A logo with text and symbols on it

Description automatically generated

**Module 12: Capstone Project Deliverable Report**

**Wannida Kruayatidee**

**ALY 6980: CAPSTONE**

**Prof. Ajit Appari**

**December 15, 2024**

**Introduction**

The Centers for Medicare and Medicaid Services (CMS) Linkable 2008–2010 Medicare Data Entrepreneurs’ Synthetic Public Use File (DE-SynPUF) provides valuable synthetic data for analyzing Medicare beneficiary demographics and healthcare usage. This dataset simulates real Medicare records while maintaining beneficiary privacy, encompassing claims data for a 5% sample of beneficiaries from 2008 to 2010.

In this analysis, I utilized 15 samples and consolidated the beneficiary data across these years into a unified dataset, focusing specifically on patients with heart disease. The data processing involves filtering patients with both primary and secondary diagnoses of heart disease. Subsequently, I link this dataset with additional files, including inpatient claims, drug prescription events, and outpatient files. These linked datasets provide a comprehensive view of healthcare usage among heart disease patients. The resulting dataset undergoes exploratory data analysis (EDA) to examine factors influencing the length of stay (LOS) in these patients. This analysis aims to reveal patterns and insights related to the LOS, as well as the impact of medication use and outpatient services on hospitalizations.

**Data Preparation**

In the data cleaning process, I began by consolidating the beneficiary and inpatient claims files, merging all 15 samples into a single comprehensive dataset. This initial step provided a unified structure for subsequent analysis. I then focused on identifying patients diagnosed with heart disease within the inpatient claims files, filtering by specific diagnosis codes relevant to heart-related conditions. These included codes for heart failure, chronic obstructive pulmonary disease (COPD), ischemic heart disease, and stroke, ensuring a precise selection of patients pertinent to the study. Following this, I merged the inpatient claims data with the beneficiary files to create a dominant dataset, which served as the foundation for further exploratory data analysis. To enhance analytical depth, I conducted feature engineering by creating a new column indicating each patient’s primary chronic disease based on their first diagnosis, enabling clearer insights into the main health concerns impacting each patient. Additionally, I calculated the number of admitting days to analyze the length of stay (LOS), a critical metric for both predictive modeling and understanding patient outcomes, age year for all patients, and did some feature engineering on creating new column of “Previous\_Payment\_6M” as to see how much claim or amount that each patient paid for the last 6 months.

Building on this foundation, I continued data cleaning and preparation with the Prescription Drug Event (PDE) files. I began by inspecting these files for missing values and confirmed that no imputation was necessary. To streamline the analysis, I compressed the PDE file and focused on patient-level data aggregated by year. This involved converting the service dates (SRVC\_DT column) from the YYYYMMDD format into a datetime format, which allowed for the extraction of specific components like the year. I then created a new YEAR column to group the data annually and aggregated key metrics such as the total quantity of drugs dispensed (QTY\_DSPNSD\_NUM), total prescription costs (TOT\_RX\_CST\_AMT), and total patient payments (PTNT\_PAY\_AMT) by patient ID (DESYNPUF\_ID) and year. This process simplified the dataset, reduced complexity, and prepared it for further analysis, such as trend identification and predictive modeling. Finally, I merged the cleaned PDE file with the master dataset, enabling a more comprehensive analysis of factors influencing the length of stay (LOS), including patterns in drug utilization, financial impacts, and their relationship to patient behavior and treatment outcomes.

Lastly, I worked on the outpatient file claims, where I selected only variables that seem relevant and influenced length of stay which consist of attending physicians, operating physicians, and other physicians. These variables could be factors that if the more patients have physicians, the likely chance they will have longer length of stays. After that I merged everything into one master file.

**Extrapolatory Data Analysis**

A pie chart with different colored circles

Description automatically generatedA group of colored boxes

Description automatically generated with medium confidenceThe analysis of admission days for heart disease patients, as illustrated in *Figure 1*, reveals consistent trends across 2008, 2009, and 2010. The median length of stay for chronic disease categories, including heart failure, COPD, ischemic heart disease, stroke, and others, remains stable at approximately five days. However, the presence of significant outliers indicates that some patients require extended hospital stays, underscoring variability in outcomes. This pattern highlights a consistent baseline for hospitalizations while drawing attention to the need for further investigation into the factors contributing to prolonged stays.

Figure 2: Percentage of First Diagnosis – Different Heart Disease vs Others

Figure 1: Admission Days for Patients with Heart Disease both First and Secondary Diagnosis

A graph of different colored bars

Description automatically generatedA deeper look into the distribution of first diagnosed diseases, presented in *Figure 2*, underscores the dominance of the "other" category, which consistently accounts for 69% to 70% of diagnoses over the three years. Heart failure, COPD, ischemic heart disease, and stroke exhibit steady proportions with minor variations, collectively making up the remaining diagnoses. These findings suggest that while heart disease-related conditions are prevalent, a significant proportion of initial diagnoses fall outside these categories, warranting additional exploration into the health conditions classified under "other."

A graph of different colored lines

Description automatically generated with medium confidenceDemographic trends provide further insights, as highlighted in *Figures 3 and 4*. Gender distribution across chronic disorders shows that females consistently represent a slightly larger proportion (55%-58%) compared to males (42%-45%) each year, indicating stable patterns in gender prevalence. Age distribution trends reveal that patients aged 65 to 85, particularly those in their 70s, account for the highest frequencies of chronic disease cases, with 2008 showing slightly higher numbers overall. These patterns emphasize the disproportionate impact of chronic conditions on older adults and women, offering a critical foundation for understanding patient demographics in this population.

Figure 3: Grouped Bar Chart of Gender Proportion of each Chronic Disorder in each year

Figure 4: Age Distribution by Year for Each Chronic Disorder

A screenshot of a graph

Description automatically generatedA diagram of different colored squares

Description automatically generatedMoreover, I delved into drug utilization and patient payment patter across chronic diseases from 2008 to 2010. *Figure 5* highlights the distribution of drug prescriptions (QTY\_DSPNSD\_NUM), with higher medians and broader ranges observed in conditions like Heart Failure and COPD, indicating greater drug utilization compared to narrower distributions in conditions like Stroke. Outliers, particularly in Heart Failure and Ischemic Heart Disease, suggest more severe cases requiring higher prescriptions.

Figure 5: Distribution of Drug Prescriptions by Disease and Year

Figure 6: Histogram for Age Distribution by Year and Disease

*Figure 6* illustrates the distribution of patient pay amounts (PTNT\_PAY\_AMT), revealing a consistent skew across years, with most patients incurring low out-of-pocket costs and a few bearings significantly higher expenses. Heart Failure shows broader payment ranges, reflecting prolonged or intensive treatments, while Stroke exhibits concentrated payments in the lower range. These patterns provide a comprehensive understanding of prescribing trends and financial burdens associated with chronic conditions over time.

A group of colorful rectangular objects

Description automatically generated with medium confidenceAdditionally, I conducted histograms of the number of physicians by first diagnosis chronic diseases which is shown in *Figure 7.* The visualization provides an overview of the total prescription costs (TOT\_RX\_CST\_AMT) across chronic diseases for 2008, 2009, and 2010, highlighting consistent patterns in median costs over time. Heart Failure and COPD display higher medians and broader interquartile ranges (IQRs), reflecting higher treatment expenses compared to narrower distributions in conditions like Stroke and "Other." Outliers are notable in all categories, especially in Heart Failure, Ischemic Heart Disease, and others, suggesting that certain patients incur significantly higher costs due to severe cases or complex treatments. The stable distributions over the years, with a slight decline in costs by 2010 for some diseases, may reflect shifts in healthcare practices or drug pricing. These trends provide valuable insights into the financial burden of chronic disease management.

Figure 7: Number of Physicians by First Diagnosis Chronic Diseases

A pie chart with different colored circles

Description automatically generatedFigure 8: Dashboard

Lastly, after completing the exploratory data analysis (EDA), I created an interactive dashboard using Dash and Plotly to offer an in-depth and dynamic exploration of data trends. This dashboard enables users to interactively examine key elements of the dataset, providing valuable insights into the prevalence of chronic diseases, patient demographics, and hospitalization patterns. It features four primary visualizations: a pie chart, box plot, bar chart, and line chart, all of which dynamically update and filter based on the chosen year and chronic disorder.

**Prediction Model**

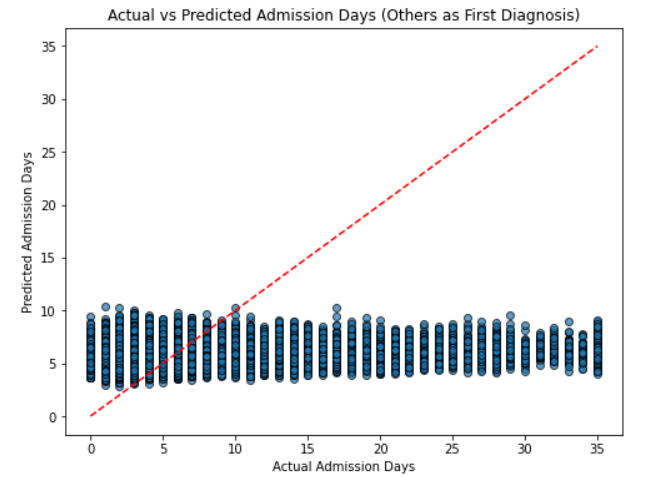
Prior to conducting prediction model, I observed the relationship between each variable towards target variable (TOTAL\_ADMSN\_DY). Upon examining the correlation matrix in *Figure 9*, A screenshot of a graph

Description automatically generated I observed that the target variable TOTAL\_ADMSN\_DY has very weak correlations with all the features, with QTY\_DSPNSD\_NUM showing the highest but still weak positive relationship at 0.17, followed by TOT\_RX\_CST\_AMT and PTNT\_PAY\_AMT, which have correlations around 0.10 to 0.13. Interestingly, some strong inter-feature relationships emerged: QTY\_DSPNSD\_NUM and TOT\_RX\_CST\_AMT exhibit a high positive correlation of 0.94, indicating that higher quantities dispensed align with higher prescription costs, while PTNT\_PAY\_AMT also correlates strongly with both at 0.72 and 0.75, respectively, showing logical dependencies between cost and payments. Additionally, the physician-related variables (AT\_PHYSN\_NPI\_CNT, OP\_PHYSN\_NPI\_CNT, and OT\_PHYSN\_NPI\_CNT) show moderate to strong correlations, particularly AT\_PHYSN\_NPI\_CNT and OP\_PHYSN\_NPI\_CNT at 0.75, suggesting these counts increase together in complex cases. Overall, while the target variable shows little direct influence from individual features, these inter-feature relationships highlight important patterns that could be further explored for multivariate analysis. Subsequently, I began to perform prediction models where I implemented 2 models: Linear Regression and Random Forest. However, before conducting and building the models, I transform categorical variables into dummies variables to enhance the capability and accuracy of the two prediction models and perceive the best outcomes.

Figure 9: Correlation Matrix

1. **Linear Regression**

For the prediction models, I split the dataset into 2 by separating patients who have heart disease as first diagnosis and another one is patients who have heart disease as their secondary diagnosis.

 A graph with a red line

Description automatically generated

Figure 10: Linear Regression Model Predicting Total Admission Days for Patients With Heart Disease as a **Secondary Diagnosis** (Left) and as a **Primary Diagnosis** (Right).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Prediction Model Result** | | | | |
| **Model** | **First Diagnosis** | **R-Squared** | **AIC** | **MSE** |
| Linear Regression | Other | 2.62% | 267263 | 31.39 |
| Heart Disease | 3.87% | 100597 | 19.39 |

Table 1: Linear Regression Model Result

Based on the results of the linear regression models, I found that both models performed poorly in predicting **admission days** for cases where "Others" and "Heart Disease" were the first diagnoses. For the "Others" model, the **R-squared** value is **2.62%**, while for "Heart Disease," it is only slightly better at **3.87%**, indicating that both models explain a very small portion of the variability in admission days. Additionally, the **Mean Squared Error (MSE)** for "Others" is **31.39**, which is significantly higher than the **19.39** observed for "Heart Disease." This suggests that the model for "Heart Disease" predictions has smaller errors on average compared to "Others." The **Akaike Information Criterion (AIC)** values further highlight this difference, with **267263** for "Others" and a notably lower **100597** for "Heart Disease," indicating that the "Heart Disease" model provides a marginally better fit.

From the scatter plots, I observed that the predictions for both models are tightly clustered around the lower range of admission days, particularly between **3 and 10 days**, regardless of the actual values. The red dashed line, representing the ideal relationship between actual and predicted values, shows that the models fail to scale predictions accurately for higher admission days, further underscoring their inability to capture meaningful patterns in the data. This clustering and underfitting behavior suggest that linear regression is insufficient for modeling this dataset, likely due to its inability to account for non-linear relationships or interactions among features. I recommend exploring more advanced models, such as Random Forests or other ensemble methods, to improve predictive accuracy and better represent the complexity of the data.

1. **Random Forest**

Similarly, I attempted to conduct random forest for better model accuracy for two subgroups: patient with heart disease as primary and secondary diagnosis chronic disease.

A graph with a red line

Description automatically generatedA graph with green and red dots

Description automatically generated

Figure 11: Random Forest Regressor Model Predicting Total Admission Days for Patients With Heart Disease as a **Secondary Diagnosis** (Left) and as a **Primary Diagnosis** (Right).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Prediction Model Result** | | | | |
| **Model** | **First Diagnosis** | **R-Squared** | **AIC** | **MSE** |
| Random Forest | Other | 27.34% | 244722 | 3.17 |
| Heart Disease | 22.42% | **93476** | 15.65 |

Table 2: Random Forest Regressor Model Result

The results of the Random Forest Regressor models in *Figure 11* and *Table 2* show a significant improvement over the linear regression models in predicting **total admission days**. When "Others" is the first diagnosis, the model achieves an **R-squared value of 27.34%**, indicating that it explains a much larger portion of the variability in admission days compared to the linear regression results. Similarly, for cases where "Heart Disease" is the first diagnosis, the R-squared value is **22.42%**, which is slightly lower but still a meaningful improvement. This suggests that the Random Forest model is better equipped to capture complex, non-linear relationships in the data. Additionally, the **Mean Squared Error (MSE)** for "Others" is much lower at **3.17**, compared to **15.65** for "Heart Disease," highlighting that the model's predictions for "Others" are considerably closer to the actual values. The **Akaike Information Criterion (AIC)** also supports this finding, with the "Others" model having a value of **244722** and the "Heart Disease" model at **93476**, suggesting that while both models perform well, the "Heart Disease" predictions may still have more variance and larger errors.

Although the model is far more capable of capturing non-linear relationships compared to linear regression, there are signs of **overfitting** and **underfitting** in specific ranges of actual admission days. For instance, when the actual admission days are **below 15**, the model tends to underfit the data, producing relatively consistent and less dispersed predictions that fail to capture variations in the lower range. Conversely, when the actual admission days exceed **15**, the Random Forest model exhibits **overfitting** tendencies. In this range, the predictions start to deviate, and the model becomes overly sensitive to the training data. This is evident from the scatter plots, where predictions for higher admission days show noticeable variability and dispersion, especially for "Heart Disease" cases. The model tries to accommodate the complexity of higher values, leading to predictions that may align too closely with noise in the data rather than generalizable patterns.

**3. Tree Visualizations**

Building on the insights gained from the Random Forest model, I further explored the predictions using a **decision tree visualization** to better understand the relationships between features and the length of stay for patients with **heart disease as a secondary chronic diagnosis**. **A diagram of a computer

Description automatically generated**

Figure 12: Decision Tree Visualization for Heart Disease as Secondary Diagnosis

Surprisingly, I observed in *Figure 12* that the most significant predictor at the root node is BENE\_RACE\_CD, which splits the data at a threshold of 3925.5. This suggests that the race code of patients plays a key role in determining their admission days (~ 5.91 days) , as it effectively divides the dataset into two distinct groups. Moving deeper into the tree, I noticed that Previous\_Payment\_6M, representing prior healthcare payments in the last 6 months, emerges as another influential feature. Patients with lower previous payments tend to have smaller predicted lengths of stay, while higher payments lead to further splits based on race code and other features, such as QTY\_DSPNSD\_NUM and TOT\_RX\_CST\_AMT. These medication-related features highlight the connection between healthcare costs and admission days, particularly in cases involving chronic conditions.

As the tree progresses, features like OP\_PHYSN\_NPI\_CNT (Operating Physician Count) and ICD9\_DGNS\_CD\_1 (first diagnosis code) appear at deeper levels, reflecting the importance of physician involvement and comorbidity details in predicting the outcome. Interestingly, splits involving BENE\_RACE\_CD recur throughout the branches, reinforcing its significance in the model. For example, in the branch where **BENE\_RACE\_CD > 3930.5**, the average length of stay rises sharply to **18.006 days**, highlighting a subgroup of patients experiencing significantly longer admissions.

*A diagram of a computer

Description automatically generated with medium confidence*

Figure 13: Decision Tree Visualization for Heart Disease as Primary Diagnosis

Following the analysis of the decision tree for patients with heart disease as a secondary diagnosis, I now turn to the decision tree for cases where heart disease is the primary chronic diagnosis. At the root of the tree, the most significant feature is **Previous\_Payment\_6M** with a threshold of **0.5**, indicating that patients' prior healthcare payments have a notable impact on determining their admission days (~ 4.81 days). If the previous payment is below this value, the tree further splits based on **BENE\_RACE\_CD**, , with an average stay of 3.142 days, reinforcing the importance of race as a key predictor and suggesting that race-related factors play a significant role in determining shorter stays. For patients with **BENE\_RACE\_CD <= 54.5**, additional splits occur on **AT\_PHYSN\_NPI\_CNT** (Attending Physician Count) and **PTNT\_PAY\_AMT** (Patient Payment Amount), which highlights that the number of physicians involved and direct patient payments are influential for shorter predicted admission days. On the right branch, where previous payments exceed the threshold, **ICD9\_DGNS\_CD\_1** (First Diagnosis Code) emerges as the next key predictor, showing the role of specific diagnosis codes in driving longer stays.

Then I delved into deeper tree, splits involving **TOT\_RX\_CST\_AMT** (Total Prescription Cost), **OP\_PHYSN\_NPI\_CNT** (Operating Physician Count), and **QTY\_DSPNSD\_NUM** (Quantity Dispensed) appear prominently, indicating the impact of treatment intensity and healthcare costs on the length of stay where values exceeding certain thresholds, such as **2438.5** for operating physicians and **2447.0** for prescription costs, are linked to prolonged stays of **6.824 to 6.978 days**. Notably, repeated splits on **BENE\_RACE\_CD** and physician-related features like **PHYSN\_NPI\_CNT** emphasize the interconnected roles of race, physician involvement, and resource utilization in predicting patient admission days. The MSE progressively decreases at each split, which reflects the tree’s ability to refine its predictions by isolating patient subgroups with shared characteristics. However, the decision tree’s limited depth (max = 4) simplifies these relationships, potentially missing more complex patterns in longer hospital stays.

Lastly, when comparing the two decision tree visualizations, I observed that the key features influencing the predicted **length of stay** differ due to the nature of heart disease as a **primary** or **secondary diagnosis**. For patients with heart disease as a **primary diagnosis**, **Previous\_Payment\_6M** emerges as the most critical predictor, likely reflecting that higher prior healthcare spending signals more severe or ongoing medical conditions, leading to longer hospital stays. Features like **ICD9\_DGNS\_CD\_1** (first diagnosis code) and physician-related variables (**AT\_PHYSN\_NPI\_CNT** and **OP\_PHYSN\_NPI\_CNT**) appear prominently, suggesting that comorbidities and the level of physician involvement play significant roles in managing complex primary heart disease cases, contributing to extended stays. On the other hand, for patients with heart disease as a **secondary diagnosis**, **BENE\_RACE\_CD** (race code) is the dominant predictor, which may indicate underlying socioeconomic or demographic factors that influence access to healthcare, treatment duration, or severity of illness. Additionally, medication-related variables such as **QTY\_DSPNSD\_NUM** (quantity dispensed) and **TOT\_RX\_CST\_AMT** (total prescription cost) become key predictors, reflecting that managing secondary heart disease often requires intensive pharmacological interventions, which can drive longer admission durations. The differences suggest that primary heart disease cases are more directly influenced by clinical complexity and resource utilization, while secondary heart disease cases are shaped by demographic, economic, and treatment-related factors; thus, patients with **heart disease as a secondary diagnosis** tend to have **longer lengths of stay** compared to those with heart disease as a **primary diagnosis**. In the secondary diagnosis model, certain groups, particularly those with higher **BENE\_RACE\_CD** values and **Previous\_Payment\_6M**, show stays reaching up to **18 days**, driven by factors like race, prior payments, and treatment intensity. In contrast, the primary diagnosis model shows shorter maximum stays, with lengths reaching approximately **6–7 days**, influenced primarily by **prescription costs**, physician involvement, and diagnosis codes.

**Conclusion and Recommendation**

In conclusion, my analysis highlights key factors influencing the **length of stay (LOS)** for patients with heart disease as either a **primary** or **secondary diagnosis**. Moreover, we also could observe that patients with **heart disease as a secondary diagnosis** tend to have **longer lengths of stay** compared to those with heart disease as a **primary diagnosis**. For patients with heart disease as a **primary diagnosis**, prior healthcare spending (**Previous\_Payment\_6M**) emerged as the most influential factor, reflecting the severity or ongoing nature of their conditions, with physician involvement (**AT\_PHYSN\_NPI\_CNT**, **OP\_PHYSN\_NPI\_CNT**) and comorbidity details (**ICD9\_DGNS\_CD\_1**) playing critical roles in extending hospital stays. On the other hand, for patients with heart disease as a **secondary diagnosis**, **BENE\_RACE\_CD** (race code) stood out as the dominant predictor, suggesting that socioeconomic or demographic factors heavily influence LOS. Additionally, medication-related features, including **QTY\_DSPNSD\_NUM** (quantity dispensed) and **TOT\_RX\_CST\_AMT** (total prescription cost), were significant drivers, underscoring the role of treatment intensity and associated costs.

In terms of predictive modelling, the **Random Forest Regressor** outperformed **Linear Regression**, as it captured the non-linear relationships and complex interactions between features more effectively. While the Random Forest models achieved R-squared values of **27.34%** for secondary diagnosis cases and **22.42%** for primary diagnosis cases, both models showed overfitting tendencies when actual admission days exceeded **15** and underfitting for shorter stays below this range.

Last but not least, based on the findings of this analysis, I recommend that healthcare providers and policymakers implement strategies that focus on both **clinical management** and **resource optimization** to address the factors influencing the length of stay (LOS) for patients with heart disease. For those with heart disease as a primary diagnosis, a focus on high-risk individuals with higher prior healthcare spending and complex comorbidities is essential, as these factors significantly influence prolonged stays. Proactively assigning specialized care teams, enhancing early intervention programs, and improving physician coordination can help streamline patient management and optimize resource use. For patients with heart disease as a secondary diagnosis, addressing socioeconomic disparities is crucial, as race-related factors and treatment costs heavily impact LOS. Implementing tailored care plans, optimizing medication management programs, and ensuring cost-effective pharmacological treatments can further enhance outcomes and reduce unnecessary stays. By leveraging these insights, healthcare providers can deliver equitable and efficient care, minimizing hospital burden while improving patient outcomes.

**Appendix**

Confusion Matrix before final presentation’s feeback

**A screenshot of a computer

Description automatically generated**

Models before final presentation’s feedback

**A red line with blue dots

Description automatically generated with medium confidence A red line with green dots

Description automatically generated**

**A graph of a graph

Description automatically generated with medium confidence A graph of blue dots

Description automatically generated**

**A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated**

**References**

*Chronic obstructive pulmonary disease and Allied conditions ICD-9 code range 490-496. ICD-9 Code CHRONIC OBSTRUCTIVE PULMONARY DISEASE AND ALLIED CONDITIONS 490-496 Codify By AAPC. (n.d.).* [*https://www.aapc.com/codes/icd9-codes*](https://www.aapc.com/codes/icd9-codes) *range/67/#:~:text=CHRONIC%20OBSTRUCTIVE%20PULMONARY%20DISEASE%20AND% 0ALLIED%20CONDITIONS%20ICD%2D9%20Code,World%20Health%20Organization%20( HO).*

Cozzolino, F., Abraha, I., Orso, M., Mengoni, A., Cerasa, M. F., Eusebi, P., Ambrosio, G., & Montedori, A. (2017, March 29). Protocol for validating cardiovascular and Cerebrovascular ICD-9-CM codes in healthcare administrative databases: The Umbria Data Value Project. BMJ open. <https://pmc.ncbi.nlm.nih.gov/articles/PMC5372118/>

*Diseases of the circulatory system ICD-9 code range (390-459)*. ICD-9 Code DISEASES OF THE CIRCULATORY SYSTEM 390-459- Codify By AAPC. (n.d.). <https://www.aapc.com/codes/icd9-codes> range/53/?srsltid=AfmBOopevh5Z7GbkMcnK57yvd\_9B1xSewMu\_CzIKNn8H4LY8Wk\_nF7ep

*Ischemic heart disease ICD-9 code range 410-414*. ICD-9 Code ISCHEMIC HEART DISEASE 410-414 Codify By AAPC. (n.d.). https://www.aapc.com/codes/icd9-codes-range/57/

*Medicare claims synthetic public use files (synpufs)*. CMS.gov. (n.d.). <https://www.cms.gov/data> research/statistics-trends-and-reports/medicare-claims-synthetic-public-use-files

Other forms of heart disease ICD-9 code range 420-429. ICD-9 Code OTHER FORMS OF HEART DISEASE 420-429- Codify By AAPC. (n.d.). <https://www.aapc.com/codes/icd9-codes> range/59/?srsltid=AfmBOoqdVnW6hCCibfhN3EtxbpYB3ygvZpoayXwTjLgdLHxiNrUoZb0t