Predictive Indicators for 30-Day Readmission

By Michael Koo

1. Introduction

Hospital readmissions, particularly those occurring within 30 days of discharge, have emerged as a significant metric in evaluating healthcare quality and patient outcomes [1]. These readmissions not only impose financial burdens on healthcare systems but also serve as a proxy for gaps in continuity of care, treatment effectiveness, and patient adherence to post-discharge protocols. Thus, accurate prediction of 30-day readmissions is vital for implementing proactive interventions aimed at mitigating avoidable hospitalizations.

In recent years, the increasing availability of large-scale healthcare datasets has opened opportunities for advanced predictive modeling in patient outcomes. The MIMIC-IV database [10], a robust collection of de-identified clinical data containing over 65,000 intensive care unit (ICU) admissions and over 200,000 emergency department (ED) patient visits, provides a unique lens into the complexities of patient care. This dataset allows for comprehensive analyses of diverse clinical, demographic, and administrative variables that influence hospital readmissions.

Existing literature has identified key predictors of hospital readmissions, including patient comorbidities, length of stay, discharge disposition, and socioeconomic factors [2]. However, these studies often rely on limited datasets or fail to integrate complex interactions among variables. This study leverages the rich and detailed data within MIMIC-IV to address these gaps. By utilizing advanced statistical methods, this study aims to elucidate the relationships between patient characteristics and 30-day readmissions, ultimately contributing to improved care strategies and resource allocation.

This paper presents a rigorous analysis of the predictors of 30-day readmissions using the MIMIC-IV dataset [10]. Through a combination of exploratory data analysis, statistical modeling, and visualizations, the goal is to identify key factors influencing readmissions and assess their relative impact. To then provide actionable insights for clinicians, hospital administrators, and policymakers to enhance patient care and reduce preventable readmissions.

2. Aims, Objectives, and Hypothesis

Hospital readmissions represent a critical quality metric, with 30-day readmissions serving as an essential focus for healthcare providers and policymakers [3]. This study seeks to identify significant predictors of 30-day readmissions using detailed clinical, demographic, and administrative data from the MIMIC-IV dataset [10].

2.1 Aims

The primary aim is to explore and quantify the relationships between patient characteristics and the likelihood of 30-day hospital readmissions. This includes examining the impact of clinical indicators, comorbidities, and discharge details on readmission rates. By identifying key predictors, the analysis aims to inform interventions that may reduce preventable readmissions and improve patient outcomes.

2.2 Objectives

1. Conduct a comprehensive analysis of patient and clinical variables associated with 30-day readmissions.
2. Employ statistical analysis to estimate the strength of associations between predictors and the binary readmission outcome while accounting for confounding factors and clinical relevance.
3. Provide actionable insights by presenting odds ratios and confidence intervals for significant predictors.
4. Validate the findings through descriptive, bivariate, and multivariable analyses.

2.3 Hypothesis

- Null Hypothesis (H_0): No significant association exists between the predictors and 30-day hospital readmissions.
- Alternative Hypothesis (H_a): One or more predictors exhibit a statistically significant association with the likelihood of 30-day hospital readmissions.

3. Study Design

3.1 Study Design

A retrospective cohort study design [4] was employed to analyze patient data from the MIMIC-IV database [10]. This approach allowed for the examination of historical clinical and demographic information to identify predictors of 30-day readmissions. The study utilized both bivariate, multivariate, and logistic regression modeling to quantify associations between predictors and the binary readmission outcome.

3.2 Inclusion Criteria

1. Patients aged 18 years or older at the time of hospital admission.
2. Hospital admissions with complete discharge data and associated clinical records.
3. Availability of all key predictor variables required for analysis.

Excluding patients under 18 limits the study to adults, ensuring consistent disease presentation and treatment. Complete discharge data and clinical records provide essential information for

outcome assessment and predictor variable analysis. The availability of all key predictor variables guarantees accurate analysis and interpretation of results.

3.3 Exclusion Criteria

1. Admissions with missing or incomplete data for the primary outcome or key predictors.
2. Pediatric admissions involving patients under 18 years of age.

Excluding cases with missing or incomplete data and pediatric admissions helps to maintain data quality and the relevance of the study to the target population of adult patients.

3.4 Data and Variable Explanations

The analysis incorporated variables representing demographic, clinical, and administrative characteristics. These variables were selected based on prior literature and their relevance to hospital readmission studies.

Health Outcome Variable:

- **30-Day Readmission:** Binary indicator (1 = readmitted within 30 days, 0 = not readmitted).

Predictors:

- **Continuous Variables**:
    - **Length of Stay**: Total days hospitalized during the index admission.
    - **Medication Count**: Number of medications prescribed at discharge.
    - **Serum Creatinine**: Biomarker reflecting kidney function.
    - **Number of Lab Tests**: Total diagnostic tests performed during admission.
    - **Number of Diagnoses**: Count of all ICD codes associated with the admission.
- **Categorical Variables**:
    - **Charlson Comorbidity Index (CCI)**: Grouped comorbidity scores representing overall patient health burden [5].
    - **Age Group**: Grouped as Young (18-35 years old), Middle (36-60 years old), and Old (<60 years old). This was polychotmized because of the similarity in health related admissions within age groups. The cut values were assigned by associated literature [6].
    - **Discharge Disposition**: Categorized by destination post-discharge (e.g., home, skilled nursing facility, acute hospital).
    - **Major Diagnostic Categories (MDCs)**: Broad body system and disease groupings based on ICD codes [7]. This was grouped due to the large amount of ICD codes and the generalization in health outcomes by major disease or body system correlation.

- **Ventilation Used**: Indicator for mechanical ventilation during admission (0 or 1).
- **Prior Admissions Within 1 Year**: Grouped as no prior admissions (0 admissions), low frequency (1-2 admissions), moderate frequency (3-4 admissions), and high frequency (5 or more admissions). This was polychotmized because any patient with 5 or more admissions is a high frequency patient, thus making this a predictor where outliers do not add value to the distribution.

The dataset underwent rigorous preprocessing to ensure data quality. Missing values were handled using imputation techniques or exclusion based on the degree of missingness and clinical association to missingness. Continuous variables were standardized where appropriate, and categorical variables were dichotomized or grouped to enhance interpretability. This design ensures a robust and comprehensive analysis, enabling insights into the predictors of 30-day readmissions while minimizing biases introduced by data limitations.

4. Methods

4.1 Data Preparation

MIMIC-IV is a large, de-identified dataset of electronic health records from the Beth Israel Deaconess Medical Center, containing data on over 65,000 ICU admissions and over 200,000 emergency department visits [6]. The MIMIC-IV dataset underwent extensive preprocessing to ensure consistency, accuracy, and usability for statistical analysis. The key preprocessing steps included:

1. **Data Cleaning:**
    1. Duplicate records and admissions outside the inclusion criteria were removed.
    2. Date and time variables were standardized for accurate computation of derived metrics such as length of stay.
2. **Handling Missing Data:**
    1. Missing values for continuous variables, such as medication count, were imputed using median values to reduce bias.
    2. Missing categorical data were assigned an “unknown” category to maintain data completeness.
3. **Variable Transformation:**
    1. Continuous variables were inspected for skewness and transformed if necessary to meet the assumptions of statistical analysis.
    2. Categorical variables, such as discharge disposition and age group, were grouped into clinically meaningful categories.
4. **Computation of Predictors:** (Each predictor was derived from specific MIMIC-IV tables as follows)

1. **Charlson Comorbidity Index (CCI) (Categorical)**: Calculated using ICD codes from the "diagnoses_icd" table, with weights applied for comorbid conditions.
2. **Length of Stay (Continuous)**: Computed as the difference between admission and discharge times from the "admissions" table.
3. **Age Group (Categorical):** Derived from the "patients" table using anchor age and categorized into Young (18-35 years old), Middle (36-60 years old), and Old (<60 years old) groups.
4. **Discharge Disposition (Categorical):** Extracted from the "admissions" table, categorized into home, skilled nursing facility, acute hospital, and other settings.
5. **Major Diagnostic Categories (MDC) (Categorical)**: Assigned based on ICD codes grouped by body systems and disease type from the "diagnoses_icd" table.
6. **Ventilation Used (Categorical)**: Binary indicator derived from the "chartevents" or "procedureevents" table for mechanical ventilation events.
7. **Medication Count (Continuous)**: Total medications prescribed during discharge from the "prescriptions" table.
8. **Serum Creatinine (Continuous)**: Laboratory results extracted from the "labevents" table, focusing on kidney function.
9. **Prior Admissions Within 1 Year (Categorical)**: Aggregated from the "admissions" table by counting admissions within 365 days prior to the index admission.
10. **Number of Lab Tests (Continuous)**: Total number of diagnostic tests ordered during admission from the "labevents" table.
11. **Number of Diagnoses (Continuous)**: Count of ICD codes recorded during admission from the "diagnoses_icd" table.

4.2 Statistical Analysis

A multistep statistical approach was applied to examine the associations between predictors and 30-day readmissions:

1. **Descriptive Analysis**:
    1. Summary statistics were computed for all variables, stratified by readmission status.
    2. Continuous variables were summarized using means, medians, and standard deviations.
    3. Categorical variables were reported as frequencies and percentages.
    4. The prevalence of 30-day readmissions was calculated as the proportion of total admissions.
2. **Bivariate Analysis**:
    1. Initial associations between predictors and 30-day readmissions were evaluated.

2. Visualizations of variable distributions were created, including histograms and boxplots for continuous variables and bar charts for categorical variables.
3. Correlation analysis was conducted using:
    1. Mean differences for continuous predictors.
    2. Pearson correlation coefficients for linear relationships.
    3. Odds ratios for categorical variables.
    4. Chi-square tests of independence for categorical predictors.
    5. Two sample t-tests for mean differences in continuous variables.
    6. Pairwise numerical variable analysis to identify relationships and collinearity.

3. **Logistic Regression Modeling**:
    1. Logistic regression was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the association between predictors and 30-day readmissions.
    2. The model included all predictors to adjust for potential confounders and to identify independent associations.

5. Data Analysis

***Tables and Figures are found at the end of the paper**

5.1 Descriptive Analysis

The study population is a dataset of 546038 patients with 11 predictor variables grouped by the health outcome 30-day readmission. **Table 1** provides a generalized overview of the distributions of the complete dataset for all predictors by providing quantitative statistics of patients grouped by readmitted within 30 days and those not readmitted. In **Table 1**, the prevalence of 30-day readmissions is 20.34%, indicating that approximately one-fifth of admissions result in readmission. Patients with no prior admissions make up the majority (50.5%), but those with higher frequencies of prior admissions (5 or more) show increased vulnerability (15.5%). The distribution of the Charlson Comorbidity Index (CCI) reveals that most patients (93.4%) have no significant comorbidities, with a small proportion presenting higher CCI scores, which are linked to increased readmission risk. Length of stay averages 4.2 days but has substantial variability (range 0–515), reflecting diverse clinical scenarios. Medication count at discharge also varies widely, with an average of 20 prescriptions, indicating differing levels of care complexity. Serum creatinine, a marker of kidney function, has a mean of 1.21 mg/dL, with some extreme outliers indicating potential severe renal impairment. Lab tests and diagnosis counts show high variability, highlighting the complexity of care for certain patients. Categorical predictors further differentiate patients. Older adults represent the largest age group (50.7%), with higher readmission rates than younger cohorts. Discharge disposition shows that most patients return home (35.6%), but those discharged to skilled nursing facilities or other healthcare settings exhibit higher readmission risks. Diagnoses are dominated by circulatory system diseases (8.4%)

and other general categories (48.7%), reflecting the broad spectrum of medical conditions treated. Mechanical ventilation is relatively rare (1.9%), but its presence signals severe illness and a higher likelihood of readmission.

5.2 Visualizations of Predictors

Continuous Predictors:

The box and violin plots (**Figures 1 and 2**) illustrate the distributions of numerical variables for patients readmitted within 30 days (readmitted_30_days = 1) and those not readmitted (readmitted_30_days = 0). Length of stay shows the most notable difference, with the readmitted group exhibiting longer stays and a wider distribution, making it a strong potential predictor of readmission. Similarly, medication count and number of diagnoses also show higher values and broader spreads for readmitted patients, suggesting moderate predictive value. Serum creatinine levels are slightly elevated among readmitted patients, indicating possible kidney function issues, while the number of lab tests ordered shows significant overlap between the two groups, indicating a weaker association with readmission. Overall, the visualizations highlight length of stay as the most distinctive variable, while the other predictors show varying levels of separation and predictive potential.

Categorical Predictors:

The bar and stacked plots (**Figures 3 and 4**) reveal notable differences in categorical variable distributions between patients readmitted within 30 days and those not readmitted. Higher CCI scores, older age groups, and discharge dispositions such as "Home with Home Health Care" and "Skilled Nursing Facility" are more prevalent among readmitted patients, indicating potential associations with readmission. Additionally, certain diagnostic categories, including "Diseases of the Circulatory System" and "Diseases of the Respiratory System," show higher proportions of readmissions, as do patients requiring mechanical ventilation or with two or more prior admissions within a year.

5.3 Bivariate Analysis

The bivariate analysis examines **Table 1**, the quantitative data for all predictors grouped by 30 day readmission, and **Table 3**, the quantitative data for bivariate analysis, to interpret associations between individual predictors in relation to readmission within 30 days. This analysis is essential in determining correlation and significance of predictors to readmission.

Continuous Predictors:

Mean Difference:

The results in **Table 3** highlight significant mean differences in numerical predictors between the readmitted and not readmitted groups. Patients who were readmitted had, on average, a 0.998-day longer hospital stay (95% CI: 0.930, 1.067), 2.4 more medications prescribed at discharge (95% CI: 2.255, 2.546), and serum creatinine levels 0.144 units higher (95% CI: 0.130, 0.159). Additionally, readmitted patients had 49.984 more lab tests ordered (95% CI: 46.701, 53.268) and 1.641 more diagnoses recorded during their hospital stay (95% CI: 1.572, 1.710). All confidence intervals exclude 0, confirming that these differences are statistically significant. These findings indicate that patients readmitted within 30 days tend to experience longer hospital stays, more complex medication regimens, elevated serum creatinine levels, increased diagnostic evaluations, and a greater number of recorded health conditions compared to those not readmitted.

Pearson Correlation

According to **Table 3**, all the correlation coefficients are positive, indicating a positive linear relationship between each numerical predictor and the readmission outcome. This means that as the value of the predictor increases, the likelihood of being readmitted within 30 days also tends to increase.

However, all the correlation coefficients are relatively small, ranging from 0.045 to 0.088. This suggests weak linear relationships between these numerical predictors and the readmission outcome. Among the predictors, the number of diagnoses has the highest correlation coefficient (0.088), indicating a relatively slightly stronger positive relationship with readmission compared to the other predictors. Serum creatinine has the lowest correlation coefficient (0.045), suggesting the weakest linear relationship.

The correlation coefficients indicate that while there are positive linear relationships between these numerical predictors and readmission, the relationships are generally weak. This means that these predictors, when considered individually and linearly, have limited ability to strongly predict readmission.

Thus, it is important to consider the possible reasons for the weak linear relationships. One possibility could be the Pearson correlation linearity statistic. Thus, if the relationships are non-linear, the correlation coefficients might underestimate the true strength of the association.

Another potential limiter could be confounding. These correlations are bivariate, meaning they only consider the relationship between one predictor and the outcome at a time. Other factors (confounders) might influence both the predictor and the outcome, leading to greater correlations. Thus, multivariate analysis is also needed to control for confounders and assess the independent effects of each predictor. Lastly, a reason for the weakness of these predictors could be the clinical significance of them. While the correlations are statistically significant, it's

important to consider their clinical significance. Even weak correlations might be clinically relevant depending on the context and the impact of the predictor on the outcome.

Altogether, the correlation analysis suggests weak positive linear relationships between the numerical predictors (length of stay, medication count, serum creatinine, number of lab tests, and number of diagnoses) and the readmission outcome. However, these relationships should be interpreted cautiously, considering the limitations of Pearson correlation and the need for further analysis to assess the true strength and clinical significance of these associations

Two Sample T-Test Statistic

The two-sample t-test results in **Table 3** indicate statistically significant differences in the mean values of all the examined numerical predictors between patients who were readmitted within 30 days and those who were not. For length of stay, the t-statistic was 41.30 with a p-value of 0.0, suggesting that readmitted patients had significantly longer average hospital stays than non-readmitted patients. Similarly, the t-test results showed statistically significant differences in mean medication count ($t=45.04$, $p=0.0$), serum creatinine levels ($t=29.04$, $p<0.001$), number of lab tests ordered ($t=44.25$, $p=0.0$), and number of diagnoses ($t=64.20$, $p=0.0$) between the two groups. In all cases, the positive t-statistics indicate that the readmitted patients had higher average values for these predictors compared to the non-readmitted patients. This data demonstrates the statistical significance of these numerical variables in distinguishing between patients who were readmitted within 30 days and those who were not, highlighting their potential as important predictors of hospital readmission risk.

Correlation Heatmap

Figure 5, the correlation heatmap shows strong positive correlations between length of stay and medication count ($r=1.0$), serum creatinine and number of lab tests ($r=1.0$), and number of lab tests and number diagnoses of ($r=1.0$), suggesting these predictors increase linearly together. There are moderate positive correlations between length of stay and serum creatinine ($r=0.82$), and medication count and number of lab tests ($r=0.71$), indicating positive but not necessarily strong linear relationships. Weaker positive correlations exist for length of stay and number of diagnoses ($r=0.48$), and medication count and number of diagnoses ($r=0.50$). A few predictors show weak negative correlations, like serum creatinine and num diagnoses ($r=-0.21$), indicating low inverse relationships. The heatmap, **Figure 5**, provides a comprehensive view of the linear associations.

Categorical Predictors:

Chi-Square Test

The **Table 3** chi-square test results show statistically significant differences in the mean values of all the examined numerical predictors. Patients who were readmitted within 30 days had a longer average length of stay ($t=41.30$, $p<0.001$), higher medication count ($t=45.04$, $p<0.001$), elevated serum creatinine levels ($t=29.04$, $p<0.001$), more laboratory tests ordered ($t=44.25$, $p<0.001$), and a greater number of diagnoses ($t=64.20$, $p<0.001$) compared to those who were not readmitted. Thereby the observations indicate that higher values for these predictors are associated with an increased risk of 30-day hospital readmission.

Odds Ratio:

The results in **Table 1 and 3** show the odds ratio measuring the odds of an event occurring in one group to the odds of it occurring in another group in each categorical predictor, and unique event, between the readmitted and not readmitted groups, along with the 95% confidence intervals for these differences.

In analyzing the data, there is a clear trend that higher CCI scores (6 and 7) are strongly associated with increased odds of readmission. Patients with these scores are significantly more likely to be readmitted, and patients with lower CCI scores (0) are associated with decreased odds of readmission. These patients are less likely to be readmitted. Also, it is interesting that CCI scores of 2, 3, and 8 do not show statistically significant associations with readmission, although there might be a slight tendency towards increased odds for CCI score 2 and 8, and decreased odds for CCI score 3. This could potentially mean that a CCI score of 8 means an increased risk of death which would also lead to a lack of readmission.

When focusing on age groups, the association between the middle age group and readmission has a statistically significant increased risk of readmission compared to the young age group. Additionally, the young age group has a slightly lower risk of readmission compared to the old age group, but the difference is relatively small and might not be clinically significant.

In determining correlation between discharge disposition, individual increased and decreased odds of readmission are found to be statistically significant. Discharge to a psych facility, against medical advice, or to an acute hospital are the strongest predictors of readmission, with significantly higher odds. Additionally, discharge to chronic/long-term acute care and home health care also increase the odds of readmission. Conversely, discharge to hospice, rehab, and home (without home health care) are associated with lower odds of readmission. This suggests that these discharge dispositions might be protective against readmission. Discharge to a skilled nursing facility shows a slight increase in readmission odds, but the magnitude of the effect is relatively small. Discharge dispositions to other facilities, assisted living, and healthcare facilities do not have statistically significant associations with readmission. It is important to note that patients who died during their hospital stay obviously have significantly lower odds of readmission (OR: 0.005, CI: 0.003, 0.009), and this is expected and serves as a validation of the data.

Determining the strongest predictive categories of major diagnostic categories include: external causes of morbidity and mortality (OR: 3.483), mental disorders (OR: 2.668), mental and behavioral disorders (OR: 1.848). These categories have the highest odds ratios, indicating a significantly increased likelihood of readmission for patients with these diagnoses. Comparatively, the relatively lower predictive categories of major diagnostic categories for readmission include: complications of pregnancy, childbirth, and the puerperium (OR: 0.168), diseases of the nervous system (OR: 0.766), symptoms, signs, and abnormal clinical and laboratory findings (OR: 0.713), injury, poisoning, and certain other consequences of external causes (OR: 0.748), diseases of the respiratory system (OR: 0.858), diseases of the circulatory system (OR: 0.769), diseases of the nervous system and sense organs (OR: 0.861), certain infectious and parasitic diseases (OR: 0.947), diseases of the skin and subcutaneous tissue (OR: 0.764), diseases of the musculoskeletal system and connective tissue (OR: 0.604), pregnancy, childbirth, puerperium, and diseases of the ear and mastoid process (OR: 0.290). The major diagnostic categories with no significant association are diseases of the genitourinary system, endocrine, nutritional, and metabolic diseases, infectious and parasitic diseases, diseases of the eye and adnexa, congenital malformations and chromosomal abnormalities, symptoms, signs, and ill-defined conditions, certain conditions originating in the perinatal period. And, slightly increased odds of readmission but could possibly be not clinically significant include: diseases of the blood and blood-forming organs and immune disorders (OR: 1.467), neoplasms (OR: 1.320), diseases of the digestive system (OR: 1.034), injury and poisoning (OR: 1.271), unknown MDC category (OR: 1.336).

The observation of the use of ventilation during the hospital stay is associated with a lower likelihood of 30-day readmission. This might seem counterintuitive, but it could be due to several factors. One factor might be patients who require ventilation might receive more intensive care and monitoring during their hospital stay, leading to better management of their conditions and reduced risk of readmission. Another factor could be patients who use ventilation might have more severe conditions, and some of them might not survive to be readmitted within 30 days. Lastly, there could be other confounding factors influencing this relationship that are not captured in this analysis.

The results of prior admissions within 1 year clearly show that as the frequency of prior admissions in the past year increases, the odds of 30-day readmission also increase. In contrast, patients with no prior admissions have the lowest risk of readmission. It is also found that patients with a high frequency of prior admissions (5 or more) have the highest risk of readmission. Overall, highlighting the importance of considering a patient's prior hospitalization history when assessing their risk of readmission.

Altogether, **Tables 1 and 3** clearly portray that several factors are found to be significantly associated with 30-day readmission risk, including CCI score, age group, discharge disposition, major diagnostic category, ventilation use, and prior admissions. High-risk factors for

readmission include higher CCI scores, discharge to psych facilities or against medical advice, certain major diagnostic categories (e.g., mental disorders, external causes), and a high frequency of prior admissions. Protective factors for readmission include lower CCI scores, discharge to hospice or rehab, and certain major diagnostic categories (e.g., diseases of the nervous system, musculoskeletal system). Ventilation use is associated with lower readmission odds, but could possibly be related to confounding or clinical significance.

5.4 Multivariable Analysis

A multivariate logistic regression model was constructed to analyze the predictors for readmission within 30 days to address confounding factors that potentially affected the bivariate analysis, and reinforce the statistical and significance of predictors found prior.

By examining **Table 2**, the analysis reveals several significant findings regarding predictors of 30-day hospital readmissions. The baseline odds of readmission, when all predictors are at their reference levels (CCI = 0, Age Group = Young, Discharge Disposition = Home, MDC = Other, Ventilation Use = 0, Prior Admissions = 0), are low (OR = 0.1336), with a narrow confidence interval (CI = 0.1293, 0.1381), suggesting a precise estimate. The Charlson Comorbidity Index (CCI) exhibits a nuanced relationship with readmissions. While higher CCI scores, such as 6 (OR = 1.2182), significantly increase readmission odds, intermediate scores, such as 2 (OR = 0.8349), indicate reduced risk compared to the reference group (CCI = 0). Non-significant results for scores like CCI = 7 highlight potential data limitations or insufficient sample size within certain categories. Age group also influences readmission risk, with older (OR = 0.8204) and middle-aged (OR = 0.9329) patients showing slightly reduced odds compared to younger patients. This trend may reflect differential healthcare utilization or discharge planning practices among age groups. Discharge disposition demonstrates strong associations, particularly for patients discharged to psychiatric facilities (OR = 4.7510), against medical advice (OR = 1.8222), or to home health care (OR = 1.2332), who face elevated readmission risks. Conversely, discharges to hospice (OR = 0.4351) or rehabilitation (OR = 0.9104) are associated with reduced odds, reflecting potentially more comprehensive post-discharge care or end-of-life considerations. Among Major Diagnostic Categories (MDCs), conditions involving mental disorders (OR = 2.7120) or external causes (OR = 3.4462) strongly predict readmissions, likely due to the complex nature of these conditions. In contrast, circulatory system diseases (OR = 0.6899) and musculoskeletal conditions (OR = 0.5854) are associated with lower odds, potentially reflecting better-managed chronic care pathways for these conditions. The numerical predictors further illuminate the drivers of readmissions. Longer hospital stays marginally reduce readmission odds (OR = 0.9913), possibly due to more thorough inpatient care. Higher medication counts at discharge modestly increase risk (OR = 1.0065), reflecting the complexity of care for patients with extensive prescriptions. Serum creatinine levels, a marker of renal function, are significantly associated with increased odds of readmission (OR = 1.0311), emphasizing the impact of renal dysfunction on post-discharge outcomes. Additionally, a higher

number of diagnoses per admission (OR = 1.0167) correlates with greater readmission risk, highlighting the burden of comorbidities and complications.

Overall, **Table 2** shows that significant predictors of readmission include high Charlson Comorbidity Index scores, discharge against advice, frequent prior admissions, and conditions related to external causes or mental health. Also, continuous variables like length of stay and medication count also show modest but significant effects. Confidence intervals are narrow for most variables, indicating precise estimates, except for rare categories like "Died" or certain MDCs.

The multivariate analysis confirms many observed associations, highlighting key predictors of 30-day readmissions while acknowledging potential confounding factors. Higher CCI scores (e.g., 6 and 7) significantly increase readmission risk, aligning with the elevated burden of comorbidities, while lower scores and protective discharge settings (e.g., hospice, rehab) reduce risk. Age group findings suggest middle-aged and older patients are less likely to be readmitted, potentially reflecting differences in healthcare utilization and support. Discharge to psychiatric facilities, against medical advice, or to acute hospitals is strongly predictive of readmission, underscoring the role of care transitions. Some results, such as reduced readmissions with ventilation use, may reflect confounding from mortality or intensive care. Similarly, the strong association of frequent prior admissions with readmission risk highlights chronic disease burden but may be influenced by unmeasured socioeconomic factors.

Overall, the multivariate analysis largely reinforces the observed associations in the data, while highlighting the role of confounding factors. Adjusting for multiple variables clarifies the true effects of predictors like CCI scores, age group, and discharge disposition, lending confidence to these findings. However, some results, such as the protective effect of ventilation use or non-significant associations for certain MDCs, suggest areas where confounding or data limitations may obscure the relationships. This analysis emphasizes the complexity of 30-day readmissions and the need for comprehensive models to disentangle direct effects from confounding influences.

6. Results

The study analyzed 546,038 patients from the MIMIC-IV database, revealing a 20.34% prevalence of 30-day readmissions. Patients who were readmitted exhibited distinct characteristics, including longer hospital stays (mean: 5.0 days compared to 4.0 days for non-readmitted patients), higher medication counts (22.0 vs. 19.6), and elevated serum creatinine levels (1.3 vs. 1.2 mg/dL). The majority of patients had no prior admissions (50.5%), but those with five or more prior admissions had the highest readmission rates at 38.4%. Analysis of the Charlson Comorbidity Index (CCI) showed that higher scores, such as 6 or 7, were strongly associated with increased readmission risk, while lower scores, like 0 or 1, were protective. Also

discharge dispositions played a significant role, with patients discharged to psychiatric facilities (52.3%) or against medical advice (34.9%) exhibiting the highest readmission rates. In contrast, those discharged to hospice (14.6%) or rehabilitation (18.4%) had the lowest rates. Among age groups, older adults (60+ years) comprised the largest cohort (50.7%), with slightly higher readmission rates than younger groups.

Bivariate analysis confirmed significant mean differences in continuous predictors. Patients who were readmitted had, on average, a 0.998-day longer hospital stay ($p < 0.001$), were prescribed 2.4 more medications at discharge ($p < 0.001$), and exhibited serum creatinine levels that were 0.144 mg/dL higher ($p < 0.001$). Pearson correlation coefficients indicated positive but weak linear relationships between predictors and readmission likelihood. For instance, the number of diagnoses showed the strongest correlation ($r = 0.088$, $p < 0.001$). Chi-square tests for categorical predictors revealed strong associations, with discharge to psychiatric facilities ($\chi^2 = 26{,}726$, $p < 0.001$) and high-frequency prior admissions being key risk factors. Conversely, protective factors included lower CCI scores and hospice discharge.

Multivariable logistic regression further highlighted significant predictors of 30-day readmissions. Higher CCI scores, such as 6 (OR = 1.218, $p < 0.001$), significantly increased readmission odds, while intermediate scores like 2 (OR = 0.835, $p = 0.001$) were associated with reduced risk. Discharge dispositions had a notable impact; patients discharged to psychiatric facilities (OR = 4.751, $p < 0.001$) or against medical advice (OR = 1.822, $p < 0.001$) were at greater risk, whereas discharge to hospice (OR = 0.435, $p < 0.001$) was protective. Older patients showed slightly reduced odds of readmission compared to younger individuals (OR = 0.820, $p < 0.001$). Continuous predictors, including medication count (OR = 1.007, $p < 0.001$) and serum creatinine levels (OR = 1.031, $p < 0.001$), had minimal correlation but are statistically significant contributors to readmission risk.

Moreover, **figures 1 and 2** highlighted longer lengths of stay and higher medication counts among readmitted patients, while **figures 3 and 4** illustrated increased readmission rates in older adults and those discharged to psychiatric or skilled nursing facilities. The correlation heatmap revealed strong positive correlations between certain predictors, such as length of stay and medication count.

Overall, the analysis identified critical predictors of 30-day readmissions, including higher CCI scores, discharge dispositions, and prior admission frequency. Protective factors, such as hospice discharge, were also evident. These findings emphasize the complexity of hospital readmissions and highlight the need for targeted interventions to reduce risk.

7. Discussion and Limitations

7.1 Discussion

This study provides a comprehensive analysis of 30-day hospital readmissions, identifying significant predictors using the MIMIC-IV database that are consistent with prior research. These statistically and clinically significant factors such as higher Charlson Comorbidity Index (CCI) scores, frequent prior admissions, and discharge to high-risk settings, including psychiatric facilities and skilled nursing facilities, were strongly associated with increased readmission odds [8]. These findings underscore the critical role of patient complexity and post-discharge care environments in predicting readmissions, while also identifying the protective effect of discharge to hospice or rehabilitation aligns with literature emphasizing the value of structured, supportive care in reducing readmissions.

Clinical implications of these findings are significant. Targeted interventions for high-risk groups, such as patients with elevated CCI scores or those discharged to psychiatric or skilled nursing facilities, could mitigate readmission risks. Enhanced post-discharge care planning, including patient education and robust follow-up protocols, may be particularly effective for these populations. Furthermore, integrating predictive models into clinical workflows could aid in identifying vulnerable patients before discharge, enabling timely interventions to prevent readmissions. For instance, patients with high medication counts at discharge may benefit from medication reconciliation services or pharmacist-led counseling to improve adherence and reduce complexity.

An unexpected finding was the lower odds observed for readmission among patients who used mechanical ventilation during their hospital stay. While counterintuitive, this may reflect the intensive monitoring and care received by ventilated patients, leading to better-managed conditions and fewer complications post-discharge. Alternatively, it is possible that higher in-hospital mortality among ventilated patients reduced the likelihood of subsequent readmissions, introducing a potential survivorship bias. This finding warrants further exploration to disentangle these effects.

7.2 Limitations

Despite the strengths of this study, including its large sample size and robust statistical methods, there are limitations that should be acknowledged. The retrospective design inherently limits causal inferences, as observed associations may be influenced by unmeasured confounding variables. For example, socioeconomic factors, health literacy, and access to outpatient care were not captured in this analysis but likely play a role in readmission risks [9]. Measurement biases, such as inaccuracies in coding or data imputation methods, may also affect the results. Additionally, generalizability is restricted to patient populations and healthcare systems similar to those represented in the MIMIC-IV database, limiting the applicability of findings to other settings. Also, it is important to note that the sample only includes ED and ICU patients and does not provide a clear picture of the whole medical admission population by excluding other departments.

8. Conclusion

In conclusion, this study highlights key predictors of 30-day hospital readmissions, including higher CCI scores, frequent prior admissions, and discharge to high-risk care settings. These findings have important implications for healthcare providers and policymakers, emphasizing the need for targeted interventions and enhanced care coordination to reduce preventable readmissions. Future research should aim to integrate additional predictors, such as social determinants of health, and validate these findings in diverse patient populations and healthcare systems. Prospective studies evaluating the effectiveness of tailored interventions for high-risk groups could further advance our understanding of how to optimize care transitions and improve patient outcomes.

9. Tables and Figures

Table 1: Quantitative Data for All Predictors Grouped by 30 Day Readmission

	Grouped by readmitted_30_days								
		Missing	**Overall**	**0**	**1**	**Mean Diff/OR**	**CI (L, U)**	**P-Value**	**Test**
n			546038	434993	111045				
cci_score, n (%)	0		510143 (100.0)	407478 (79.9)	102665 (20.1)	0.827	(0.807, 0.849)	<0.001	Chi-squared (warning: expected count < 5)
	1		29895 (100.0)	23160 (77.5)	6735 (22.5)	REF			
	2		2240 (100.0)	1771 (79.1)	469 (20.9)	1.038	(0.937, 1.149)		
	3		26 (100.0)	23 (88.5)	3 (11.5)	0.511	(0.153, 1.702)		
	6		3522 (100.0)	2414 (68.5)	1108 (31.5)	1.806	(1.682, 1.940)		
	7		200 (100.0)	138 (69.0)	62 (31.0)	1.76	(1.304, 2.376)		
	8		12 (100.0)	9 (75.0)	3 (25.0)	1.306	(0.354, 4.823)		
length_of_stay, mean (SD)		0	4.2 (7.2)	4.0 (6.9)	5.0 (8.2)	0.998	(0.930, 1.067)	<0.001	Two Sample T-test
age_group, n (%)	Middle		187306 (100.0)	146139 (78.0)	41167 (22.0)	REF		<0.001	Chi-squared
	Old		277130 (100.0)	221988 (80.1)	55142 (19.9)	0.946	(0.934, 0.959)		
	Young		81602 (100.0)	66866 (81.9)	14736 (18.1)	0.842	(0.826, 0.859)		
discharge_disposition, n (%)	ACUTE HOSPITAL		2334 (100.0)	1583 (67.8)	751 (32.2)	1.864	(1.709, 2.034)	<0.001	Chi-squared

	AGAINST ADVICE		3393 (100.0)	2208 (65.1)	1185 (34.9)	2.114	(1.970, 2.270)		
	ASSISTED LIVING		622 (100.0)	492 (79.1)	130 (20.9)	1.035	(0.853, 1.256)		
	CHRONIC/LONG TERM ACUTE CARE		8125 (100.0)	5817 (71.6)	2308 (28.4)	1.566	(1.492, 1.644)		
	DIED		11722 (100.0)	11706 (99.9)	16 (0.1)	0.005	(0.003, 0.009)		
	HEALTHCARE FACILITY		50 (100.0)	42 (84.0)	8 (16.0)	0.746	(0.350, 1.589)		
	HOME		194205 (100.0)	159810 (82.3)	34395 (17.7)	REF			
	HOME HEALTH CARE		99305 (100.0)	74061 (74.6)	25244 (25.4)	1.434	(1.411, 1.457)		
	HOSPICE		5397 (100.0)	4608 (85.4)	789 (14.6)	0.668	(0.620, 0.721)		
	OTHER FACILITY		1592 (100.0)	1260 (79.1)	332 (20.9)	1.032	(0.915, 1.165)		
	PSYCH FACILITY		2965 (100.0)	1413 (47.7)	1552 (52.3)	4.349	(4.046, 4.676)		
	REHAB		13845 (100.0)	11295 (81.6)	2550 (18.4)	0.882	(0.844, 0.921)		
	SKILLED NURSING FACILITY		52658 (100.0)	40900 (77.7)	11758 (22.3)	1.141	(1.117, 1.166)		
	UNKNOWN		149825 (100.0)	119798 (80.0)	30027 (20.0)	0.975	(0.961, 0.990)		
mdc_category, n (%)	Certain Conditions Originating in the Perinatal Period		3 (100.0)	3 (100.0)		0.653	(0.033, 13.034)	<0.001	Chi-squared (warning: expected count < 5)
	Certain Infectious and Parasitic Diseases		12472 (100.0)	10041 (80.5)	2431 (19.5)	0.947	(0.906, 0.991)		
	Complications of Pregnancy, Childbirth, and the Puerperium		195 (100.0)	187 (95.9)	8 (4.1)	0.168	(0.083, 0.340)		
	Congenital Malformations and Chromosomal Abnormalities		983 (100.0)	805 (81.9)	178 (18.1)	0.866	(0.736, 1.019)		
	Diseases of the Blood and Blood-forming Organs and Immune Disorders		3288 (100.0)	2394 (72.8)	894 (27.2)	1.467	(1.358, 1.584)		
	Diseases of the Circulatory System		45956 (100.0)	38287 (83.3)	7669 (16.7)	0.769	(0.749, 0.788)		
	Diseases of the Digestive System		31424 (100.0)	24872 (79.1)	6552 (20.9)	1.034	(1.005, 1.063)		
	Diseases of the Ear and Mastoid Process		897 (100.0)	835 (93.1)	62 (6.9)	0.29	(0.224, 0.376)		
	Diseases of the Eye and Adnexa		78 (100.0)	66 (84.6)	12 (15.4)	0.712	(0.385, 1.317)		
	Diseases of the Genitourinary System		10796 (100.0)	8643 (80.1)	2153 (19.9)	0.975	(0.930, 1.023)		

	Diseases of the Musculoskeletal System and Connective Tissue		12789 (100.0)	11066 (86.5)	1723 (13.5)	0.604	(0.574, 0.635)		
	Diseases of the Nervous System		7788 (100.0)	6510 (83.6)	1278 (16.4)	0.766	(0.721, 0.814)		
	Diseases of the Nervous System and Sense Organs		4163 (100.0)	3412 (82.0)	751 (18.0)	0.861	(0.796, 0.932)		
	Diseases of the Respiratory System		16944 (100.0)	13889 (82.0)	3055 (18.0)	0.858	(0.824, 0.892)		
	Diseases of the Skin and Subcutaneous Tissue		4400 (100.0)	3681 (83.7)	719 (16.3)	0.764	(0.705, 0.827)		
	Endocrine, Nutritional, and Metabolic Diseases		9668 (100.0)	7700 (79.6)	1968 (20.4)	1.001	(0.952, 1.052)		
	External Causes of Morbidity and Mortality		6607 (100.0)	3530 (53.4)	3077 (46.6)	3.483	(3.317, 3.658)		
	Factors Influencing Health Status and Contact with Health Services		6242 (100.0)	3283 (52.6)	2959 (47.4)	3.6	(3.424, 3.785)		
	Infectious and Parasitic Diseases		2130 (100.0)	1700 (79.8)	430 (20.2)	0.991	(0.891, 1.102)		
	Injury and Poisoning		1376 (100.0)	1039 (75.5)	337 (24.5)	1.271	(1.124, 1.438)		
	Injury, Poisoning, and Certain Other Consequences of External Causes		29989 (100.0)	25119 (83.8)	4870 (16.2)	0.748	(0.725, 0.772)		
	Mental Disorders		3542 (100.0)	2114 (59.7)	1428 (40.3)	2.668	(2.494, 2.854)		
	Mental and Behavioural Disorders		19295 (100.0)	13217 (68.5)	6078 (31.5)	1.848	(1.791, 1.906)		
	Neoplasms		17886 (100.0)	13412 (75.0)	4474 (25.0)	1.32	(1.275, 1.366)		
	Other		266190 (100.0)	212467 (79.8)	53723 (20.2)	REF			
	Pregnancy, Childbirth, and the Puerperium		11870 (100.0)	10674 (89.9)	1196 (10.1)	0.433	(0.408, 0.460)		
	Symptoms, Signs, and Abnormal Clinical and Laboratory Findings		18459 (100.0)	15591 (84.5)	2868 (15.5)	0.713	(0.685, 0.743)		
	Symptoms, Signs, and Ill-defined Conditions		77 (100.0)	60 (77.9)	17 (22.1)	1.11	(0.648, 1.902)		
	Unknown		531 (100.0)	396 (74.6)	135 (25.4)	1.336	(1.099, 1.624)		
ventilation_used, n (%)	0		535561 (100.0)	426067 (79.6)	109494 (20.4)	REF		<0.001	Chi-squared
	1		10477 (100.0)	8926 (85.2)	1551 (14.8)	0.676	(0.640, 0.714)		
medication_count, mean (SD)		0	20.1 (15.9)	19.6 (15.6)	22.0 (16.9)	2.4	(2.255, 2.546)	<0.001	Two Sample T-test
serum_creatinine, mean (SD)		130364	1.2 (1.3)	1.2 (1.2)	1.3 (1.5)	0.144	(0.130, 0.159)	<0.001	Two Sample T-test

prior_admissions_1yr, n (%)	High frequency (5 or more admissions)		84453 (100.0)	52005 (61.6)	32448 (38.4)	3.04	(2.993, 3.089)	<0.001	Chi-squared
	Low frequency (1-2 admissions)		137810 (100.0)	109509 (79.5)	28301 (20.5)	1.017	(1.001, 1.032)		
	Moderate frequency (3-4 admissions)		48019 (100.0)	34767 (72.4)	13252 (27.6)	1.56	(1.527, 1.593)		
	No prior admissions (0 admissions)		275756 (100.0)	238712 (86.6)	37044 (13.4)	REF			
num_lab_tests, mean (SD)		0	154.9 (336.6)	144.8 (318.4)	194.8 (397.3)	49.984	(46.701, 53.268)	<0.001	Two Sample T-test
num_diagnoses, mean (SD)		0	11.7 (7.6)	11.3 (7.5)	13.0 (8.0)	1.641	(1.572, 1.710)	<0.001	Two Sample T-test

Table 2: Odds Ratio Data for The Logistic Regression Model

Predictor	OR	Lower CI	Upper CI	P value	Z stat
Intercept	0.1336	0.1293	0.1381	0.0000	-120.0859
cci_score[T.1]	0.9704	0.9399	1.0019	0.0656	-1.8410
cci_score[T.2]	0.8349	0.7481	0.9318	0.0013	-3.2214
cci_score[T.3]	0.3287	0.0955	1.1310	0.0776	-1.7647
cci_score[T.6]	1.2182	1.1168	1.3287	0.0000	4.4538
cci_score[T.7]	1.0398	0.7510	1.4397	0.8139	0.2353
cci_score[T.8]	0.7911	0.0934	6.6984	0.8297	-0.2150
C(age_group, Treatment('Young'))[T.Middle]	0.9329	0.9056	0.9611	0.0000	-4.5754
C(age_group, Treatment('Young'))[T.Old]	0.8204	0.7961	0.8455	0.0000	-12.9015
C(discharge_disposition, Treatment('HOME'))[T.ACUTE HOSPITAL]	1.3496	1.2190	1.4943	0.0000	5.7731
C(discharge_disposition, Treatment('HOME'))[T.AGAINST ADVICE]	1.8222	1.6820	1.9741	0.0000	14.6860
C(discharge_disposition, Treatment('HOME'))[T.ASSISTED LIVING]	0.9913	0.8060	1.2191	0.9337	-0.0832
C(discharge_disposition, Treatment('HOME'))[T.CHRONIC/LONG TERM ACUTE CARE]	1.0527	0.9955	1.1132	0.0717	1.8008
C(discharge_disposition, Treatment('HOME'))[T.DIED]	0.0028	0.0017	0.0048	0.0000	-21.6173
C(discharge_disposition, Treatment('HOME'))[T.HEALTHCARE FACILITY]	0.7961	0.3653	1.7350	0.5662	-0.5736
C(discharge_disposition, Treatment('HOME'))[T.HOME HEALTH CARE]	1.2332	1.2073	1.2597	0.0000	19.3313
C(discharge_disposition, Treatment('HOME'))[T.HOSPICE]	0.4351	0.4008	0.4724	0.0000	-19.8222
C(discharge_disposition, Treatment('HOME'))[T.OTHER FACILITY]	0.9108	0.7994	1.0377	0.1605	-1.4035
C(discharge_disposition, Treatment('HOME'))[T.PSYCH FACILITY]	4.7510	4.3837	5.1491	0.0000	37.9552
C(discharge_disposition, Treatment('HOME'))[T.REHAB]	0.9104	0.8674	0.9556	0.0001	-3.7982
C(discharge_disposition, Treatment('HOME'))[T.SKILLED NURSING FACILITY]	1.0643	1.0352	1.0942	0.0000	4.4025

C(discharge_disposition, Treatment('HOME'))[T.UNKNOWN]	0.9868	0.9622	1.0120	0.3018	-1.0326
C(mdc_category, Treatment('Other'))[T.Certain Conditions Originating in the Perinatal Period]	0.0000	0.0000	inf	0.9989	-0.0014
C(mdc_category, Treatment('Other'))[T.Certain Infectious and Parasitic Diseases]	0.7177	0.6828	0.7544	0.0000	-13.0457
C(mdc_category, Treatment('Other'))[T.Complications of Pregnancy, Childbirth, and the Puerperium]	0.2836	0.0677	1.1877	0.0846	-1.7246
C(mdc_category, Treatment('Other'))[T.Congenital Malformations and Chromosomal Abnormalities]	0.8616	0.7269	1.0212	0.0858	-1.7179
C(mdc_category, Treatment('Other'))[T.Diseases of the Blood and Blood-forming Organs and Immune Disorders]	1.0419	0.9585	1.1326	0.3347	0.9647
C(mdc_category, Treatment('Other'))[T.Diseases of the Circulatory System]	0.6899	0.6700	0.7104	0.0000	-24.8885
C(mdc_category, Treatment('Other'))[T.Diseases of the Digestive System]	0.8924	0.8645	0.9212	0.0000	-7.0212
C(mdc_category, Treatment('Other'))[T.Diseases of the Ear and Mastoid Process]	0.3387	0.2573	0.4459	0.0000	-7.7175
C(mdc_category, Treatment('Other'))[T.Diseases of the Eye and Adnexa]	0.5094	0.2477	1.0474	0.0667	-1.8340
C(mdc_category, Treatment('Other'))[T.Diseases of the Genitourinary System]	0.8042	0.7623	0.8485	0.0000	-7.9646
C(mdc_category, Treatment('Other'))[T.Diseases of the Musculoskeletal System and Connective Tissue]	0.5854	0.5523	0.6205	0.0000	-18.0376
C(mdc_category, Treatment('Other'))[T.Diseases of the Nervous System]	0.6967	0.6523	0.7442	0.0000	-10.7495
C(mdc_category, Treatment('Other'))[T.Diseases of the Nervous System and Sense Organs]	0.7822	0.7180	0.8523	0.0000	-5.6138
C(mdc_category, Treatment('Other'))[T.Diseases of the Respiratory System]	0.6965	0.6664	0.7280	0.0000	-16.0410
C(mdc_category, Treatment('Other'))[T.Diseases of the Skin and Subcutaneous Tissue]	0.6481	0.5901	0.7118	0.0000	-9.0716
C(mdc_category, Treatment('Other'))[T.Endocrine, Nutritional, and Metabolic Diseases]	0.6873	0.6494	0.7275	0.0000	-12.9349
C(mdc_category, Treatment('Other'))[T.External Causes of Morbidity and Mortality]	3.4462	3.2487	3.6557	0.0000	41.0854
C(mdc_category, Treatment('Other'))[T.Factors Influencing Health Status and Contact with Health Services]	3.0712	2.8872	3.2669	0.0000	35.6014
C(mdc_category, Treatment('Other'))[T.Infectious and Parasitic Diseases]	0.8401	0.7488	0.9426	0.0030	-2.9674
C(mdc_category, Treatment('Other'))[T.Injury and Poisoning]	1.0530	0.9228	1.2015	0.4431	0.7670
C(mdc_category, Treatment('Other'))[T.Injury, Poisoning, and Certain Other Consequences of External Causes]	0.6633	0.6388	0.6887	0.0000	-21.4126
C(mdc_category, Treatment('Other'))[T.Mental Disorders]	2.7120	2.2881	3.2144	0.0000	11.5056
C(mdc_category, Treatment('Other'))[T.Mental and Behavioural Disorders]	1.2616	1.1831	1.3454	0.0000	7.0871
C(mdc_category, Treatment('Other'))[T.Neoplasms]	1.3721	1.3187	1.4276	0.0000	15.6303
C(mdc_category, Treatment('Other'))[T.Pregnancy, Childbirth, and the Puerperium]	0.6144	0.5543	0.6811	0.0000	-9.2715
C(mdc_category, Treatment('Other'))[T.Symptoms, Signs, and Abnormal Clinical and Laboratory Findings]	0.7254	0.6878	0.7650	0.0000	-11.8272
C(mdc_category, Treatment('Other'))[T.Symptoms, Signs, and Ill-defined Conditions]	0.5722	0.3230	1.0135	0.0556	-1.9139
C(mdc_category, Treatment('Other'))[T.Unknown]	1.5927	1.1502	2.2055	0.0051	2.8027

ventilation_used[T.1]	0.6874	0.6433	0.7346	0.0000	-11.0710
C(prior_admissions_1yr, Treatment('No prior admissions (0 admissions)'))[T.High frequency (5 or more admissions)]	3.0973	3.0299	3.1661	0.0000	100.8248
C(prior_admissions_1yr, Treatment('No prior admissions (0 admissions)'))[T.Low frequency (1-2 admissions)]	1.5445	1.5141	1.5754	0.0000	42.8977
C(prior_admissions_1yr, Treatment('No prior admissions (0 admissions)'))[T.Moderate frequency (3-4 admissions)]	2.0866	2.0322	2.1425	0.0000	54.5710
length_of_stay	0.9913	0.9893	0.9932	0.0000	-8.8849
medication_count	1.0065	1.0056	1.0074	0.0000	13.9499
serum_creatinine	1.0311	1.0254	1.0368	0.0000	10.8373
num_lab_tests	1.0004	1.0004	1.0005	0.0000	18.1081
num_diagnoses	1.0167	1.0153	1.0182	0.0000	22.5785

Table 3: Quantitative Data for Bivariate Analysis

Variable	OR (CI)	Chi-square (P-value) (DoF)	Mean Diff (CI)	Pearson Correlation (CI)	T-Test (P-Value)
length_of_stay			0.998 (0.93, 1.067)	0.248 (0.237, 0.260)	41.303 (0)
medication_count			2.4 (2.255, 2.556)	0.122 (0.117, 0.128)	45.041 (0)
serum_creatinine			0.144 (0.13, 0.129)	0.122 (0.115, 0.13)	29.038 (3.342E-185)
num_lab_tests			49.984 (46.701, 53.268)	0.345 (0.33, 0.361)	44.249 (0)
num_diagnoses			1.641 (1.572, 1.71)	0.145 (0.141, 0.149)	64.2 (0)
cci_score_1	REF	387.712 (1.22E-80) (6)			
cci_score_0	0.827 (0.807, 0.849)				
cci_score_2	1.038 (0.937, 1.149)				
cci_score_6	1.806 (1.682, 1.940)				
cci_score_7	1.760 (1.304, 2.376)				
cci_score_3	0.511 (0.153, 1.702)				
cci_score_8	1.306 (0.354, 4.823)				
age_group_Middle	REF	606.078 (2.46E-132) (2)			
age_group_Young	0.842 (0.826, 0.859)				
age_group_Old	0.946 (0.934, 0.959)				
discharge_disposition_HOME	REF	8490.961 (0) (13)			
discharge_disposition_HOSPICE	0.668 (0.620, 0.721)				
discharge_disposition_UNKNOWN	0.975 (0.961, 0.990)				
discharge_disposition_HOME HEALTH CARE	1.434 (1.411, 1.457)				
discharge_disposition_SKILLED NURSING FACILITY	1.141 (1.117, 1.166)				
discharge_disposition_REHAB	0.882 (0.844, 0.921)				
discharge_disposition_DIED	0.005 (0.003, 0.009)				
discharge_disposition_CHRONIC/LONG TERM ACUTE CARE	1.566 (1.492, 1.644)				
discharge_disposition_PSYCH FACILITY	4.349 (4.046, 4.676)				
discharge_disposition_ACUTE HOSPITAL	1.864 (1.709, 2.034)				
discharge_disposition_AGAINST ADVICE	2.114 (1.970, 2.270)				

discharge_disposition_OTHER FACILITY	1.032 (0.915, 1.165)				
discharge_disposition_ASSISTED LIVING	1.035 (0.853, 1.256)				
discharge_disposition_HEALTHCARE FACILITY	0.746 (0.350, 1.589)				
mdc_category_Other	REF	10780.123 (0) (28)			
mdc_category_Diseases of the Nervous System	0.766 (0.721, 0.814)				
mdc_category_Symptoms, Signs, and Abnormal Clinical and Laboratory Findings	0.713 (0.685, 0.743)				
mdc_category_Injury, Poisoning, and Certain Other Consequences of External Causes	0.748 (0.725, 0.772)				
mdc_category_Diseases of the Blood and Blood-forming Organs and Immune Disorders	1.467 (1.358, 1.584)				
mdc_category_Diseases of the Respiratory System	0.858 (0.824, 0.892)				
mdc_category_Neoplasms	1.320 (1.275, 1.366)				
mdc_category_Mental and Behavioural Disorders	1.848 (1.791, 1.906)				
mdc_category_Diseases of the Circulatory System	0.769 (0.749, 0.788)				
mdc_category_External Causes of Morbidity and Mortality	3.483 (3.317, 3.658)				
mdc_category_Complications of Pregnancy, Childbirth, and the Puerperium	0.168 (0.083, 0.340)				
mdc_category_Diseases of the Genitourinary System	0.975 (0.930, 1.023)				
mdc_category_Certain Infectious and Parasitic Diseases	0.947 (0.906, 0.991)				
mdc_category_Diseases of the Digestive System	1.034 (1.005, 1.063)				
mdc_category_Mental Disorders	2.668 (2.494, 2.854)				
mdc_category_Diseases of the Nervous System and Sense Organs	0.861 (0.796, 0.932)				
mdc_category_Diseases of the Skin and Subcutaneous Tissue	0.764 (0.705, 0.827)				
mdc_category_Endocrine, Nutritional, and Metabolic Diseases	1.001 (0.952, 1.052)				
mdc_category_Diseases of the Musculoskeletal System and Connective Tissue	0.604 (0.574, 0.635)				
mdc_category_Pregnancy, Childbirth, and the Puerperium	0.433 (0.408, 0.460)				
mdc_category_Factors Influencing Health Status and Contact with Health Services	3.600 (3.424, 3.785)				
mdc_category_Infectious and Parasitic Diseases	0.991 (0.891, 1.102)				
mdc_category_Injury and Poisoning	1.271 (1.124, 1.438)				
mdc_category_Diseases of the Eye and Adnexa	0.712 (0.385, 1.317)				
mdc_category_Diseases of the Ear and Mastoid Process	0.290 (0.224, 0.376)				
mdc_category_Unknown	1.336 (1.099, 1.624)				
mdc_category_Congenital Malformations and Chromosomal Abnormalities	0.866 (0.736, 1.019)				
mdc_category_Symptoms, Signs, and Ill-defined Conditions	1.110 (0.648, 1.902)				
mdc_category_Certain Conditions Originating in the Perinatal Period	0.653 (0.033, 13.034)				
ventilation_used_0	REF	201.480 (9.93E-46) (1)			
ventilation_used_1	0.676 (0.640, 0.714)				
prior_admissions_1yr_No prior admissions (0 admissions)	REF	26726.042 (0) (3)			
prior_admissions_1yr_Low frequency (1-2 admissions)	1.017 (1.001, 1.032)				
prior_admissions_1yr_Moderate frequency (3-4 admissions)	1.560 (1.527, 1.593)				

prior_admissions_1yr_High frequency (5 or more admissions)	3.040 (2.993, 3.089)				

Figure 1: Box Plots for Continuous Predictors

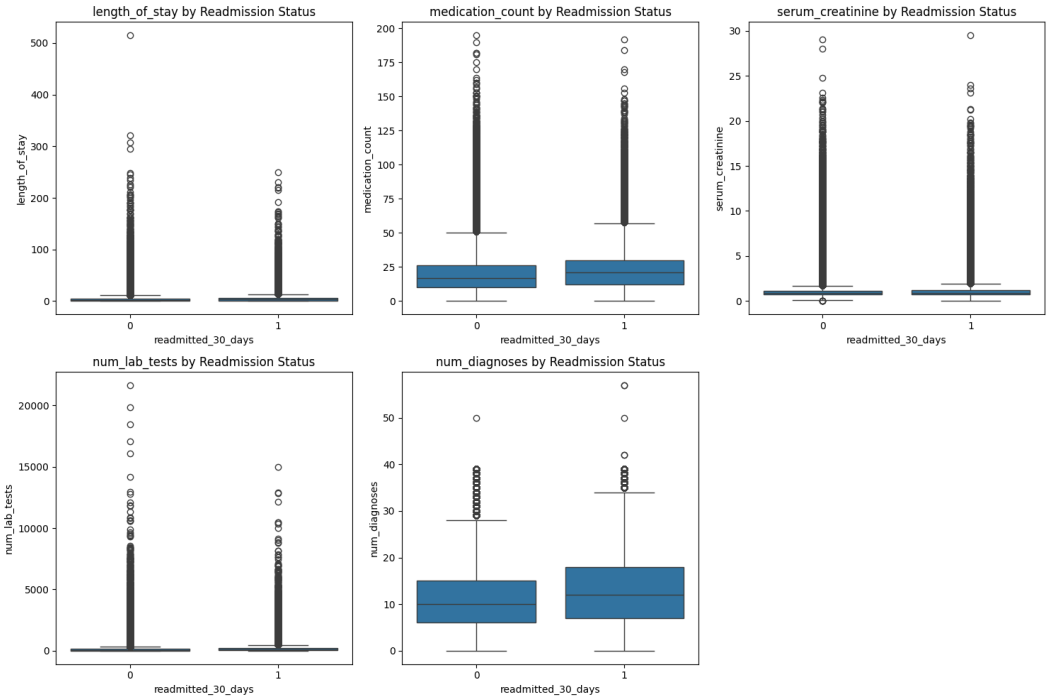


Figure 2: Violin Plots for Continuous Predictors

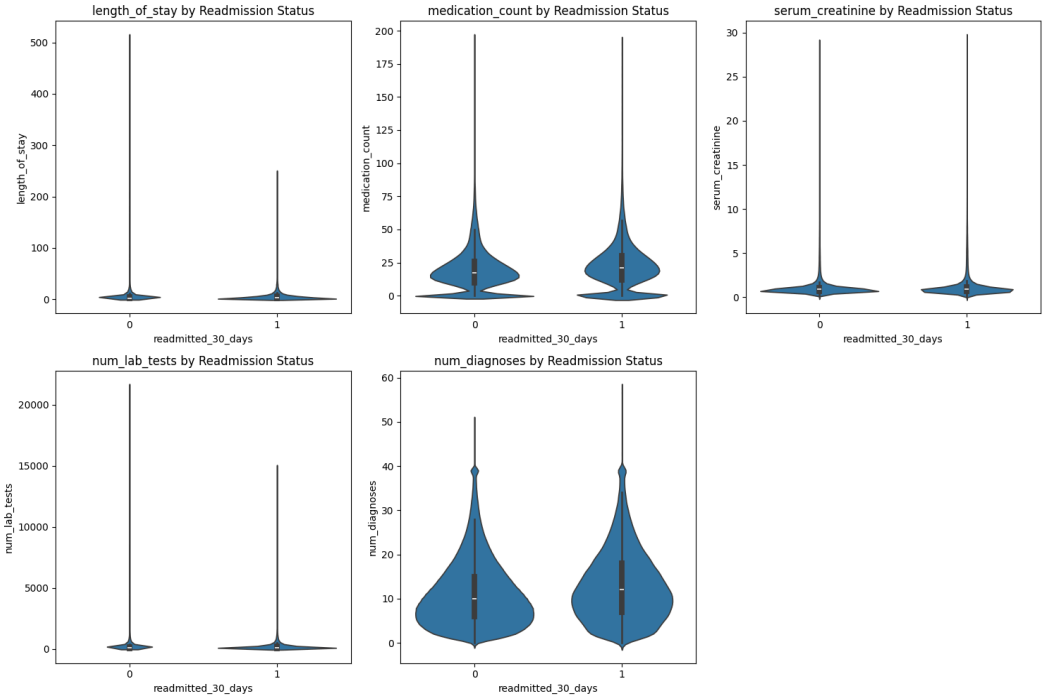


Figure 3: Bar Plot for Categorical Predictors

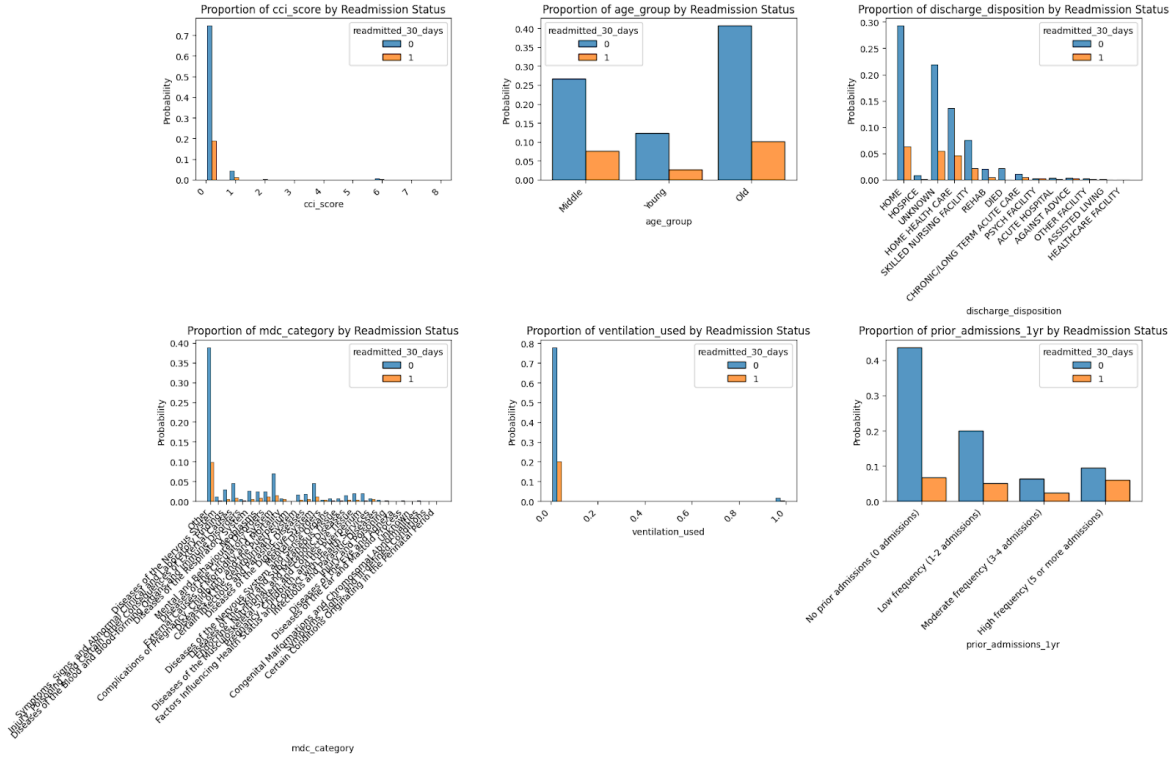


Figure 4: Stacked Bar Plot for Categorical Predictors

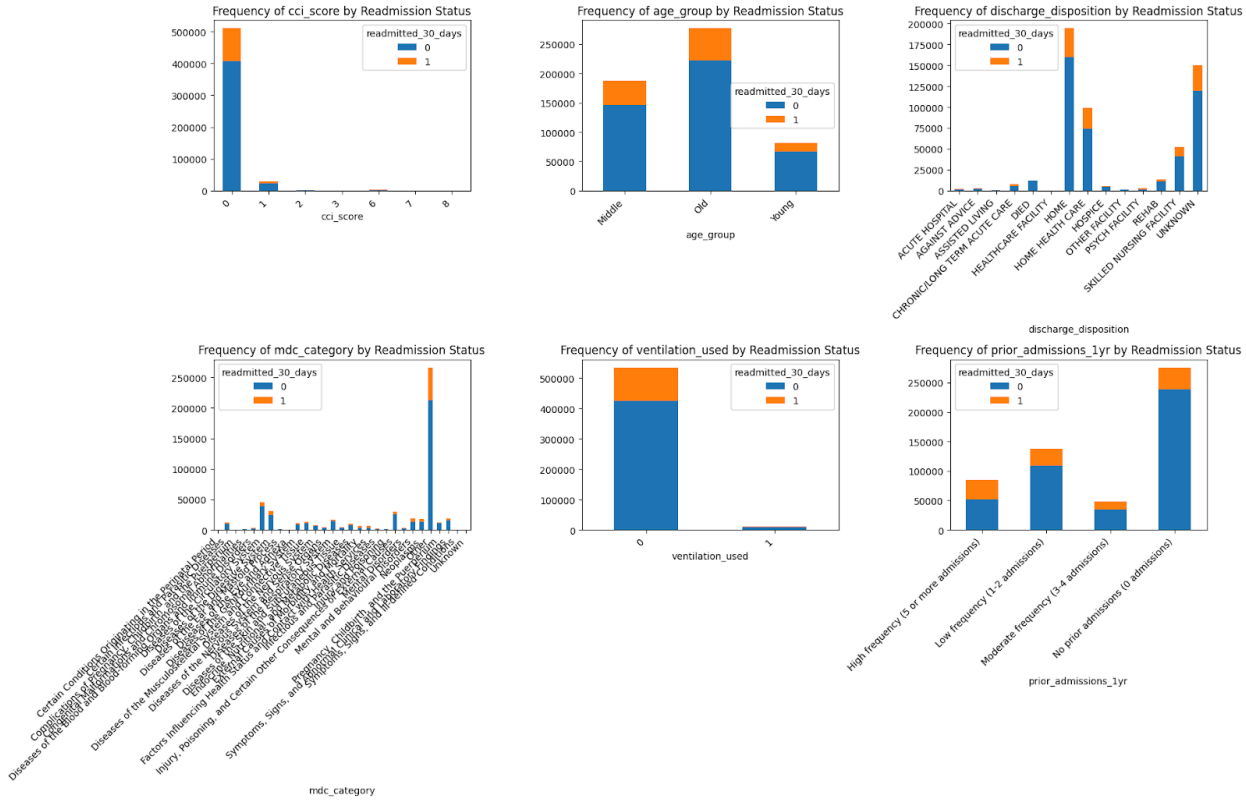
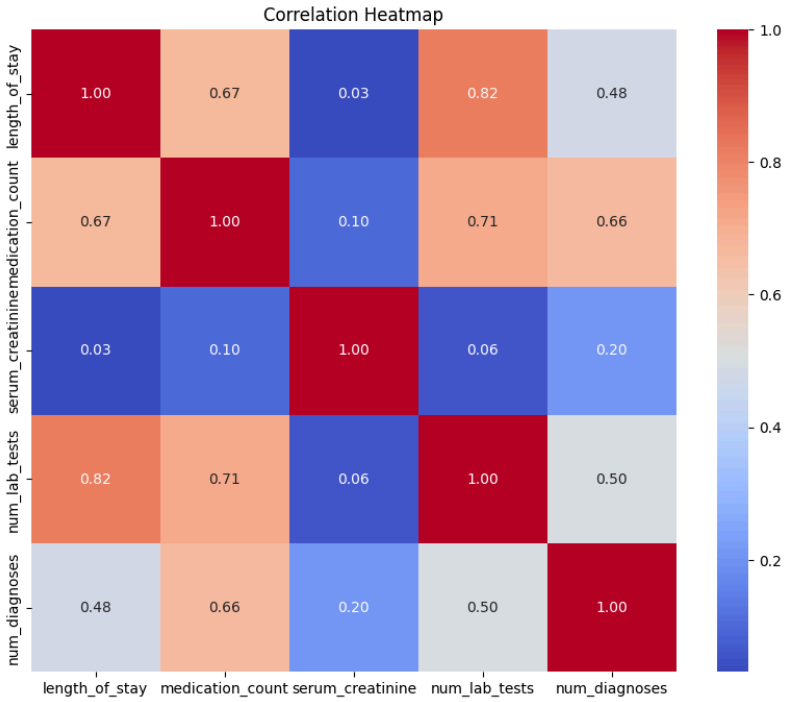


Figure 5: Correlation Heatmap for Categorical Predictors



Appendix

[1] Dharmarajan, K., Hsieh, A. F., Lin, Z., Bueno, H., Ross, J. S., Horwitz, L. I., Barreto-Filho, J. A., Kim, N., Bernheim, S. M., Suter, L. G., Drye, E. E., & 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. https://doi.org/10.1001/jama.2012.216476

[2] Glans, Maria et al. (2020). Risk factors for hospital readmission in older adults within 30 days of discharge – a comparative retrospective study. *BMC Geriatrics*, 20(1). https://doi.org/10.1186/s12877-020-01867-3

[3] McIlvennan, C. K., Eapen, Z. J., & Allen, L. A. (2015). Hospital readmissions reduction program. *Circulation*, *131*(20), 1796–1803. https://doi.org/10.1161/CIRCULATIONAHA.114.010270

[4] Setia M. S. (2016). Methodology Series Module 1: Cohort Studies. *Indian journal of dermatology*, *61*(1), 21–25. https://doi.org/10.4103/0019-5154.174011

[5] Tuty Kuswardhani, R. A., Henrina, J., Pranata, R., Anthonius Lim, M., Lawrensia, S., & Suastika, K. (2020). Charlson comorbidity index and a composite of poor outcomes in COVID-19 patients: A systematic review and meta-analysis. *Diabetes & metabolic syndrome*, *14*(6), 2103–2109. https://doi.org/10.1016/j.dsx.2020.10.022

[6] Gupta, M., Gallamoza, B., Cutrona, N., Dhakal, P., Poulain, R., & Beheshti, R. (2022). An Extensive Data Processing Pipeline for MIMIC-IV. *Proceedings of machine learning research*, *193*, 311–325.

[7] Hensche, M. K., Barrett, M. L., Jiang, H. J., Liang, L., & Heslin, K. C. (2020, June 12). Impact of the ICD-10-CM/PCS transition on readmission rates, 2015-2017 [Online]. Agency for Healthcare Research and Quality. Retrieved from https://www.hcupus.ahrq.gov/datainnovations/icd10_resources.jsp

[8] Kwok, C.S., Martinez, S.C., Pancholy, S. *et al.* Effect of Comorbidity On Unplanned Readmissions After Percutaneous Coronary Intervention (From The Nationwide Readmission Database). *Sci Rep* **8**, 11156 (2018). https://doi.org/10.1038/s41598-018-29303-y

[9] Murray, F., Allen, M., Clark, C. M., Daly, C. J., & Jacobs, D. M. (2021). Socio-demographic and -economic factors associated with 30-day readmission for conditions targeted by the hospital readmissions reduction program: a population-based study. BMC public health, 21(1), 1922. https://doi.org/10.1186/s12889-021-11987-z

[10] Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L. A., & Mark, R. (2024). MIMIC-IV (version 3.1). PhysioNet. https://doi.org/10.13026/kpb9-mt58.

[11] Johnson, A.E.W., Bulgarelli, L., Shen, L. et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 10, 1 (2023). https://doi.org/10.1038/s41597-022-01899-x