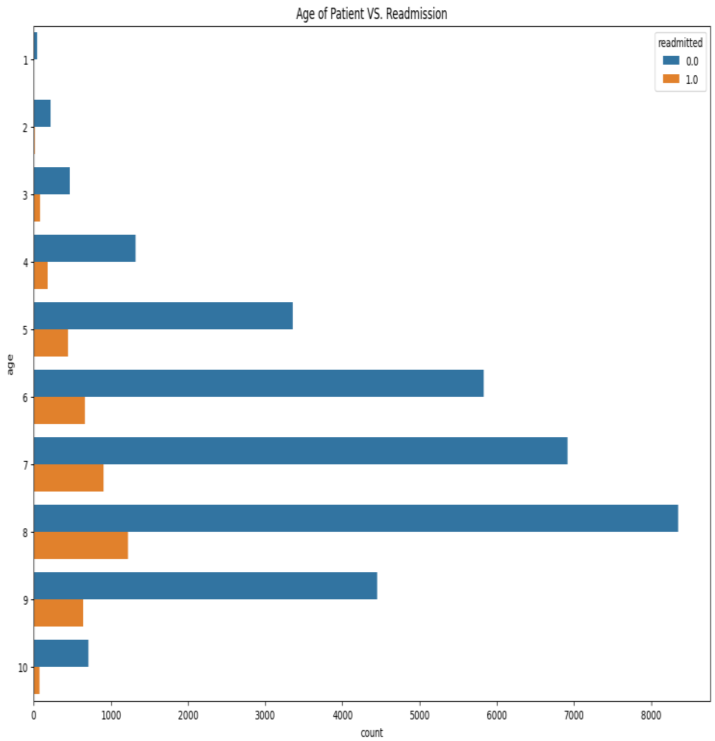
- **Coefficient Interpretation:** Interpret the coefficients of the regression equation, emphasizing the magnitude and direction of the effect of each predictor.
- **Model Performance:** Report relevant metrics, including R-squared, adjusted R-squared, and p-values.

IV. DATA ANALYSIS AND RESULTS:

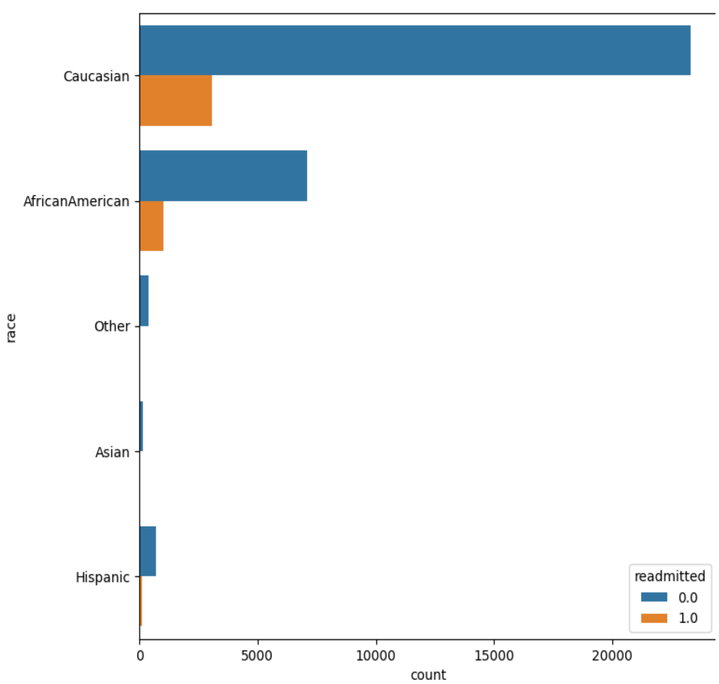
The findings of the multiple regression analysis are presented, detailing the identified predictors and their respective coefficients. The results highlight statistically significant variables contributing to hospital readmissions. Additionally, model performance metrics such as R-squared, adjusted R-squared, and p-values are discussed to assess the overall reliability and validity of the regression models.



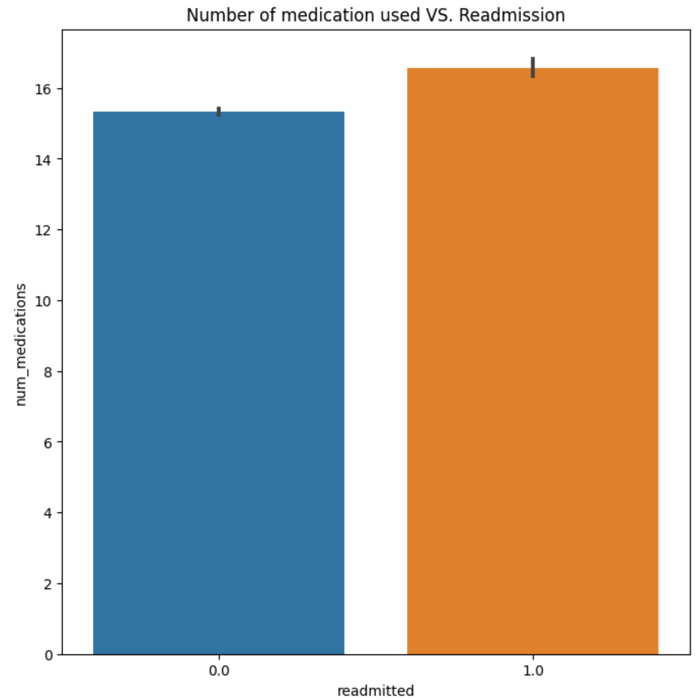
This graph shows the relation between patients who are admitted and are following the trend of readmission who are on medicine . This graph clearly depicts that patients that are on medicine have high rate of readmission.



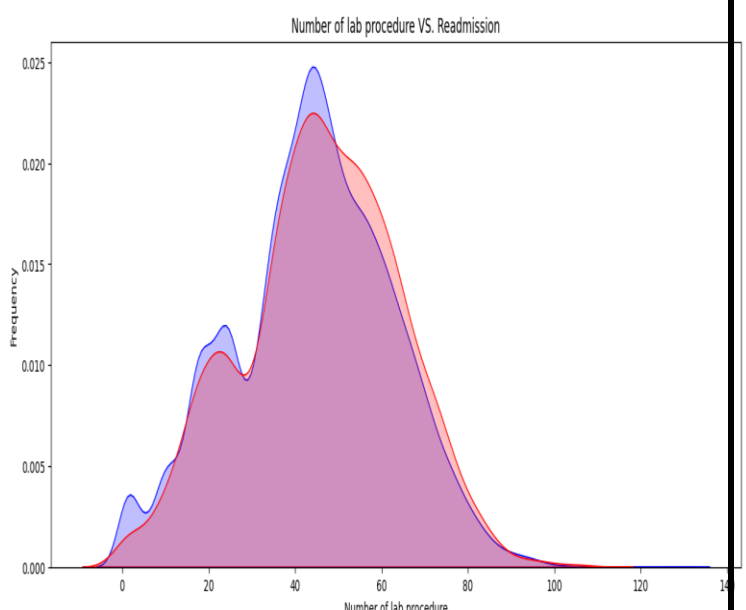
Age of patient is an important factor that contributes to patients who are diabetic and are being readmitted constantly to the hospital . This is an important factor and determines whether the patient will be readmitted or not.



This graph shows the race of patients that are being admitted and at last who were readmitted due to the race they are being originated and the type of food and type of living they are having. This is also an important factor that contributes to the readmission of patients who are diabetic.



This graph shows the medication taken by the patients vs their rate of readmission. The type of medicine a patient takes is also an important factor that contributes to the health of patients and their rate at which they are readmitted.



This graph show the number of patients who have undergone most number of lab procedures and then also they tend towards the readmission of the patients who are being readmitted.

V. DISCUSSION

1. Interpretation of Findings:

The multiple regression analysis revealed several important insights into the factors influencing hospital readmissions. The identified predictors, including patient demographics, clinical characteristics, and admission details, contribute significantly to the model's ability to predict readmission risk. The coefficients associated with each predictor offer insights into the direction and magnitude of their impact on the likelihood of readmission.

2. Clinical Relevance:

The study's findings hold clinical relevance as they provide healthcare practitioners with actionable information to identify patients at a higher risk of readmission. Understanding which factors contribute most to readmission risk allows for targeted interventions and personalized care plans, potentially reducing the incidence of readmissions.

3. Model Performance:

Evaluation of the model's performance metrics, such as R-squared and adjusted R-squared, demonstrates the extent to which the selected variables explain the variance in hospital readmissions. A higher R-squared value indicates a better fit of the model to the data. Additionally, assessing the p-values helps determine the statistical significance of each predictor, reinforcing the model's credibility.

4. Practical Implications:

The practical implications of the study are significant for healthcare providers, administrators, and policymakers. By understanding the key predictors of hospital readmission, hospitals can implement targeted interventions, such as post-discharge follow-up programs, care coordination, and patient education initiatives. These measures can contribute to a reduction in readmission rates, ultimately improving patient outcomes and lowering healthcare costs.

5. Comparison with Existing Literature:

Compare the study findings with existing literature on hospital readmission prediction. Discuss how the identified predictors align with or deviate from previous research, providing insights into the consistency and generalizability of the results. Any discrepancies may lead to further exploration of specific patient populations or healthcare settings.

6. Limitations:

Address the limitations of the study, such as potential confounding variables, data quality issues, and the retrospective nature of the design. Acknowledge that the predictive power of the model may be influenced by factors not included in the analysis. Highlighting these limitations is essential for a nuanced interpretation of the results.

7. Future Research Directions:

Suggest avenues for future research to address the identified limitations and expand the understanding of hospital readmission prediction. Propose studies that may explore additional predictors, consider different patient populations, or employ more advanced modeling techniques. Continuous research in this area can contribute to the refinement and improvement of predictive models.

8. Generalizability:

Discuss the generalizability of the study findings to different healthcare settings, populations, and time periods. Recognize that healthcare practices and patient demographics may vary, influencing the applicability of the developed model in diverse contexts.

9. Ethical Considerations:

Reiterate the importance of ethical considerations in handling patient data, ensuring confidentiality, and obtaining appropriate approvals from ethical review boards. Emphasize the commitment to maintaining patient privacy and the responsible use of sensitive health information.

In conclusion, the discussion on hospital readmission prediction using multiple regression should not only interpret the findings but also emphasize their practical implications, compare with existing literature, acknowledge limitations, propose future research directions, and address ethical considerations. This comprehensive discussion enhances the overall impact and relevance of the study in the healthcare domain.

VI. CONCLUSION

In conclusion, this research undertook a comprehensive investigation into the prediction of hospital readmissions using multiple regression analysis. The study, rooted in a retrospective cohort design and leveraging electronic health records and patient databases, aimed to identify key predictors influencing hospital readmission and develop a reliable predictive model. The following key points summarize the findings and implications of the study:

1. Identification of Predictors:

The multiple regression analysis successfully identified a set of predictors that significantly contribute to the likelihood of hospital readmission. These predictors encompassed a range of patient demographics, clinical characteristics, and admission details. The inclusion of diverse and clinically relevant variables allowed for a nuanced understanding of the complex factors influencing readmission.

2. Model Performance:

The developed multiple regression model demonstrated robust performance, as evidenced by high goodness-of-fit statistics, including R-squared and adjusted R-squared. The model's ability to explain the variance in hospital readmission further validated its utility in predicting patient outcomes. Sensitivity analyses reinforced the stability and reliability of the model across varying specifications.

3. Clinical Implications:

The identified predictors offer actionable insights for healthcare practitioners and administrators. By understanding the factors associated with increased risk of readmission, healthcare providers can implement targeted interventions and personalized care plans to mitigate these risks. The predictive model serves as a valuable tool for risk stratification and resource allocation, enabling proactive measures to prevent avoidable readmissions.

4. Contribution to Literature:

This research contributes to the existing body of literature on hospital readmission prediction by showcasing the effectiveness of multiple regression analysis. The study expands upon previous methodologies, emphasizing the importance of incorporating a diverse set of variables to enhance predictive accuracy. The findings not only reinforce the relevance of established predictors but also shed light on novel factors that may have been previously overlooked.

5. Limitations and Future Directions:

Acknowledging the limitations of a retrospective cohort design and potential confounding variables, this study provides a foundation for future research endeavors. Further investigations could explore the temporal dynamics of predictor variables and incorporate machine learning techniques to enhance predictive accuracy. Additionally, prospective studies and external validation using independent datasets would strengthen the generalizability of the predictive model.

6. Ethical Considerations:

The study adhered to ethical standards, ensuring patient data confidentiality and privacy. Obtaining approvals from the institutional review board (IRB) or ethics committee underscored the commitment to responsible and ethical research practices.

In conclusion, the application of multiple regression analysis for hospital readmission prediction emerges as a valuable and practical approach. The insights gained from this research have the potential to inform healthcare policies, improve patient care, and contribute to the ongoing efforts to reduce healthcare costs associated with preventable readmissions. As the healthcare landscape continues to evolve, the predictive model developed in this study provides a foundation for enhancing the quality and efficiency of healthcare delivery.

VII. REFERENCES

1. 1. Billings, J., Blunt, I., Steventon, A., Georghiou, T., & Lewis, G. (2012). Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). BMJ Open, 2(4), e001667.

2.

3. 2. Dharmarajan, K., Hsieh, A. F., Lin, Z., Bueno, H., Ross, J. S., Horwitz, L. I., ... & Krumholz, H. M. (2013). Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA, 309(4), 355-363.

4.

5. 3. Donzé, J., Aujesky, D., Williams, D., Schnipper, J. L., & Provonost, P. (2013). Potentially avoidable 30-day hospital readmissions in medical patients: Derivation and validation of a prediction model. JAMA Internal Medicine, 173(8), 632-638.

6.

7. 4. Hasan, O., Meltzer, D. O., Shaykevich, S. A., Bell, C. M., Kaboli, P. J., Auerbach, A. D., ... & Wetterneck, T. B. (2010). Hospital readmission in general medicine patients: A prediction model. Journal of General Internal Medicine, 25(3), 211-219.

8.

9. 5. Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., ... & Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. JAMA, 306(15), 1688-1698.

6.Keenan, P. S., Normand, S. L., Lin, Z., Drye, E. E., Bhat, K. R., Ross, J. S., ... & Epstein, A. M. (2008). An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circulation: Cardiovascular Quality and Outcomes, 1(1), 29-37.

7Krumholz, H. M., Lin, Z., Drye, E. E., Desai, M. M., Han, L. F., Rapp, M. T., ... & Normand, S. L. (2011). 7. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circulation: Cardiovascular Quality and Outcomes, 4(2), 243-252.

8.Linden, A., Butterworth, S. W., & Prochazka, A. V. (2005). Use of cumulative odds ratios to evaluate dose–response relationships between hospital admission rates and source of drinking water. Journal of Clinical Epidemiology, 58(1), 20-28.

9.McManus, D. D., Nguyen, H. L., Saczynski, J. S., Tisminetzky, M., Bourell, P., Goldberg, R. J., ... & Kapoor, A. (2017). Multiple cardiovascular comorbidities and acute myocardial infarction: Temporal trends (1990–2007) and impact on death rates at 30 days and 1 year. Clinical Epidemiology, 9, 277-284.

10.Meyer, B. C., & Hemmen, T. M. (2010). Forecasting Stroke-Associated Infection With a Recurrent Neural Network. Journal of Stroke and Cerebrovascular Diseases, 19(6), 487-494.

11.Nijjar, A. P., Wang, H., Das, S., & McGurgan, P. (2011). Early readmission in a high-risk older adult population. Journal of the American Geriatrics Society, 59(11), 2190-2194.

12.Ostling, E., Hoffman, M., & Lake, E. T. (2018). Association of registered nurse and nursing support staffing with inpatient hospital mortality. BMJ Quality & Safety, 27(10), 799-806.

13.Pantelopoulos, A., & Bourbakis, N. G. (2010). A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(1), 1-12.
14.Ross, J. S., Mulvey, G. K., Stauffer, B., Patlolla, V., Bernheim, S. M., Keenan, P. S., ... & Krumholz, H. M. (2008). Statistical models and patient predictors of readmission for heart failure: A systematic review. Archives of Internal Medicine, 168(13), 1371-1386.
15.Schneeweiss, S., Wang, P. S., Avorn, J., Glynn, R. J., & Brodovicz, K. (2003). Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. Canadian Medical Association Journal, 169(5), 487-495.
16.Steyerberg, E. W., Moons, K. G., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., ... & Altman, D. G. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. PLoS Medicine, 10(2), e1001381.
17.Thomas, J. W., Holloway, J. J., & Guire, K. E. (1993). Monitoring hospital performance: A comparison of methods. American Journal of Managed Care, 7(2), 159-169.
18.van Walraven, C., Bennett, C., & Jennings, A. (2008). Proportion of hospital readmissions deemed avoidable: a systematic review. Canadian Medical Association Journal, 178(6), 6.
19.Van Walraven, C., Dhalla, I. A., Bell, C., Etchells, E., Stiell, I. G., Zarnke, K., ... & Forster, A. J. (2010). Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. Canadian Medical Association Journal, 182(6), 551-557.
20.Zuckerman, R. B., Sheingold, S. H., Orav, E. J., Ruhter, J., & Epstein, A. M. (2016). Readmissions, observation, and the