THE USE OF BAYESIAN PRINCIPLES TO PREDICT THE OPTIMAL REVISIT INTERVAL AND THE CAUSAL RELATIONSHIP BETWEEN REVISIT INTERVALS AND CHRONIC KIDNEY DISEASE FOR MEDICARE PATIENTS WITH TYPE II DIABETES

by

Annie Enikunda Lasway
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Health Services Research

_____	Farrokh Alemi, Ph.D., Chair
_____	Panagiota Kitsantas, Ph.D., Committee Member
_____	Tracey Koehlmoos, PhD., Committee Member
_____	Robert Pitsko, Ph.D., Committee Member
_____	PJ Maddox, Ph.D., Department Chair
_____	Germaine M. Louis, Ph.D., Dean and Professor, College of Health & Human Services
Date: _____	Spring Semester 2020 George Mason University Fairfax, VA

The Use of Bayesian Principles to Predict the Optimal Revisit Interval and the Causal Relationship between Revisit Intervals and Chronic Kidney Disease for Medicare Patients with Type II Diabetes

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Annie Enikunda Lasway
Master of Public Health
George Mason University, 2016
Bachelor of Science
University of Maryland, 2013

Director: Farrokh Alemi, Director, Program in Health Service Research
Department of Health Administration and Policy

Spring Semester 2020
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my father, Major General Benjamin Msuya (rtd.) who is battling Diabetes and Chronic Kidney Disease (CKD) and is on Dialysis 3 days a week. You have been the inspiration for my research, and it is my hopes that this study will help many other diabetic patients who are suffering from CKD. To my mother, Deborah Msuya – who has taught me what unconditional love is, who has demonstrated resilience and who is the epitome of excellence. To my siblings Erick, Eunice, Roderick and Bisala for their unwavering love, support and for being my cheerleaders. To my loving husband Eric N. Lasway, who has supported me and our household while focus on my research. Eric has uplifted, motivated and encouraged me throughout the entire process. To my son, Sebastian Phocus Lasway – my motivation, the reason I push through and endure, my heart.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Farrokh Alemi for making me the researcher and expert I am today. To my committee members, Dr. Panagiota Kitsantas, Dr.Tracey Koehlmoos and Dr. Robert Pitsko for their advice, mentorship and unwavering support. My friends, relatives, and supporters who have been with me every step of the way. Thank you for your patience, support and understanding as I conducted my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Affordable Care Act ..... ACA
American Diabetes Association ..... ADA
Centers for Medicare and Medicaid Services ..... CMS
Chronic Kidney Disease ..... CKD
Chronic Obstructive Pulmonary Disease ..... COPD
Confidence Interval CI
Electronic Health Record ..... EHR
Excerpta Medica Database ..... EMBASE
Hemoglobin A1c HbA1c
Institutional Review Boards ..... IRB
Least Absolute Shrinkage and Selection Operator ..... LASSO
Likelihood Ratio ..... LR
Limited Dataset ..... LDS
Long RVI ..... lRVI
Markov Blanket ..... MB
National Institutes for Health ..... NIH
National Institute for Health and Care Excellence ..... NICE
National Provider Identifier ..... NPI
Odds Ratio ..... OR
Parents of the Markov Blanket ..... pMB
PubMed Central ..... PubMed
Revisit Interval ..... RVI
Short RVI ..... sRVI
Standard Query Language ..... SQL
United States ..... US

ABSTRACT

THE USE OF BAYESIAN PRINCIPLES TO PREDICT THE OPTIMAL REVISIT INTERVAL AND THE CAUSAL RELATIONSHIP BETWEEN REVISIT INTERVALS AND CHRONIC KIDNEY DISEASE FOR MEDICARE PATIENTS WITH TYPE II DIABETES

Annie Enikunda Lasway MPH

George Mason University, 2020

Dissertation Director: Dr. Farrokh Alemi

"Evidence-based follow-up intervals have the potential to reduce healthcare costs per person and improve access without compromising or restricting care"

In prediction modeling and causality, one of the main goals is to build an algorithm that best represents a dataset. This process first involves the task of selecting features that would best describe the response variable. This paper aims to address the issue of feature selection in causal models by using Bayesian Network principals and the Least Absolute Shrinkage and Selection Operator (LASSO) methodology executed in Standard Query Language (SQL). LASSO is an algorithm-based method and it yet to be executed solely through SQL.

In demonstrating the effective use of this method, this study applies it to 2014 to 2016 outpatient healthcare utilization data from the Centers for Medicare and Medicaid Services (CMS) in order to predict the optimal revisit interval (RVI) and to determine the causal relationship between RVI and Chronic Kidney Disease (CKD) for patients with Type II Diabetes. CKD and diabetes are the 9^{th} and 7^{th} leading causes of death in the United States (respectively), and therefore the cohort of interest in this study.

In this study, Likelihood Ratios were used to determine feature importance as they relate to CKD. The RVIs were then calculated as the number of days between two consecutive appointments by the same patient and the same provider. From there causality was derived from determining correlations, sequences, mechanism and counterfactuals related to the relevant features. The optimal RVIs were then deduced from the probabilities of CKD occurring given the presence of specific comorbidities. The probabilities were calculated by analyzing the set of comorbidities that patients had prior to a set date, and the prevalence of CKD after the set encounter date.

Results showed that there were 136 million outpatient observations for patients with Type II Diabetes. This resulted from approximately 800,000 distinct patients. The average RVI was 39.45 days, with a median of 91 days with a maximum of 363 days and a standard deviation of 64.3 days. Table 9 includes data on optimal RVI based on various comorbidities. If a patient had a probability of developing CKD that is above 0.5, then their optimal RVI was shorter, compared to those patients with probabilities that are below 0.5. Blood toxicity, orthopedic injuries, anemia and other diseases of the connective tissue were the leading causes and predictors of CKD. The biophysical mechanisms between CKD as a result of kidney overuse due to filtering toxins in the blood, drugs and medications is well known; however, patients who present with a history of the comorbidities should potentially be screened early for CKD as the Likelihood of occurrence may be higher in those patients. Optimal RVI can help ensure that patients with these risk factors are seen before their disease progresses.

This method is executable solely in SQL and therefore can be used directly in an Electronic Health Record (EHR) as a decision-making tool for providers. Since it does not involve exporting data from an EHR into statistical tools, patient data is protected, and the process is less time consuming. This method can potentially enable providers identify patients who are at higher risk of developing CKD and be able to allot an optimal RVI in patients need to be seen. Ultimately this can help improve health outcomes for diabetic patients and be leveraged for use with other chronic diseases such as hypertension.

Key words: Bayesian Networks, Causality, Revisits, Diabetes,
Chronic Kidney Disease, Causality, high-dimensional networks, LASSO Regression

CHAPTER 1: INTRODUCTION

The introductory chapter starts with providing background of the problem. This involves introducing the challenges with feature selection, and the problems related to diabetes, Chronic Kidney Disease (CKD), and Revisit Interval (RVI). It provides the aims of the study and an overview of the methods and definitions of the main terms. The significance of this study and the conceptual framework is presented as well.

Background

Feature selection in predictive modeling leans heavily on subject matter expertise.[1] Bayesian Network principals can be applied to well-known statistical algorithms in order to determine feature importance. This study accomplished this by applying the principals to the Least Absolute Shrinkage and Selection Operator (LASSO) methodology in Standard Query Language (SQL). The study used outpatient healthcare utilization data from the Centers for Medicare and Medicaid Services (CMS) from 2014 – 2016 to determine the important features needed to predict the optimal RVI and to determine the causal relationship between RVI and CKD for Medicare patients with Type II Diabetes. RVI is defined as the number of days between two consecutive appointments by the same patient and the same provider.

The method of applying Bayesian principles to execute LASSO Regression in SQL involved determining the Likelihood Ratios of the prevalence each diagnosis in the

dataset with a positive outcome, namely CKD, versus the presence of the same diagnosis with a negative outcome. Features with the highest Likelihood Ratios were then analyzed further in the predictive and causal model.

Diabetes and CKD is of importance in the United States (US) because Diabetes is the 9th leading cause of death [2], while CKD is the 7th leading cause of death in the US as of 2019 according to the National Institute of Kidney and Digestive and CKD.[3]

The physiological mechanisms in which diabetes leads to CKD is clinically known. Patients who present with comorbidities should potentially be screened early for CKD as the Likelihood of occurrence may be higher with specific comorbidities. Optimal RVI can help ensure that patients with these risk factors are seen before their disease progresses. The screening can involve running the model directly in an Electronic Health Record (EHR) as a decision-making tool for providers. Using SQL enables us to execute the model directly in an EHR without having to export data and feed it into a statistical tool in order to execute the mode. The predictions will enhance provider's decision-making when determining the optimal RVI for each diabetic patient based on their unique set of comorbidities. Ultimately this can help improve health outcomes for diabetic patients and be leveraged for use with other chronic diseases such as hypertension.

There are two commonly used processes of feature selection known as Forward selection and Backward selection.[4] LASSO Regression fundamentally uses Backward selection by assigning a value known as lambda to each coefficient. [5] The larger the lambda the higher the importance of the coefficient.[6] Those variables with little to no lambda strength are eliminated as they are considered irrelevant.[7]

There have been other methods similar to LASSO Regression that have been formed as an extension of LASSO Regression such as Elastic Net that was proposed by Zou and Hastic in 2005. [8] Elastic Net combines LASSO and Ridge Regression to stabilize the selection of grouped variables.[9] These algorithms are available through various statistical packages and software. R statistical software has a built-in function to implement LASSO Regression. Packages such as Glmnet algorithm that uses cyclical coordinate descent to penalize the maximum likelihoods for LASSO Regression, Ridge and Elastic Net (combination of LASSO and Ridge Regression). Least Angle Regression (Lars) uses Forward Selection to implement LASSO Regression. Again, the disadvantage of these algorithms are that in order to be used, data must be exported from an EHR, then imported into the statistical software in order to conduct an analysis. Based on the size of the data, this can be time consuming and costly. There are many studies that try to optimize and standardize this process for any kind of data, but this is difficult to do.[10]

The proposed new methodology of executing LASSO Regression in SQL is used as a first step in causal analysis as it utilizes casual Bayesian Network concepts. Specifically, feature selection in SQL utilizes Likelihood Ratios in order to determine the parents of the Markov Blanket (pMB).[11] A Markov Blanket (MB) is a set of variables that have the highest impact on the outcome of interest.[12] The 'parents' of the Markov Blanket are variables that occur prior to the outcome of interest.[13] In other words, the pMB consist of covariates that have an effect size that is greater than the predetermined threshold and have a statistically significant impact on the outcome. This is particularly important in causal modelling (see aim #3 below) where the sequence of events is critical

in ensuring that the predictor variables occur prior to the occurrence of the outcome of interest. [14] Identifying the pMB (sequencing) is done after correlations have been determined through Likelihood Ratios in SQL. Again, the use of SQL is important in that the model can be implemented directly in an EHR using SQL as a decision-making tool.

This study built upon a well-known feature selection method in predictive and causal modeling by Shojaie and Michailidis' known as LASSO Regression and Bayesian principals.[15] The Bayesian principals leveraged here are that of identifying the parents of the Markov Blanket (pMB). LASSO Regression is an algorithm-based method that is currently executable by the use of statistical tools. The study offers a new way of executing this method by the use of SQL queries which is integral in developing decision-making tools that can be built directly in EHRs. The purpose of this study was to demonstrate the use of LASSO Regression solely in SQL in order to accomplish feature selection related to a major public health issue. Specifically, this was demonstrated by using Bayesian principals to determine the optimal RVI and the causal relationship between revisits and CKD for Medicare patients with Type II Diabetes. The specific aims of the study were as follows:

Aim #1: To demonstrate the application of LASSO Regression in feature selection using SQL.

Aim #2: To predict the optimal RVI for Medicare patients with Type II Diabetes.

Aim #3: To determine the causal relationships between CKD and RVI for Medicare patients with Type II Diabetes.

Aim 1: To Demonstrate the Effective use of Conducting Feature Selection by using the Least Absolute Shrinkage and Selection Operator (LASSO) executed in Standard Query Language (SQL)

Feature selection is an important topic in data mining, especially for high-dimensional datasets. It is also a crucial and challenging task in statistical and probabilistic modeling. Historically, feature selection has been made based on the knowledge and experience of the researcher. Feature selection is even more important in high-dimensional datasets where defining the set of attributes that will most likely have the largest impact is difficult.[16] To build and interpret a model that takes into consideration all variables is also difficult. This is because in selecting features, the best subsets contain the least number of dimensions that most contribute to the accuracy of the model.[17] All other features are then discarded as being unimportant and not relevant to the model. In doing this, you remain with a subset of input variables with the most predictive information. This is a crucial stage of data preprocessing as the first step in pattern recognition, data mining and in determining causality.[18]

Aim 2: To Predict Chronic Kidney Disease (CKD) and the Optimal Revisit Interval (RVI) for Medicare Patients with Type II Diabetes

In demonstrating the effective use feature selection methodology, we apply the method to real data to predict the optimal RVI. RVI is defined in this study as the number of days between two outpatient consecutive appointments by the same patient and the same provider. Short RVI (sRVI) is defined as an interval that is shorter than the average, while long RVI (lRVI) is defined as the interval in days that is longer than the average.

This section discusses the importance of optimizing RVI specifically in patients who suffer from CKD (outcome of interest) as a complication of Type II Diabetes. CKD is used as an example of an adverse effect that could result from Type II diabetes in patients who tend to have a long RVI (lRVI). RVI are defined as long if the number of days between two consecutive appointments are above the average.

Overly frequent revisits related to diabetes set the stage for critical complications including CKD, overtreatment and unnecessary changes in regimen, which increases the risk of hypoglycemia.[19] In a US cohort of adults with stable and controlled Type II diabetes, more than 60% received overly frequent Hemoglobin A1c (HbA1c) testing, a practice associated with potential overtreatment with hypoglycemic medications, while contributing to the growing problem of waste in healthcare and increased patient burden in diabetes management.[20]

Patients with Type II Diabetes and CKD usually require long-term care that involves multiple revisits for provider-based evaluations, testing and pharmacological care after diagnosis. Despite diabetes affecting approximately 30 million people in the US, the optimal time in which a patient needs to return to the clinic for an outpatient follow-up visit as a preventive measure for CKD is not standardized in the US healthcare system.[21] Some providers consistently assign shorter follow-up visit intervals compared with other providers. This practice variation exists despite training and practicing at the same institution. The time in which a patient is asked to return to a clinic is currently driven by factors such as prescription medication cycles, provider experience and individual preferences, practice policies, and patient health status.[22] These intervals tend

to vary by provider, and facility, even when treating patients with similar demographics and health complexities.[23] This causes reduced optimization of access to care if patients that need to be seen sooner so that the diabetes will not lead to CKD are unable to be seen due to unavailability of appointments, while patients whose visits can be delayed may be seen sooner than needed. Too frequent and unnecessary visits lead to increased administrative burden and costs, and takes up appointment space for patients who need to be seen sooner.[24] This may lead to under- and over-treatment.[25] Intervening to reduce this variation in practice is challenging because research is not currently available on the optimal RVI for patients with chronic illnesses such as diabetes and hypertension which is customized for patients.

Guidelines from the American Diabetes Association (ADA) recommends that HbA1c be tested every 6 months in diabetic patients who have glycemic control and at times this is what drives the RVI for patients with controlled Type II Diabetes.[26] In comparing these guidelines with those around the world, we see that the National Institute for Health and Care Excellence (NICE) in the United Kingdom recommends testing every 6 months.[27] Like the US, guidelines are provided mainly based on expert opinion. NICE also recommends that practitioners adopt an individualized care approach to diabetes care that is tailored to the specific needs of each patient. NICE suggests considering comorbidities, polypharmacy risk and patient preferences.[28] The focus of this study is outpatient appointment visits that are related to diabetes. The following section takes a deeper dive into the issues involving diabetes in our nation and how it relates to RVI.

Aim 3: To Examine the Probable Causal Relationship between Revisit Intervals (RVI) and Chronic Kidney Disease (CKD) in Diabetic Medicare Patients

According to the National Institutes of Health (NIH), there are approximately 30.3 million people in the United States that are living with Diabetes.[29] The ADA reports that the economic cost of Diabetes in the US is 327 billion.[30] This includes $237 billion in direct medical costs and $90 billion in reduced productivity. Where might there be more room for patients in this system? Notably, a substantial portion of outpatient office visits are follow-up visits. According to the National Health Statistics Report for 2009, there were nearly 1 billion office visits in 2009, 30% of which were for routine follow-up of a chronic problem and an additional 26% of which were for preventive care or follow-up of an acute condition.[31]A 2010 Commonwealth Fund study of 11 industrialized countries found waiting times were longer in the United States than in all the other countries except Canada, Norway, and Sweden.[32] Therefore, this study seeks to examine how any causal relationships between CKD and RVI for Medicare patients with Type II Diabetes.

Significance of the Study

Feature selection is the process of selecting a reduced number of explanatory variables to describe an outcome variable.[33] Feature selection is used to reduce overfitting, reduce the scope of the study to enable algorithms to work faster, make it possible to handle high-dimensional data, and make the model easier to interpret but dropping variables that are redundant and irrelevant to the study.[34] LASSO Regression

helped to increase model interpretability by eliminating irrelevant variables that are not associated with the response variable and therefore reducing over-fitting.[35] This was the main focus of the LASSO Regression in SQL, because shrinking by dropping coefficients reduced variance without a substantial increase of bias.[36] This is particularly helpful with datasets that have a small number of observations and large number of variables. EHR-based screening has many advantages. Conducting predictive modeling within the EHR reduces the need to import the data into statistical software and therefore overcomes privacy concerns, saves time and reduces costs. It also does not require approval from Institutional Review Boards (IRB) since it is an operational improvement process geared towards practice management rather than research.[37]

There are approximately 1 billion outpatient revisits each year in the US; however, there is limited existing documentation on evidence based RVIs for even the most common and costly conditions. [38] RVI are of an even greater importance after the implementation of the Affordable Care Act (ACA) in 2010 and Medicaid Expansion in 2014 that led to increased access to care for approximately 30 million US citizens and lawful permanent residents who were previously uninsured.[39] This led to an increase in workforce shortages particularly in primary care.[40]As studies and initiatives for increasing workforce shortages in order to meet the demand, such as expanding the scope of practice for Nurse Practitioners, are underway, this study attempts to tackle the issue by optimizing the availability of currently available appointments.[41] This problem is confounded by increasing patient demand in an aging population and slow growth in

physician supply, which lags behind other countries on a per capita basis, and is further exacerbated by economic disparities.[42]

The aim of the study was to provide a comprehensive approach that allows for the identification of feasible and optimal revisit monitoring strategy and casual relationships between diagnoses based on patient characteristics and comorbidities. This involves combining multi-state models for dynamic prediction with patient level health and demographic models to structure the algorithm. The method enables researchers to remove confounding in EHR without accessing to statistical software. This study focuses on Diabetic Medicare patients because this condition is a serious and common chronic disease that results from complex risk factors and its complications constitute a major worldwide public health problem, affecting almost all populations in both developed and developing countries with high rates of diabetes-related morbidity, specifically CKD.[43]

Conceptual Framework

In Bayesian Network Analysis, directed separation for prediction and causality can be achieved through blocking Back-Door paths and by identifying the Markov Blanket of a Network.[44] A Markov Blanket (MB) of a Network is a set of features that shield other features from the rest of the Network.[45] In other words they are a set of the most relevant features in a dataset. LASSO Regression can be used to determine the parents of the Markov Blanket.

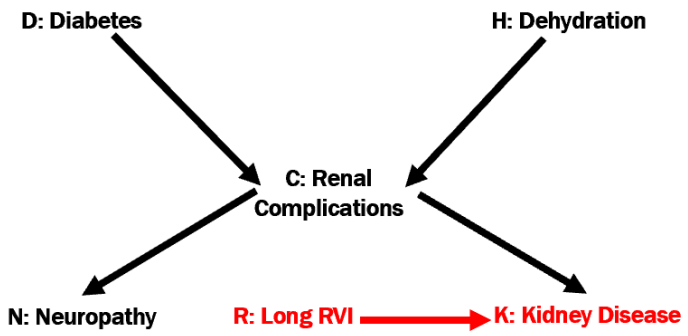


Figure 1:Blocking Back-Door Pathways

In Figure 1 above we see that Pearl's *do*-Operations enabled us to select features that should be conditioned on in order to determine the causal impact.[46] In Figure 1 Dehydration represents other comorbidities. The minimum viable set of features must be identified and conditioned through stratification, but stratification is not realistic in large dimensional datasets.[47] Therefore, LASSO Regression can be used to identify the parents of the Markov Blanket so that other covariates in the data can be dropped because it makes other variables irrelevant. Parents of the Markov Blanket are those covariates that are statistically significant and have a large impact on treatment that is being evaluated for its effect on the outcome.[48] LASSO regression is a regularization method for automatically penalizing extra features. It can make a regression simpler in terms of the number of features it uses. It sets the coefficients of the irrelevant features to zero.

Equation 1: Least Absolute Shrinkage Selection Operator (LASSO)

$$\text{Let } y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + b$$

x_1 through x_4 are all the available covariates in the dataset, and that m_1 through m_4 are the coefficients of regression.

Table 1: Identifying Parents in the Markov Blanket

Identifying parents in the Markov Blanket		
	LASSO Regression	PostgreSQL
Drop covariates that occur after the index date	Set coefficients of the covariates that occur after the target outcome to zero	Remove all covariates that occur after the onset of the target outcome
Determining the effect of the outcome	Add covariates one at a time into the model and determine the effect of the outcome	Stratify the data and determine the Likelihood Ratios and the inverse Likelihood Ratios for each covariate in the dataset for both RVI and CKD.
Determining Significance	Use the penalty term	Calculate Odds Ratio and a 95 percent Confidence Interval
Identify parents of the Markov Blanket	Automatic feature selection	Drop each individual covariate from the dataset and determine whether it changes the Odds Ratio beyond the Confidence Interval. This satisfies the counterfactual criterion for causality

LASSO Regression will add each covariate into the model, one at a time and if the newly added covariate does not improve the fit of the model enough to outweigh the penalty term of the including the covariate, then it will not be included.[49]

Study Assumptions

We used the Naïve Bayes Formula even though the assumption of independence is not verified. We used the prediction rule when the predictors are independent of each other. When predictors co-occur the assumption of independence is violated. In parsing the covariates in the data prior to calculating the Likelihood Ratio, we expected to reach an accurate conclusion.

Markov Models are beneficial in other similar cases where a decision problem involves risk that may occur over time and when the timing of events is important to the study. The computational complexity of such calculations are difficult and unrealistic. Therefore, the second assumption that Markov Models make is that patients are always in the same finite state of health known as discrete Markov States.

A Markov blanket makes the Markovian assumption that all you need to know in order to make a prediction about one node is encoded in the neighboring nodes it depends on. In a sense, a Markov blanket extends a two-dimensional Markov chain into a folded, three-dimensional field, and everything that affects a given node must first pass through that blanket, which channels and translates information through a layer.

In summary we used the Naïve Bayes methods even though the assumption of independence was not verified.[50] We assumed that the data points are independent of each other between and within groups and that each National Provider Identifier (NPI) belongs to a provider of a separate practice. Dropping features that are not in the causal path may open other Back-Door paths; however, this study did not analyze newly opened paths that were created.[51] Another assumption that Markov Models make is that patients

are always in the same finite state of health known as discrete Markov States.[52] A Markov blanket makes the Markovian assumption that all you need to know in order to make a prediction about one node is encoded in the neighboring nodes it depends on.[53] Lastly, in datasets with many variables but small number of cases LASSO selects most of the variables before it saturates.[54]

Chapter 1 Summary

This chapter provided an introduction to the study. It provided background information and the context around the study, then discussed the three study aims which were (1) to introduce a new method of conducting feature selection by the use of LASSO methodology executed in SQL; (2) to predict CKD and the Optimal RVI for Medicare patients with Type II Diabetes, and (3) to examine the causal relationship between RVI and CKD in Diabetic Medicare patients. The chapter discussed the significance of the study and the conceptual framework used. Lastly, this introductory chapter provided the study assumptions.

CHAPTER 2: LITERATURE REVIEW

This chapter first takes a deep dive into the literature involving patient Revisit Intervals (RVI). An overview of the feature selection methods used in the study as well as the conceptual framework is then discussed in subsequent subsections of this chapter. The chapter subsections that follow provides information on Chronic Kidney Disease (CKD) and RVI Likelihood estimations as well as the impact of covariates on CKD.

Currently, revisits are scheduled by providers based on heuristics and experience, with large variability and little empirical evidence. Yet, evidence suggests that RVI can be safely lengthened for many patients without decrements in quality or outcomes.[55] Longer RVI for diabetic patients who need to be seen sooner can lead to complications related to high blood sugar which can affect various cells and organs in the body. Complications may include kidney and eye damage, which could result in blindness, or an increased risk for heart disease or stroke.[56]

Revisits in patients with Type II Diabetes

A systematic review was conducted based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[57] The inclusion criteria used for the literature review involved focusing on studies involving diabetes, CKD, RVI, Markovian Methods, Pearl's Do-Operator, Least Absolute Shrinkage and Selection

Operator (LASSO) Regression, Bayesian Causality, Predictive Analytics and feature selection. Key words used in the search included Bayesian Networks, Causal Directed Acyclic Graphs, Revisits, Diabetes, Chronic Kidney Disease, Causality, high-dimensional networks, LASSO Regression. Full versions as well as their abbreviations were used. Bayesian principles were first introduced in the 1960s by Judea Pearl;[58] therefore, the search first focused on studies in the United States (US) from the years 1960 to 2019. Then the search evolved to studies internationally with the same narrow focus. Search engines used include PubMed, PubMed Central (PMC) which is hosted by the National Institutes of Health (NIH) and Europe PMC; A proprietary database called Excerpta Medica Database (EMBASE), Cochrane Library which is best known for its systematic review and UpToDate which provides detailed reviews of various clinical topics.

There were 76 studies that included evidence-based guidelines on RVI for various conditions, out of those there were two that met the inclusion criteria of this study. There were no specific studies that analyzed RVI on Medicare patients with Type II Diabetes; however, there are notably 8 studies that recommended specific follow-up times (Table 1).[59, 60] Table 2 illustrates a review of studies that provide evidence-based guidelines for RVI for the top 5 chronic conditions that make up the greatest number of outpatient visits in 2010. These 5 conditions accounted for approximately $281 billion in healthcare expenditures in 2010.[61] The chronic conditions were hypertension, arthritis, mental disorders, back problems, and Chronic Obstructive Pulmonary Disease (COPD)/Asthma.[62]

A study by Quinn et al (2010) recommended RVI of 1 -2 months until the Hypertension was managed, then 3-6 months once the diabetes was controlled.[63] Keenan (2009) recommended over 6 months of monitoring intervals for diabetic patients.[64] Van den Bent (2008) determined that intervals for the chronic progressive disease, epilepsy, should be based anywhere from 1 month to 1 year.[65] Schulberg (1998) noted that for chronic depression patients should be seen between 6 to 8 weeks. Other studies looked at the optimal RVI for melanoma and found the optimal period to be 2 weeks (Frencken, 2009).[66] Previous studies have determined provider decision-making regarding RVIs using mostly provider surveys.[67-9] There was only been one study that examined predictors of RVI assignment as the primary study focus in actual practice.[68] No studies were found that utilized a PostgreSQL to determine parents of the Markov Blanket. One Study uses the Stratified Covariate Balancing to predict the prognosis of patients from their diagnostic history, but it does not focus on predicting RVI.[69]

Table 2: Existing Evidence-Based Guidelines

Guidelines	Search Yield	Number of Studies Meeting Inclusion Criteria	Studies Meeting Inclusion Criteria	Evidence Based	Recommended Follow-up Interval
HTN evidence-based guidelines and follow-up	76	2	Quinn, 2010[7]	No	1-2 months initially until well controlled; 3-6 months when stable
			Keenan, 2009[8]	Yes	>6-month monitoring intervals
Van den Bemt, 2008[10]				**No**	**Reviewed guidelines recommend follow-up intervals of 1 month to 1 year based on either disease severity or the local healthcare system**
Mental disorders evidence-based guidelines and follow-up (depression)	93	1	Schulberg, 1998[11]	No	Weekly or biweekly visits during the initial 6-8 weeks of therapy
Back pain evidence-based guidelines and follow-up	21	0	None	N/A	N/A
Skin disorders evidence-based guidelines and follow-up (melanoma)	85	2	Francken, 2009[12]	Yes	2 weeks
			Francken, 2008[13]	No	Yearly for 10 years
NMSC evidence-based guidelines and follow- up	1	0	None	N/A	N/A
Melanoma evidence-based guidelines and follow-up	19	2	Francken, 2007[14]	Yes	No specific time period recommended; a reduction in the frequency of follow-up visits may be safe and effective
			Einwachter-Thompson, 2008[15]	Yes	No follow-up for thin melanomas (<0.5 mm)

COPD indicates chronic obstructive pulmonary disease; HTN, hypertension; NMSC, nonmelanoma skin cancer.

Previous research found that female providers assign significantly shorter RVIs than their male counterparts, regardless of patient demographics and management.[70] Female providers seem to focus more on preventive care compared to male providers.[71,72] Provider experience was also one of the most powerful predictors of RVI allocation.[73,74,75] Providers who have a greater number of total patients may have less free time available in their schedule and may therefore postpone revisits. Seeing patients more

frequently may also benefit providers financially if they are compensated per visit. These findings implied that a significant amount of RVI variation is due to modifiable factors and can be impacted upon by feedback to providers about their individual practice patterns and cannot be explained by the size of the patient population or financial incentives.

A randomized clinical trial on patients with Hypertension examined at blood pressure control, patient satisfaction and whether or not they adhered to their medications.[76] This study looked at health outcomes for patients at 3 months and 6 months for 3 years. The findings showed that RVI of every 3 months had equivalent outcomes as RVI every 6 months.[77] Previous studies showed that the mean RVI was 12.4 weeks (range 1–42 weeks) and was similar for patients with diabetes and hypertension which accounted for 35.7% of the variance in RVI assignment. The identity of the physician was the largest contributor to the variance, accounting for 14.7%.[78]

The existence of great variation in recommended revisit intervals suggests that physicians are uncertain about what interval is best. Educational interventions can successfully retrain providers to extend the return visit interval and reduce the scheduling of routine and perhaps unnecessary appointments.[79]

The following subsection discusses the first step in predicting RVI which is feature selection. It provides an account for the feature selection method used in this study and how it leverages Bayesian principles and Least Absolute Shrinkage and Selection Operator (LASSO) Regression.

Methods for Feature Selection

There are many different types of predictors in electronic health records, e.g. diagnoses, treatment, medications, or demographics to predict health outcomes for a patient. Bayesian methods of determining these predictors are computationally very intensive and therefore not particularly appropriate for high-dimensional settings.[80] Therefore evidence-based attribute selection enables us to shrink the universe of attributes to those that are relevant to the model. In this study we used diagnosis codes to determine the Likelihood of CKD and having a short RVI. Diagnosis codes are more predictive than laboratory results since lab work can be highly influenced by medication. For example, a diabetic patient's lab results may show up as normal after they have taken medication.

Generally, researchers have relied on clinicians to select diagnoses that are known to affect the target outcome. The approach that we prefer is to use all diagnoses in order to ensure that all potential predictors are considered.[81]This provides a more accurate prediction model than if we were to remove some diagnoses or if we were to batch the diagnoses into homogenous categories before determining the predictions.[82]

The LASSO method was first formulated by Robert Tibshirani in 1996.[83]The goal of this process is to minimize the prediction error. This is a powerful method of conducting regularization and feature selection. It puts a limitation on the sum of the absolute values of the model parameters. The limitation is essentially a threshold (upper

bound) in which the sum cannot exceed. This is done by penalizing some coefficients and by shrinking them to zero. The LASSO minimizes the sum of squared errors, with an upper bound on the sum of the absolute values of the model parameters. The tuning parameter known as lambda (λ) determines the strength of the penalty. When λ is large the coefficients are pushed closer to zero therefore reducing dimensionality in the data. Λ can range from zero to infinity. In other words, the larger λ is, the greater number of coefficients are shrunk to zero. λ is inversely proportional to the upper bound for the sum of the coefficients, *t*. When λ is zero, then we have an Ordinary Least Squares (OLS) Regression.

Bu¨hlmann and van de Geer [84] defined LASSO Regression as follows:

If *y* is the outcome variable, *x* the predictor variable, *t* is the upper bound for the sum of the coefficients. Then his optimization problem is equivalent to the following parameter estimation: -

Equation 2: Bu¨hlmann and van de Geer

$$\beta(\lambda) = \left(\frac{\|\mathbf{Y} - \mathbf{X}\beta\|_2^2}{n} + \lambda \|\beta\|_1 \right)$$

Where $\lambda \geq 0$ (Penalty strength).

Y is the continuous parameter

X is the design matrix

β is the parameter vector

The above process is known as regularization. Coefficients that are not shrunk to zero are then advanced into the model.

The southern California's great thinker of causality, Judea Pearl[85] defined the concept of causality based on associations between nodes but only under specific criteria can true causal relationships be inferred.[86] The criteria are(1) Association- there must be an association between the predictor variable and the outcome of interest; (2) Mechanism - there must exist a biochemical pathway for disease progression; (3) Sequence - the predictor variable must occur before the outcome occurs; and lastly (4) Counterfactual – Rubin and Holland[87] introduced the counterfactual of potential outcomes and the expected outcomes.[88] Advancements in Bayesian Network research is tied closely to causality and the seminal work of Judea Pearl.[89] Methods such as Directed Cyclic Graphs and Bayesian Networks provide techniques that enable the establishment of causation from association when working with non-experimental data. Variable selection is even more important in high-dimensional datasets and it is often difficult to determine which variables are relevant.[90] In high dimensional observational data, the causal impact of treatment can be optimized and achieved through blocking Back-Door paths in order to identify the Markov Blanket.[91] By holding comorbidities '*ceterus paribus'*, meaning constant, we can isolate the treatment effect. The observed differences are then identified as the treatment effect.

The Markov blanket of treatment is a group of covariates that blocks the effect of other covariates on treatment.[92]A number of studies have shown that the Markov blanket can be used to decrease high-dimensional data to its relevant variables.[93] Markov blankets include direct causes that are the parents and co-parents, as well as the effects which are the children. Parents in the Markov Blanket can be determined by analyzing

independent variables that occur before treatment, this removes covariates in the causal path from treatment to outcome which tend to be the complications associated with treatment.[94] LASSO Regression is one of the methods that can be used to identify the pMB. It reduces covariates in a model to only those that are statistically significant and have a large effect size. Afterwards the remaining covariates that are outside the pMB are dropped.

Markov Blanket is a learning algorithm that is used to find a Bayesian Network that defines an outcome node. Regression, Decision Trees, Support Vector Machines and Neural Networks are alternatives to other supervised methods that use similar discriminative models. However, Markov Blanket uses a generative approach. In data learning that involves the use of Markov Blanket, the direction of the arcs are not initially considered causal until more information is known about the nodes. A Bayesian network arcs represent statistical dependence between different features and can be automatically elicited from the data by Bayesian network.[95] Markov blankets include direct causes that are the parents and co-parents, as well as the effects which are the children. pMB can be determined by analyzing independent variables that occur before treatment, and this removes covariates in the causal path from treatment to outcome which tend to be the complications associated with treatment.[96]

Markov Blanket using Likelihood Estimations

Pearl used the do-operator to simplify the composition of a causal model by removing an impacted variable from its normal causes. This is the foundation of the

conceptual framework used in the study design.[97] Pearl used the *do*-operator to set the value of the manipulated variable, [*do*(X = x)]which is similar to setting the variable of X to be equal to the value x. This undoes one or more causal relations, basically by cutting off x from variables that would usually cause it.

Karl Friston stated that the organizing principle of life is that entities contained within a Markov blanket seek to maintain homeostasis by minimizing uncertainty through their Markov blanket.[98]The term Markov blanket, as coined by Judea Pearl[99], plays an important role in determining the treatment effect. The Markov Blanket of the outcome is determined by analyzing the sequencing of events before the outcome of interest occurs. In a Bayesian Network the Markov Blanket is the set of all Parents, Children and Co-Parents of a node. Several investigators have shown that Markov Blankets could reduce the number of stratifications by 3 to 4,250-fold, depending on the size of the data. [100] It can be shown that a node is conditionally independent of all other nodes given values for the nodes in its Markov blanket. Hence, if a node is absent from the class attribute's Markov blanket, its value is completely irrelevant to the classification of the outcome. Mediators were not included in the stratification. This is important because it provides an understanding of the variables that have a major impact on the variable of interest. Therefore, other covariates that are outside the blanket can be removed from the analysis.

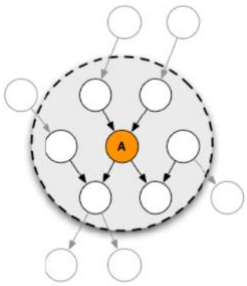


Figure 2: Node A of the Markov Blanket

In a Bayesian Networks, the probability of some nodes depends on other nodes upstream from them, which are sometimes causal. The Markov Blanket of CKD (A) contains all variables, which if we know their states, will shield node A from the rest of the network (Figure 1). These means the Markov Blanket of a node is the only knowledge needed to predict the behavior of node A. Determining the Markov Blanket provides us the relevant predictor variables, which are particularly helpful when there are many variables such as in this claims data sample.

The Markov Blanket of an attribute or node comprises of all nodes that make an outcome but that are conditionally independent of all the other nodes in the model. The Parents are used for separating the data coming from their ascendants, while children separate the information coming from the descendants. Co-parents are used to separating information from the ascendants of the children. The outcome is conditionally dependent on children and marginally independent of spouses. Parents in the Markov blanket consist

of covariates that have an effect size that is greater than the predetermined threshold and have a statistically significant impact on the outcome.

Our proposed method considers every single comorbidity, and it also involves calculating a 95% Confidence Interval in PostgreSQL. This method is demonstrated in this study as a predictive model that establishes the optimal RVI for patients with Diabetes. This paper provides an alternative method of identifying pMB using Likelihood Estimations that runs the probability of each state (pA - which are positive and negative outcomes) as they correspond to the frequency in the sample (Figure 2).

Figure 3: Likelihood Ratio

This new method incorporates patient comorbidities and specific characteristics to determine the optimal RVI directly in SQL. It involves combining multi-state models for dynamic prediction with patient level health and demographic models to structure the algorithm. It enables researchers to remove confounding in EHR when predicting the risk for chronic conditions without needing to access any statistical software.

Conceptual Framework

Judea Pearl highlights in his book of Causality "*I no longer see no greater impediment into scientific progress that than prevailing practice of focusing all our*

mathematical resources on probabilistic and statistical inferences while leaving causal considerations to the mercy of intuition and good judgment."[101] Rubin and Holland[102], who introduced the counterfactual of potential outcomes and the expected outcomes is Pearl's Do-Operator Pearl[103] used the *do*-operator to simplify the composition of a causal model by removing an impacted variable from its normal causes. This is the foundation of the conceptual framework used in the study design.

Pearl used the do-operator to set the value of the manipulated variable, $do(X = x)$ which is similar to setting the variable of X to be equal to the value x, which *undoes* one or more causal relations, basically by cutting off x from variables that would usually cause it. The purpose of this process is to set all other interventions related to x as irrelevant by setting the value of x to equal x.

Determining the Markov Blanket provides us the relevant predictor variables, which are particularly helpful when there is a large number of variables such as in this claims data sample. In other studies, involving early prediction of disease progression such as Multiple Sclerosis, inferences were drawn from the selection of an appropriate model structure starting from a large and comprehensive model, and then systematically simplifying it through the removal of those variables which were conditionally not relevant; however, the study the authors used Cox Regression to identify the variables.[104] Other studies have used other various types of Regression techniques to accomplish the same goal.[105] Varol et al used Markov Blanket based feature selection algorithms and wrapper based feature selection algorithms to identify and select relevant features, [106, 107, 108, 109, 110] and others used this technique for text classification.[111]The HITON method was

used by other studies to reduce the number of variables in the prediction models by three orders of magnitude compared to the initial variable group while improving or maintaining accuracy. HITON works by inducing the Markov Blanket of the variable to be classified or predicted.[112] The wrapper method in the Varol et al (year) is similar to the methodology used in this paper; however, the method of identifying irrelevant variables is different (reference). Other studies describe a family of algorithms TIE* that can discover all Markov boundaries in a distribution;[113] while Alemi et. al (year) used the Stratified Covariate Balancing technique;[114] Li et al also used inverse Regression to determine the Markov Blanket;[115] others used Recursive Bayesian Networks[116] and Representative Sets[117].

The pMB in this study is determined with the use of Likelihood and Odds Ratios.[118] The inverse Likelihood Ratios are calculated as well in order to verify that the covariates in the Markov blanket do not interact with any remaining covariates to have a statistically significant and that have a large effect on treatment.[119] Likelihood Ratios are a ratio of two conditional probabilities.[120] It is a probability concept that is commonly used to develop predictive data algorithms.[121] In probability theory, Likelihood Ratios are used to indicate how useful one element is in predicting the occurrence of an event.[122] Likelihood Ratios measure the association between each predictor and an outcome variable.[123] It is the ratio of the probability of a predictor in the presence of the outcome, to the probability of the predictor when the outcome is not present.[124] This ratio provides information on the number of times the odds of an outcome changes in the presence of a risk factor or predictor.[125] For example, a Likelihood Ratio of 40 means that the odds of

the outcome changes 40 times in the presence of the risk factor. The inverse Likelihood Ratio in which the ratio is close to zero, demonstrates that the outcome is unlikely to occur in the presence of the predictor variable.[126] These theoretically infers that in the presence of the predictor, the outcome will not occur.[127] This mainly occurs more often in smaller datasets, but in large datasets we may see Likelihood Ratio that are near to zero but not zero.[128]

Under the assumption of independence, we calculated the Likelihood Ratio associated with each diagnosis as the ratio of prevalence of the diagnosis among patients with CKD and those without CKD patients, and the same for Long and Short RVI.[129] The Likelihood Ratios for CKD and RVIs were then calculated as follows:

Equation 3: Odds of CKD

$$\text{Change in the Odds of CKD} = \prod_{\substack{Patient's \\ Diagnoses}} \mathbf{LR_{Diagnosis}}$$

The *pai* notation indicates that the calculation is performed on each individual diagnosis code.

Estimating CKD via the Likelihood Ratio Method

LR(dx) = dx among diabetic patients with CKD / prevalence of dx amongst patients without CKD

= p(Dx|CKD) / p(Dx|No CKD)

= (# of patients with CKD & a specific Dx/ # of total patients with CKD)

(# of patients without CKD & a specific Dx / # of total patients without CKD)

Equation 4: Likelihood Ratio

$$\text{Change in the Odds of RVI} = \prod_{\substack{Patient's \\ Diagnoses}} LR_{Diagnosis}$$

Revisit (RVI) Likelihood Ratio

The Likelihood Ratios of each individual diagnosis is also determined against the long and short RVI. The originating appointments or indexes were selected at random across each distinct patient and the number of days between the originating appointment and the follow-up appointment was determined for each random appointment index (seed). In this study, the RVI is defined as the number of days between two appointments. RVI per each unique patient with Diabetes, was calculated and the average RVI per each patient and for the overall sample was determined. The RVI per unique patient was determined as the number of days between two service appointments for each patient who had Diabetes, starting with the randomly identified index date. The average RVI across the sample was then calculated and found to be 33.88 days. Long RVI was defined as those encounters with an RVI that is greater or equal than then 33.88 days across the sample, and short RVI was defined as encounters with RVI that was less than 33.88 days. Likelihood Ratios for each diagnosis as they pertain to Long RVI and Short RVI were then calculated. Table 12 illustrates the distribution of RVI in the data.

LR(dx) = dx among diabetic patients with Long RVI / dx amongst patients with Short RVI

= p(Dx|Long RVI) / p(Dx|Short RVI)

= (# of patients with Long RVI & a specific Dx/ # of total patients)

(# of patients with Short RVI & a specific Dx / # of total patients)

Impact of Covariates on Chronic Kidney Disease

Each unique combination of covariates is considered one stratum or subgroup. In the above process individual diagnoses that occur post-index are parsed (Table 3) and ordered in order to determine the counts and percentages of each strata.

Table 3: Example of Unique Strata

ID	*Date*	*Next_Date*	*Difference*	*Strata*
10000000015	*2015-11-24*	*2015-11-25*	*1*	*E118, I2510, J449*
10000000053	*2016-04-04*	*2016-05-17*	*43*	*E1142, B351, M79674, M79675*
10000000099	*2012-04-03*		*0*	*K824, I4891, R 7989, E119*

The covariates were used to define the strata. Within each stratum the effect of RVI and other covariates on CKD for diabetic Medicare patients were examined. Therefore, in order to determine the comparative effectiveness of RVI, the data table was then divided into cases and controls. Covariates used are Date of Service Index selected randomly and diagnoses that occur after the index Date. This allowed cases and controls to be compared while keeping covariates constant and the average impact of Short RVI on CKD to be calculated.

This process in table 4 and 5 below organized the claims into a partial factorial design where cases (X=1) and controls (X=0) are analyzed at different factorial

combinations of covariates. Therefore, the cases within this study were diabetic Medicare Patients who received Short RVI (a_i) and the controls were diabetic Medicare patients with short RVI (b_i). Each stratum reports the impact of cases and controls on an outcome, *Y=1* (CKD).

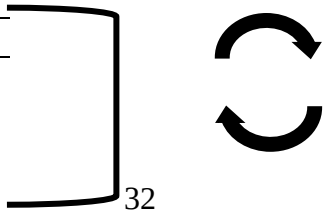
Table 4: Outcomes Strata

Outcomes in ith stratum, $i = 1,\ldots, k$		
	$Y = 1$	$Y = 0$
Cases ($X = 1$)	a_i	b_i
Controls ($X = 0$)	c_i	d_i

In highly dimensional observational data, we can optimize the number of variables that need to be conditioned through Propensity Score Matching. Alternatively, stratification can be used to block the Back-Door pathways (Table 5) and determine the causal effect by keeping the condition in the stratum constant while observing the difference within the stratum. Stratification involves dividing the data into subgroups called strata in which each case has one or more confounding variables that are held constant, so that the effect of long RVI on CKD can be determined without the influence of confounders.

Table 5: Conditioning of Strata

Same level of covariates 1- *k*		
	Desired Outcome	Undesired Outcome
Cases	a_i	b_i
Controls	c_i	d_i



Repeat for all Stratum in order to block the Back-Door path

Same level of covariates 1- *k*		
	Desired Outcome	Undesired Outcome
Cases	a_{i2}	b_{i2}
Controls	c_{i2}	d_{i2}

Observed differences may be due to the treatment or could be due to other explanations which we need to block against in order to isolate the effect of the treatment. This concept was first introduced in calculus by Judea Pearl (Equation 5). Pearl developed a mathematical methodology that discredits Back-Door paths in order to determine causality.

Equation 5: Pearl's Do Operator

$$P(y \mid do(x)) = \sum_{z} P(y \mid z, x)P(z),$$

Judea Pearl used calculus to explain causality.[130] The operator do(), marks an action or an intervention in the model.[131] The causal calculus uses Bayesian conditioning, p(y|x), where x is observed feature, and causal conditioning, p(y|do(x)), where an action is taken to force a specific value x.[132] The objective is to generate probabilistic formulas for the effect of interventions in terms of the observed probabilities.[133] However, the application of this methodology is difficult because the counterfactual variables Y(0) and Y(l) are unobservable.

Chapter 2 Summary

This chapter provided an account of the literature review of the main concepts in this study. Specifically, it provided an analysis of previous studies conducted on RVI for

patients with Type II Diabetes, methods of feature selection, as well as the on the Bayesian Methodologies used for predictive analytics. The chapter went through the conceptual framework, background on the methods of estimating the Likelihood of CKD and RVI as well as the impact of covariates on CKD.

CHAPTER 3: METHODOLOGY

Chapter 3 provides a detailed account of the study methodology. Includes information on the data source, a description of the data and how it was procured, and the importance and advantages of using this dataset. The chapter provides the Institutional Review Board (IRB) approval, measures taken, methods description as well as the step by step process for each of the three aims.

Data Source

The dataset contains data from the Center for Medicare and Medicaid Services (CMS) Limited Dataset (LDS), from 2014 – 2016 that was made available for free to George Mason University faculty and students for free through the Health Administration and Policy, Discovery Science and Health Informatics – Virtual Environment. The data was gathered primarily for the use of CMS operations related to revenue and utilization.[134] It contains deidentified information on patient demographics, Medicare status, procedure and diagnosis codes as well as provider identification numbers. It contains 135 million outpatient claims and 823,584 distinct patients. For this purpose of this study, we included only patients with diabetes. 82.9% of the sample data subjects were white, 57.2% were female, and 39% were over the age of 75 years. A Data Use Agreement was required for use of CMS data. This was signed by the Principle

Investigators, on 22 October 2018. Category 4 exemption approval to conduct the study was obtained from IRB in March 1, 2019 by the Office of Research Development, Integrity, and Assurance. IRB number is [1337403-1].

Health Services utilization data, commonly referred to as claims data, are derived from reimbursement information or the payment of bills. As a general rule, those pieces of information that are required to determine payment/reimbursement will be of higher quality than other information reported on a claim. Also included in the available CMS data are enrollment data, which are the basis for determining whose bills are qualified to be paid by Medicare. Demographic data, such as age, date of birth, race, place of residence and date of death, are also included in these administrative datasets and are considered largely reliable and valid. Files containing this type of information about all enrolled Medicare beneficiaries are known as "denominator" files.[135]

CMS Health Services Utilization (Claims) Data is the largest and most robust dataset in the United States. LDS files contains a 5% random-sample of claims-level data, which contain greater beneficiary-level information. This dataset facilitates a more robust patient- and encounter-level analyses and assessment of downstream outcomes.[136] It is estimated that over 98% of adults age 65 and over are enrolled in Medicare, making Medicare data one of the richest sources of utilization information in the country.[137] Over 45 million beneficiaries enrolled in the Medicare program today[138], allowing for detailed sub-group analysis with reduced concerns about loss of statistical power. It contains an aggregate of data for most of the population. Claims data for the

population group that is under the age of 65 is housed in various payer databases. Furthermore, 99% of deaths in the US among persons age 65 and older are accounted for by the Medicare program.[139] The inclusion/exclusion criteria will involve using the entire 2016 dataset, narrowing the data to only outpatient claims data, and limiting it to include only diabetic patients who were over the age of 65.

The Revisit Intervals (RVIs) were then calculated as the number of days between two consecutive appointments by the same patient and the same provider. From there causality was derived from determining correlations, sequences, mechanism and counterfactuals related to the relevant features. The optimal RVIs were then deduced from the probabilities of Chronic Kidney Disease (CKD) occurring given the presence of specific comorbidities. The probabilities were calculated by analyzing the set of comorbidities that patients had prior to a set date, and the prevalence of CKD after the set encounter date.

Measures

The Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) codes for Diabetes were Clinical Classification Software (CCS) codes 49, 50. Other endocrine disorders were 51, therefore these were included in the study. For CKD CCS codes 158, 161 were included. We excluded CCS codes 157 for acute cases and 33 for neoplasm.

Age was categorized into 6 categories – less than 65, between 65 and 69, 70-74, 75-79, 80-84 and greater than 84. Race will be divided into 4 categories – black, white, Asian and other. Ethnicity will be described as Hispanic, non-Hispanic and unknown; and gender will be described as male or female. A table of the variable names, codes and descriptions will be included.

The diagnosis code assignment for patients with diabetes and CKD was based on the Agency of Healthcare Research and Quality's (AHRQ) Clinical Classification Software (CCS) diagnosis categories. The International Classification of Diseases volume 9 (ICD 9) and ICD-10 (Clinical Modification) codes for kidney conditions were CCS category is 156 -163, and this gave us a total of 510 codes. The diagnosis codes which were tautological to diabetes and CKD were removed from these groupings.

Methods

The Least Absolute Shrinkage and Selection Operator (LASSO) method involves determining the parent Markov Blanket (pMB). Data was divided into natural existing strata. Each stratum has a combination of covariates that are specific to each encounter. Then the following steps were taken in order to identify the pMB: -

An appointment index date was randomly selected and the average RVI was then calculated as the number of days between the index date a subsequent encounter date for the same patient. All comorbidities that occurred after the index date were excluded from the analysis because they are tautological predictors of CKD. Only those diagnoses that occurred prior to the index date were used. Within each stratum, encounters with CKD

and those without CKD were identified as cases if they had a RVI that was less or equal to the average RVI (Short) and as controls if they had a RVI greater than the average RVI (long). In determining pMB, we used a Filter Method of selecting features by first calculating the prevalence of a risk factor in the presence of CKD / the prevalence of a risk factor in the absence of CKD. This provided us with the Likelihood Ratio for CKD. Then we calculated the prevalence of a risk factor in the presence of short RVI / the prevalence of a risk factor in the presence of a long RVI. This provides the Likelihood Ratio for RVI.

$= p(Dx|Outcome^{+}) / p(Dx|Outcome^{-})$

$= (\#\ of\ patients\ with\ Outcome^{+}\ \&\ a\ specific\ Dx / \#\ of\ total\ patients) /$

$(\#\ of\ patients\ with\ Outcome^{-}\ \&\ a\ specific\ Dx / \#\ of\ total\ patients)$

Similar to the Wrapper Method, the results for these two categories were then combined ranked by Likelihood Ratio. A 95% Confidence Interval was calculated in Standard Query Language (SQL) and used to determine the usefulness of the covariates based on their statistical significance. This provided a subset of features which were used in the prediction model. To select the features for the subsets' we used backward selection. We started with the full dataset and dropped each individual variable one by one and to determine their impact. The select subset are the pMB – the results of which are provided below.

The proposed methodology involved determining the Likelihood Ratio that CKD would be expected in a patient with a risk factor compared to the Likelihood Ratio of a

patient having CKD without the risk factor. Likelihood Ratio is used to assess the chances that a patient would have CKD given a RVI based on patient comorbidities.

Aim 1: To Demonstrate the Effective use of Conducting Feature Selection by using the Least Absolute Shrinkage and Selection Operator (LASSO) executed in Standard Query Language (SQL)

Least Absolute Shrinkage and Selection Operator (LASSO) is a regularization method, and thus provides a way reducing overfitting by using less complicated functions. As discussed later, this can be done manually by examining significance of the coefficients and discarding those variables whose coefficients are not significant. One way to do this is by dropping less important variables, after checking the impact they have on the outcome. However, this can become tedious when conducting analyses in large high-dimensional data with many covariates. In searching for the combination of covariates there could potentially be 2^k binary covariates which is computationally labor-intensive. Therefore, in determining the covariates that have a significant impact on the CKD we determine a Likelihood Ratios threshold that leads to a manageable number of covariates. Like LASSO Regression, LASSO executed through SQL provided a way of selecting significant variables by reducing the number coefficients of unimportant predictors without the use of *p*-values. The Markov Blanket renders the node independent of the rest of the network; the joint distribution of the variables (strata) in the Markov Blanket of a node is enough knowledge for calculating the distribution of the node. Each stratum holds a unique combination of confounding variables constant.

The Likelihood Ratio in the contingency table illustrates the number of times the outcome and the risk factors co-occur. In this table the Likelihood Ratio is calculated as (a/ (a + c)) / (b/(b +d)

Prior Odds = Is the odds of the outcome occurring before considering the risk factor.

Probability of the outcome occurring / Probability of the outcome not occurring

= *p*(Outcome) / 1-*p*(Outcome) = (a + c) / (b + d)

Posterior Odds = Is the odds of the outcome occurring before considering the risk factor.

= Probability of the outcome occurring in the presence of a predictor / Probability of the outcome not occurring in the presence of a predictor

= *p*(Outcome with Predictor) / 1-*p*(Not Outcome with Predictor)

= LR of Predictor * Prior Odds of the Outcome (Bayes Formula)

Conditional probabilities were used to reduce the number of comorbidities needed in the predictive analysis. This is represented by a Contingency Table that illustrates counts of joint observations in the cells (Table 6). Contingency tables provided an account of two or more variables that occur together, such as CKD and RVI. In Table 6, *a* and *c* represented counts of when Attribute 1 is present, while *b* and *d* represented the absence of Attribute 1. The presence of risk factors is represented by *a* and *b*, while the absence of risk factors (Attribute 2) are represented by *c* and *d.* The total number of times that both events co-occur is given in *a*; whereas the *d* provides an account of the number of times neither of the events co-occur. The total number of times Attribute 1 is present is

calculated as a + b, and the total number of times that the risk factor is present (Attribute 2) is calculated as a + c.

A positive Likelihood Ratio (LR+) is = [a/(a+c)] / [b/(b+d)], while a negative Likelihood Ratio is calculated as = [c/(a+c)] / [d/(b+d)]. The universe of probabilities are provided as the total of a + b + c + d.

Table 6: Condition and Risk Factor Counts

		Attribute 1 (Condition)		
		Present	**Absent**	**Total**
Attribute 2 (Risk Factors)	**Present**	a	b	a+b
	Absent	c	d	c+d
	Total	a+c	b+d	a+b+c+d

Likelihood Ratios measure how informative a predictor is. Any comorbidity that have a Likelihood Ratio that is greater than 10 or less than 0.1 has a large effect of the outcome. Comorbidities with Likelihood Ratios that are between 5 and 10 or 0.1 and 0.2 cause a moderate effect on the outcome. Those that are less than 2 or greater than 0.5 have a small effect on the outcome. A Likelihood Ratio that is equal to 1 has no effect on the outcome.

Aim 2: To Predict Chronic Kidney Disease (CKD) and the Optimal Revisit Interval (RVI) for Medicare Patients with Type II Diabetes

The second aim is to provide a comprehensive approach that allows for the identification of feasible and optimal revisit monitoring strategy based on patient

characteristics and comorbidities. In this study we wanted to predict the odds of a patient having CKD from the patients' diagnoses and RVI by the use of Naïve Bayes. We calculated the Likelihood Ratios associated with each diagnosis. In handling thousands of diagnoses this method provides a simpler way of calculating Likelihood Ratio directly in the patients' electronic medical record, that would otherwise be difficult to do by using statistical tools such as Regression.

$$\text{Change in the Odds of CKD} = \prod_{\substack{Patient's \\ Diagnoses}} \mathbf{LR}_{\mathbf{Diagnosis}}$$

Only predictors that occurred before the outcome were used, while those that occur prior to the outcome were discarded. All diagnoses were included as potential predictors of CKD.

We randomly selected an appointment date for each patient and generate an index number for each case. Based on the assumption of independence, the Likelihood Ratio of each diagnosis is calculated as the ratio of prevalence of the diagnosis among patients with CKD to those without CKD.

LR = (Prevalence of predictors when the CKD is present)/(Prevalence of predictor when the CKD is absent)

This calculation involved determining the number of patients with CKD (a + c), the counts of patients with diagnosis amongst those with CKD and the risk factor (a); the counts of patients without CKD (b + d), and those without CKD and with the risk factor (b).

LR = p(Dx|CKD) / p(Dx|No CKD)

= (a / (a+b)) / (b / (b + d)) (Table 1)

Equation 6: Common Odds Ratio Calculations

Estimate of Common Odds Ratio of Impact of X on Y

$$\pi_i = \frac{a_i + d_i}{n_i} \quad Q_i = \frac{b_i + c_i}{n_i}$$

$$R_i = \frac{a_i d_i}{n_i} \quad S_i = \frac{b_i c_i}{n_i}$$

$$V = \frac{\sum_i \pi_i R_i}{2\left(\sum_i R_i\right)^2} + \frac{\sum_i Q_i S_i}{2\left(\sum_i S_i\right)^2} + \frac{\sum_i \left(\pi_i S_i + Q_i R_i\right)}{2\left(\sum_i R_i\right)\left(\sum_i S_i\right)}$$

$$\widehat{\mathrm{OR}} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$$

$$95\%\text{C.I.} = \exp\left(\mathrm{Log}\left(\widehat{\mathrm{OR}}\right) \pm Z_{.025}\sqrt{V}\right)$$

$$\pi_i = \frac{a_i + d_i}{n_i}, \quad Q_i = \frac{b_i + c_i}{n_i}, \quad R_i = \frac{a_i d_i}{n_i}, \quad S_i = \frac{b_i c_i}{n_i}, \quad \widehat{OR} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$$

$$V = \frac{\sum_i \pi_i R_i}{2(\sum_i R_i)^2} + \frac{\sum_i Q_i S_i}{2(\sum_i S_i)^2} + \frac{\sum_i (\pi_i S_i + Q_i R_i)}{2(\sum_i R_i)(\sum_i S_i)}$$

The value of V turned out to be 0.000006185. The Odds Ratio was calculated as follows:

$$\widehat{OR} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$$

Aim 3: To Examine the Probable Causal Relationship between Revisit Intervals (RVI) and Chronic Kidney Disease (CKD) in Diabetic Medicare Patients

Drawing upon Pearl's 2009 '*do*-operation' which involved determining the causal effect of an external intervention, Pearl's *do*-operator provides a connection between causal effects and a series of randomized experiments which can be applied to various variables and truncated Markov Factorization[140] by the use of a backdoor adjustment.[141] This alternative method of conditioning covariates in the Back-Door path overcomes one of the most difficult part of

controlling for confounders in causal analysis. It provided a procedure that can be used directly within an EHR. In this predictive model, we looked to determine the risk of a future event (CKD), and so only predictors that occur before the outcome are used. Therefore, events that occurred after the Index Date were then ignored to avoid stratifying events on the causal path from RVI to CKD. Including these variables will distort the estimated impact of long RVI on CKD. Under assumption of independence, the Likelihood association of each individual diagnosis is calculated as the ratio of prevalence of the diagnosis among diabetic patients with CKD versus those without.

In determining the causal effect of Long RVI on CKD, conditioning occurs in O, C or both O and C in Figure 3. For the causal path for Long RVI (D) on CKD (Y). There is a non-causal relationship from C to O, C to D and O to Y which are Back-Door paths. In order to determine the causal effect of D on Y, we must first block the Backdoors. Given that Z s the set of variables conditioned on, then Pearl's Criterion can be used to ensure that all Back-Door paths that are between the causal variable and the outcome variable are blocked after conditioning on Z. This can be done if it satisfies one of the following: (1) It contains a chain that includes a mediated variable Z. (2) Contains an inverted fork of mutual causation where the middle attribute and all its descendants are not in Z. This condition is met at node Y. (3) Includes a fork of mutual dependence (D < C > O), where the middle variable, that the other two variables are dependent upon are in Z. This condition is met at C.

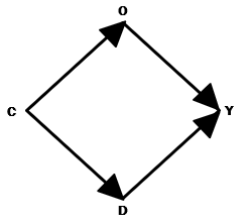


Figure 4: Back-Door Pathways

In the Figure 3, the Back-Door path is D < C > O > Y and it contains a mediated path C > O > Y, and a fork of mutual dependence D < C > O. Therefore, in order to condition on Z, we can choose to condition on C, O or both C&O. This will block all Back-Door paths between the causal variable and the outcome variable. The second condition is that no variable in Z are descendants of the causal variable or is directly positioned in the causal path.

The minimum viable Z must be identified and conditioned through stratification, but stratification is not realistic in large dimensional datasets. Therefore, LASSO executed in SQL can be used to identify the pMB. pMB are those covariates that are statistically significant and have a large impact on treatment that is being evaluated for its effect on the outcome.

Process Steps

1) Randomization: An appointment index date was randomly selected for each patient and the average RVI was then calculated as the number of days between the index date a subsequent encounter date for the same patient and the same provider.

2) Remove Tautological Predictors: All diagnoses that occurred after the index date were excluded from the analysis because they are tautological predictors of CKD.

3) Age Factor: A new variable labeled DxAge is created from each diagnosis code and its associated age category. This is to capture the occurrence of a diagnosis between various age groups.

4) Average RVI: The average RVI was calculated across the entire sample. Long RVI is defined as the RVI that is 1 standard deviation above the average. Short RVI was defined those that were below the average. Using standard deviations instead of values below and above the cut-off point increases the sensitivity of the model.

5) Calculate LR: The Likelihood Ratio of each individual DxAge against CKD and the LR of each individual DxAge against RVI is calculated.

Based on the assumption of independence, the Likelihood Ratios were calculated as the prevalence of the risk factor in the presence of a positive outcome over the prevalence of the same risk factor in the absence of the outcome.

= p(Dx|Outcome+) / p(Dx|Outcome-)

= (# of patients with Outcome+ & a specific Dx/ # of total patients) / ---------Equation 2

(# of patients with Outcome- & a specific Dx / # of total patients)

Conditional probabilities were used to reduce the number of comorbidities needed in the predictive analysis. This is represented by a Contingency Table that illustrates counts of joint observations in the cells (Table 4). Contingency tables provide an account of two or more variables that occur together, such as CKD and RVI. In Table 4, *a* and *c* represent counts of when Attribute 1 is present, while *b* and *d* represent the absence of Attribute 1. The presence of risk factors is represented by *a* and *b,* while the absence of risk factors (Attribute 2) are represented by *c* and *d.* The total number of times that both events co-occur is given in *a*; whereas the *d* provides an account of the number of times neither of the events co-occur. The total number of times Attribute 1 is present is calculated as a + b, and the total number of times that the risk factor is present (Attribute 2) is calculated as a + c. The universe of probabilities were provided as the total of a + b + c + d.

Table 4: Counts of Conditional Probabilities

		Attribute 1 (CKD or RVI)		
		Present	**Absent**	**Total**
Attribute 2 (Diagnosis)	**Present**	a	b	a+b
	Absent	c	d	c+d
	Total	a+c	b+d	a+b+c+d

Based on the assumption of independence, the Likelihood Ratio of each diagnosis is calculated as the ratio of prevalence of the diagnosis among patients with CKD to those without CKD (Equation 7).

This calculation involves determining the number of encounters with the diagnosis and with CKD (*a*), the counts of encounters with diagnosis amongst those without CKD (*b*); The total number of encounters with CKD (*a* + *c*); the total number of encounters without CKD (*b* + *d*)

Equation 7: Likelihood Ratio of CKD

LR = *p*(DxAge|Kidney|Disease) / *p*(DxAge|No Kidney Disease)

= (*a* / (*a* + *b*)) / (*b* / (*b* + *d*)) (Table 4)

6) Covariate Selection: A Likelihood Ratio cut-off point was determined in order to select only the DxAge variables with large Likelihood Ratio and that would yield a group of codes that were between approximately 30 and 40 codes, so that they were manageable. Values greater than the cut-off increase the probability of disease (+LR). DxAge codes that co-occur in both the list of high Likelihood Ratio for both CKD and RVI are then populated into a separate list of covariates. The inverse LR (-LR) were those Likelihood Ratios that are between 0 and 0.15 decrease the probability of disease (-LR). These are also captured in a separate table that we will name CovariateTable.

7) Stratification: The list of covariates were then stratified (Table 7). Stratification allows for the examination of the common impact across strata. Stratification is done by

dividing the data into subgroups called strata. Each stratum is a unique combination of covariates that can be identified by using the 'Group by' clause on the CovariateTable in SQL.

Table 7: Strata

ID	Date	Next_Date	Difference	Strata
10000000015	*2015-11-24*	*2015-11-25*	*1*	*E118, I2510, J449*
10000000053	*2016-04-04*	*2016-05-17*	*43*	*E1142, B351, M79674, M79675*
10000000099	*2012-04-03*		*0*	*K824, I4891, R 7989, E119*

5) Sensitivity Analysis: The OR(hat) and a 95% Confidence Interval (CI) was then calculated in PostgreSQL and used to determine the usefulness of the covariates based on their size and statistical significance. OR(hat) was calculated as $\widehat{OR} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$ based on the results of the contingency table. The CI for the association is calculated based on variance of the measure across strata. Therefore *Var* (*v*) as $V = \frac{\sum_i \pi_i R_i}{2(\sum_i R_i)^2} + \frac{\sum_i Q_i S_i}{2(\sum_i S_i)^2} + \frac{\sum_i (\pi_i S_i + Q_i R_i)}{2(\sum_i R_i)(\sum_i S_i)}$ is first calculated, where $\pi_i = \frac{a_i + d_i}{n_i}$, $Q_i = \frac{b_i + c_i}{n_i}$, $R_i = \frac{a_i d_i}{n_i}$, $S_i = \frac{b_i c_i}{n_i}$. Then a 95% CI is calculated in PostgreSQL as $95\%\ C.I. = exp\left(Log\left(\widehat{OR}\right) \pm Z_{.025}\sqrt{V}\right)$. Each individual stratum is then removed from the sample and the OR(hat) is recalculated. Those with low sensitivity are then dropped, keeping only those with high sensitivity.

Chapter 3 Summary

This chapter provided an account of the methodology used in this study in order to accomplish each of the three study aims. It discussed the data source and provided a description of the data as well as the inclusion and exclusion criteria. It then went into the study measures and methods. The chapter also provided the process steps. Details Standard Query Language (SQL) codes are provided in the appendix.

CHAPTER 4: RESULTS

The following are the results of the analysis as it pertains to the 3 aims. The focus on Aim 1 was to showcase a new method of conducting Least Absolute Shrinkage and Selection Operator (LASSO) Regression to accomplish feature selection by the use of Standard Query Language (SQL), while in Aim 2 was to predict Chronic Kidney Disease (CKD) and the Optimal Revisit Interval (RVI) for Medicare Patients with Type II Diabetes. Finally, in aim 3 was to examine the causal relationships between CKD and RVI for Medicare patients with Type II Diabetes was determined.

Aim 1: To Demonstrate the Effective use of Conducting Feature Selection by using the Least Absolute Shrinkage and Selection Operator (LASSO) executed in Standard Query Language (SQL)

In the sample, the odds for developing CKD for Medicare diabetic patients with long RVI were 1.056. If a patient had a probability of developing CKD that was above 0.5, then their optimal RVI was shorter, compared to those patients with probabilities that were below 0.5. Specific data on patients who presented with a diagnosis code of orthopedic injuries such as those of the hips, diagnosis code S79819A, had a probability of 0.077 for CKD and an optimal RVI of 305 – 306 days.

In determining the relevant features and the causal relationship between features and CKD, we used Likelihood Ratios. Based on the distribution of the Likelihood Ratios

we see that the median was around the value of 40 for CKD. This was used as the threshold for the inclusion criteria (in addition to outpatient encounters, diabetics and in primary care settings). Overall, we identified 23 diagnoses that had a CKD Likelihood Ratio that was greater than 40, and 10 diagnoses that had a RVI Likelihood Ratio that was greater than 5. Dropping those individual diagnoses from the dataset did not change the common Odds Ratio.

The diagnosis with the highest Likelihood Ratios was T56894A - toxic effect of other metals, undetermined, initial encounter- with a Likelihood Ratio of 208.90, followed by S79819A- other specified injuries of unspecified hip, initial encounter-- with a Likelihood Ratio of 165.92. The diagnosis with the highest count of occurrence was D631- anemia in CKD, with a count of 26,708, followed by 7105- Other specified systemic involvement of connective tissue-- with a count of 295, and S79819A - Other specified injuries of unspecified hip.

Diagnosis codes with a count that was in the top 5 and also a Likelihood Ratio that was in the top 5 were:

➢ T56894A- Toxic effect of other metals, undetermined, initial encounter, Count – 120, LR – 208.90

➢ S79819A- Other specified injuries of unspecified hip, initial encounter, Count – 167, LR – 165.92

➢ 7105 - Other specified systemic involvement of connective tissue, Count – 295, LR 43.26

Three were 6 diagnosis codes related to infection (T83510D, T80211D, B9621, T83510D, 37231, 0529); 5 related to mechanical failure of devices (S79819A, T8241XA, T83028A, T85611A, T82318A); 4 that were cancer-related (17322, 17311, D2362, 1740); 3 that were associated with fractures (S79819A, S72351D, S14104D); 2 related to toxic metals (T56894A, T5694XA); and other.

Aim 2: To Predict Chronic Kidney Disease (CKD) and the Optimal Revisit Interval (RVI) for Medicare Patients with Type II Diabetes

In determining the optimal RVI for Medicare patients with Type II Diabetes, we first calculated the average RVI. The average RVI for diabetic Medicare patients was 33.88 days. From the average, the long versus short RVI for each comorbidity was then calculated. Table 8 displays the counts of covariates captured in encounters with short RVI and whether or not they have CKD. It also captures the same for long RVI with and without CKD. In the dataset, there were 123,613 cases of CKD with a RVI that was greater than 33.88 days; 14,300,915 cases did not have CKD, but had a RVI greater than 33.88. There were 580,524 cases of patients with CKD and a RVI of less or equal to 33.88 days, and 37,103,508 cases did not have CKD but had a RVI of less or equal to 33.88.

Table 8: Table of Results

Patient Characteristics		With Kidney Disease 704,137 claims $y = 1$	Without Kidney Disease 66,409,475 claims $y = 0$
Same *'n'* Covariates for Cases and Controls	Short RVI (Cases) Greater or equal to 33.88 52,689,084 cases ($x = 1$)	*a* *580,524 claims*	*b* *52,108,560 claims*
	Long RVI (Controls) Less than 33.88 14,424,528 Cases ($x = 0$)	*c* *123,613 Claims*	*d* *14,300,915 Claims*

Table 9 provides the results of the optimal RVI for Medicare patients. The first column in Table 9 lists of the covariates per stratum. The column that follows provides the probability of CKD occurring given that a patient has the specific comorbidity. The following columns include the maximum and minimum values in days for the predicted optimal RVI for each individual stratum. In clinics and in hospitals, the same diagnosis indicates different levels of severity of illness, treatment, and outcomes. For this reason, different strata were developed for each comorbidity or combination of comorbidities.

Table 9 includes data on optimal RVI based on various comorbidities. If a patient had a probability of developing CKD that is above 0.5, then their optimal RVI was shorter, compared to those patients with probabilities that are below 0.5. For example, if a diabetic patient presents with a comorbidity ICD10 code of 'Other specified injuries of unspecified hip, initial encounter' (S79819A), then their optimal RVI is between 305 and 306 for that specific strata. Another example is if a patient presents with 'Malignant

neoplasm of nipple and areola of female breast' (174.0) then their optimal RVI is between 213 and 349.

Table 9: Optimal Revisit Interval (RVI) Predictions

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.793	52	123
Dx1740	0.730	213	349
Dx74332	0.933	130	329
T83510d	0.366	268	306
Dx17311	0.434	147	200
Z578	0.519	89	222
Dx74332,Dx17311	0.074	313	360
T83510d,Dx17311	0.997	21	101
S79819A,Dx17311	0.865	199	342
Dx1740,Dx17311	0.365	154	216
Dx17311 ,Dx17311	0.691	335	336
Z578,Dx17311	0.383	338	350
Z578,Dx17311 , Dx37232	0.343	209	342
S79819A,Dx17311 , Dx37232	0.708	27	285
Dx1740,Dx17311 , Dx37232	0.238	105	263
Dx74332,Dx17311 , Dx37232	0.110	171	278
T83510d,Dx17311 , Dx37232	0.797	65	123
Dx17311 ,Dx17311 , Dx37232	0.444	34	73
Dx74332,Dx17311, Dx7433	0.009	179	253
T83510d,Dx17311, Dx7433	0.875	165	201
Dx17311 ,Dx17311, Dx7433	0.184	97	326
Dx1740,Dx17311, S14104d	0.008	55	352
Dx74332,Dx17311, S14104d	0.629	313	316
Z578,Dx17311, S14104d	0.260	213	255
S79819A, T83510D,Dx17311, S14104d	0.143	320	364
S79819A,Dx17311, S14104d	0.021	325	347
T83510d,Dx17311, S14104d	0.589	348	352
Dx17311 ,Dx17311, S14104d	0.868	194	275
T83510d,Dx17311, Z578	0.917	315	326
S79819A,Dx17311, Z578	0.763	219	226

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740,Dx17311, Z578	0.226	216	292
Dx74332,Dx17311, Z578	0.693	344	364
Dx17311 ,Dx17311, Z578	0.488	21	50
Z578,Dx17311, Z578	0.930	297	314
S79819A, T83510D,Dx17311, Z578	0.369	284	293
Dx17311, Z578 ,Dx17311, Z578	0.972	149	170
Dx36842 ,Dx17311, Z578	0.805	263	290
T80211d,Dx17311, Z578	0.089	152	323
s14104d ,Dx17311, Z578	0.354	248	270
T50992a,Dx17311, Z578	0.438	8	188
T381x5a,Dx17311, Z578	0.246	123	292
Dx74332 ,Dx17311, Z578	0.039	67	298
Dx7105,Dx17311, Z578	0.187	251	332
Z578,Dx1741	0.436	40	163
S79819A,Dx1741	0.927	28	126
Dx1740,Dx1741	0.683	314	342
Dx7105,Dx1741	0.924	54	228
T83510d,Dx1741	0.304	56	114
Dx17311, Z578 ,Dx1741	0.226	82	189
Dx36842 ,Dx1741	0.828	9	23
T381x5a,Dx1741	0.550	187	198
s14104d ,Dx1741	0.426	349	364
Dx17311 ,Dx1741	0.871	96	167
Dx74332,Dx1741	0.675	183	280
T50992a,Dx1741	0.944	59	67
S79819A, T83510D,Dx1741	0.940	283	324
T80211d,Dx1741	0.158	45	360
s31125a,Dx1741	0.253	56	144
Dx17311, S14104d ,Dx36842	0.484	37	174
Dx1740,Dx36842	0.157	300	327
Dx7105,Dx36842	0.356	326	338
T83510d,Dx36842	0.169	334	345
S79819A, T83510D,Dx36842	0.544	60	282
Dx17311, Z578 ,Dx36842	0.405	73	335
Dx36842 ,Dx36842	0.635	284	338
T80211d,Dx36842	0.732	247	324
Z578,Dx36842	0.568	321	326
s31125a,Dx36842	0.164	47	230

Covariates	Probability	Minimum RVI	Maximum RVI
s14104d ,Dx36842	0.323	34	236
Dx17311 ,Dx36842	0.435	348	357
Dx74332,Dx36842	0.378	237	343
S79819A,Dx36842	0.045	202	268
T381x5a,Dx36842	0.969	219	243
m10361,Dx36842	0.160	261	351
S79819A,Dx36848	0.889	168	222
Dx17311, S14104d ,Dx36848	0.152	161	189
Dx529,Dx36848	0.183	307	318
Dx7105 ,Dx36848	0.708	46	274
Dx7105,Dx36848	0.399	363	364
T83510d,Dx36848	0.115	291	297
Z578,Dx36848	0.266	20	326
Dx17311, Z578 ,Dx36848	0.565	133	152
Dx1740,Dx36848	0.353	97	145
s14104d ,Dx36848	0.786	262	359
Dx74332,Dx36848	0.412	210	269
m10361 ,Dx36848	0.950	84	169
m10361,Dx37231	0.643	81	165
Z578,Dx37231	0.253	221	327
T83510D,Dx37231	0.584	364	365
Dx1740,Dx37231	0.357	13	241
Dx74332,Dx37231	0.384	163	224
S79819A,Dx37231	0.488	117	191
Dx17311, S14104d ,Dx37231	0.003	345	358
Dx529,Dx37231	0.410	285	333
T83510d,Dx530	0.912	153	332
Z578,Dx530	0.843	260	265
Dx17311, Z578 ,Dx530	0.135	365	#NUM!
Dx1740,Dx530	0.383	8	131
Dx74332,Dx530	0.977	247	349
Z578,Dx7106	0.621	121	159
Dx17311 ,Dx7106	0.916	207	214
Dx1740,Dx7106	0.306	141	280
Dx74332,Dx7106	0.079	103	136
T83510D,Dx7106	0.273	273	308
S79819A,Dx7106	0.300	303	341
Dx17311, S14104d ,Dx7106	0.229	144	210

Covariates	Probability	Minimum RVI	Maximum RVI
Dx529,Dx7106	0.620	243	327
Dx17311, Z578 ,Dx7106	0.118	111	334
Dx36842 ,Dx7106	0.744	107	163
T381x5a,Dx7106	0.285	106	224
s14104d ,Dx7106	0.190	156	269
T83510d,Dx74333	0.462	70	200
Z578,Dx74333	0.158	117	206
Dx17311, Z578 ,Dx74333	0.517	11	309
Dx1740,Dx74333	0.899	347	360
Dx74332,Dx74333	0.318	61	358
Dx17311 ,Dx74333	0.954	250	275
S79819A,Dx74333	0.867	216	323
Dx17311, S14104d ,Dx74333	0.714	1	213
Dx37231 ,Dx74333	0.618	196	216
Dx529,Dx74333	0.097	137	239
Dx36842 ,Dx74333	0.785	98	147
T381x5a,Dx74333	0.526	85	351
s14104d ,Dx74333	0.547	115	161
S79819A, T83510D,Dx74333	0.977	152	208
T80211d,Dx74333	0.353	207	333
Dx36847,Dx74333	0.611	213	317
Dx1740,m10362	0.905	218	280
Dx74332,m10362	0.983	317	340
T83510d,m10362	0.746	132	202
S79819A,s14104d	0.597	64	280
Dx17311, S14104d ,s14104d	0.300	153	285
Dx36847,s14104d	0.490	295	312
T83510d,s14104d	0.684	197	287
Z578,s14104d	0.676	333	347
Dx17311, Z578 ,s14104d	0.673	201	331
Dx36842 ,s14104d	0.217	94	159
T381x5a,s14104d	0.199	203	336
Dx1740,s14104d	0.935	51	305
s14104d ,s14104d	0.572	227	268
Dx74332,s14104d	0.941	116	363
S79819A, T83510D,s14104d	0.949	112	167
T80211d,s14104d	0.859	146	348
Dx17311 ,s14104d	0.077	250	362

Covariates	Probability	Minimum RVI	Maximum RVI
T83510D,s31125a	0.528	165	315
Dx1740,s31125a	0.609	341	342
Dx74332,s31125a	0.636	118	222
S79819A,s31125a	0.082	53	173
Dx17311 ,s31125a	0.505	335	345
Dx36847,s31125a	0.774	210	263
Z578,s31125a	0.263	94	352
S79819A, T83510D,S79819A	0.908	294	317
Dx17311, Z578 ,S79819A	0.721	96	256
Dx36842 ,S79819A	0.424	68	138
T80211d,S79819A	0.379	258	286
Z578,S79819A	0.492	342	364
Dx1740,S79819A	0.417	165	234
T381x5a,S79819A	0.214	67	235
s14104d ,S79819A	0.581	95	136
Dx17311 ,S79819A	0.921	318	364
Dx74332,S79819A	0.275	183	185
T83510D,S79819A	0.165	17	243
S79819A,S79819A	0.039	137	336
Dx36847,S79819A	0.074	218	252
Dx17311, Dx7433 ,S79819A	0.887	276	289
T83510d,S79819A, T83510D	0.361	287	294
Z578,S79819A, T83510D	0.834	23	191
S79819A, T83510D,S79819A, T83510D	0.201	296	354
Dx17311, Z578 ,S79819A, T83510D	0.222	336	352
Dx36842 ,S79819A, T83510D	0.080	115	196
T80211d,S79819A, T83510D	0.646	225	364
Dx1740,S79819A, T83510D	0.668	192	296
T381x5a,S79819A, T83510D	0.272	94	197
s14104d ,S79819A, T83510D	0.589	137	161
Dx17311 ,S79819A, T83510D	0.110	136	200
Dx74332,S79819A, T83510D	0.515	165	293
S79819A,S79819A, T83510D	0.345	274	334
Dx17311, Dx7433 ,S79819A, T83510D	0.355	214	250
Dx17311 , Dx37231,S79819A, T83510D	0.013	4	219
Dx74332,T381x5a	0.671	360	362

Covariates	Probability	Minimum RVI	Maximum RVI
Z578,T381x5a	0.097	178	196
Dx17311 ,T381x5a	0.875	113	159
Dx1740,T381x5a	0.989	344	365
T83510D,T381x5a	0.964	277	357
S79819A, T83510D,T381x5a	0.974	11	97
Dx17311 , Dx37231,T381x5a	0.662	54	331
Dx36842 ,T381x5a	0.657	332	352
Dx17311, Z578 ,T381x5a	0.249	20	100
T80211d,T381x5a	0.884	132	196
s14104d ,T381x5a	0.485	302	339
S79819A,T381x5a	0.370	129	224
S79819A, T83510D,T50992a	0.826	320	352
T83510D,T50992a	0.847	58	265
Dx1740,T50992a	0.608	167	219
Dx74332,T50992a	0.069	204	229
S79819A,T50992a	0.531	158	185
Dx17311 , Dx37231,T50992a	0.585	173	256
Dx36842 ,T50992a	0.013	38	319
T83510d,T5694xa	0.731	197	249
Z578,T5694xa	0.004	359	365
S79819A, T83510D,T5694xa	0.422	43	113
Dx17311, Z578 ,T5694xa	0.941	288	298
Dx36842 ,T5694xa	0.984	96	332
T80211d,T5694xa	0.680	224	337
Dx1740,T5694xa	0.038	328	348
s14104d ,T5694xa	0.948	77	135
Dx17311 ,T5694xa	0.588	132	218
Dx74332,T5694xa	0.412	266	311
Dx17311 , Dx37231,T5694xa	0.922	326	349
s14104d ,T80211d	0.377	181	361
Z578,T80211d	0.673	195	203
Dx17311 ,T80211d	0.031	185	210
Dx1740,T80211d	0.508	294	330
Dx74332,T80211d	0.591	335	342
T83510D,T80211d	0.605	328	340
S79819A, T83510D,T80211d	0.896	189	315
Dx17311 , Dx37231,T80211d	0.082	312	359
Dx36842 ,T80211d	0.734	44	327

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578 ,T80211d	0.114	100	139
T80211d,T80211d	0.310	297	315
Dx36842 ,T83510d	0.394	101	189
T83510d,T83510d	0.669	315	342
Z578,T83510d	0.045	140	356
Dx17311 ,T83510d	0.859	148	208
Dx1740,T83510d	0.465	56	306
Dx74332,T83510d	0.486	39	44
S79819A,T83510d	0.963	64	356
S79819A, T83510D,T83510d	0.884	68	105
Dx17311 , Dx37231,T83510d	0.418	175	197
Dx17311, Z578 ,T83510d	0.547	192	282
T80211d,T83510d	0.262	12	192
T5694xa,T83510D	0.909	248	353
S79819A, T83510D,Z579	0.920	92	355
Dx17311 , Dx37231,Z579	0.180	236	321
Dx36842 ,Z579	0.756	217	334
T83510d,Z579	0.550	252	289
Z578,Z579	0.823	286	295
Dx17311, Z578 ,Z579	0.909	244	359
T80211d,Z579	0.653	221	350
Dx1740,Z579	0.263	328	330
Dx74332,Z579	0.270	291	321
T5694xa,Z579	0.617	194	264
Dx17311 ,Z579	0.056	188	194
Dx17311 , Dx37231 ,Z579	0.400	345	347
S79819A,Z579	0.291	133	188
Z579	0.670	90	176

For each comorbidity and its repetition, referred to as predictor, a Likelihood Ratio was calculated using the prevalence of each covariate among patients with and without CKD. A ratio of 3 for example, indicated that the predictor triples the risk of CKD. A ratio of 0.5 indicated that the odds of CKD are reduced by half. This is showed

in the individual stratum in Table 9. The full list covariates per strata, its probability and minimum and maximum values can be found in the appendix.

Aim 3: To Examine the Probable Causal Relationship between Revisit Intervals (RVI) and Chronic Kidney Disease (CKD) in Diabetic Medicare Patients

The diagnosis codes that had the highest Likelihood Ratios on for both RVI and CKD were then analyzed for their causal effects. These are referred to as the covariates. The Cochran–Mantel–Haenszel (CMH) was used to measure the strength of the causal associations between covariates and CKD, after stratification (Figure 4).[142]

The results show(Figure 4) that orthopedic conditions and fractures, specifically hip and femur fractures had the highest CMH index, followed by spinal cord injuries, gout, neoplasms such as melanomas, and blood toxicity such as poisoning due to long-term use of medications. These covariates showed to have the highest causal relationship to CKD.

Covariates with the highest Likelihood Ratios were individually dropped in order to determine their causal impact on the outcome. Those that had resulted into an increase in homogeneity in the dataset, but without a significant change in the common odds ratio remained in the final group of covariates as having a causal impact on CKD. A

homogeneous association implies that the conditional association between any two features given the third one is the same at each stratum.

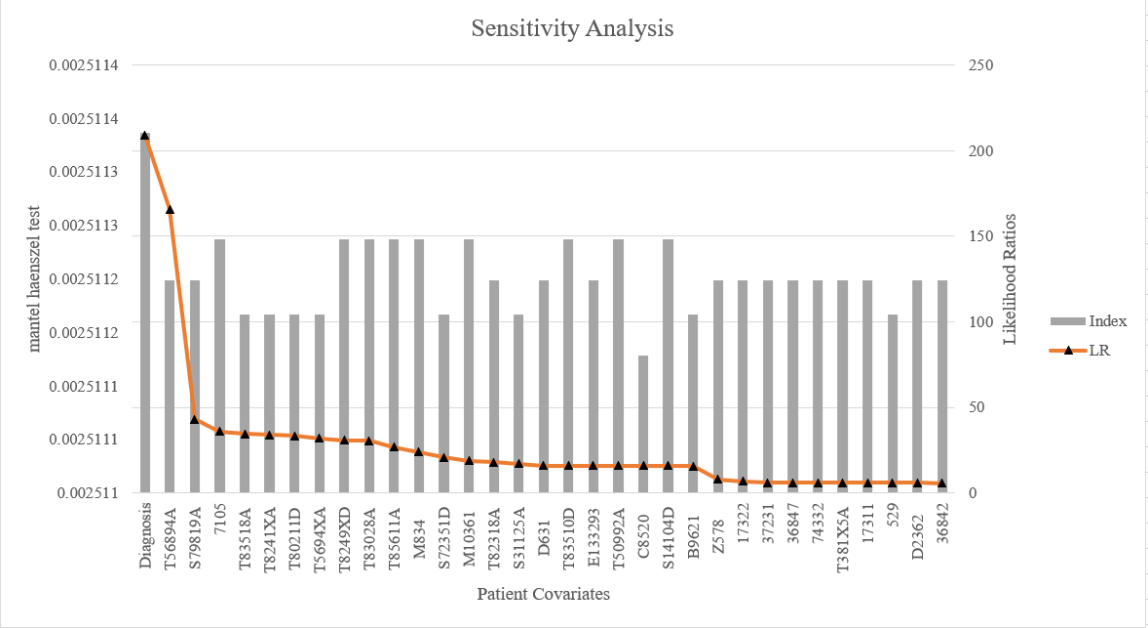


Figure 5: Cochrane Mantel-Haenszel Analysis

For the entire set of covariates, the CMH index was 0.002511338 (Table 10). Once the diagnosis code for toxic effects of metals (T56894A) was removed, the CMH index changed from 0.002511338 to 0.002511338 with little to no change in the common odds ratio. The results associated with dropping the toxic metals were similar to that of the diagnosis code for hip injuries (S79819A). This was repeated for each covariate and the results of the individual codes are presented in Table 10. In Table 10, the first column represents the covariate (diagnosis codes that have a high Likelihood Ratio for both RVI and CKD), the second column shows the description of the code, the third column

displays the Likelihood Ratio, and the fourth column shows the CMH Index. The columns that follow include the changes in the CMH Index as a result of dropping covariates. These results show that covariates which had the highest Likelihood ratios were also the ones that had the highest increase in homogeneity.

Table 10: The Impact of Individual Covariates

Diagnosis Code	Description	Likelihood Ratio	Index	Variation	Weighted Variation
T56894A	Toxic effect of other metals, undetermined, initial encounter	208.9010417	0.002511337	-1.07E-07	-107
S79819A	Other specified injuries of unspecified hip, initial encounter	165.921875	0.002511199	3.1E-08	31
7105	Eosinophilia myalgia syndrome	43.25732899	0.002511199	3.1E-08	31
T83518A	Infection and inflammatory reaction due to other urinary catheter, initial encounter	35.984375	0.002511237	-7E-09	-7
T80211D	Bloodstream infection due to central venous catheter, subsequent encounter.	33.984375	0.002511167	6.3E-08	63
T5694XA	Toxic effect of unspecified metal, undetermined, initial encounter (Toxic effect of unspecified metal, undetermined, initial encounter)	33.51259774	0.002511167	6.3E-08	63
M834	Aluminum bone disease	26.984375	0.002511237	-7E-09	-7
S72351D	Displaced comminuted fracture of shaft of right femur, subsequent encounter for closed fracture with routine healing	23.98784722	0.002511237	-7E-09	-7
M10361	Gout due to renal impairment, right knee	20.98958333	0.002511167	6.3E-08	63
T82318A	Breakdown (mechanical) of other vascular grafts	18.98958333	0.002511237	-7E-09	-7

Diagnosis Code	Description	Likelihood Ratio	Index	Variation	Weighted Variation
S31125A	Laceration of abdominal wall with foreign body, periumbilical region without penetration into peritoneal cavity, initial encounter	17.98958333	0.002511199	3.1E-08	31
E133293	Other specified diabetes mellitus with mild non-proliferative diabetic retinopathy without macular edema, bilateral	15.99479167	0.002511237	-7E-09	-7
C8520	mediastinal (thymic) large B-cell lymphoma, unspecified site	15.99479167	0.002511237	-7E-09	-7
B9621	Shiga toxin-producing Escherichia coli [E. coli] (STEC) O157 as the cause of diseases classified elsewhere	15.99479167	0.002511237	-7E-09	-7
T50992A	Poisoning by other drug/meds/biol substance, self-harm, initial encounter	15.99479167	0.002511199	3.1E-08	31
S14104D	Unspecified injury at C4 level of cervical spinal cord, subsequent encounter	15.99479167	0.002511129	1.01E-07	101
Z578	Occupational exposure to other risk factors	15.70036477	0.002511167	6.3E-08	63
17322	Squamous cell carcinoma of skin of ear and external auditory canal	7.994791667	0.002511199	3.1E-08	31
37231	Endovascular Revascularization (Open or Percutaneous, Transcatheter)	6.994791667	0.002511199	3.1E-08	31
36847	Procedures Heteronymous bilateral field defects.	5.997395833	0.002511199	3.1E-08	31
74332	Heteronymous hemianopsia	5.994791667	0.002511199	3.1E-08	31
T381X5A	Capsular cataract	5.994791667	0.002511199	3.1E-08	31
17311	Adverse effect of thyroid hormones and substitutes, initial encounter	5.994791667	0.002511199	3.1E-08	31
529	Surgical Pathology	5.994791667	0.002511199	3.1E-08	31
36842	Disorders of visual pathways in (due to) neoplasm, unspecified side	5.994791667	0.002511199	3.1E-08	31
	Scotoma of blind spot area	5.994791667	0.002511199	3.1E-08	31

Diagnosis Code	Description	Likelihood Ratio	Index	Variation	Weighted Variation
D2362	Other benign neoplasm of skin of left upper limb, including shoulder	5.994791667	0.002511167	6.3E-08	63
1740	Malignant neo nipple (Malignant neoplasm of nipple and areola of female breast)	5.663194444	0.002511199	3.1E-08	31

The features with the highest impact were then analyzed separately through the same method of dropping individual variables from the dataset. The ones that had the largest change in homogeneity (Figure 5) were toxic effect of metals (T56894A), followed by Aluminum bone disease (M834), femur fracture (S72351D), mechanical breakdown of vascular grafts (T82318A), large b-cell lymphoma (C852), and Escherichia Coli (B9621).

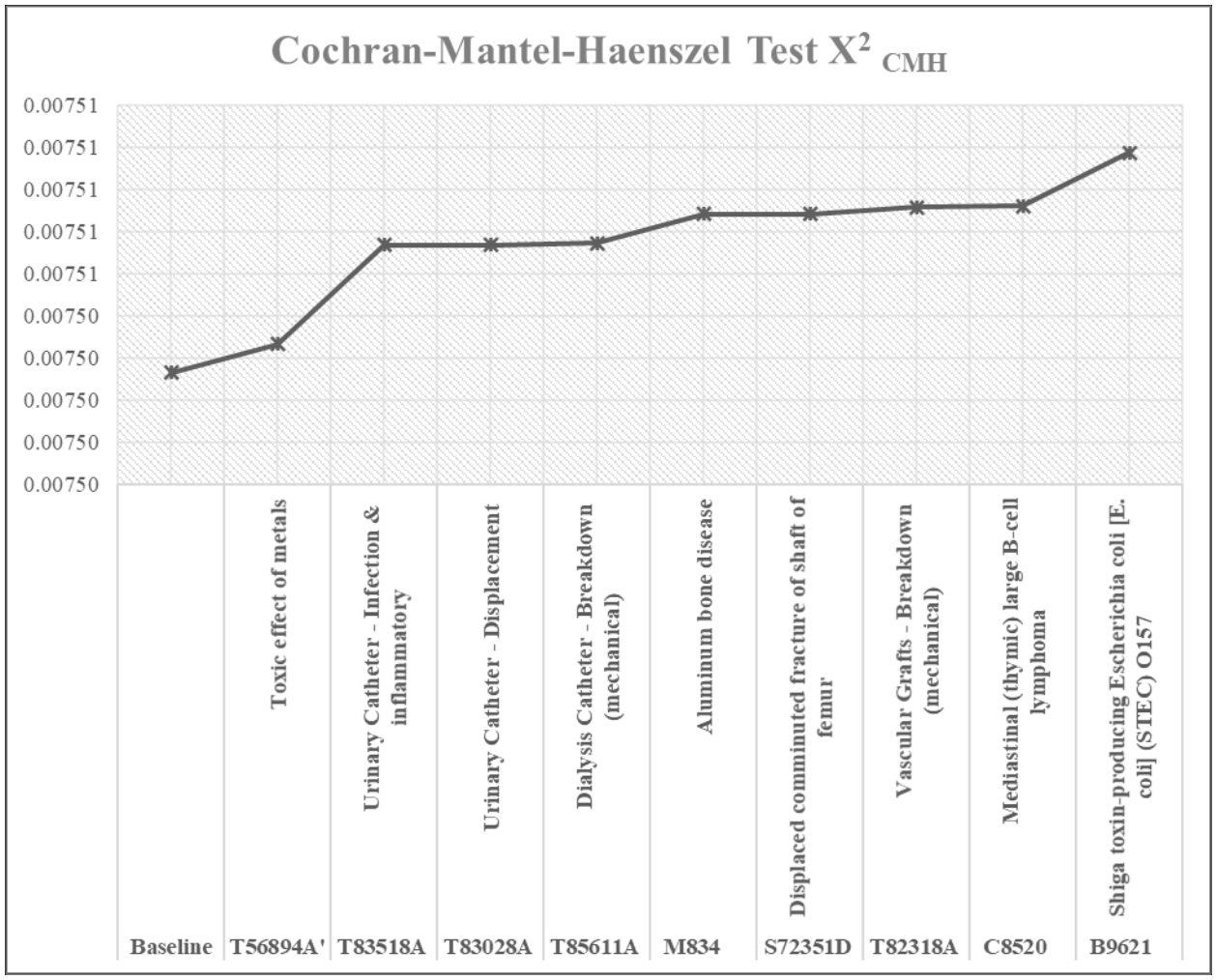


Figure 6: Narrowed CMH Covariates

CHAPTER 5: DISCUSSION

Chapter 5 discusses the main findings and it compares the general findings with those from previous studies. It reviews the main applications of the study results and presents suggestions for future research. The discussion is organized by the three aims of the study.

Overview

Overall, this study has shown that outpatient diagnoses can provide probable predictions of adverse effects of Chronic Kidney Disease (CKD) in patients with diabetes. We relied on a comprehensive set of diagnoses, including approximately 18,000 distinct diagnoses. We predicted future CKD in more than 800,000 distinct patients over a period of 3 years. The data was randomly divided into before and after a notional date in which the predictions were being made. The notional date served as the date in which predictions were made. The encounters before the date were used as the learning data that made predictions on encounters that occured after the date. The data was then narrowed down to only those patients with diabetes and who were over the age of 65 years old. All occurrences of CKD prior to the date were removed from the dataset. Long and short Revisit Intervals (RVI) in the before data was determined. Long RVI were defined as those RVI which were above the average RVI. While short RVI were defined as RVI which were below the average. The probability of occurrence of CKD given a long RVI

can assist providers in determining the optimal time by calculating the maximum and minimum values in the given observations.

The methodology summarized above can be executed in SQL directly in an Electronic Health Record (EHR)-based screening in clinical settings. This methodology is relatively easy to implement since it uses Standard Query Language (SQL) queries to execute. It does not rely on effort of either the patient or the clinician and therefore can be widely and consistently implemented across clinics. Our proposed EHR-based screening was found to be more comprehensive; it included all diagnosis codes, as opposed to selecting diagnoses that are considered by the researcher to be of most relevance based on prior knowledge and experience. As more data is entered, the methodology can continue to 'learn' and narrow the RVI. Overall, it provides a starting point and a tool to assist providers in making decisions around the appropriate RVI for patients with CKD . This methodology has the potential to benefit patients in terms of access and positive health outcomes. Additional research should be conducted with other chronic conditions, and this method should be tested in a clinical setting.

Aim 1: To Demonstrate the Effective use of Conducting Feature Selection by using the Least Absolute Shrinkage and Selection Operator (LASSO) executed in Standard Query Language (SQL)

Other methods of feature selection and predictive analytics such as Random Forest and Decision Trees could have been used; however, the Least Absolute Shrinkage and Selection Operator (LASSO) was chosen because the data from the Centers for

Medicare and Medicaid Services (CMS), Limited Database (LDS) is a sparse massive dataset. With LASSO, a researcher can include all the features in a high-dimensional dataset and then analyze their importance through Bayesian methodologies by determining Likelihood Ratios.

By executing LASSO through the use of Standard Query Language (SQL), the need of accessing statistical tools was overcame and packages since this method was implemented directly inside an Electronic Health Record (EHR). Thus, it can be part of automated methods of analyzing data and making predictions within an EHR as a decision support tool.

This method has shown to be simple and reliable since it runs on widely used and standard programing language, and is therefore implementable through an array of software such as SAS ®, PostgreSQL, Microsoft SQL Server etc. We found this method to be beneficial in predictive analytics and causal research that involves high-dimensional datasets which have very large numbers of features and classes.

As a result, we see that feature selection in high-dimensional dataset where there are thousands of variables for example, does not have to be based on expert prior knowledge of the features alone. Researchers can use this aforementioned methodology to determine feature importance by analyzing the Likelihood Ratios of all features.

Aim 2: To Predict Chronic Kidney Disease (CKD) and the Optimal Revisit Interval (RVI) for Medicare Patients with Type II Diabetes

The method of prediction used in this study is similar to the time-varying hazard model[143,144] because as patients have new encounters, predictions occur over time as the encounters and the predictions change. The American Diabetes Association provides medical practice guidelines for hemoglobin A1C testing as 6 months.[145] Providers may use this RVI for diabetic patients though they may need to be seen sooner if they have various comorbidities. However, these guidelines do not provide insights on optimizing RVI for diabetic patients. This study found that the average RVI across the data was 33.88 days with a wide variation, as compared to the 6-month standard from the American Diabetes Association. The over-use of RVI may be wasteful for it increases resource utilization without any additional health benefit to the patient. Providers agree that RVI for patients with severe chronic conditions such as those who are hypertensive or those with acute diabetes should be shorter[146] and they agree on the recommended RVI for Hemoglobin A1c testing.[147] Furthermore, the findings of this study showed that RVI for common chronic diseases such as hypertension vary greatly which is in agreement with previous studies.[148, 149, 150, 151, 152, 153, 154] For example, family medicine providers recommend short RVI as compared to internal medicine providers.[155] Studies also show that female providers also recommend shorter RVI than their male counterparts.[156, 157, 158]

Previous studies that focused on RVI for other chronic diseases found that patients who were hypertensive should be seen 1-2 months initially then 3-6 months once their hypertension was controlled.[159] The present study provides recommendations for the optimal RVI for CKD; however, it does not provide customized recommendations based on the patient's individual set of comorbidities.

The results can be categorized into 5 main categories:

Orthopedic Conditions: Patients with CKD are known to be more susceptible to fractures.[160] These results are consistent with other clinical studies that shows the relationship involving the biophysical pathways between orthopedic conditions and renal failure. Clinical studies show that fractures with patients with End-Stage Renal Disease (ESRD) at a significant risk of mortality and the prevalence is higher within the aging population.[161,162] Patients suffering from complications related to operative procedures were also found to have a relatively higher incidence of renal dysfunction.[163,164]

Hip injuries: In the present study, patients with hip fractures had a probability of 0.585 for CKD, and a suggested RVI of 71 – 102 days. Other studies involving RVI did not propose optimal RVI for diabetic patients with hip fractures; however, there were several studies that described the clinical manifestation of CKD from orthopedic fractures. Hip fractures are a common injury for patients over 65 years old.[165] Renal disease is an important adverse event in patients with hip fractures.[166] According to Bennet et al., 1 in 3 men with fractured hips developed renal disease.[167,168] Moreover, the increased prevalence of osteoporosis at the hip is expected to lead to a tripling of the number of hip fractures worldwide by 2050.[169]

Spinal Cord Injuries: The findings of this study indicate that patients with spinal cord injuries and a probability of 0.786 for CKD, had a suggested RVI of 46-193 days. Other studies that looked at predicting RVI focused mainly on chronic conditions and not specifically on spinal cord injuries. But there were a few that illustrated the mechanisms in which patients with spinal cord conditions can develop CKD as an indirect result of it.

CKD is a high predictor of mortality amongst patients with Spinal Cord Injuries and Disorders (SCI/D).[170] This is because patients with SCI/D tend to be even more vulnerable, costly and complex than the rest of the patient population.[171] Patients with SCI/D are more likely to develop CKD due to being at a higher risk for bladder dysfunctions, nephrolithiasis, and other chronic infections.[172, 173,174, 175,176]

Gout: In this study, patients with Gout and a probability of 0.393 for CKD, and a suggested RVI of 95 – 121 days. As stated in the literature review in Chapter 2, there was no specific studies that provided evidence on the optimal RVI for diabetic patients with gout. However, studies do show that gout is associated with considerable co-morbidity including hypertension and diabetes mellitus.[177] Studies show that gout is associated with chronic kidney disease.[178] Patients with gout should be actively screened for CKD and its consequences.[179] Gout together with hypertension, is one of the major medical manifestations of lead nephropathy.[180,181] Revisits targeted towards testing lead urinary excretion after Ethylenediamine Tetra-acetic Acid (EDTA)-lead mobilization testing may help differentiate the diagnosis.[182]

Neoplasms: Findings show that patients with a melanoma and a probability of 0.18 for CKD had a suggested RVI of 126-127 days. Literature review (Chapter 2) showed a study that determined the optimal RVI for patients with melanomas that were less than 0.5mm in diameter. The results of this study concluded that there was no optimal time period recommended that may be safe and effective. [183] CKD and cancer are interconnected in both directions.[184] Cancer can lead to CKD indirectly or directly through the adverse effects of cancer therapy.[185] Cancer can lead to CKD through various

channels such a Paraneoplastic nephropathy, chemotherapy radiation, and other toxins.[186] There are a variety of carcinomas that are more commonly seen in patients with CKD than the general population. Some cutaneous diseases are clearly unique to this population. It is important for patients and physicians to recognize the manifestations of skin disease in renal disease to minimize and even prevent much of the morbidity associated with these conditions. Optimizing the RVI for this patient population can help assist with that. These conditions include benign neoplasm of skin of left upper limb or shoulder; Basal and squamous cell carcinoma of skin.

Toxicity and Poisoning: In this study, patients with toxicity and a probability of 0.47 for CKD, had a suggested RVI of 87 – 89 days. Previous studies on optimizing RVI did not focus on the optimal intervals for blood toxicity and poisoning, but there were a few studies that discussed the clinical manifestation of CKD as a result of toxicity and poisoning. There was a report on nephrotoxicity that was attributable to metals such as lead was published in 1863 by Lancereaux.[187] The study noted substantial atrophy of the renal cortex and tubular fibrosis in the kidney in subjects that were exposed to metal toxicity. In the late 1920s, an epidemic of chronic nephritis in Queensland, Australia due to childhood lead poisoning helped shed some light into a larger spectrum of lead-induced nephropathy.[188] Subsequently, reports of lead nephropathy appeared among blue collar workers who worked as distillers of alcoholic beverages in the southeast United States (US) and among industrial lead workers.[189, 190]

This study provided a method of determining the optimal RVI based on specific patient comorbidities. It also showed that without taking into account patient

comorbidities, the optimal RVI was 33.88 days, though this is specific to the Medicare population. Keenan (2009) recommended a RVI of over 6 months of monitoring intervals for diabetic patients.[191] Van den Bent (2008) determined that intervals for the chronic progressive disease, epilepsy, should be based anywhere from 1 month to 1 year.[192] Schulberg (1998) noted that for chronic depression patients should be seen between 6 to 8 weeks. Other studies looked at the optimal RVI for melanoma and found the optimal period to be 2 weeks (Frencken, 2009).[193] Though all these studies provide helpful guidelines for RVI, they are not customized recommendations. This study provided a method of predicting RVI based on individual comorbidities and the probability of CKD.

This study also focused on determining the diagnoses that have the greatest impact on RVI for patients with CKD in order to optimize routine RVI for primary care. The results of which, could help maximize access to care for diabetic patients and therefore inform and influence practice management and policy standards related to RVI.

Aim 3: To Examine the Probable Causal Relationship between Revisit Intervals (RVI) and Chronic Kidney Disease (CKD) in Diabetic Medicare Patients

Overall, blood toxicity, neoplasm, orthopedic and mechanical injuries, and Aluminum Bone Disease were the leading causes of CKD in the present study. The biophysical mechanisms between CKD as a result of kidney overuse due to filtering toxins in the blood, drugs and medications is well known; however, patients who present with a history of the comorbidities should potentially be screened early for CKD as the Likelihood of occurrence may be higher in those patients. Determining the main causes

of a chronic illness for each patient can help ensure that patients with the risk factors are seen before their disease progresses.

In general, researcher experience and expertise would determine covariates that are deemed causal, and therefore results are already directionally steered from the beginning of the study.[194] We see that researchers at times perform studies on data collected through carefully designed experiments where solid prior causal knowledge is of vital importance.[195]

In this study, the concept of causality was first based on associations between nodes but only under specific criteria can true causal relationships be inferred. Bayes formula helps us predict the odds of CKD occurring from the Likelihood Ratios associated with other covariates. The results were derived from the probability distribution P($y|t$) which results from a mixture of the causal effect P($y|do(t)$) and the statistical associations produced by the back-door path t ← x → y, where x is the confounder. Here we see that neither x → t nor x → y is the causal effect we wanted to estimate.[196] Knowing the direct and indirect causes of CKD can help ensure that patients with these risk factors are seen before their disease progresses.

Study Limitations

The average RVI was calculated as the number of days between two consecutive appointments by the same patient, by the same provider. At times a patient may see a different provider in the same practice. The appointment with another provider in the

same practice should be factored into the RVI calculation. For the purposes of this study, we assumed that the providers, as identified by their National Provider Identifier (NPIs) are part of a different practices.

Dropping variables that are not in the causal path may open up other Back-Door paths; however, this study did not review new paths that were created. Stratification can lead to other combinations and interactions that may have not been accounted for in this study.

When variables are highly correlated with each other, LASSO tends to select one variable from strata and ignore the others. Also, in datasets with many variables but small number of cases, LASSO selects most of the variables before it saturates. Another limitation is that the study findings may be specific to the Medicare population who on average tend to be over 65 years old. It is possible that the findings will be different when the methodology is applied to other population groups.

Conclusions

This study has shown that outpatient diagnoses can provide probable predictions of adverse effects of chronic conditions such as CKD. We relied on a comprehensive set of diagnoses, including approximately 18,000 distinct repeated diagnoses. We predicted future CKD in more than 800,000 distinct patients over 3 years. The data was randomly divided into before and after a notional date in which the predictions are being made. The data was narrowed down to only patients with Diabetes who were over the age of 65 years old. All occurrences of CKD prior to the notional date were removed from the

dataset. The probability of occurrence of CKD given a long RVI by the use of LASSO executed in SQL can assist providers in determining the optimal time by calculating the maximum and minimum values in the given observations.

This methodology can be used in an EHR-based screening in clinical settings. This method is relatively easy to implement since it uses SQL queries to execute. It does not rely on effort of either the patient or the clinician and therefore can be widely and consistently implemented across clinics. Patients are not asked to complete surveys that later must be integrated into the EHR or fill out separate consent forms as a result of this. Our proposed EHR-based screening is more comprehensive; it includes all diagnoses, as opposed to selected health diagnoses. This allows the screening tool to be relevant to a wider set of patients. As more data is entered, the algorithm can continue to 'learn' and narrow the RVI, but it is starting point and a tool to assist providers in making their decisions. With continued use and proven benefit of the algorithm as it relates to access and patient outcomes, there will be an increase benefits realization. Additional research should be conducted with other chronic conditions and the algorithm should be tested in a clinical setting.

In robust and highly dimensional datasets such as those in EHR and utilization/claims data, we see that being able to effectively shrink the number of features based on their relative importance is a crucial step. This study provided a methodology of executing this directly in an EHR in order to facilitate customized, evidence-based decision-making. Operational decision making plays a key role in provider productivity, appointment capacity, and in turn quality. Despite the important influence of ambulatory

appointment revisit intervals (RVI) on access to care, physicians receive no formal training in this area and research indicates that there is significant practice variation.[197] Determining the optimal time in which patients need to be seen, as well as the most probable causes of adverse effects can help tailor medical treatment to patient characteristics using decision making tools.

APPENDIX

The appendix provides steps and Standard Query Language (SQL) codes that were used to run the calculations and to generate tables and results. The calculations were made on outpatient data from the Centers for Medicare and Medicaid Services (CMS) Limited Dataset and executed in SQL.

1) Create New Alias

The codes below provide the renaming method of the features from the Centers for Medicare and Medicaid Services (CMS) Limited Dataset in order to make them more intuitive for analysis.

```
Table: Diabetes
Drop table Diabetes;
Select dsysrtky as ID, claimno as Claim, dob_dt as Age, thru_dt as Date, gndr_cd as Gender, Race_cd as Race,
carr_clm_blg_npi_num as NPI,
icd_dgns_cd1 as I1,
icd_dgns_cd2 as I2,
icd_dgns_cd3 as I3,
icd_dgns_cd4 as I4,
icd_dgns_cd5 as I5,
icd_dgns_cd6 as I6,
```

icd_dgns_cd7 as I7,
icd_dgns_cd8 as I8,
icd_dgns_cd9 as I9,
icd_dgns_cd10 as I10,
icd_dgns_cd11 as I11,
icd_dgns_cd12 as I12
into Diabetes
from Car_clm2016;
SELECT 45,673,594
Table D0 – 276 codes

2) Create a Sample of all Medicare patients with Diabetes

The codes below provides the procedure used in order to include patients with diabetes for each individual table.

Table: D1
Drop table D1;
Select d.*
Into D1
From Diabetes d, D0 o
Where d.i1 = o.dx;

Repeated for all columns through i12.

Results:
I1 = SELECT 1789479
I2 = SELECT 1068898
I3 = SELECT 796353
I4 = SELECT 525594
I5 = SELECT 200343
I6 = SELECT 117490

I7 = SELECT 70828
I8 = SELECT 43727
I9= SELECT 24579
I10= SELECT 15807
I11 = SELECT 10730
I12 = SELECT 6685

3) Merging all Diabetes Columns

The Standard Query Language (SQL) codes below provides the procedures used to merge the individual columns and tables of patients with diabetes into one standard dataset.

```
Table: DiabetesTable
Drop table DiabetesTable;
Select * into DiabetesTable
From (
(Select * from D1) union all
(Select * from D2) union all
(Select * from D3) union all
(Select * from D4) union all
(Select * from D5) union all
(Select * from D6) union all
(Select * from D7) union all
(Select * from D8) union all
(Select * from D9) union all
(Select * from D10) union all
(Select * from D11) union all
(Select * from D12)) as tmp;
SELECT 4,670,513
```

4) Drop Ages Lower Than 65 and Unknowns

The procedures below illustrate the methods in which patients who are younger than 65 years old, and those who the age category is NULL is excluded from the sample.

Table: DiabetesTableOld
0 = Unknown
1 = <65
2 = 65 Thru 69
3 = 70 Thru 74
4 = 75 Thru 79
5 = 80 Thru 84
6 = >84

```
Drop Table DiabetesTableOld;
Select * Into DiabetesTableOld
From DiabetesTable
Where Age <> '1';
SELECT 3813850
```

5) Calculating the Revisit Interval (RVI)

The SQL codes below provides procedures of calculating the Revisit Intervals (RVI) in the dataset. This first includes generating a unique identifier, calculating the RVI as the number of days between two consecutive outpatient appointments, by the same patient and the same provider. Lastly an index date is then generated randomly throughout the encounters for each patient.

Table: UniqueID

```
Drop table UniqueID;
Select distinct(ID), claim
Into UniqueID
From DiabetesTableOld;
```

SELECT 3571362

Table: UniqueNPI
Drop table UniqueNPI;
Select Distinct(NPI)
Into UniqueNPI
From DiabetesTableOld;
SELECT 104484

Drop table RVI;
Select distinct(d.ID), d.NPI, d.Claim, d.Age, d.Gender, d.Race, d.Date, d.i1, d.i2, d.i3, d.i4, d.i5, d.i6, d.i7, d.i8, d.i9, d.i10, d.i11, d.i12,
LEAD(d.Date) OVER (PARTITION BY ID, NPI ORDER BY d.Date) as Next_Date,
LEAD(d.Date) OVER (PARTITION BY ID, NPI ORDER BY d.Date) - Date as Difference
Into RVI
From DiabetesTableOld d
Group by d.NPI, d.ID, d.Claim, d.Age, d.Gender, d.Race, d.Date, d.i1, d.i2, d.i3, d.i4, d.i5, d.i6, d.i7, d.i8, d.i9, d.i10, d.i11, d.i12;
SELECT 3571362

<u>Randomization:</u>
Randomized Table: RD
Drop Table RD;
Select * Into RD From (Select ID, Claim, Date From RVI Order by Random()) as rd;
SELECT 3571362

<u>Determine an Index Date:</u>
Index Date Table: RandomDate
Drop Table RandomDate;
Select Distinct on (ID) ID, Claim, Date
Into RandomDate

From RVI;
SELECT 546262

Select ID, Claim, Date From RandomDate Group by ID, Claim, Date Order by ID Limit 30;

```
ldsbase=> Select ID, Claim, Date From RandomDate Group by ID, Claim, Date Order by ID Limit 30;
    id     | claim |    date
-----------+-------+------------
 100000015 |    12 | 2016-03-28
 100000053 |    57 | 2016-12-14
 100000099 |   115 | 2016-12-05
 100000241 |   184 | 2016-07-29
 100000285 |   187 | 2016-02-08
 100000905 |   579 | 2016-10-15
 100000909 |   592 | 2016-07-11
 100001359 |   829 | 2016-12-16
 100001399 |   873 | 2016-02-11
 100001501 |   927 | 2016-08-29
 100001871 |   988 | 2016-03-25
 100001909 |  1011 | 2016-05-10
 100001913 |  1058 | 2016-06-06
 100001925 |  1094 | 2016-01-08
 100001989 |  1191 | 2016-09-21
 100002065 |  1216 | 2016-08-02
 100002329 |  1308 | 2016-02-19
 100002449 |  1507 | 2016-03-24
 100002729 |  1660 | 2016-03-10
 100002745 |  1726 | 2016-01-07
 100002829 |  1813 | 2016-06-15
 100003415 |  1962 | 2016-07-22
 100003615 |  2120 | 2016-01-04
 100003935 |  2172 | 2016-09-13
 100004311 |  2460 | 2016-03-30
 100004369 |  2547 | 2016-05-26
 100004701 |  2677 | 2016-05-10
```

6) Drop Variables after the Index Date

The section shows how tautological predicators are then removed from the data by dropping all diagnosis that occur after the index date.

Matching Random Claims to Master List Table: RandomDate1
Drop Table RandomDate1;
Select rv.*
Into RandomDate1
From RVI rv, RandomDate rd
Where rv.Claim = rd.Claim;
SELECT 645,351

Dropping Post-Index Variables Table: RandomDate2
Drop Table RandomDate2;
Select rv.*
Into RandomDate2
From RVI rv, RandomDate rd
Where rv.id = rd.id and rv.Date <= rd.Date;
SELECT 2081632

7) Calculating the Average Revisit Interval (RVI) Across the Random Index Sample

This section provides the SQL codes for determining the average RVI across the entire sample based on the randomized index date.

Difference Table: RandomDate2
Select Sum(Difference) as DaySums,
Sum(Difference) / 2081632.0 as AverageRVI
From RandomDate2;

```
ldsbase-> From RandomDate2;
 daysums  |     averagervi
----------+--------------------
 82125623 | 39.4525175439270726
(1 row)
```

8) Determining the Median Revisit Interval (RVI) Across the Random Index Sample

The codes below provides information on determining the median RVI across the sample.

Select max(difference) as Median from (select difference, ntile(2) over (order by difference) AS bucket from randomdate2) as t where bucket = 1 group by bucket;

```
ldsbase=> select max(difference) as Median from (select difference, ntile(2) over (order by difference) AS bucket from randomdate2) as t where bucket =
 1 group by bucket;
 median
--------
     91
(1 row)
```

9) Determining the Maximum RVI Across the Random Index Sample

Here the maximum Revisit Intervals are calculated across the strata based on the randomized index date.

Select stddev(difference) from randomdate2;

```
ldsbase=> select max(difference) from randomdate2;
 max
-----
 363
(1 row)
```

Select max(difference) from Randomdate2;

```
ldsbase=> select stddev(difference) from randomdate2;
       stddev
--------------------
 64.3122452336158123
(1 row)
```

10) Create a Diagnosis Table for all CKD Codes

This section provides steps used to create a table that includes all patients with CKD that occurs after the randomized index date.

```
Drop table K0;
Create table K0 (Dx VARCHAR Not NULL);
Insert into K0 (Dx)
Values
    (  'N181'    )       ,
    (  'N182'    )       ,
    (  'N183'    )       ,
    (  'N184'    )       ,
    (  'N185'    )       ,
    (  'N186'    )       ,
    (  'N189'    )       ,
    (  'R880'    )       ,
    (  'Z4901'   )       ,
    (  'Z4902'   )       ,
    (  'Z4931'   )       ,
    (  'Z4932'   )       ,
    (  'Z9115'   )       ,
    (  'Z940'    )       ,
    (  'Z992'    )       ,
    (  'N131'    )       ,
    (  'N132'    )       ,
    (  'N1330'   )       ,
    (  'N1339'   )       ,
    (  'N134'    )       ,
    (  'N135'    )       ,
```

```
(  'N1370'  )  ,
(  'N1371'  )  ,
(  'N13721' )  ,
(  'N13722' )  ,
(  'N13729' )  ,
(  'N13731' )  ,
(  'N13732' )  ,
(  'N13739' )  ,
(  'N138'   )  ,
(  'N159'   )  ,
(  'N16'    )  ,
(  'N250'   )  ,
(  'N251'   )  ,
(  'N2581'  )  ,
(  'N2589'  )  ,
(  'N259'   )  ,
(  'N270'   )  ,
(  'N271'   )  ,
(  'N279'   )  ,
(  'N280'   )  ,
(  'N281'   )  ,
(  'N2881'  )  ,
(  'N2882'  )  ,
(  'N2883'  )  ,
(  'N2884'  )  ,
(  'N2885'  )  ,
(  'N2886'  )  ,
(  'N2889'  )  ,
(  'N289'   )  ,
```

```
(   'N29'       )       ,
(   'R802'      )       ,
(   'Z87441'    )       ;

INSERT 0 53
```

Kidney Diagnosis Code Table: K0

```
Drop table K1;
Select x.*
Into K1
From RandomDate2 x, K0 o
Where x.i1 = o.dx;
```

Repeated for all columns through i12.
K1= 50723
K2= 61368
K3= 58983
K4= 47673
K5= 24878
K6= 17650
K7= 12140
K8= 8146
K9= 4431
K10= 2943
K11= 1923
K12= 1276

Diabetics with CKD Table: KDN

Drop Table KDN;
Select * Into KDN
From (
(Select * from K1) union all
(Select * from K2) union all
(Select * from K3) union all
(Select * from K4) union all
(Select * from K5) union all
(Select * from K6) union all
(Select * from K7) union all
(Select * from K8) union all
(Select * from K9) union all
(Select * from K10) union all
(Select * from K11) union all
(Select * from K12)) as tmp;
SELECT 292134
Distinct Patients: 86,182

```
ldsbase=> select count(distinct(id)) from kdn;
 count
-------
 86182
(1 row)
```

11) Created Table for Patients Without Kidney Related Conditions

Below includes the SQL codes used to create tables of patients without CKD

Diabetics without CKD: NoNK

Drop table NoNK;
Select *
into NoNK
From RandomDate2

except
(Select * From KDN);
SELECT 1877943

Parse Tables:

Parsed Diagnosis Codes in Master Table: Parsed

```
Drop table Parsed;
Select Parsed.* Into Parsed From (
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i1 as Dx
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i2
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i3
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i4
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i5
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i6
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i7
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i8
```

From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i9
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i10
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i11
From RandomDate2
Union all
Select id, npi, claim, age, gender, race, date, next_date, Coalesce (Difference,0) as Difference, i12
From RandomDate2) as Parsed;
SELECT 24979584

id	npi	claim	age	gender	race	date	next_date	difference	dx
100000015	1023049236	12	4	1	1	2016-03-28	2016-05-24	57	E118
100000015	1801874573	4	4	1	1	2016-02-19	2016-03-02	12	I252
100000015	1801874573	8	4	1	1	2016-03-02	2016-03-18	16	R079
100000015	1801874573	11	4	1	1	2016-03-18		0	I10
100000053	1164646733	44	5	1	2	2016-04-04		0	E1142
100000053	1205868528	56	5	1	2	2016-11-30	2016-12-14	14	I10
100000053	1205868528	57	5	1	2	2016-12-14		0	I10
100000053	1417030412	46	5	1	2	2016-05-17	2016-08-23	98	M2041
100000053	1417030412	50	5	1	2	2016-08-23	2016-10-18	56	M2041
100000053	1417030412	53	5	1	2	2016-10-18		0	M2041
100000099	1265412001	115	6	1	1	2016-12-05		0	K824
100000241	1801990494	184	5	2	2	2016-07-29		0	E1140
100000285	1649236357	187	5	1	1	2016-02-08	2016-05-09	91	B351
100000905	1063447316	565	5	2	1	2016-09-19	2016-09-20	1	N200
100000905	1063447316	567	5	2	1	2016-09-20	2016-10-12	22	N200
100000905	1063447316	577	5	2	1	2016-10-12		0	A419
100000905	1427041268	564	5	2	1	2016-09-18	2016-10-19	31	N132
100000905	1538194980	576	5	2	1	2016-10-12	2016-10-15	3	N10
100000905	1538194980	579	5	2	1	2016-10-15		0	J189
100000905	1588606123	549	5	2	1	2016-04-26	2016-10-27	184	N390

(20 rows)

12) Add a Seriel Primay Key to Master Table Parsed:

Alter Table Parsed Add Column Identifier Serial Primary Key;

id	npi	claim	age	gender	race	date	next_date	difference	dx	identifier
100000015	1023049236	12	4	1	1	2016-03-28	2016-05-24	57	E118	1
100000015	1801874573	4	4	1	1	2016-02-19	2016-03-02	12	I252	2
100000015	1801874573	8	4	1	1	2016-03-02	2016-03-18	16	R079	3
100000015	1801874573	11	4	1	1	2016-03-18		0	I10	4
100000053	1164646733	44	5	1	2	2016-04-04		0	E1142	5

(5 rows)

Parsed CKD Table: ParsedKDN

Drop Table ParsedKDN;
Select p.* Into ParsedKDN From Parsed p, K0 k Where p.Dx = k.Dx;
SELECT 225795

```
ldsbase=> select * from ParsedKDN limit 5;
```

id	npi	claim	age	gender	race	date	next_date	difference	dx	identifier
100000905	1427041268	564	5	2	1	2016-09-18	2016-10-19	31	N132	17
100002829	1912006842	1833	4	1	4	2016-08-21		0	N186	76
100010677	1366763260	6416	5	1	1	2016-08-18	2016-08-30	12	N289	212
100011915	1417204074	7261	5	2	1	2016-12-05		0	N182	231
100016631	1881659506	10292	4	2	1	2016-09-01	2016-10-01	30	N186	342

(5 rows)

11) Parse NoNK

Parsed Diabetics without CKD Table: ParsedNoNK
Drop table ParsedNoNK;
Select *
Into ParsedNoNK
From Parsed
Except
(Select * From ParsedKDN);
SELECT 24753789

12) Calculating the Likelihood Ratios for Kidney Conditions and Each Diagnosis Code

Below are steps used to calculate the Likelihood Ratios for patients with Kidney conditions

```
Drop table Num;
Select distinct(p.dx), (round(cast(count(distinct(p.id)) as decimal)/ 24979584,10)) as N
Into Num3
From Parsed p, ParsedKDN k
Where p.dx = k.dx
Group by p.dx
Limit 10;
SELECT 50
```

dx	n
N3010	0.0000057380
N37	0.0000000576
N181	0.0001021329
N159	0.0000031857
78862	0.0000034735
58389	0.0000016504
N142	0.0000003262
78891	0.0000014585
Z87442	0.0000169838
59381	0.0000015928

(10 rows)

```
Drop table Den1;
Select distinct(dx), (round(cast(count(dx) as decimal)/ 24979584,10)) as D
Into Den1
From ParsedNoNK
Group by dx;
SELECT 18789
```

dx	d
	0.6420518241
37762	0.0000001151
Q188	0.0000000192
Y92098	0.0000002495
S299XXS	0.0000000192
H60592	0.0000001343
B0053	0.0000000576
9299	0.0000000768
30270	0.0000008060
N770	0.0000000192

(10 rows)

```
Drop Table LRK;
Select p.Dx, (n.N/d.D) as LR
Into LRK
From Parsed p, Num n, Den d
Where p.dx = n.dx or p.dx = d.dx
Group by p.Dx, n.N, d.D
Order by p.Dx asc;

Select p.Dx, (n.N/d.D) as LR
From Parsed p, Num n, Den d
Where p.dx = n.dx or p.dx = d.dx
Group by p.Dx, n.N, d.D
Order by p.Dx asc;
```

```
ldsbase=> select distinct(dx), lr from LRK limit 10;
   dx    |          lr
---------+----------------------
 734     | 0.01631279759012729570
 0380    | 0.05193882384787169250
 D530    | 0.02501302761855132882
 C8268   | 1.00000000000000000000
 2863    | 0.08004168837936425221
 S93432A | 0.05885959534028203556
 5695    | 0.26330224904004388371
 71680   | 0.05407689057879171948
 N403    | 0.83028580451602715932
 E342    | 0.20010422094841063054
(10 rows)
```

Order by desc:

13) Calculate the Likelihood Ratios for each Diagnosis Related to Long and Short RVI was then Determined

Below are the SQL codes used to calculate the average, median, long and short Likelihood Ratios across the strata.

Drop table NumR;
Select p.dx, (round(cast(count(p.id) as decimal)/ 52108560,10)) as N
Into NumR
From Parsed p
Where p.Difference >= 39.45
Group by p.dx;
21132

Drop table DenR;
Select p.dx, (round(cast(count(p.id) as decimal)/ 52108560,10)) as D
Into DenR
From Parsed p
Where p.Difference < 39.45
Group by p.dx;
28128

```
Drop Table LRR;
Select p.Dx, (n.N/d.D) as LRR
Into LRR
From Parsed p, NumR n, DenR d
Where (p.Dx = n.Dx) and (p.Dx = d.Dx)
Group by p.Dx, n, d
Order by p.Dx desc;
19331

Select * From LRR r
Group by r.Dx, r.lrr
Order by r.lrr desc;
```

```
ldsbase=> Select * From LRR r
ldsbase-> Group by r.Dx, r.lrr
ldsbase-> Order by r.lrr desc;
   dx    |          lrr
---------+-----------------------
 17322   |     7.9947916666666667
 37231   |     6.9947916666666667
 36847   |     5.9973958333333333
 0529    |     5.9947916666666667
 17311   |     5.9947916666666667
 36842   |     5.9947916666666667
 74332   |     5.9947916666666667
 D2362   |     5.9947916666666667
 T381X5A |     5.9947916666666667
 1740    |     5.6631944444444444
 05441   |     5.0000000000000000
 1961    |     5.0000000000000000
 36474   |     5.0000000000000000
 37245   |     5.0000000000000000
 37762   |     5.0000000000000000
 52513   |     5.0000000000000000
 7232    |     5.0000000000000000
 73329   |     5.0000000000000000
 7517    |     5.0000000000000000
 79415   |     5.0000000000000000
 9174    |     5.0000000000000000
```

14) Determining the Cross-Strata Odds Ratio

The SQL codes below provide information on the queries made to calculate the odds ratio across the strata.

```
Drop table LRK15;
Select distinct(dx), lr
Into LRK15
From LRK r
Where lr > 15;
23

Drop table LRR5;
```

```
Select distinct(dx), lrr
Into LRR5
From LRR r
Where lrr > 5;
10

Drop table LRK1.5;
Select distinct(dx), lr
Into LR1.5
From LRK r
Where lr <= 1.5;

Drop table LRR5;
Select distinct(dx), lrr
Into LRR5
From LRR r
Where lrr > 5;
10


Select * Into CombinedLR From LRK15
Union all (Select * From LRR5);
33
```

```
ldsbase=> select * from combinedlr;
   dx    |          lr
---------+----------------------
 T50992A |  15.9947916666666667
 T8249XD |  31.9843750000000000
 E133293 |  15.9947916666666667
 C8520   |  15.9947916666666667
 M10361  |  20.9895833333333333
 T83518A |  35.9843750000000000
 S72351D |  23.9878472222222222
 T5694XA |  33.5125977410947003
 T80211D |  33.9843750000000000
 S79819A | 165.9218750000000000
 T83510D |  15.9947916666666667
 D631    |  17.2508438535231444
 M834    |  26.9843750000000000
 T83028A |  30.9843750000000000
 S14104D |  15.9947916666666667
 T82318A |  18.9895833333333333
 7105    |  43.2573289902280130
 B9621   |  15.9947916666666667
 T85611A |  30.4843750000000000
 T56894A | 208.9010416666666667
 T8241XA |  34.6493055555555556
 Z578    |  15.7003647733194372
 S31125A |  17.9895833333333333
 D2362   |   5.9947916666666667
 17311   |   5.9947916666666667
 1740    |   5.6631944444444444
 36842   |   5.9947916666666667
 T381X5A |   5.9947916666666667
 37231   |   6.9947916666666667
 0529    |   5.9947916666666667
 17322   |   7.9947916666666667
 74332   |   5.9947916666666667
 36847   |   5.9973958333333333
(33 rows)
```

15) Determining the Number of Cases and Controls Per Strata

The queries below provide the codes used to calculate the number cases and controls per each strata and the images are the results from the queries.

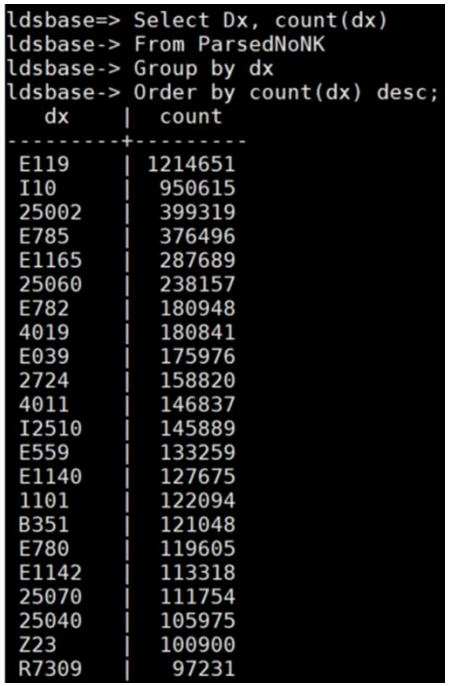
```
Select Dx, count(dx)
From ParsedKDN
Group by dx
Order by count(dx) desc;
```

dx	count
E119	259410
I10	257996
N183	169114
E1122	108502
E785	106295
N186	105400
N390	90349
E1165	83904
N179	72035
N189	71723

```
Select Dx, count(dx)
From Parsed
Group by dx
Order by count(dx) desc;
```

```
ldsbase=> Select Dx, count(dx)
ldsbase-> From Parsed
ldsbase-> Group by dx
ldsbase-> Order by count(dx) desc;
   dx    |  count
---------+---------
 E119    | 1452213
 I10     | 1186850
 E785    |  473730
 25002   |  416656
 E1165   |  370979
 25060   |  251250
 E782    |  225722
 E039    |  223675
 I2510   |  194724
 4019    |  193824
 E559    |  178033
 2724    |  172147
 E1140   |  160380
 4011    |  155965
 E780    |  145216
 E1142   |  140105
 N183    |  133699
 25040   |  129941
 B351    |  124853
 1101    |  124803
 Z23     |  124472
 25070   |  117320
 R7309   |  116138
 D649    |  113648
 E1122   |  106926
 K219    |  101899
 R7301   |   96923
```

Select Dx, count(dx)
From ParsedNoNK
Group by dx
Order by count(dx) desc;



```
ldsbase=> Select Dx, count(dx)
ldsbase-> From ParsedNoNK
ldsbase-> Group by dx
ldsbase-> Order by count(dx) desc;
   dx    |  count
---------+---------
 E119    | 1214651
 I10     |  950615
 25002   |  399319
 E785    |  376496
 E1165   |  287689
 25060   |  238157
 E782    |  180948
 4019    |  180841
 E039    |  175976
 2724    |  158820
 4011    |  146837
 I2510   |  145889
 E559    |  133259
 E1140   |  127675
 1101    |  122094
 B351    |  121048
 E780    |  119605
 E1142   |  113318
 25070   |  111754
 25040   |  105975
 Z23     |  100900
 R7309   |   97231
```

Drop Table pA;
Select identifier, id, age, gender, race, date, next_date, difference, dx, count(id) as a
Into pA
From ParsedKDN
Where Difference > 39.45
Group by identifier, id, age, gender, race, date, next_date, difference, dx;
123613

```
ldsbase=> select * from pA limit 10;
 identifier |    id     | age | gender | race |    date    | next_date  | difference |  dx   | a
------------+-----------+-----+--------+------+------------+------------+------------+-------+---
   20627287 | 489986357 |   3 | 1      | 1    | 2015-11-25 | 2015-12-30 |         35 | N186  | 1
   36198599 | 491485689 |   4 | 1      | 1    | 2016-06-29 | 2016-10-11 |        104 | R319  | 1
   26216063 | 489346623 |   2 | 2      | 2    | 2015-12-04 | 2016-01-08 |         35 | N184  | 1
   28519335 | 492487509 |   4 | 2      | 1    | 2016-06-17 | 2016-09-09 |         84 | N183  | 1
   20253198 | 100600831 |   5 | 2      | 2    | 2016-07-25 | 2016-10-24 |         91 | N181  | 1
   30082225 | 100815379 |   4 | 2      | 1    | 2014-12-02 | 2015-04-06 |        125 | 59654 | 1
   22048771 | 486961967 |   3 | 1      | 5    | 2016-07-29 | 2016-10-03 |         66 | N186  | 1
   18166206 | 471622909 |   2 | 2      | 1    | 2016-02-04 | 2016-05-05 |         91 | N186  | 1
   35304975 | 104456097 |   4 | 2      | 1    | 2015-08-11 | 2015-10-27 |         77 | 59651 | 1
   36054461 | 493610055 |   4 | 2      | 5    | 2016-06-13 | 2016-08-08 |         56 | N183  | 1
(10 rows)
```

Drop Table pB;
Select identifier, id, age, gender, race, date, next_date, difference, dx, count(id) as b
Into pB
From ParsedNoNK
Where Difference > 33.88
Group by identifier, id, age, gender, race, date, next_date, difference, dx;
SELECT 14300915

```
ldsbase=> select * from pB limit 10;
 identifier |    id     | age | gender | race |    date    | next_date  | difference | dx | b
------------+-----------+-----+--------+------+------------+------------+------------+----+---
          7 | 466382797 |   2 | 1      | 1    | 2016-01-06 | 2016-07-06 |        182 |    | 1
         11 | 151925529 |   6 | 2      | 1    | 2014-09-19 | 2014-10-29 |         40 |    | 1
         13 | 181761251 |   5 | 1      | 1    | 2016-06-03 | 2016-07-21 |         48 |    | 1
         18 | 162611667 |   1 | 1      | 2    | 2015-10-28 | 2016-01-12 |         76 |    | 1
         23 | 490325885 |   4 | 2      | 2    | 2016-04-20 | 2016-06-06 |         47 |    | 1
         27 | 162845927 |   5 | 2      | 1    | 2016-05-26 | 2016-11-17 |        175 |    | 1
         40 | 471355999 |   1 | 1      | 1    | 2015-04-17 | 2015-07-17 |         91 |    | 1
         42 | 131591607 |   5 | 1      | 1    | 2015-10-08 | 2015-11-11 |         34 |    | 1
         43 | 487999317 |   1 | 2      | 2    | 2015-01-14 | 2015-04-06 |         82 |    | 1
         45 | 157279087 |   5 | 2      | 1    | 2016-06-08 | 2016-08-03 |         56 |    | 1
(10 rows)
```

Drop Table pC;
Select identifier, id, age, gender, race, date, next_date, difference, dx, count(id) as c

Into pC
From ParsedKDN
Where Difference <=39.45
Group by identifier, id, age, gender, race, date, next_date, difference, dx;
SELECT 580524

identifier	id	age	gender	race	date	next_date	difference	dx	c
26660029	479630873	2	2	2	2015-12-29	2016-01-31	33	N189	1
23495033	464755313	2	1	0	2016-09-18	2016-09-19	1	N183	1
48900154	498357231	1	1	2	2015-12-02	2015-12-14	12	R809	1
31442114	100174961	5	2	2	2016-11-04	2016-12-05	31	N189	1
20357487	496460643	4	1	1	2016-03-29	2016-04-19	21	N183	1
26049337	493104539	4	2	1	2016-07-01	2016-07-01	0	N184	1
27642686	164604657	5	2	1	2016-12-14		0	N184	1
23180461	482382491	1	2	2	2015-11-29	2015-12-12	13	Z992	1
31356548	153856519	1	1	6	2016-01-22	2016-01-22	0	Z992	1
32572903	472766373	2	1	1	2016-06-29	2016-06-30	1	N183	1

(10 rows)

Drop Table pD;
Select identifier, id, age, gender, race, date, next_date, difference, dx, count(id) as d
Into pD
From ParsedNoNK
Where Difference <= 39.45
Group by identifier, id, age, gender, race, date, next_date, difference, dx;
SELECT 37103508

```
ldsbase=> select * from pD limit 10;
 identifier |    id     | age | gender | race |    date    | next_date  | difference |  dx  | d
------------+-----------+-----+--------+------+------------+------------+------------+------+---
          1 | 188593821 |   6 | 1      | 1    | 2015-11-24 |            |          0 |      | 1
          2 | 104287607 |   4 | 1      | 3    | 2016-01-27 | 2016-01-28 |          1 |      | 1
          3 | 468459827 |   2 | 2      | 0    | 2016-08-27 | 2016-08-27 |          0 |      | 1
          4 | 472502217 |   2 | 2      | 1    | 2014-01-13 | 2014-01-27 |         14 |      | 1
          5 | 485007279 |   3 | 1      | 5    | 2016-06-03 | 2016-06-04 |          1 |      | 1
          6 | 474629197 |   1 | 1      | 1    | 2016-05-12 | 2016-05-19 |          7 |      | 1
          8 | 489802589 |   1 | 2      | 1    | 2014-10-27 | 2014-11-11 |         15 |      | 1
          9 | 491555189 |   1 | 2      | 2    | 2014-08-21 | 2014-09-17 |         27 | 2875 | 1
         10 | 160598581 |   5 | 2      | 1    | 2015-01-12 | 2015-02-12 |         31 |      | 1
         12 | 485253479 |   3 | 1      | 1    | 2016-06-07 | 2016-06-07 |          0 |      | 1
(10 rows)
```

"Strata"
Create a table of all subjects:

Drop table pCalculation;
Drop table pCalculation1;
Select pCalculation1.* Into pCalculation1 From
(Select p.identifier, p.id, p.difference, p.age, p.Dx, p.date, p.next_date,
Coalesce (a,0) as a,
Coalesce (b,0) as b,
Coalesce (c,0) as c,
Coalesce (d,0) as d
From Parsed P
Full outer join pA a on a.identifier = p.identifier
Full outer join pB b on b.identifier = p.identifier
Full outer join pC c on c.identifier = p.identifier
Full outer join pD d on d.identifier = p.identifier) as pCalculation1;
SELECT 52,108,560

```
ldsbase=> select * from pcalculation limit 10;
 identifier |    id     | difference |  dx   |    date    | next_date  | a | b | c | d
------------+-----------+------------+-------+------------+------------+---+---+---+---
   16441497 | 136422969 |          3 | E119  | 2016-02-16 | 2016-02-19 | 0 | 0 | 0 | 1
   16441556 | 150271065 |          0 | F419  | 2016-11-03 | 2016-11-03 | 0 | 0 | 0 | 1
   16441641 | 473522185 |          0 | E349  | 2015-12-15 | 2015-12-15 | 0 | 0 | 0 | 1
   16441688 | 486668839 |         10 | D649  | 2016-01-19 | 2016-01-29 | 0 | 0 | 0 | 1
   16441689 | 466081495 |          0 | N179  | 2016-11-14 | 2016-11-14 | 0 | 0 | 1 | 0
   16441742 | 486303119 |        110 | 700   | 2015-07-02 | 2015-10-20 | 0 | 1 | 0 | 0
   16441777 | 474292459 |          0 | 25060 | 2015-06-08 |            | 0 | 0 | 0 | 1
   16441808 | 470654297 |        210 | E119  | 2015-11-18 | 2016-06-15 | 0 | 1 | 0 | 0
   16441876 | 497192397 |          5 | E1151 | 2016-06-10 | 2016-06-15 | 0 | 0 | 0 | 1
   16441931 | 493445771 |        112 | 25040 | 2015-05-28 | 2015-09-17 | 0 | 1 | 0 | 0
(10 rows)
```

Select id, sum(a) as a, sum(b) as b, sum(c) as c, sum(d) as d
Into Grouped
From (select Identifier, id, min(a) as a, min(b) as b, min(c) as c, min(d) as d From pCalculation1 Group by Identifier, id, a) as tmp
Group by id
Order by id;
SELECT 659857

```
ldsbase=> Select id, sum(a) as a, sum(b) as b, sum(c) as c, sum(d) as d
ldsbase-> Into Grouped
ldsbase-> From (select Identifier, id, min(a) as a, min(b) as b, min(c) as c, min(d) as d From pCalculation1 Grou
p by Identifier, id, a) as tmp
ldsbase-> Group by id
ldsbase-> Order by id;
SELECT 659857
ldsbase=> select * from Grouped limit 50;
    id     | a | b  | c  |  d
-----------+---+----+----+-----
 100000015 | 0 | 12 |  0 |  12
 100000053 | 0 | 24 |  0 |  12
 100000099 | 0 |  0 |  0 |  12
 100000221 | 0 | 48 |  0 |  12
 100000241 | 0 |  0 |  1 |  11
 100000285 | 0 | 48 |  0 |  48
 100000905 | 0 | 72 |  7 | 149
 100000909 | 0 | 24 |  0 |  36
 100001193 | 0 | 12 |  0 |  12
 100001359 | 0 | 84 |  0 | 300
 100001399 | 0 | 84 |  0 | 168
 100001501 | 0 |  0 |  0 |  12
 100001871 | 0 | 12 |  0 |   0
 100001909 | 0 | 12 |  0 |  24
 100001913 | 0 |  0 |  1 |  11
 100001925 | 0 | 12 |  0 |  24
```

Select id, a, b, c, d, (a + b + c + d) as ni Into Grouped1 From Grouped Group by id, a, b, c, d;

```
ldsbase=> Select id, a, b, c, d, (a + b + c + d) as ni Into Grouped1 From Grouped Group by id, a, b, c, d;
SELECT 659857
ldsbase=> select * from Grouped1 limit 10;
    id     | a | b  | c |  d  | ni
-----------+---+----+---+-----+-----
 100000015 | 0 | 12 | 0 |  12 |  24
 100000053 | 0 | 24 | 0 |  12 |  36
 100000099 | 0 |  0 | 0 |  12 |  12
 100000221 | 0 | 48 | 0 |  12 |  60
 100000241 | 0 |  0 | 1 |  11 |  12
 100000285 | 0 | 48 | 0 |  48 |  96
 100000905 | 0 | 72 | 7 | 149 | 228
 100000909 | 0 | 24 | 0 |  36 |  60
 100001193 | 0 | 12 | 0 |  12 |  24
 100001359 | 0 | 84 | 0 | 300 | 384
(10 rows)
```

Select
Sum(a + b) as LongRVI,

Sum(c + d) as ShortRVI,
Sum(a + c) as KidneyDisease,
Sum(b + d) as NoDisease
From Grouped3
Group by a, b, c, d;

```
ldsbase=> Select
ldsbase-> Sum(a + b) as LongRVI,
ldsbase-> Sum(c + d) as ShortRVI,
ldsbase-> Sum(a + c) as KidneyDisease,
ldsbase-> Sum(b + d) as NoDisease
ldsbase-> From Grouped3
ldsbase-> Group by a, b, c, d;
 longrvi | shortrvi | kidneydisease | nodisease
---------+----------+---------------+-----------
       0 |      384 |             4 |       380
      96 |      444 |            17 |       523
      24 |      108 |            14 |       118
     576 |     2208 |            24 |      2760
      36 |      264 |            26 |       274
     108 |      120 |             7 |       221
      12 |       60 |            15 |        57
       0 |      384 |             3 |       381
       0 |     5040 |             0 |      5040
      12 |      300 |             6 |       306
     180 |      576 |             9 |       747
     336 |      384 |            16 |       704
      36 |      264 |            17 |       283
     144 |      432 |            14 |       562
     168 |      576 |            12 |       732
      36 |       48 |            18 |        66
     192 |      552 |            16 |       728
```

Select
Sum(a + b) as LongRVI,
Sum(c + d) as ShortRVI,
Sum(a + c) as KidneyDisease,

Sum(b + d) as NoDisease
Into Cases1
From Grouped3
Group by a, b, c, d;
SELECT 22386

```
ldsbase=> select * from Cases1 limit 10;
 longrvi | shortrvi | kidneydisease | nodisease
---------+----------+---------------+-----------
       0 |      384 |             4 |       380
      96 |      444 |            17 |       523
      24 |      108 |            14 |       118
     576 |     2208 |            24 |      2760
      36 |      264 |            26 |       274
     108 |      120 |             7 |       221
      12 |       60 |            15 |        57
       0 |      384 |             3 |       381
       0 |     5040 |             0 |      5040
      12 |      300 |             6 |       306
(10 rows)
```

15) Determining the Confidence Interval

Below provides the SQL codes used to calculate the confidence interval for the data.

If a homogenous common odds-ratio, exists, then its statistical significance is tested as:

Estimate of Common Odds Ratio of Impact of X on Y

$\pi_i = \frac{a_i + d_i}{n_i}$, $Q_i = \frac{b_i + c_i}{n_i}$, $R_i = \frac{a_i d_i}{n_i}$, $S_i = \frac{b_i c_i}{n_i}$

Select id, a, b, c, d, ni,

(a + d) as OnePai,
(b + c) as OneQi,
(a * d) as OneRi,
(b * c) as OneSi
Into Grouped2
From Grouped1
Group by id, a, b, c, d, ni;
SELECT 659857

```
ldsbase=> Select id, a, b, c, d, ni,
ldsbase-> (a + d) as OnePai,
ldsbase-> (b + c) as OneQi,
ldsbase-> (a * d) as OneRi,
ldsbase-> (b * c) as OneSi
ldsbase-> Into Grouped2
ldsbase-> From Grouped1
ldsbase-> Group by id, a, b, c, d, ni;
SELECT 659857
ldsbase=> select * from Grouped2 limit 10;
    id     | a | b  | c |  d  | ni  | onepai | oneqi | oneri | onesi
-----------+---+----+---+-----+-----+--------+-------+-------+-------
 100000015 | 0 | 12 | 0 |  12 |  24 |     12 |    12 |     0 |     0
 100000053 | 0 | 24 | 0 |  12 |  36 |     12 |    24 |     0 |     0
 100000099 | 0 |  0 | 0 |  12 |  12 |     12 |     0 |     0 |     0
 100000221 | 0 | 48 | 0 |  12 |  60 |     12 |    48 |     0 |     0
 100000241 | 0 |  0 | 1 |  11 |  12 |     11 |     1 |     0 |     0
 100000285 | 0 | 48 | 0 |  48 |  96 |     48 |    48 |     0 |     0
 100000905 | 0 | 72 | 7 | 149 | 228 |    149 |    79 |     0 |   504
 100000909 | 0 | 24 | 0 |  36 |  60 |     36 |    24 |     0 |     0
 100001193 | 0 | 12 | 0 |  12 |  24 |     12 |    12 |     0 |     0
 100001359 | 0 | 84 | 0 | 300 | 384 |    300 |    84 |     0 |     0
(10 rows)
```

Select id, a, b, c, d, ni, (OnePai/ni) as Pai, (OneQi/ni) as Qi, (OneRi/ni) as Ri, (OneSi/ni) as Si Into Grouped3 From Grouped2 Group by id, a, b, c, d, ni, OnePai, OneQi, OneRi, OneSi;
SELECT 659857

```
ldsbase=> select * from Grouped3 limit 10;
    id     | a | b  | c |  d  | ni  |          pai          |          qi           |          ri           |
         si
-----------+---+----+---+-----+-----+-----------------------+-----------------------+-----------------------+-
----------------------
 100000015 | 0 | 12 | 0 |  12 |  24 | 0.50000000000000000000 | 0.50000000000000000000 | 0.00000000000000000000 |
0.00000000000000000000
 100000053 | 0 | 24 | 0 |  12 |  36 | 0.33333333333333333333 | 0.66666666666666666667 | 0.00000000000000000000 |
0.00000000000000000000
 100000099 | 0 |  0 | 0 |  12 |  12 | 1.00000000000000000000 | 0.00000000000000000000 | 0.00000000000000000000 |
0.00000000000000000000
 100000221 | 0 | 48 | 0 |  12 |  60 | 0.20000000000000000000 | 0.80000000000000000000 | 0.00000000000000000000 |
0.00000000000000000000
 100000241 | 0 |  0 | 1 |  11 |  12 | 0.91666666666666666667 | 0.08333333333333333333 | 0.00000000000000000000 |
0.00000000000000000000
 100000285 | 0 | 48 | 0 |  48 |  96 | 0.50000000000000000000 | 0.50000000000000000000 | 0.00000000000000000000 |
0.00000000000000000000
 100000905 | 0 | 72 | 7 | 149 | 228 | 0.65350877192982456140 | 0.34649122807017543860 | 0.00000000000000000000 |
   2.2105263157894737
 100000909 | 0 | 24 | 0 |  36 |  60 | 0.60000000000000000000 | 0.40000000000000000000 | 0.00000000000000000000 |
0.00000000000000000000
 100001193 | 0 | 12 | 0 |  12 |  24 | 0.50000000000000000000 | 0.50000000000000000000 | 0.00000000000000000000 |
```

```
ldsbase=> select id, a, b, c, d, (a + d)/(a + b + c + d) as Pai, (b + c)/ (a + b + c + d) as Qi, (b*c)/(a + b + c + d) as Si, (a*d)/(a + b + c + d) as Ri from Grouped group
by id, a, b, c, d order by id;
```

id	a	b	c	d	pai	qi	si	ri
100000015	0	12	0	12	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000
100000053	0	24	0	12	0.3333333333333333333	0.6666666666666666667	0.0000000000000000000	0.0000000000000000000
100000099	0	0	0	12	1.0000000000000000000	0.0000000000000000000	0.0000000000000000000	0.0000000000000000000
100000221	0	48	0	12	0.2000000000000000000	0.8000000000000000000	0.0000000000000000000	0.0000000000000000000
100000241	0	0	1	11	0.9166666666666666667	0.0833333333333333333	0.0000000000000000000	0.0000000000000000000
100000285	0	48	0	48	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000
100000905	0	72	7	149	0.6535087719298245614	0.3464912280701754386	2.2105263157894737	0.0000000000000000000
100000909	0	24	0	36	0.6000000000000000000	0.4000000000000000000	0.0000000000000000000	0.0000000000000000000
100001193	0	12	0	12	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000
100001359	0	84	0	300	0.7812500000000000000	0.2187500000000000000	0.0000000000000000000	0.0000000000000000000
100001399	0	84	0	168	0.6666666666666666667	0.3333333333333333333	0.0000000000000000000	0.0000000000000000000
100001501	0	0	0	12	1.0000000000000000000	0.0000000000000000000	0.0000000000000000000	0.0000000000000000000
100001871	0	12	0	0	0.0000000000000000000	1.0000000000000000000	0.0000000000000000000	0.0000000000000000000
100001909	0	12	0	24	0.6666666666666666667	0.3333333333333333333	0.0000000000000000000	0.0000000000000000000
100001913	0	0	1	11	0.9166666666666666667	0.0833333333333333333	0.0000000000000000000	0.0000000000000000000
100001925	0	12	0	24	0.6666666666666666667	0.3333333333333333333	0.0000000000000000000	0.0000000000000000000
100001953	0	0	0	12	1.0000000000000000000	0.0000000000000000000	0.0000000000000000000	0.0000000000000000000
100001989	0	60	2	202	0.7651515151515151515	0.2348484848484848484	0.4545454545454545454	0.0000000000000000000
100001991	0	60	0	36	0.3750000000000000000	0.6250000000000000000	0.0000000000000000000	0.0000000000000000000
100002033	0	24	0	96	0.8000000000000000000	0.2000000000000000000	0.0000000000000000000	0.0000000000000000000
100002153	0	36	0	36	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000
100002329	0	12	0	12	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000
100002449	0	12	1	47	0.7833333333333333333	0.2166666666666666667	0.2000000000000000000	0.0000000000000000000
100002515	0	0	0	12	1.0000000000000000000	0.0000000000000000000	0.0000000000000000000	0.0000000000000000000
100002729	0	0	1	11	0.9166666666666666667	0.0833333333333333333	0.0000000000000000000	0.0000000000000000000
100002745	2	70	0	156	0.6929824561403508771	0.3070175438596491228	0.0000000000000000000	1.3684210526315789
100002829	0	12	1	11	0.4583333333333333333	0.5416666666666666667	0.5000000000000000000	0.0000000000000000000
100003415	2	94	11	133	0.5625000000000000000	0.4375000000000000000	4.3083333333333333	1.1083333333333333
100003615	0	24	0	24	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000
100003935	2	22	2	22	0.5000000000000000000	0.5000000000000000000	0.9166666666666666667	0.9166666666666666667
100004311	0	24	0	12	0.3333333333333333333	0.6666666666666666667	0.0000000000000000000	0.0000000000000000000
100004459	0	0	0	48	1.0000000000000000000	0.0000000000000000000	0.0000000000000000000	0.0000000000000000000
100004701	0	12	0	12	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000
100004805	0	24	0	24	0.5000000000000000000	0.5000000000000000000	0.0000000000000000000	0.0000000000000000000

$$\pi_i = \frac{a_i+d_i}{n_i} \quad , \quad Q_i = \frac{b_i+c_i}{n_i} \quad , \quad R_i = \frac{a_i d_i}{n_i} \quad , \quad S_i = \frac{b_i c_i}{n_i}$$

Select id, a, b, c, d, (a + d)/(a + b + c + d) as Pai, (b + c)/ (a + b + c + d) as Qi, (b*c)/(a + b + c + d) as Si, (a*d)/(a + b + c + d) as Ri Into Round2 From Grouped group by id, a, b, c, d order by id;

Select sum(Ri) as SumRi, sum(si) as SumSi From Round2;

```
ldsbase=> Select sum(Ri) as SumRi, sum(si) as SumSi From Round2;
           sumri              |              sumsi
------------------------------+------------------------------
 68258.824704871164807489796 | 105873.710899606952406728 58
(1 row)
```

$$\widehat{OR} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$$

```
ldsbase=> Select( sum(Ri)/sum(Si) ) as OR From Round2;
           or
------------------------
 0.64471929929420329257
(1 row)
```

$$V = \frac{\sum_i \pi_i R_i}{2(\sum_i R_i)^2} + \frac{\sum_i Q_i S_i}{2(\sum_i S_i)^2} + \frac{\sum_i (\pi_i S_i + Q_i R_i)}{2(\sum_i R_i)(\sum_i S_i)}$$

Drop table V;
Select id, pai, qi, ri, si,
((Sum(Pai * Ri))/ (2 * (sum(Ri))* (sum(Ri))) +
((Sum(Qi * Si)))/ (2 * (sum(Si)) * (sum(Si))) +
(Sum((Pai * Si) + (Qi * Ri)))/ (2 * sum(Ri) * sum(Si))) As V
From Round2
Group by id, pai, qi, ri, si;

```
ldsbase-> (Sum(Pai * Ri))/ (2 * (sum(Ri))* (sum(Ri)))
ldsbase-> From Grouped3;
                  ?column?
-------------------------------------------
 0.000004829769076441400038954486797789787 1
(1 row)

ldsbase=>
ldsbase=> Select
ldsbase-> ((Sum(Qi * Si)))/ (2 * (sum(Si)) * (sum(Si))) From Grouped3;
                  ?column?
-------------------------------------------
 0.000001666076650871720624517358376555770 8
(1 row)

ldsbase=> Select
ldsbase-> (Sum((Pai * Si) + (Qi * Ri)))/ (2 * sum(Ri) * sum(Si))
ldsbase-> From Grouped3;
                  ?column?
-------------------------------------------
 0.000006349633340298652130819403413837305 9
(1 row)
```

V = 0.0000128454790

$$\widehat{OR} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$$

Select
(Sum(Ri)) / (Sum(Si)) as OR
From Grouped3;

```
ldsbase=> Select
ldsbase-> (Sum(Ri)) / (Sum(Si)) as OR
ldsbase-> From Grouped3;
          or
-----------------------
 0.6447192992942032 9257
(1 row)
```

$$95\%\ C.I. = exp\left(Log\left(\widehat{OR}\right) \pm Z_{.025}\sqrt{V}\right)$$

Note: Z (0.025) = 1.96

Select

Exp(Log(0.64471929929) + (1.96 * SQRT(0.0000128454790))) as UpperL,

Exp(Log(0.64471929929) - (1.96 * SQRT(0.0000128454790))) as LowerL;

```
ldsbase=> Select
ldsbase-> Exp(Log(0.64471929929) + (1.96 * SQRT(0.0000128454790))) as UpperL,
ldsbase-> Exp(Log(0.64471929929) - (1.96 * SQRT(0.0000128454790))) as LowerL;
         upperl          |         lowerl
-------------------------+-------------------------
 0.832264838270622810466 | 0.820653680537017811439
(1 row)
```

16) Dropping Each Diagnosis Code with a Likelihood Ratio > 40 and Re-calculating the Confidence Interval

The following calculations were used to calculate the sensitivity analysis for the narrowed list of diagnosis (those with a Likelihood Ratio of above 40).

Confidence level is 95%: If the confidence interval does not contain the null hypothesis

value, the results are statistically significant.

<u>Baseline:</u>

Select sum(a) as a, sum(b) as b, sum(c) as c, sum(d) as d from pCalculation1;

```
ldsbase=> Select sum(a) as a, sum(b) as b, sum(c) as c, sum(d) as d from pCalculation1;
   a    |    b     |   c    |    d
--------+----------+--------+----------
 123613 | 14300915 | 580524 | 37103508
(1 row)
```

```
ldsbase=> Select sum(a) as a, sum(b) as b, sum(c) as c, sum(d) as d,
ldsbase-> (sum(a) * sum(d)) / (sum(b) * sum(c)) as OR7592
ldsbase-> From pDrop;
   a    |    b    |   c    |    d     |         or7592
--------+---------+--------+----------+------------------------
 192984 | 8544144 | 944688 | 26917500 | 0.643574855185955338793
(1 row)
```

```
ldsbase=> select * from Grouped Group by id, a, b, c, d order by id limit 1000;
    id     | a |  b  | c  |  d
-----------+---+-----+----+-----
 100000015 | 0 |  12 |  0 |  12
 100000053 | 0 |  24 |  0 |  12
 100000099 | 0 |   0 |  0 |  12
 100000221 | 0 |  48 |  0 |  12
 100000241 | 0 |   0 |  1 |  11
 100000285 | 0 |  48 |  0 |  48
 100000905 | 0 |  72 |  7 | 149
 100000909 | 0 |  24 |  0 |  36
 100001193 | 0 |  12 |  0 |  12
 100001359 | 0 |  84 |  0 | 300
 100001399 | 0 |  84 |  0 | 168
 100001501 | 0 |   0 |  0 |  12
 100001871 | 0 |  12 |  0 |   0
 100001909 | 0 |  12 |  0 |  24
 100001913 | 0 |   0 |  1 |  11
 100001925 | 0 |  12 |  0 |  24
 100001953 | 0 |   0 |  0 |  12
 100001989 | 0 |  60 |  2 | 202
 100001991 | 0 |  60 |  0 |  36
 100002033 | 0 |  24 |  0 |  96
 100002153 | 0 |  36 |  0 |  36
 100002329 | 0 |  12 |  0 |  12
 100002449 | 0 |  12 |  1 |  47
 100002515 | 0 |   0 |  0 |  12
 100002729 | 0 |   0 |  1 |  11
 100002745 | 2 |  70 |  0 | 156
 100002829 | 0 |  12 |  1 |  11
 100003415 | 2 |  94 | 11 | 133
```

$\pi_i = \frac{a_i+d_i}{n_i}$, $Q_i = \frac{b_i+c_i}{n_i}$, $R_i = \frac{a_i d_i}{n_i}$, $S_i = \frac{b_i c_i}{n_i}$

$$\widehat{OR} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$$

```
Drop table DxDrop;
Select *, Replace(Dx,'36847',' ')
Into DxDrop
From pCalculation1;
```

Select * From DxDrop where Replace like ' ' limit 20;

```
ldsbase=> Select * From DxDrop where Replace like ' ' limit 20;
 identifier |    id     | difference | age |   dx    |    date    | next_date  | a | b | c | d | replace
------------+-----------+------------+-----+---------+------------+------------+---+---+---+---+---------
   48099327 | 488666197 |          0 |   1 | T50992A | 2016-11-07 | 2016-11-07 | 0 | 0 | 0 | 1 |
   19128801 | 483432607 |         33 |   1 | T50992A | 2016-05-19 | 2016-06-21 | 0 | 0 | 0 | 1 |
   47566143 | 488666197 |          0 |   1 | T50992A | 2016-11-07 | 2016-11-07 | 0 | 0 | 0 | 1 |
   46873138 | 488666197 |          1 |   1 | T50992A | 2016-11-11 | 2016-11-12 | 0 | 0 | 0 | 1 |
   25337508 | 188975039 |          1 |   6 | T50992A | 2015-11-25 | 2015-11-26 | 0 | 0 | 0 | 1 |
   47756634 | 488666197 |          0 |   1 | T50992A | 2016-11-11 | 2016-11-11 | 0 | 0 | 0 | 1 |
   47320958 | 488666197 |          0 |   1 | T50992A | 2016-11-07 | 2016-11-07 | 0 | 0 | 0 | 1 |
(7 rows)
```

Select Replace From DxDrop where Replace = 'E133293';

```
ldsbase=> Select Replace From DxDrop where Replace = 'T50992A';
 replace
---------
(0 rows)
```

```
Drop Table GroupedDx;
Select id, sum(a) as a, sum(b) as b, sum(c) as c, sum(d) as d
Into GroupedDx
From (select Identifier, id, min(a) as a, min(b) as b, min(c) as c, min(d) as d From DxDrop Group by Identifier, id) as tmp
Group by id
Order by id;
```

```
ldsbase=> select * from GroupedDx Group by id, a, b, c, d order by id limit 20;
    id     | a | b  | c |  d
-----------+---+----+---+-----
 100000015 | 0 | 12 | 0 |  12
 100000053 | 0 | 24 | 0 |  12
 100000099 | 0 |  0 | 0 |  12
 100000221 | 0 | 48 | 0 |  12
 100000241 | 0 |  0 | 1 |  11
 100000285 | 0 | 48 | 0 |  48
 100000905 | 0 | 72 | 7 | 149
 100000909 | 0 | 24 | 0 |  36
 100001193 | 0 | 12 | 0 |  12
 100001359 | 0 | 84 | 0 | 300
 100001399 | 0 | 84 | 0 | 168
 100001501 | 0 |  0 | 0 |  12
 100001871 | 0 | 12 | 0 |   0
 100001909 | 0 | 12 | 0 |  24
 100001913 | 0 |  0 | 1 |  11
 100001925 | 0 | 12 | 0 |  24
 100001953 | 0 |  0 | 0 |  12
 100001989 | 0 | 60 | 2 | 202
 100001991 | 0 | 60 | 0 |  36
 100002033 | 0 | 24 | 0 |  96
(20 rows)
```

Drop Table Final;
Select a, b, c, d, (a*d) as ad, (b*c) as bc, (a+b+c+d) as ni
Into Final
From GroupedDx;

```
ldsbase=> select * from Final limit 20;
 a | b  | c |  d  | ad | bc  | ni
---+----+---+-----+----+-----+-----
 0 | 12 | 0 |  12 |  0 |   0 |  24
 0 | 24 | 0 |  12 |  0 |   0 |  36
 0 |  0 | 0 |  12 |  0 |   0 |  12
 0 | 48 | 0 |  12 |  0 |   0 |  60
 0 |  0 | 1 |  11 |  0 |   0 |  12
 0 | 48 | 0 |  48 |  0 |   0 |  96
 0 | 72 | 7 | 149 |  0 | 504 | 228
 0 | 24 | 0 |  36 |  0 |   0 |  60
 0 | 12 | 0 |  12 |  0 |   0 |  24
 0 | 84 | 0 | 300 |  0 |   0 | 384
 0 | 84 | 0 | 168 |  0 |   0 | 252
 0 |  0 | 0 |  12 |  0 |   0 |  12
 0 | 12 | 0 |   0 |  0 |   0 |  12
 0 | 12 | 0 |  24 |  0 |   0 |  36
 0 |  0 | 1 |  11 |  0 |   0 |  12
 0 | 12 | 0 |  24 |  0 |   0 |  36
 0 |  0 | 0 |  12 |  0 |   0 |  12
 0 | 60 | 2 | 202 |  0 | 120 | 264
 0 | 60 | 0 |  36 |  0 |   0 |  96
 0 | 24 | 0 |  96 |  0 |   0 | 120
(20 rows)
```

```
Drop Table Final1;
Select
(Round(Cast(ad/ni as decimal),30)) as numerator,
(Round(Cast(bc/ni as decimal),30)) as denominator
Into Final1
From Final;
```

```
ldsbase=> Select(sum(numerator))/(sum(denominator)) as OR From Final1;
                or
---------------------------------
 0.644719299294203292574126471987
(1 row)
```

```
ldsbase=> select * from Final1 limit 20;
            numerator             |            denominator
----------------------------------+----------------------------------
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 2.210526315789473700000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.454545454545454545450000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
 0.000000000000000000000000000000 | 0.000000000000000000000000000000
(20 rows)
```

Select sum(numerator) as N,
sum(denominator) as D
From Final1;

```
ldsbase=> select sum(numerator) as n, sum(denominator) as d from Final1;
                  n                   |                   d
--------------------------------------+--------------------------------------
 68258.824704871648074897960000000000 | 105873.710899606952406728580000000000
(1 row)
```

$$\widehat{OR} = \frac{\sum_i a_i d_i / n_i}{\sum_i b_i c_i / n_i}$$

Select(sum(numerator))/(sum(denominator)) as OR From Final1;

```
ldsbase=> Select(sum(numerator))/(sum(denominator)) as OR From Final1;
                or
-----------------------------------
 0.64471929929420329257412647l987
(1 row)
```

17) Data Dictionary

Below is a list of the data dictionary from CMS LDS dataset from 2014 – 2016 that includes outpatient utilization data.

Variable	Description	Possible Values
Outpatient Base Claim File		
DSYSRTKY	This field contains the key to link data for each beneficiary across all claim files.	
CLAIMNO	The unique number used to identify a unique claim.	

PROVIDER	This variable is the provider identification number. The first two digits indicate the state where the provider is located, using the SSA state codes; the middle two characters indicate the type of provider; and the last two digits are used as a counter for the number of providers within that state and type of provider (i.e., this is a unique but not necessarily sequential number).	
THRU_DT	The last day on the billing statement covering services rendered to the beneficiary (a.k.a 'Statement Covers Thru Date').	
RIC_CD	A code defining the type of claim record being processed.	M = Part B DMEPOS O = Part B physician/supplier U = Both Part A and B institutional HHA V = Part A institutional (IP, SNF, HOS, or HHA) W = Part B institutional claim record (HOP, HHA)

CLM_TYPE	The code used to identify the type of claim record being processed in NCH.	10 = Home Health Agency 20 = Non swing bed SNF 30 = Swing bed SNF 40 = Hospital Outpatient 50 = Hospice 60 = Inpatient 71 = Local carrier non-DMEPOS 72 = Local carrier DMEPOS 81 = Regional carrier non-DMEPOS 82 = Regional carrier DMEPOS
QUERY_CD	Code indicating the type of claim record being processed with respect to payment (debit/credit indicator; interim/final indicator).	0 - Credit adjustment 1 - Interim bill 3 - Final bill 5 - Debit adjustment
FAC_TYPE	The first digit of the type of bill (TOB1) submitted on an institutional claim used to identify the type of facility that provided care to the beneficiary.	1 = Hospital 2 = Skilled Nursing Facility (SNF) 3 = Home Health Agency (HHA) 4 = Religious Non-medical (hospital) 6 = Intermediate Care 7 = Clinic services or hospital-based renal dialysis facility 8 = Ambulatory Surgery Center (ASC) or other

		special facility (e.g. Hospice)
TYPESRVC	The second digit of the type of bill (TOB2) submitted on an institutional claim record to indicate the classification of the type of service provided to the beneficiary.	For facility type codes 1-6: 1 = Inpatient 2 = Inpatient or Home Health (covered on Part B) 3 = Outpatient (or HHA - covered on Part A) 4 = Other (Part B) - Includes HHA medical services 5 = Intermediate Care - Level I 6 = Intermediate Care - Level II 7 = Subacute Inpatient (revenue code 019x required) 8 = Swing Bed For facility type code 7 (clinics): 1 = Rural Health Clinic 2 = Hospital based or indep renal dialysis facility 3 = Free-standing provider

		based FQHC 4 = Other Rehab Facility (ORF) 5 = Comprehensive Rehab Center (CORF) 6 = Community Mental Health Center (CMHC) 7 = Federally Qualified Health Center (FQHC) For facility type code 8 (special facility): 1 = Hospice (non-hospital based) 2 = Hospice (hospital based) 3 = Ambulatory Surgical Center (ASC) in hospital OPT 4 = Freestanding birthing center 5 = Critical Access Hospital - OPT services

FREQ_CD	The third digit of the type of bill (TOB3) submitted on an institutional claim record to indicate the sequence of a claim in the beneficiary's current episode of care.	0 = Non-payment / zero claim 1 = Admit thru discharge claim 2 = Interim - first claim 3 = Interim - continuing claim 4 = Interim - last claim 5 = Late charges only claim 7 = Replacement of prior claim 8 = Void / cancel prior claim 9 = Final claim (HH PPS = process as debit/credit to RAP claim) G = Common Working File (CWF) adjustment claim H = CMS generated adjustment claim I = Misc adjustment claim (from QIO, etc) J = Other adjustment request M = Medicare secondary payer (MSP) adjustment P = Adjustment required by QIO

FI_NUM	The identification number assigned by CMS to a fiscal intermediary authorized to process institutional claim records.	
NOPAY_CD	The reason that no Medicare payment is made for services on an institutional claim.	
PMT_AMT	The Medicare claim payment amount. For hospital services, this amount does not include the claim pass-through per diem payments made by Medicare. To obtain the total amount paid by Medicare for the claim, the pass-through amount (which is the daily per diem amount) must be multiplied by the number of Medicare-covered days (i.e., multiply the CLM_PASS_THRU_PER_DIEM_AMT by the CLM_UTLZTN_DAY_CNT), and then added to the claim payment amount (this field). For non-hospital services (SNF, home health, hospice, and hospital outpatient) and for other non-institutional services (Carrier and DME), this variable equals the total actual Medicare payment amount, and pass-through amounts do not	

	apply. For Part B non-institutional services (Carrier and DME), this variable equals the sum of all the line item-level Medicare payments (variable called the LINE_NCH_PMT_AMT).	
PRPAYAMT	The amount of a payment made on behalf of a Medicare beneficiary by a primary payer other than Medicare, that the provider is applying to covered Medicare charges on a non-institutional claim.	

PRPAY_CD	The code on an institutional claim, specifying a federal non-Medicare program or other source that has primary responsibility for the payment of the Medicare beneficiary's health insurance bills. The presence of a primary payer code indicates that some other payer besides Medicare covered at least some portion of the charges.	A = Working aged bene/spouse with employer group health plan (EGHP) B = End stage renal disease (ESRD) beneficiary in the 18 month coordination period with an EGHP C = Conditional payment by Medicare; future reimbursement expected D = Automobile no-fault E = Worker's Compensation F = Public Health Service or other federal agency (other than Dept of Veterans Affairs) G = Working disabled bene (under age 65 with LGHP) H = Black Lung I = Dept of Veterans Affairs L = Any liability insurance M = Override code: EGHP services involved N = Override code: non-EGHP services involved W = Worker's Compensation Medicare Set-Aside Arrangement (WCMSA) Blank = Medicare is primary payer

PRSTATE	The two position SSA state code where provider facility is located.	
ORGNPINM	On an institutional claim, the National Provider Identifier (NPI) number assigned to uniquely identify the institutional provider certified by Medicare to provide services to the beneficiary.	
SRVC_LOC_NPI_NUM	The National Provider Identifier (NPI) of the location where the services were provided.	
AT_UPIN	On an institutional claim, the unique physician identification number (UPIN) of the physician who would normally be expected to certify and recertify the medical necessity of the services rendered and/or who has primary responsibility for the beneficiary's medical care and treatment (attending physician).	
AT_NPI	On an institutional claim, the national provider identifier (NPI) number assigned to uniquely identify the physician who has overall responsibility for the beneficiary's care and treatment.	

AT_PHYSN_SPCLTY_CD	This variable is the code used to identify the CMS specialty code corresponding to the attending physician.	00 = Carrier wide 01 = General practice 02 = General surgery 03 = Allergy/immunology 04 = Otolaryngology 05 = Anesthesiology 06 = Cardiology 07 = Dermatology 08 = Family practice 09 = Interventional Pain Management (IPM) (eff. 4/1/03) 10 = Gastroenterology 11 = Internal medicine 12 = Osteopathic manipulative therapy 13 = Neurology 14 = Neurosurgery 15 = Speech / language pathology 16 = Obstetrics/gynecology 17 = Hospice and Palliative Care 18 = Ophthalmology 19 = Oral surgery (dentists only) 20 = Orthopedic surgery 21 = Cardiac Electrophysiology 22 = Pathology

		24 = Plastic and reconstructive surgery 25 = Physical medicine and rehabilitation 26 = Psychiatry 27 = General Psychiatry 28 = Colorectal surgery (formerly proctology) 29 = Pulmonary disease 30 = Diagnostic radiology 31 = Intensive cardiac rehabilitation 32 = Anesthesiologist Assistants (eff. 4/1/03—previously grouped with Certified Registered Nurse Anesthetists (CRNA)) 33 = Thoracic surgery 34 = Urology 35 = Chiropractic 36 = Nuclear medicine 37 = Pediatric medicine 38 = Geriatric medicine 39 = Nephrology 40 = Hand surgery 41 = Optometrist 42 = Certified nurse midwife 43 = Certified Registered Nurse Anesthetist (CRNA) (Anesthesiologist Assistants

		were removed from this specialty 4/1/03) 44 = Infectious disease 45 = Mammography screening center 46 = Endocrinology 47 = Independent Diagnostic Testing Facility (IDTF) 48 = Podiatry 49 = Ambulatory surgical center (formerly miscellaneous) 50 = Nurse practitioner 51 = Medical supply company with certified orthotist (certified by American Board for Certification in Prosthetics and Orthotics) 52 = Medical supply company with certified prosthetist (certified by American Board for Certification in Prosthetics and Orthotics) 53 = Medical supply company with certified prosthetist-orthotist (certified by American Board for Certification in

		Prosthetics and Orthotics) 54 = Medical supply company for DMERC (and not included in 51-53) 55 = Individual certified orthotist 56 = Individual certified prosthetist 57 = Individual certified prosthetist-orthotist 58 = Medical supply company with registered pharmacist 59 = Ambulance service supplier, (e.g., private ambulance companies, funeral homes, etc.) 60 = Public health or welfare agencies (federal, state, and local) 61 = Voluntary health or charitable agencies (e.g. National Cancer Society, National Heart Association, Catholic Charities) 62 = Psychologist (billing independently) 63 = Portable X-ray supplier 64 = Audiologist (billing independently)

		65 = Physical therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 66 = Rheumatology 67 = Occupational therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 68 = Clinical psychologist 69 = Clinical laboratory (billing independently) 70 = Multispecialty clinic or group practice 71 = Registered Dietician/Nutrition Professional (eff. 1/1/02) 72 = Pain Management (eff. 1/1/02) 73 = Mass Immunization Roster Biller 74 = Radiation Therapy Centers (prior to 4/2003 this included Independent Diagnostic Testing Facilities (IDTF) 75 = Slide Preparation Facilities (added to differentiate them from Independent Diagnostic

		Testing Facilities (IDTFs -- eff. 4/1/03) 76 = Peripheral vascular disease 77 = Vascular surgery 78 = Cardiac surgery 79 = Addiction medicine 80 = Licensed clinical social worker 81 = Critical care (intensivists) 82 = Hematology 83 = Hematology/oncology 84 = Preventive medicine 85 = Maxillofacial surgery 86 = Neuropsychiatry 87 = All other suppliers (e.g. drug and department stores) 88 = Unknown supplier/provider specialty 89 = Certified clinical nurse specialist 90 = Medical oncology 91 = Surgical oncology 92 = Radiation oncology 93 = Emergency medicine 94 = Interventional radiology 95 = Competitive

		Acquisition Program (CAP) Vendor (eff. 07/01/06). Prior to 07/01/06, known as Independent physiological laboratory 96 = Optician 97 = Physician assistant 98 = Gynecologist/oncologist 99 = Unknown physician specialty A0 = Hospital (DMERCs only) A1 = SNF (DMERCs only) A2 = Intermediate care nursing facility (DMERCs only) A3 = Nursing facility, other (DMERCs only) A4 = Home Health Agency (DMERCs only) A5 = Pharmacy (DMERC) A6 = Medical supply company with respiratory therapist (DMERCs only) A7 = Department store (DMERC) A8 = Grocery store (DMERC)

		A9 = Indian Health Service (IHS), tribe and tribal organizations (non-hospital or non-hospital based facilities, eff. 1/2005) B1 = Supplier of oxygen and/or oxygen related equipment (eff. 10/2/07) B2 = Pedorthic Personnel (eff. 10/2/07) B3 = Medical Supply Company with pedorthic personnel (eff. 10/2/07) B4 = Does not meet definition of health care provider (e.g., Rehabilitation agency, organ procurement organizations, histocompatibility labs) (eff. 10/2/07) B5 = Ocularist C0 = Sleep medicine C1 = Centralized flu C2 = Indirect payment procedure C3 = Interventional cardiology C5 = Dentist (eff. 7/2016)

OP_UPIN	On an institutional claim, the unique physician identification number (UPIN) of the physician who performed the principal procedure. This element is used by the provider to identify the operating physician who performed the surgical procedure.	
OP_NPI	On an institutional claim, the National Provider Identifier (NPI) number assigned to uniquely identify the physician with the primary responsibility for performing the surgical procedure(s).	

OP_PHYSN_SPCLTY_CD	The code used to identify the CMS specialty code corresponding to the operating physician. The Affordable Care Act (ACA) provides for incentive payments for physicians and non-physician practitioners with specific primary specialty designations. In order to determine if the physician or non-physicians is eligible for the incentive payment, the specialty code, NPI and name must be carried on the claims.	00 = Carrier wide 01 = General practice 02 = General surgery 03 = Allergy/immunology 04 = Otolaryngology 05 = Anesthesiology 06 = Cardiology 07 = Dermatology 08 = Family practice 09 = Interventional Pain Management (IPM) (eff. 4/1/03) 10 = Gastroenterology 11 = Internal medicine 12 = Osteopathic manipulative therapy 13 = Neurology 14 = Neurosurgery 15 = Speech / language pathology 16 = Obstetrics/gynecology 17 = Hospice and Palliative Care 18 = Ophthalmology 19 = Oral surgery (dentists only) 20 = Orthopedic surgery 21 = Cardiac Electrophysiology 22 = Pathology

		24 = Plastic and reconstructive surgery 25 = Physical medicine and rehabilitation 26 = Psychiatry 27 = General Psychiatry 28 = Colorectal surgery (formerly proctology) 29 = Pulmonary disease 30 = Diagnostic radiology 31 = Intensive cardiac rehabilitation 32 = Anesthesiologist Assistants (eff. 4/1/03—previously grouped with Certified Registered Nurse Anesthetists (CRNA)) 33 = Thoracic surgery 34 = Urology 35 = Chiropractic 36 = Nuclear medicine 37 = Pediatric medicine 38 = Geriatric medicine 39 = Nephrology 40 = Hand surgery 41 = Optometrist 42 = Certified nurse midwife 43 = Certified Registered Nurse Anesthetist (CRNA) (Anesthesiologist Assistants

		were removed from this specialty 4/1/03) 44 = Infectious disease 45 = Mammography screening center 46 = Endocrinology 47 = Independent Diagnostic Testing Facility (IDTF) 48 = Podiatry 49 = Ambulatory surgical center (formerly miscellaneous) 50 = Nurse practitioner 51 = Medical supply company with certified orthotist (certified by American Board for Certification in Prosthetics and Orthotics) 52 = Medical supply company with certified prosthetist (certified by American Board for Certification in Prosthetics and Orthotics) 53 = Medical supply company with certified prosthetist-orthotist (certified by American Board for Certification in

		Prosthetics and Orthotics) 54 = Medical supply company for DMERC (and not included in 51-53) 55 = Individual certified orthotist 56 = Individual certified prosthetist 57 = Individual certified prosthetist-orthotist 58 = Medical supply company with registered pharmacist 59 = Ambulance service supplier, (e.g., private ambulance companies, funeral homes, etc.) 60 = Public health or welfare agencies (federal, state, and local) 61 = Voluntary health or charitable agencies (e.g. National Cancer Society, National Heart Association, Catholic Charities) 62 = Psychologist (billing independently) 63 = Portable X-ray supplier 64 = Audiologist (billing independently)

		65 = Physical therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 66 = Rheumatology 67 = Occupational therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 68 = Clinical psychologist 69 = Clinical laboratory (billing independently) 70 = Multispecialty clinic or group practice 71 = Registered Dietician/Nutrition Professional (eff. 1/1/02) 72 = Pain Management (eff. 1/1/02) 73 = Mass Immunization Roster Biller 74 = Radiation Therapy Centers (prior to 4/2003 this included Independent Diagnostic Testing Facilities (IDTF) 75 = Slide Preparation Facilities (added to differentiate them from Independent Diagnostic

		Testing Facilities (IDTFs -- eff. 4/1/03) 76 = Peripheral vascular disease 77 = Vascular surgery 78 = Cardiac surgery 79 = Addiction medicine 80 = Licensed clinical social worker 81 = Critical care (intensivists) 82 = Hematology 83 = Hematology/oncology 84 = Preventive medicine 85 = Maxillofacial surgery 86 = Neuropsychiatry 87 = All other suppliers (e.g. drug and department stores) 88 = Unknown supplier/provider specialty 89 = Certified clinical nurse specialist 90 = Medical oncology 91 = Surgical oncology 92 = Radiation oncology 93 = Emergency medicine 94 = Interventional radiology 95 = Competitive

		Acquisition Program (CAP) Vendor (eff. 07/01/06). Prior to 07/01/06, known as Independent physiological laboratory 96 = Optician 97 = Physician assistant 98 = Gynecologist/oncologist 99 = Unknown physician specialty A0 = Hospital (DMERCs only) A1 = SNF (DMERCs only) A2 = Intermediate care nursing facility (DMERCs only) A3 = Nursing facility, other (DMERCs only) A4 = Home Health Agency (DMERCs only) A5 = Pharmacy (DMERC) A6 = Medical supply company with respiratory therapist (DMERCs only) A7 = Department store (DMERC) A8 = Grocery store (DMERC)

		A9 = Indian Health Service (IHS), tribe and tribal organizations (non-hospital or non-hospital based facilities, eff. 1/2005) B1 = Supplier of oxygen and/or oxygen related equipment (eff. 10/2/07) B2 = Pedorthic Personnel (eff. 10/2/07) B3 = Medical Supply Company with pedorthic personnel (eff. 10/2/07) B4 = Does not meet definition of health care provider (e.g., Rehabilitation agency, organ procurement organizations, histocompatibility labs) (eff. 10/2/07) B5 = Ocularist C0 = Sleep medicine C1 = Centralized flu C2 = Indirect payment procedure C3 = Interventional cardiology C5 = Dentist (eff. 7/2016)

OT_UPIN	On an institutional claim, the unique physician identification number (UPIN) of the other physician associated with the institutional claim.	
OT_NPI	On an institutional claim, the National Provider Identifier (NPI) number assigned to uniquely identify the other physician associated with the institutional claim.	

OT_PHYSN_SPCLTY_CD	The code used to identify the CMS specialty code corresponding to the other physician.	00 = Carrier wide 01 = General practice 02 = General surgery 03 = Allergy/immunology 04 = Otolaryngology 05 = Anesthesiology 06 = Cardiology 07 = Dermatology 08 = Family practice 09 = Interventional Pain Management (IPM) (eff. 4/1/03) 10 = Gastroenterology 11 = Internal medicine 12 = Osteopathic manipulative therapy 13 = Neurology 14 = Neurosurgery 15 = Speech / language pathology 16 = Obstetrics/gynecology 17 = Hospice and Palliative Care 18 = Ophthalmology 19 = Oral surgery (dentists only) 20 = Orthopedic surgery 21 = Cardiac Electrophysiology 22 = Pathology

		24 = Plastic and reconstructive surgery 25 = Physical medicine and rehabilitation 26 = Psychiatry 27 = General Psychiatry 28 = Colorectal surgery (formerly proctology) 29 = Pulmonary disease 30 = Diagnostic radiology 31 = Intensive cardiac rehabilitation 32 = Anesthesiologist Assistants (eff. 4/1/03—previously grouped with Certified Registered Nurse Anesthetists (CRNA)) 33 = Thoracic surgery 34 = Urology 35 = Chiropractic 36 = Nuclear medicine 37 = Pediatric medicine 38 = Geriatric medicine 39 = Nephrology 40 = Hand surgery 41 = Optometrist 42 = Certified nurse midwife 43 = Certified Registered Nurse Anesthetist (CRNA) (Anesthesiologist Assistants

		were removed from this specialty 4/1/03) 44 = Infectious disease 45 = Mammography screening center 46 = Endocrinology 47 = Independent Diagnostic Testing Facility (IDTF) 48 = Podiatry 49 = Ambulatory surgical center (formerly miscellaneous) 50 = Nurse practitioner 51 = Medical supply company with certified orthotist (certified by American Board for Certification in Prosthetics and Orthotics) 52 = Medical supply company with certified prosthetist (certified by American Board for Certification in Prosthetics and Orthotics) 53 = Medical supply company with certified prosthetist-orthotist (certified by American Board for Certification in

		Prosthetics and Orthotics) 54 = Medical supply company for DMERC (and not included in 51-53) 55 = Individual certified orthotist 56 = Individual certified prosthetist 57 = Individual certified prosthetist-orthotist 58 = Medical supply company with registered pharmacist 59 = Ambulance service supplier, (e.g., private ambulance companies, funeral homes, etc.) 60 = Public health or welfare agencies (federal, state, and local) 61 = Voluntary health or charitable agencies (e.g. National Cancer Society, National Heart Association, Catholic Charities) 62 = Psychologist (billing independently) 63 = Portable X-ray supplier 64 = Audiologist (billing independently)

		65 = Physical therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 66 = Rheumatology 67 = Occupational therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 68 = Clinical psychologist 69 = Clinical laboratory (billing independently) 70 = Multispecialty clinic or group practice 71 = Registered Dietician/Nutrition Professional (eff. 1/1/02) 72 = Pain Management (eff. 1/1/02) 73 = Mass Immunization Roster Biller 74 = Radiation Therapy Centers (prior to 4/2003 this included Independent Diagnostic Testing Facilities (IDTF) 75 = Slide Preparation Facilities (added to differentiate them from Independent Diagnostic

		Testing Facilities (IDTFs -- eff. 4/1/03) 76 = Peripheral vascular disease 77 = Vascular surgery 78 = Cardiac surgery 79 = Addiction medicine 80 = Licensed clinical social worker 81 = Critical care (intensivists) 82 = Hematology 83 = Hematology/oncology 84 = Preventive medicine 85 = Maxillofacial surgery 86 = Neuropsychiatry 87 = All other suppliers (e.g. drug and department stores) 88 = Unknown supplier/provider specialty 89 = Certified clinical nurse specialist 90 = Medical oncology 91 = Surgical oncology 92 = Radiation oncology 93 = Emergency medicine 94 = Interventional radiology 95 = Competitive

		Acquisition Program (CAP) Vendor (eff. 07/01/06). Prior to 07/01/06, known as Independent physiological laboratory 96 = Optician 97 = Physician assistant 98 = Gynecologist/oncologist 99 = Unknown physician specialty A0 = Hospital (DMERCs only) A1 = SNF (DMERCs only) A2 = Intermediate care nursing facility (DMERCs only) A3 = Nursing facility, other (DMERCs only) A4 = Home Health Agency (DMERCs only) A5 = Pharmacy (DMERC) A6 = Medical supply company with respiratory therapist (DMERCs only) A7 = Department store (DMERC) A8 = Grocery store (DMERC)

		A9 = Indian Health Service (IHS), tribe and tribal organizations (non-hospital or non-hospital based facilities, eff. 1/2005) B1 = Supplier of oxygen and/or oxygen related equipment (eff. 10/2/07) B2 = Pedorthic Personnel (eff. 10/2/07) B3 = Medical Supply Company with pedorthic personnel (eff. 10/2/07) B4 = Does not meet definition of health care provider (e.g., Rehabilitation agency, organ procurement organizations, histocompatibility labs) (eff. 10/2/07) B5 = Ocularist C0 = Sleep medicine C1 = Centralized flu C2 = Indirect payment procedure C3 = Interventional cardiology C5 = Dentist (eff. 7/2016)

RNDRNG_PHYSN_NPI	This variable is the National Provider Identifier (NPI) for the physician who rendered the services. NPIs replaced UPINs as the standard provider identifiers beginning in 2007. The UPIN is almost never populated after 2009.	
RNDRNG_PHYSN_SPCLTY_CD	The code used to identify the CMS specialty code of the rendering physician/practitioner.	00 = Carrier wide 01 = General practice 02 = General surgery 03 = Allergy/immunology 04 = Otolaryngology 05 = Anesthesiology 06 = Cardiology 07 = Dermatology 08 = Family practice 09 = Interventional Pain Management (IPM) (eff. 4/1/03) 10 = Gastroenterology 11 = Internal medicine 12 = Osteopathic manipulative therapy 13 = Neurology 14 = Neurosurgery 15 = Speech / language pathology 16 = Obstetrics/gynecology 17 = Hospice and

		Palliative Care 18 = Ophthalmology 19 = Oral surgery (dentists only) 20 = Orthopedic surgery 21 = Cardiac Electrophysiology 22 = Pathology 24 = Plastic and reconstructive surgery 25 = Physical medicine and rehabilitation 26 = Psychiatry 27 = General Psychiatry 28 = Colorectal surgery (formerly proctology) 29 = Pulmonary disease 30 = Diagnostic radiology 31 = Intensive cardiac rehabilitation 32 = Anesthesiologist Assistants (eff. 4/1/03—previously grouped with Certified Registered Nurse Anesthetists (CRNA)) 33 = Thoracic surgery 34 = Urology 35 = Chiropractic 36 = Nuclear medicine 37 = Pediatric medicine 38 = Geriatric medicine

		39 = Nephrology 40 = Hand surgery 41 = Optometrist 42 = Certified nurse midwife 43 = Certified Registered Nurse Anesthetist (CRNA) (Anesthesiologist Assistants were removed from this specialty 4/1/03) 44 = Infectious disease 45 = Mammography screening center 46 = Endocrinology 47 = Independent Diagnostic Testing Facility (IDTF) 48 = Podiatry 49 = Ambulatory surgical center (formerly miscellaneous) 50 = Nurse practitioner 51 = Medical supply company with certified orthotist (certified by American Board for Certification in Prosthetics and Orthotics) 52 = Medical supply company with certified prosthetist (certified by

		American Board for Certification in Prosthetics and Orthotics) 53 = Medical supply company with certified prosthetist-orthotist (certified by American Board for Certification in Prosthetics and Orthotics) 54 = Medical supply company for DMERC (and not included in 51-53) 55 = Individual certified orthotist 56 = Individual certified prosthetist 57 = Individual certified prosthetist-orthotist 58 = Medical supply company with registered pharmacist 59 = Ambulance service supplier, (e.g., private ambulance companies, funeral homes, etc.) 60 = Public health or welfare agencies (federal, state, and local) 61 = Voluntary health or charitable agencies (e.g. National Cancer Society,

		National Heart Association, Catholic Charities) 62 = Psychologist (billing independently) 63 = Portable X-ray supplier 64 = Audiologist (billing independently) 65 = Physical therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 66 = Rheumatology 67 = Occupational therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 68 = Clinical psychologist 69 = Clinical laboratory (billing independently) 70 = Multispecialty clinic or group practice 71 = Registered Dietician/Nutrition Professional (eff. 1/1/02) 72 = Pain Management (eff. 1/1/02) 73 = Mass Immunization Roster Biller 74 = Radiation Therapy

		Centers (prior to 4/2003 this included Independent Diagnostic Testing Facilities (IDTF) 75 = Slide Preparation Facilities (added to differentiate them from Independent Diagnostic Testing Facilities (IDTFs -- eff. 4/1/03) 76 = Peripheral vascular disease 77 = Vascular surgery 78 = Cardiac surgery 79 = Addiction medicine 80 = Licensed clinical social worker 81 = Critical care (intensivists) 82 = Hematology 83 = Hematology/oncology 84 = Preventive medicine 85 = Maxillofacial surgery 86 = Neuropsychiatry 87 = All other suppliers (e.g. drug and department stores) 88 = Unknown supplier/provider specialty 89 = Certified clinical

		nurse specialist 90 = Medical oncology 91 = Surgical oncology 92 = Radiation oncology 93 = Emergency medicine 94 = Interventional radiology 95 = Competitive Acquisition Program (CAP) Vendor (eff. 07/01/06). Prior to 07/01/06, known as Independent physiological laboratory 96 = Optician 97 = Physician assistant 98 = Gynecologist/oncologist 99 = Unknown physician specialty A0 = Hospital (DMERCs only) A1 = SNF (DMERCs only) A2 = Intermediate care nursing facility (DMERCs only) A3 = Nursing facility, other (DMERCs only) A4 = Home Health Agency (DMERCs only) A5 = Pharmacy

		(DMERC) A6 = Medical supply company with respiratory therapist (DMERCs only) A7 = Department store (DMERC) A8 = Grocery store (DMERC) A9 = Indian Health Service (IHS), tribe and tribal organizations (non-hospital or non-hospital based facilities, eff. 1/2005) B1 = Supplier of oxygen and/or oxygen related equipment (eff. 10/2/07) B2 = Pedorthic Personnel (eff. 10/2/07) B3 = Medical Supply Company with pedorthic personnel (eff. 10/2/07) B4 = Does not meet definition of health care provider (e.g., Rehabilitation agency, organ procurement organizations, histocompatibility labs) (eff. 10/2/07) B5 = Ocularist C0 = Sleep medicine C1 = Centralized flu

		C2 = Indirect payment procedure C3 = Interventional cardiology C5 = Dentist (eff. 7/2016)
RFR_PHYSN_NPI	The national provider identifier (NPI) number assigned to uniquely identify the referring physician.	
RFR_PHYSN_SPCLTY_CD	The code used to identify the CMS specialty code of the referring physician/practitioner.	00 = Carrier wide 01 = General practice 02 = General surgery 03 = Allergy/immunology 04 = Otolaryngology 05 = Anesthesiology 06 = Cardiology 07 = Dermatology 08 = Family practice 09 = Interventional Pain Management (IPM) (eff. 4/1/03) 10 = Gastroenterology 11 = Internal medicine 12 = Osteopathic manipulative therapy

		13 = Neurology 14 = Neurosurgery 15 = Speech / language pathology 16 = Obstetrics/gynecology 17 = Hospice and Palliative Care 18 = Ophthalmology 19 = Oral surgery (dentists only) 20 = Orthopedic surgery 21 = Cardiac Electrophysiology 22 = Pathology 24 = Plastic and reconstructive surgery 25 = Physical medicine and rehabilitation 26 = Psychiatry 27 = General Psychiatry 28 = Colorectal surgery (formerly proctology) 29 = Pulmonary disease 30 = Diagnostic radiology 31 = Intensive cardiac rehabilitation 32 = Anesthesiologist Assistants (eff. 4/1/03—previously grouped with Certified Registered Nurse

		Anesthetists (CRNA)) 33 = Thoracic surgery 34 = Urology 35 = Chiropractic 36 = Nuclear medicine 37 = Pediatric medicine 38 = Geriatric medicine 39 = Nephrology 40 = Hand surgery 41 = Optometrist 42 = Certified nurse midwife 43 = Certified Registered Nurse Anesthetist (CRNA) (Anesthesiologist Assistants were removed from this specialty 4/1/03) 44 = Infectious disease 45 = Mammography screening center 46 = Endocrinology 47 = Independent Diagnostic Testing Facility (IDTF) 48 = Podiatry 49 = Ambulatory surgical center (formerly miscellaneous) 50 = Nurse practitioner 51 = Medical supply company with certified

		orthotist (certified by American Board for Certification in Prosthetics and Orthotics) 52 = Medical supply company with certified prosthetist (certified by American Board for Certification in Prosthetics and Orthotics) 53 = Medical supply company with certified prosthetist-orthotist (certified by American Board for Certification in Prosthetics and Orthotics) 54 = Medical supply company for DMERC (and not included in 51-53) 55 = Individual certified orthotist 56 = Individual certified prosthetist 57 = Individual certified prosthetist-orthotist 58 = Medical supply company with registered pharmacist 59 = Ambulance service supplier, (e.g., private ambulance companies,

		funeral homes, etc.) 60 = Public health or welfare agencies (federal, state, and local) 61 = Voluntary health or charitable agencies (e.g. National Cancer Society, National Heart Association, Catholic Charities) 62 = Psychologist (billing independently) 63 = Portable X-ray supplier 64 = Audiologist (billing independently) 65 = Physical therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 66 = Rheumatology 67 = Occupational therapist (private practice added 4/1/03) (independently practicing removed 4/1/03) 68 = Clinical psychologist 69 = Clinical laboratory (billing independently) 70 = Multispecialty clinic or group practice 71 = Registered

		Dietician/Nutrition Professional (eff. 1/1/02) 72 = Pain Management (eff. 1/1/02) 73 = Mass Immunization Roster Biller 74 = Radiation Therapy Centers (prior to 4/2003 this included Independent Diagnostic Testing Facilities (IDTF) 75 = Slide Preparation Facilities (added to differentiate them from Independent Diagnostic Testing Facilities (IDTFs -- eff. 4/1/03) 76 = Peripheral vascular disease 77 = Vascular surgery 78 = Cardiac surgery 79 = Addiction medicine 80 = Licensed clinical social worker 81 = Critical care (intensivists) 82 = Hematology 83 = Hematology/oncology 84 = Preventive medicine 85 = Maxillofacial surgery

		86 = Neuropsychiatry 87 = All other suppliers (e.g. drug and department stores) 88 = Unknown supplier/provider specialty 89 = Certified clinical nurse specialist 90 = Medical oncology 91 = Surgical oncology 92 = Radiation oncology 93 = Emergency medicine 94 = Interventional radiology 95 = Competitive Acquisition Program (CAP) Vendor (eff. 07/01/06). Prior to 07/01/06, known as Independent physiological laboratory 96 = Optician 97 = Physician assistant 98 = Gynecologist/oncologist 99 = Unknown physician specialty A0 = Hospital (DMERCs only) A1 = SNF (DMERCs only) A2 = Intermediate care

		nursing facility (DMERCs only) A3 = Nursing facility, other (DMERCs only) A4 = Home Health Agency (DMERCs only) A5 = Pharmacy (DMERC) A6 = Medical supply company with respiratory therapist (DMERCs only) A7 = Department store (DMERC) A8 = Grocery store (DMERC) A9 = Indian Health Service (IHS), tribe and tribal organizations (non-hospital or non-hospital based facilities, eff. 1/2005) B1 = Supplier of oxygen and/or oxygen related equipment (eff. 10/2/07) B2 = Pedorthic Personnel (eff. 10/2/07) B3 = Medical Supply Company with pedorthic personnel (eff. 10/2/07) B4 = Does not meet definition of health care provider (e.g., Rehabilitation

		agency, organ procurement organizations, histocompatibility labs) (eff. 10/2/07) B5 = Ocularist C0 = Sleep medicine C1 = Centralized flu C2 = Indirect payment procedure C3 = Interventional cardiology C5 = Dentist (eff. 7/2016)
MCOPDSW	A switch indicating whether or not a Managed Care Organization (MCO) has paid the provider for an institutional claim.	Blank = MCO has not paid the provider 0 = MCO has not paid the provider 1 = MCO has paid the provider for a claim

STUS_CD	The code used to identify the status of the patient as of the CLM_THRU_DT.	0 = Unknown value (but present in data) 01 = Discharged to home/self care 02 = Discharged / transferred to short term hospital 03 = Discharged / transferred to SNF 04 = Discharged / transferred to intermediate care 05 = Discharged / transferred to other IPT care 06 = Discharged / transferred to HHA home care 07 = Left against medical advice or discontinue care 08 = Discharged / transferred to home IV drug care 09 = Admitted as an inpatient to hospital after OPT 20 = Expired (did not recover - Christian Science) 21 = Discharged / transferred to court /law enforce 30 = Still a patient

		40 = Expired at home (hospice claims only) 41 = Expired in facility (hospice claims only) 42 = Expired place unknown (hospice claims only) 43 = Discharged / transferred to federal hospital 50 = Hospice - home 51 = Hospice - medical facility 61 = Discharged / transferred to swing bed internally 62 = Discharged / transferred to IPT Rehab 63 = Discharged / transferred to to LTC 64 = Discharged / transferred to Medicaid facility 65 = Discharged / transferred to Psychiatric Hospital 66 = Discharged / transferred to CAH 70 = Discharged / transferred to other misc facility

		71 = Discharged / transferred to other OPT services 72 = Discharged / transferred internally for OPT svcs
TOT_CHRG	The total charges for all services included on the institutional claim. This field is redundant with revenue center code 0001/total charges.	

BLDDEDAM	The amount of money for which the intermediary determined the beneficiary is liable for the blood deductible.	
PCCHGAMT	The amount of physician and other professional charges covered under Medicare Part B. For IP claims, this amount is not reflected in any of the other Part A claim fields (i.e., it is not a portion of the Medicare payment for the hospitalization).	
PRNCPAL_DGNS_CD	The diagnosis code identifying the diagnosis, condition, problem or other reason for the admission/encounter/visit shown in the medical record to be chiefly responsible for the services provided. This data is also redundantly stored as the first occurrence of the diagnosis code (variable called ICD_DGNS_CD1).	
ICD DGNS CD1 to CD25	The diagnosis code identifying the beneficiary's principal or other diagnosis (including E code).	

FST_DGNS_E_CD	The code used to identify the first external cause of injury, poisoning, or other adverse effect. This diagnosis E code is also stored as the first occurrence of the diagnosis E code trailer.	
ICD DGNS E CD1 to CD12	The code used to identify the external cause of injury, poisoning, or other adverse affect.	
ICD_PRCDR_CD1 to CD12	The code that indicates the principal or other procedure performed during the period covered by the institutional claim.	
PRCDR_DT1 to DT25	On an institutional claim, the date on which the principal or other procedure was performed.	
RSN_VISIT_CD1 to CD3	The diagnosis code used to identify the patient's reason for the Hospital Outpatient visit.	
PTB_DED	The amount of money for which the intermediary or carrier has determined that the beneficiary is liable for the Part B cash deductible on the claim.	
PTB_COIN	The amount of money for which the intermediary has determined that the beneficiary is liable for Part B coinsurance on the institutional claim.	
PRVDRPMT	The amount paid, from the Medicare Trust Fund, to the provider for the services reported on the Outpatient claim.	

BENEPMT	The total payments made, from the Medicare Trust Fund, to the beneficiary for the services reported on the Outpatient claim (sum of line payment amounts to the beneficiary.)	
DOB_DT	The beneficiary's date of birth, coded as a range.	0 = Unknown 1 = <65 2 = 65 Thru 69 3 = 70 Thru 74 4 = 75 Thru 79 5 = 80 Thru 84 6 = >84
GNDR_CD	The sex of a beneficiary.	0 = Unknown 1 = Male 2 = Female
RACE_CD	The race of a beneficiary.	0 = Unknown 1 = White 2 = Black 3 = Other 4 = Asian 5 = Hispanic 6 = North American Native
CNTY_CD	The 3-digit SSA standard county code of a beneficiary's residence.	

STATE_CD	The 2-digit SSA standard state code of a beneficiary's residence.	
CWF_BENE_MDCR_STUS_CD	The CWF-derived reason for a beneficiary's entitlement to Medicare benefits, as of the reference date (CLM_THRU_DT).	10 = Aged without ESRD 11 = Aged with ESRD 20 = Disabled without ESRD 21 = Disabled with ESRD 31 = ESRD only
ACTIONCD	The type of action requested by the intermediary to be taken on an institutional claim.	1 = Original debit action 5 = Force action code 3 (secondary debit adjustment) 8 = Benefits refused
BLDFRNSH	Number of whole pints of blood furnished to the beneficiary, as reported on the carrier claim (non-DMERC).	
CLM_TRTMT_AUTHRZTN_NUM	The number assigned by the medical reviewer and reported by the provider to identify the medical review (treatment authorization) action taken after review of the beneficiary's case. It designates that treatment covered by the bill has been authorized by the payer.	

CLM_PRCR_RTRN_CD	The code used to identify various prospective payment system (PPS) payment adjustment types. This code identifies the payment return code or the error return code for every claim type calculated by the PRICER tool.	The meaning of the values varies by type of bill (TOB) **********TOB 81X or 82X********** Hospice Payment Return Codes: 00 = Home rate returned Hospice Error Return Codes: 10 = Bad units 20 = Bad units2 < 8 30 = Bad MSA code 40 = Bad hospice wage index from MSA file 50 = Bad bene wage index from MSA file 51 = Bad provider number
CLM_OP_TRANS_TYPE_CD	The code derived by CMS based on the type of bill and provider number to identify the outpatient transaction type.	A = Outpatient Psychiatric Hospital B = Outpatient tuberculosis (TB) Hospital C = Outpatient General Care Hospital D = Outpatient Skilled Nursing Facility (SNF) E = Home Health Agency F = Comprehensive Health Care G = Clinical Rehab Agency H = Rural Health Clinic I = Satellite Dialysis Facility

		J = Limited Care Facility 0 = Christian Science SNF 1 = Psychiatric Hospital Facility 2 = TB Hospital Facility 3 = General Care Hospital 4 = Regular SNF Spaces = Home Health/Hospice
CLM_OP_ESRD_MTHD_CD	This variable contains the code denoting the method of reimbursement selected by the beneficiary receiving End Stage Renal Disease (ESRD) services for home dialysis (i.e. whether home supplies are purchased through a facility or from a supplier.)	0 = Not ESRD 1 = Method 1 - Home supplies purchased through a facility 2 = Method 2 - Home supplies purchased from a supplier
CLM_NEXT_GNRTN_ACO_IND_CD1-5	The field identifies the claims that qualify for specific claims processing edits related to benefit enhancement through the Next Generation (NG) Accountable Care Organization (ACO).	0 = Base record (no enhancements) 1 = Population Based Payments (PBP) 2 = Telehealth 3 = Post Discharge Home Health Visits 4 = 3-Day SNF Waiver 5 = Capitation
ACO_ID_NUM	The field identifies the Accountable Care Organization (ACO) Identification Number.	

```
ldsbase=> select distinct(dx), lr from LRK order by lr desc;
   dx    |           lr
---------+-----------------------
 T56894A |   208.9010416666666667
 S79819A |   165.9218750000000000
 7105    |    43.2573289902280130
 T83518A |    35.9843750000000000
 T8241XA |    34.6493055555555556
 T80211D |    33.9843750000000000
 T5694XA |    33.5125977410947003
 T8249XD |    31.9843750000000000
 T83028A |    30.9843750000000000
 T85611A |    30.4843750000000000
 M834    |    26.9843750000000000
 S72351D |    23.9878472222222222
 M10361  |    20.9895833333333333
 T82318A |    18.9895833333333333
 S31125A |    17.9895833333333333
 D631    |    17.2508438535231444
 B9621   |    15.9947916666666667
 C8520   |    15.9947916666666667
 E133293 |    15.9947916666666667
 S14104D |    15.9947916666666667
 T50992A |    15.9947916666666667
 T83510D |    15.9947916666666667
 Z578    |    15.7003647733194372
 E7209   |    14.9947916666666667
 Y846    |    14.3927083333333333
 C8387   |    13.9947916666666667
 M4980   |    13.9947916666666667
 T508X5D |    13.9947916666666667
```

18) Optimal Revisit Intervals

The table below provides the comprehensive list of the comorbidities and combination of comorbidities that have been identified as having the highest risk factors for CKD for Medicare patients with Type II Diabetes. Each row represents an individual stratum based on the patients comorbidity and demographics.

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.077669903	305	306
Dx1740	0.077669903	305	306
Dx74332	0.077669903	305	306
T83510d	0.077669903	305	306
S79819A	0.084210526	281	283
Dx1740	0.084210526	281	283
Dx74332	0.084210526	281	283
T83510d	0.084210526	281	283
S79819A	0.089219331	265	266
Dx1740	0.089219331	265	266
Dx74332	0.089219331	265	266
T83510d	0.089219331	265	266
S79819A	0.091254753	259	269
Dx1740	0.091254753	259	269
Dx74332	0.091254753	259	269
T83510d	0.091254753	259	269
S79819A	0.093023256	254	256
Dx1740	0.093023256	254	256
Dx74332	0.093023256	254	256
T83510d	0.093023256	254	256
S79819A	0.09375	252	253
Dx1740	0.09375	252	253
Dx74332	0.09375	252	253
T83510d	0.09375	252	253
S79819A	0.096	246	252
Dx1740	0.096	246	252
Dx74332	0.096	246	252
T83510d	0.096	246	252
S79819A	0.10041841	235	265
Dx1740	0.10041841	235	265
Dx74332	0.10041841	235	265
T83510d	0.10041841	235	265
S79819A	0.102564103	230	239

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.102564103	230	239
Dx74332	0.102564103	230	239
T83510d b	0.102564103	230	239
S79819A	0.103004292	229	259
Dx1740	0.103004292	229	259
Dx74332	0.103004292	229	259
T83510d	0.103004292	229	259
Dx17311	0.15720524	213	218
Dx1740	0.15720524	213	218
Dx74332	0.15720524	213	218
Z578	0.15720524	213	218
S79819A	0.105263158	224	300
Dx1740	0.105263158	224	300
Dx74332	0.105263158	224	300
T83510d	0.105263158	224	300
S79819A	0.10619469	222	225
Dx1740	0.10619469	222	225
Dx74332	0.10619469	222	225
T83510d	0.10619469	222	225
S79819A	0.107142857	220	282
Dx1740	0.107142857	220	282
Dx74332	0.107142857	220	282
T83510d	0.107142857	220	282
Dx17311	0.162162162	206	287
Dx1740	0.162162162	206	287
Dx74332	0.162162162	206	287
Z578	0.162162162	206	287
S79819A	0.108108108	218	225
Dx1740	0.108108108	218	225
Dx74332	0.108108108	218	225
T83510d	0.108108108	218	225
S79819A	0.108597285	217	262
Dx1740	0.108597285	217	262
Dx74332	0.108597285	217	262
T83510d	0.108597285	217	262
S79819A	0.109589041	215	221
Dx1740	0.109589041	215	221
Dx74332	0.109589041	215	221
T83510d	0.109589041	215	221
S79819A	0.110091743	214	216

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.110091743	214	216
Dx74332	0.110091743	214	216
T83510d	0.110091743	214	216
Dx17311	0.16744186	199	218
Dx1740	0.16744186	199	218
Dx74332	0.16744186	199	218
Z578	0.16744186	199	218
S79819A	0.111627907	211	239
Dx1740	0.111627907	211	239
Dx74332	0.111627907	211	239
T83510d	0.111627907	211	239
S79819A	0.112149533	210	212
Dx1740	0.112149533	210	212
Dx74332	0.112149533	210	212
T83510d	0.112149533	210	212
S79819A	0.112676056	209	210
Dx1740	0.112676056	209	210
Dx74332	0.112676056	209	210
T83510d	0.112676056	209	210
S79819A	0.113207547	208	303
Dx1740	0.113207547	208	303
Dx74332	0.113207547	208	303
T83510d	0.113207547	208	303
S79819A	0.113744076	207	238
Dx1740	0.113744076	207	238
Dx74332	0.113744076	207	238
T83510d	0.113744076	207	238
S79819A	0.114285714	206	210
Dx1740	0.114285714	206	210
Dx74332	0.114285714	206	210
T83510d	0.114285714	206	210
S79819A	0.114832536	205	220
Dx1740	0.114832536	205	220
Dx74332	0.114832536	205	220
T83510d	0.114832536	205	220
S79819A	0.115942029	203	226
Dx1740	0.115942029	203	226
Dx74332	0.115942029	203	226
T83510d	0.115942029	203	226
Dx17311	0.175609756	189	198

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.175609756	189	198
Dx74332	0.175609756	189	198
Z578	0.175609756	189	198
S79819A	0.117073171	201	202
Dx1740	0.117073171	201	202
Dx74332	0.117073171	201	202
T83510d	0.117073171	201	202
S79819A	0.117647059	200	298
Dx1740	0.117647059	200	298
Dx74332	0.117647059	200	298
T83510d	0.117647059	200	298
Dx17311	0.177339901	187	198
Dx1740	0.177339901	187	198
Dx74332	0.177339901	187	198
Z578	0.177339901	187	198
S79819A	0.118226601	199	206
Dx1740	0.118226601	199	206
Dx74332	0.118226601	199	206
T83510d	0.118226601	199	206
Dx17311	0.179104478	185	204
Dx1740	0.179104478	185	204
Dx74332	0.179104478	185	204
Z578	0.179104478	185	204
S79819A	0.12	196	197
Dx1740	0.12	196	197
Dx74332	0.12	196	197
T83510d	0.12	196	197
Dx17311	0.180904523	183	213
Dx1740	0.180904523	183	213
Dx74332	0.180904523	183	213
Z578	0.180904523	183	213
S79819A	0.120603015	195	232
Dx1740	0.120603015	195	232
Dx74332	0.120603015	195	232
T83510d	0.120603015	195	232
Dx17311	0.181818182	182	187
Dx1740	0.181818182	182	187
Dx74332	0.181818182	182	187
Z578	0.181818182	182	187
S79819A	0.121212121	194	211

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.121212121	194	211
Dx74332	0.121212121	194	211
T83510d	0.121212121	194	211
Dx17311	0.182741117	181	183
Dx1740	0.182741117	181	183
Dx74332	0.182741117	181	183
Z578	0.182741117	181	183
S79819A, T83510D	0.12	193	200
S79819A	0.121827411	193	195
Dx1740	0.121827411	193	195
Dx74332	0.121827411	193	195
T83510d	0.121827411	193	195
S79819A	0.12244898	192	193
Dx1740	0.12244898	192	193
Dx74332	0.12244898	192	193
T83510d	0.12244898	192	193
Dx17311	0.184615385	179	180
Dx1740	0.184615385	179	180
Dx74332	0.184615385	179	180
Z578	0.184615385	179	180
S79819A	0.123076923	191	200
Dx1740	0.123076923	191	200
Dx74332	0.123076923	191	200
T83510d	0.123076923	191	200
S79819A	0.12371134	190	315
Dx1740	0.12371134	190	315
Dx74332	0.12371134	190	315
T83510d	0.12371134	190	315
S79819A	0.124352332	189	191
Dx1740	0.124352332	189	191
Dx74332	0.124352332	189	191
T83510d	0.124352332	189	191
S79819A	0.125	188	190
Dx1740	0.125	188	190
Dx74332	0.125	188	190
T83510d	0.125	188	190
S79819A	0.12565445	187	189
Dx1740	0.12565445	187	189
Dx74332	0.12565445	187	189
T83510d	0.12565445	187	189

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.126315789	186	187
Dx1740	0.126315789	186	187
Dx74332	0.126315789	186	187
T83510d	0.126315789	186	187
S79819A	0.126984127	185	191
Dx1740	0.126984127	185	191
Dx74332	0.126984127	185	191
T83510d	0.126984127	185	191
S79819A	0.127659574	184	185
Dx1740	0.127659574	184	185
Dx74332	0.127659574	184	185
T83510d	0.127659574	184	185
Dx17311	0.192513369	171	184
Dx1740	0.192513369	171	184
Dx74332	0.192513369	171	184
Z578	0.192513369	171	184
S79819A	0.128342246	183	197
Dx1740	0.128342246	183	197
Dx74332	0.128342246	183	197
T83510d	0.128342246	183	197
S79819A, T83510D	0.122994652	184	270
Dx17311, Z578	0.122994652	184	270
Dx36842	0.122994652	184	270
T80211d	0.122994652	184	270
Z578	0.122994652	184	270
Dx17311, Z578	0.258064516	158	208
Dx1740	0.258064516	158	208
s14104d	0.258064516	158	208
Z578	0.258064516	158	208
S79819A	0.129032258	182	183
Dx1740	0.129032258	182	183
Dx74332	0.129032258	182	183
T83510d	0.129032258	182	183
S79819A, T83510D	0.11827957	184	185
Dx17311, Z578	0.11827957	184	185
Dx36842	0.11827957	184	185
T80211d	0.11827957	184	185
Z578	0.11827957	184	185
S79819A	0.12972973	181	182
Dx1740	0.12972973	181	182

Covariates	Probability	Minimum RVI	Maximum RVI
Dx74332	0.12972973	181	182
T83510d	0.12972973	181	182
S79819A, T83510D	0.124324324	182	189
Dx17311, Z578	0.124324324	182	189
Dx36842	0.124324324	182	189
T80211d	0.124324324	182	189
Z578	0.124324324	182	189
S79819A, T83510D	0.13	180	188
S79819A	0.130434783	180	181
Dx1740	0.130434783	180	181
Dx74332	0.130434783	180	181
T83510d	0.130434783	180	181
S79819A	0.131147541	179	189
Dx1740	0.131147541	179	189
Dx74332	0.131147541	179	189
T83510d	0.131147541	179	189
Dx17311	0.197802198	166	180
Dx1740	0.197802198	166	180
Dx74332	0.197802198	166	180
Z578	0.197802198	166	180
S79819A	0.131868132	178	182
Dx1740	0.131868132	178	182
Dx74332	0.131868132	178	182
T83510d	0.131868132	178	182
Dx17311	0.198895028	165	175
Dx1740	0.198895028	165	175
Dx74332	0.198895028	165	175
Z578	0.198895028	165	175
S79819A	0.132596685	177	181
Dx1740	0.132596685	177	181
Dx74332	0.132596685	177	181
T83510d	0.132596685	177	181
S79819A	0.133333333	176	181
Dx1740	0.133333333	176	181
Dx74332	0.133333333	176	181
T83510d	0.133333333	176	181
S79819A	0.134078212	175	176
Dx1740	0.134078212	175	176
Dx74332	0.134078212	175	176
T83510d	0.134078212	175	176

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.134831461	174	178
Dx1740	0.134831461	174	178
Dx74332	0.134831461	174	178
T83510d	0.134831461	174	178
Dx17311	0.203389831	161	328
Dx1740	0.203389831	161	328
Dx74332	0.203389831	161	328
Z578	0.203389831	161	328
S79819A	0.13559322	173	179
Dx1740	0.13559322	173	179
Dx74332	0.13559322	173	179
T83510d	0.13559322	173	179
Dx17311, Z578	0.596590909	91	246
Dx36842	0.596590909	91	246
T50992a	0.596590909	91	246
Z578	0.596590909	91	246
S79819A	0.136363636	172	197
Dx1740	0.136363636	172	197
Dx74332	0.136363636	172	197
T83510d	0.136363636	172	197
Dx17311, Z578	0.342857143	135	194
Dx1740	0.342857143	135	194
T381x5a	0.342857143	135	194
Z578	0.342857143	135	194
S79819A	0.137142857	171	173
Dx1740	0.137142857	171	173
Dx74332	0.137142857	171	173
T83510d	0.137142857	171	173
S79819A	0.137931034	170	314
Dx1740	0.137931034	170	314
Dx74332	0.137931034	170	314
T83510d	0.137931034	170	314
S79819A	0.138728324	169	171
Dx1740	0.138728324	169	171
Dx74332	0.138728324	169	171
T83510d	0.138728324	169	171
S79819A, T83510D	0.132947977	170	272
Dx17311, Z578	0.132947977	170	272
Dx36842	0.132947977	170	272
T80211d	0.132947977	170	272

Covariates	Probability	Minimum RVI	Maximum RVI
Z578	0.132947977	170	272
S79819A	0.139534884	168	170
Dx1740	0.139534884	168	170
Dx74332	0.139534884	168	170
T83510d	0.139534884	168	170
Dx17311, Z578	0.350877193	131	184
Dx1740	0.350877193	131	184
T381x5a	0.350877193	131	184
Z578	0.350877193	131	184
S79819A	0.140350877	167	169
Dx1740	0.140350877	167	169
Dx74332	0.140350877	167	169
T83510d	0.140350877	167	169
Dx17311	0.213017751	153	197
Dx1740	0.213017751	153	197
Dx74332	0.213017751	153	197
Z578	0.213017751	153	197
S79819A	0.142011834	165	168
Dx1740	0.142011834	165	168
Dx74332	0.142011834	165	168
T83510d	0.142011834	165	168
Dx17311, Z578	0.571428571	92	189
Dx36842	0.571428571	92	189
T381x5a	0.571428571	92	189
Z578	0.571428571	92	189
S79819A	0.142857143	164	173
Dx1740	0.142857143	164	173
Dx74332	0.142857143	164	173
T83510d	0.142857143	164	173
Dx17311, Z578	0.28742515	139	194
Dx1740	0.28742515	139	194
s14104d	0.28742515	139	194
Z578	0.28742515	139	194
Dx17311, Z578	0.361445783	126	181
Dx1740	0.361445783	126	181
T381x5a	0.361445783	126	181
Z578	0.361445783	126	181
Dx17311	0.21686747	150	240
Dx1740	0.21686747	150	240
Dx74332	0.21686747	150	240

Covariates	Probability	Minimum RVI	Maximum RVI
Z578	0.21686747	150	240
S79819A, T83510D	0.012048193	184	185
Dx17311, Z578	0.012048193	184	185
Dx36842	0.012048193	184	185
T80211d	0.012048193	184	185
Z578	0.012048193	184	185
S79819A	0.145454545	161	162
Dx1740	0.145454545	161	162
Dx74332	0.145454545	161	162
T83510d	0.145454545	161	162
S79819A, T83510D	0.006060606	184	270
Dx17311, Z578	0.006060606	184	270
Dx36842	0.006060606	184	270
T80211d	0.006060606	184	270
Z578	0.006060606	184	270
Dx17311, Z578	0.585365854	88	102
Dx36842	0.585365854	88	102
T381x5a	0.585365854	88	102
Z578	0.585365854	88	102
Dx17311, Z578	0.365853659	124	154
Dx1740	0.365853659	124	154
T381x5a	0.365853659	124	154
Z578	0.365853659	124	154
Dx17311	0.219512195	148	191
Dx1740	0.219512195	148	191
Dx74332	0.219512195	148	191
Z578	0.219512195	148	191
S79819A	0.146341463	160	161
Dx1740	0.146341463	160	161
Dx74332	0.146341463	160	161
T83510d	0.146341463	160	161
Dx17311, Z578	0.809815951	51	91
Dx36842	0.809815951	51	91
T50992a	0.809815951	51	91
Z578	0.809815951	51	91
Dx17311, Z578	0.717791411	66	221
Dx36842	0.717791411	66	221
T50992a	0.717791411	66	221
Z578	0.717791411	66	221
Dx17311	0.220858896	147	186

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.220858896	147	186
Dx74332	0.220858896	147	186
Z578	0.220858896	147	186
S79819A	0.147239264	159	161
Dx1740	0.147239264	159	161
Dx74332	0.147239264	159	161
T83510d	0.147239264	159	161
S79819A, T83510D	0.006134969	182	189
Dx17311, Z578	0.006134969	182	189
Dx36842	0.006134969	182	189
T80211d	0.006134969	182	189
Z578	0.006134969	182	189
Dx17311, Z578	0.740740741	62	203
Dx36842	0.740740741	62	203
T50992a	0.740740741	62	203
Z578	0.740740741	62	203
Dx17311	0.222222222	146	177
Dx1740	0.222222222	146	177
Dx74332	0.222222222	146	177
Z578	0.222222222	146	177
S79819A	0.148148148	158	162
Dx1740	0.148148148	158	162
Dx74332	0.148148148	158	162
T83510d	0.148148148	158	162
Dx17311, Z578	0.52173913	97	189
Dx36842	0.52173913	97	189
T381x5a	0.52173913	97	189
Z578	0.52173913	97	189
Dx17311, Z578	0.372670807	121	186
Dx1740	0.372670807	121	186
T381x5a	0.372670807	121	186
Z578	0.372670807	121	186
Dx17311, Z578	0.298136646	133	142
Dx1740	0.298136646	133	142
s14104d	0.298136646	133	142
Z578	0.298136646	133	142
S79819A	0.149068323	157	163
Dx1740	0.149068323	157	163
Dx74332	0.149068323	157	163
T83510d	0.149068323	157	163

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.15	156	158
Dx1740	0.15	156	158
Dx7105	0.15	156	158
T83510d	0.15	156	158
S79819A	0.150943396	155	169
Dx1740	0.150943396	155	169
Dx7105	0.150943396	155	169
T83510d	0.150943396	155	169
Dx17311, Z578	0.53164557	94	187
Dx36842	0.53164557	94	187
T381x5a	0.53164557	94	187
Z578	0.53164557	94	187
Dx17311, Z578	0.379746835	118	140
Dx1740	0.379746835	118	140
T381x5a	0.379746835	118	140
Z578	0.379746835	118	140
Dx17311	0.227848101	142	238
Dx1740	0.227848101	142	238
Dx74332	0.227848101	142	238
Z578	0.227848101	142	238
S79819A	0.151898734	154	161
Dx1740	0.151898734	154	161
Dx7105	0.151898734	154	161
T83510d	0.151898734	154	161
S79819A	0.152866242	153	154
Dx1740	0.152866242	153	154
Dx7105	0.152866242	153	154
T83510d	0.152866242	153	154
S79819A	0.153846154	152	183
Dx1740	0.153846154	152	183
Dx7105	0.153846154	152	183
T83510d	0.153846154	152	183
Dx17311, Z578	0.541935484	91	181
Dx36842	0.541935484	91	181
T381x5a	0.541935484	91	181
Z578	0.541935484	91	181
Dx17311	0.232258065	139	208
Dx1740	0.232258065	139	208
Dx74332	0.232258065	139	208
Z578	0.232258065	139	208

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.545454545	90	98
Dx36842	0.545454545	90	98
T381x5a	0.545454545	90	98
Z578	0.545454545	90	98
Dx17311, Z578	0.38961039	114	189
Dx1740	0.38961039	114	189
T381x5a	0.38961039	114	189
Z578	0.38961039	114	189
Dx17311, Z578	0.311688312	126	140
Dx1740	0.311688312	126	140
s14104d	0.311688312	126	140
Z578	0.311688312	126	140
S79819A	0.155844156	150	160
Dx1740	0.155844156	150	160
Dx7105	0.155844156	150	160
T83510d	0.155844156	150	160
Dx17311, Z578	0.470588235	101	133
Dx36842	0.470588235	101	133
T381x5a	0.470588235	101	133
Z578	0.470588235	101	133
Dx17311, Z578	0.31372549	125	134
Dx1740	0.31372549	125	134
s14104d	0.31372549	125	134
Z578	0.31372549	125	134
Dx17311	0.235294118	137	166
Dx1740	0.235294118	137	166
Dx74332	0.235294118	137	166
Z578	0.235294118	137	166
S79819A	0.156862745	149	198
Dx1740	0.156862745	149	198
Dx7105	0.156862745	149	198
T83510d	0.156862745	149	198
Dx17311, Z578	0.473684211	100	173
Dx36842	0.473684211	100	173
T381x5a	0.473684211	100	173
Z578	0.473684211	100	173
Dx17311, Z578	0.394736842	112	187
Dx1740	0.394736842	112	187
T381x5a	0.394736842	112	187
Z578	0.394736842	112	187

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.315789474	124	128
Dx1740	0.315789474	124	128
s14104d	0.315789474	124	128
Z578	0.315789474	124	128
Dx17311	0.236842105	136	175
Dx1740	0.236842105	136	175
Dx74332	0.236842105	136	175
Z578	0.236842105	136	175
S79819A	0.157894737	148	156
Dx1740	0.157894737	148	156
Dx7105	0.157894737	148	156
T83510d	0.157894737	148	156
Dx17311, Z578	0.715231788	63	194
Dx36842	0.715231788	63	194
T50992a	0.715231788	63	194
Z578	0.715231788	63	194
S79819A, T83510D	0.24	135	182
Dx17311	0.238410596	135	182
Dx1740	0.238410596	135	182
Dx74332	0.238410596	135	182
Z578	0.238410596	135	182
S79819A	0.158940397	147	152
Dx1740	0.158940397	147	152
Dx7105	0.158940397	147	152
T83510d	0.158940397	147	152
S79819A, T83510D	0.006622517	170	272
Dx17311, Z578	0.006622517	170	272
Dx36842	0.006622517	170	272
T80211d	0.006622517	170	272
Z578	0.006622517	170	272
Dx17311	0.24	134	182
Dx1740	0.24	134	182
Dx74332	0.24	134	182
Z578	0.24	134	182
S79819A	0.16	146	175
Dx1740	0.16	146	175
Dx7105	0.16	146	175
T83510d	0.16	146	175
Dx17311, Z578	0.483221477	97	133
Dx36842	0.483221477	97	133

Covariates	Probability	Minimum RVI	Maximum RVI
T381x5a	0.483221477	97	133
Z578	0.483221477	97	133
Dx17311, Z578	0.322147651	121	153
Dx1740	0.322147651	121	153
s14104d	0.322147651	121	153
Z578	0.322147651	121	153
Dx17311	0.241610738	133	141
Dx1740	0.241610738	133	141
Dx74332	0.241610738	133	141
Z578	0.241610738	133	141
S79819A	0.161073826	145	153
Dx1740	0.161073826	145	153
Dx7105	0.161073826	145	153
T83510d	0.161073826	145	153
Dx17311, Z578	0.648648649	72	150
Dx36842	0.648648649	72	150
T381x5a	0.648648649	72	150
Z578	0.648648649	72	150
Dx17311, Z578	0.486486486	96	132
Dx36842	0.486486486	96	132
T381x5a	0.486486486	96	132
Z578	0.486486486	96	132
Dx17311, Z578	0.324324324	120	121
Dx1740	0.324324324	120	121
s14104d	0.324324324	120	121
Z578	0.324324324	120	121
Dx17311	0.243243243	132	133
Dx1740	0.243243243	132	133
Dx74332	0.243243243	132	133
Z578	0.243243243	132	133
Dx17311, Z578	0.326530612	119	126
Dx1740	0.326530612	119	126
s14104d	0.326530612	119	126
Z578	0.326530612	119	126
S79819A	0.163265306	143	147
Dx1740	0.163265306	143	147
Dx7105	0.163265306	143	147
T83510d	0.163265306	143	147
Dx17311, Z578	0.657534247	70	104
Dx36842	0.657534247	70	104

Covariates	Probability	Minimum RVI	Maximum RVI
T381x5a	0.657534247	70	104
Z578	0.657534247	70	104
Dx17311, Z578	0.493150685	94	98
Dx36842	0.493150685	94	98
T381x5a	0.493150685	94	98
Z578	0.493150685	94	98
Dx17311, Z578	0.328767123	118	123
Dx1740	0.328767123	118	123
s14104d	0.328767123	118	123
Z578	0.328767123	118	123
Dx17311	0.246575342	130	180
Dx1740	0.246575342	130	180
Dx74332	0.246575342	130	180
Z578	0.246575342	130	180
S79819A	0.164383562	142	145
Dx1740	0.164383562	142	145
Dx7105	0.164383562	142	145
T83510d	0.164383562	142	145
Dx17311, Z578	0.744827586	57	105
Dx36842	0.744827586	57	105
T50992a	0.744827586	57	105
Z578	0.744827586	57	105
Dx17311, Z578	0.413793103	105	238
Dx1740	0.413793103	105	238
T381x5a	0.413793103	105	238
Z578	0.413793103	105	238
Dx17311, Z578	0.331034483	117	126
Dx1740	0.331034483	117	126
s14104d	0.331034483	117	126
Z578	0.331034483	117	126
Dx17311	0.248275862	129	151
Dx1740	0.248275862	129	151
Dx74332	0.248275862	129	151
Z578	0.248275862	129	151
S79819A	0.234482759	131	183
Dx1740	0.234482759	131	183
Dx74332	0.234482759	131	183
T83510d	0.234482759	131	183
S79819A	0.165517241	141	142
Dx1740	0.165517241	141	142

Covariates	Probability	Minimum RVI	Maximum RVI
Dx7105	0.165517241	141	142
T83510d	0.165517241	141	142
Dx17311, Z578	0.5	92	100
Dx36842	0.5	92	100
T381x5a	0.5	92	100
Z578	0.5	92	100
S79819A	0.166666667	140	141
Dx1740	0.166666667	140	141
Dx7105	0.166666667	140	141
T83510d	0.166666667	140	141
Dx17311, Z578	0.503496503	91	98
Dx36842	0.503496503	91	98
T381x5a	0.503496503	91	98
Z578	0.503496503	91	98
Dx17311, Z578	0.41958042	103	239
Dx1740	0.41958042	103	239
T381x5a	0.41958042	103	239
Z578	0.41958042	103	239
S79819A	0.244755245	128	187
Dx1740	0.244755245	128	187
Dx74332	0.244755245	128	187
T83510d	0.244755245	128	187
S79819A	0.167832168	139	140
Dx1740	0.167832168	139	140
Dx7105	0.167832168	139	140
T83510d	0.167832168	139	140
Dx17311, Z578	0.507042254	90	147
Dx36842	0.507042254	90	147
T381x5a	0.507042254	90	147
Z578	0.507042254	90	147
Dx17311, Z578	0.338028169	114	125
Dx1740	0.338028169	114	125
s14104d	0.338028169	114	125
Z578	0.338028169	114	125
Dx17311	0.253521127	126	143
Dx1740	0.253521127	126	143
Dx74332	0.253521127	126	143
Z578	0.253521127	126	143
S79819A	0.169014085	138	140
Dx1740	0.169014085	138	140

Covariates	Probability	Minimum RVI	Maximum RVI
Dx7105	0.169014085	138	140
T83510d	0.169014085	138	140
Dx17311, Z578	0.680851064	65	133
Dx36842	0.680851064	65	133
T381x5a	0.680851064	65	133
Z578	0.680851064	65	133
Dx17311, Z578	0.595744681	77	133
Dx36842	0.595744681	77	133
T381x5a	0.595744681	77	133
Z578	0.595744681	77	133
Dx17311, Z578	0.510638298	89	105
Dx36842	0.510638298	89	105
T381x5a	0.510638298	89	105
Z578	0.510638298	89	105
S79819A	0.170212766	137	144
Dx1740	0.170212766	137	144
Dx7105	0.170212766	137	144
T83510d	0.170212766	137	144
Dx17311, Z578	0.685714286	64	190
Dx36842	0.685714286	64	190
T381x5a	0.685714286	64	190
Z578	0.685714286	64	190
Dx17311, Z578	0.514285714	88	109
Dx36842	0.514285714	88	109
T381x5a	0.514285714	88	109
Z578	0.514285714	88	109
Dx17311, Z578	0.5	90	91
Dx1740	0.5	90	91
T381x5a	0.5	90	91
Z578	0.5	90	91
Dx17311, Z578	0.342857143	112	120
Dx1740	0.342857143	112	120
s14104d	0.342857143	112	120
Z578	0.342857143	112	120
Dx17311	0.257142857	124	155
Dx1740	0.257142857	124	155
Dx74332	0.257142857	124	155
Z578	0.257142857	124	155
S79819A	0.171428571	136	154
Dx1740	0.171428571	136	154

Covariates	Probability	Minimum RVI	Maximum RVI
Dx7105	0.171428571	136	154
T83510d	0.171428571	136	154
Dx17311, Z578	0.690647482	63	280
Dx36842	0.690647482	63	280
T381x5a	0.690647482	63	280
Z578	0.690647482	63	280
Dx17311, Z578	0.517985612	87	119
Dx36842	0.517985612	87	119
T381x5a	0.517985612	87	119
Z578	0.517985612	87	119
Dx17311, Z578	0.431654676	99	126
Dx1740	0.431654676	99	126
T381x5a	0.431654676	99	126
Z578	0.431654676	99	126
Dx17311	0.258992806	123	148
Dx1740	0.258992806	123	148
Dx74332	0.258992806	123	148
Z578	0.258992806	123	148
Dx17311, Z578	0.608695652	74	123
Dx36842	0.608695652	74	123
T381x5a	0.608695652	74	123
Z578	0.608695652	74	123
Dx17311, Z578	0.434782609	98	206
Dx1740	0.434782609	98	206
T381x5a	0.434782609	98	206
Z578	0.434782609	98	206
Dx17311, Z578	0.347826087	110	119
Dx1740	0.347826087	110	119
s14104d	0.347826087	110	119
Z578	0.347826087	110	119
Dx17311	0.260869565	122	161
Dx1740	0.260869565	122	161
Dx74332	0.260869565	122	161
Z578	0.260869565	122	161
S79819A	0.173913043	134	145
Dx1740	0.173913043	134	145
Dx7105	0.173913043	134	145
T83510d	0.173913043	134	145
Dx17311, Z578	0.525547445	85	126
Dx36842	0.525547445	85	126

Covariates	Probability	Minimum RVI	Maximum RVI
T381x5a	0.525547445	85	126
Z578	0.525547445	85	126
Dx17311, Z578	0.437956204	97	120
Dx1740	0.437956204	97	120
T381x5a	0.437956204	97	120
Z578	0.437956204	97	120
Dx17311	0.262773723	121	124
Dx1740	0.262773723	121	124
Dx74332	0.262773723	121	124
Z578	0.262773723	121	124
S79819A	0.175182482	133	134
Dx1740	0.175182482	133	134
Dx7105	0.175182482	133	134
T83510d	0.175182482	133	134
S79819A, T83510D	0.167883212	134	189
Dx17311, Z578	0.167883212	134	189
Dx36842	0.167883212	134	189
T80211d	0.167883212	134	189
Z578	0.167883212	134	189
Dx17311, Z578	0.705882353	60	189
Dx36842	0.705882353	60	189
T381x5a	0.705882353	60	189
Z578	0.705882353	60	189
Dx17311	0.264705882	120	121
Dx1740	0.264705882	120	121
Dx74332	0.264705882	120	121
Z578	0.264705882	120	121
S79819A	0.176470588	132	140
Dx1740	0.176470588	132	140
Dx7105	0.176470588	132	140
T83510d	0.176470588	132	140
Dx17311, Z578	0.533333333	83	154
Dx36842	0.533333333	83	154
T381x5a	0.533333333	83	154
Z578	0.533333333	83	154
Dx17311, Z578	0.355555556	107	120
Dx1740	0.355555556	107	120
s14104d	0.355555556	107	120
Z578	0.355555556	107	120
Dx17311	0.266666667	119	126

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.266666667	119	126
Dx74332	0.266666667	119	126
Z578	0.266666667	119	126
S79819A, T83510D	0.18	131	154
S79819A	0.177777778	131	134
Dx1740	0.177777778	131	134
Dx7105	0.177777778	131	134
T83510d	0.177777778	131	134
Dx17311, Z578	0.626865672	70	146
Dx36842	0.626865672	70	146
T381x5a	0.626865672	70	146
Z578	0.626865672	70	146
Dx17311, Z578	0.537313433	82	202
Dx36842	0.537313433	82	202
T381x5a	0.537313433	82	202
Z578	0.537313433	82	202
Dx17311	0.268656716	118	141
Dx1740	0.268656716	118	141
Dx74332	0.268656716	118	141
Z578	0.268656716	118	141
S79819A	0.26119403	119	182
Dx1740	0.26119403	119	182
Dx74332	0.26119403	119	182
T83510d	0.26119403	119	182
S79819A	0.179104478	130	131
Dx1740	0.179104478	130	131
Dx7105	0.179104478	130	131
T83510d	0.179104478	130	131
Dx17311, Z578	0.812030075	45	208
Dx36842	0.812030075	45	208
T50992a	0.812030075	45	208
Z578	0.812030075	45	208
Dx17311, Z578	0.526315789	83	111
Dx1740	0.526315789	83	111
T381x5a	0.526315789	83	111
Z578	0.526315789	83	111
Dx17311, Z578	0.360902256	105	119
Dx1740	0.360902256	105	119
s14104d	0.360902256	105	119
Z578	0.360902256	105	119

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311	0.270676692	117	125
Dx1740	0.270676692	117	125
Dx74332	0.270676692	117	125
Z578	0.270676692	117	125
S79819A	0.255639098	119	240
Dx1740	0.255639098	119	240
Dx74332	0.255639098	119	240
T83510d	0.255639098	119	240
S79819A	0.180451128	129	130
Dx1740	0.180451128	129	130
Dx7105	0.180451128	129	130
T83510d	0.180451128	129	130
Dx17311, Z578	0.545454545	80	229
Dx36842	0.545454545	80	229
T381x5a	0.545454545	80	229
Z578	0.545454545	80	229
Dx17311, Z578	0.454545455	92	129
Dx1740	0.454545455	92	129
T381x5a	0.454545455	92	129
Z578	0.454545455	92	129
Dx17311	0.272727273	116	131
Dx1740	0.272727273	116	131
Dx74332	0.272727273	116	131
Z578	0.272727273	116	131
S79819A	0.181818182	128	130
Dx1740	0.181818182	128	130
Dx7105	0.181818182	128	130
T83510d	0.181818182	128	130
Dx17311, Z578	0.458015267	91	105
Dx1740	0.458015267	91	105
T381x5a	0.458015267	91	105
Z578	0.458015267	91	105
Dx17311, Z578	0.450381679	92	107
Dx1740	0.450381679	92	107
s14104d	0.450381679	92	107
Z578	0.450381679	92	107
Dx17311, Z578	0.366412214	103	119
Dx1740	0.366412214	103	119
s14104d	0.366412214	103	119
Z578	0.366412214	103	119

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311	0.27480916	115	122
Dx1740	0.27480916	115	122
Dx74332	0.27480916	115	122
Z578	0.27480916	115	122
S79819A, T83510D	0.18	127	189
S79819A	0.183206107	127	129
Dx1740	0.183206107	127	129
Dx7105	0.183206107	127	129
T83510d	0.183206107	127	129
Dx17311, Z578	0.461538462	90	106
Dx1740	0.461538462	90	106
s31125a	0.461538462	90	106
Z578	0.461538462	90	106
Dx17311, Z578	0.446153846	92	222
Dx1740	0.446153846	92	222
s14104d	0.446153846	92	222
Z578	0.446153846	92	222
Dx17311, Z578	0.369230769	102	115
Dx1740	0.369230769	102	115
s14104d	0.369230769	102	115
Z578	0.369230769	102	115
Dx17311	0.276923077	114	124
Dx1740	0.276923077	114	124
Dx74332	0.276923077	114	124
Z578	0.276923077	114	124
S79819A, T83510D	0.18	126	135
S79819A	0.184615385	126	127
Dx1740	0.184615385	126	127
Dx7105	0.184615385	126	127
T83510d	0.184615385	126	127
Dx17311, Z578	0.558139535	77	92
Dx1740	0.558139535	77	92
Dx36842	0.558139535	77	105
T381x5a	0.558139535	77	92
Z578	0.558139535	77	92
Dx17311, Z578	0.465116279	89	126
Dx1740	0.465116279	89	126
s31125a	0.465116279	89	126
Z578	0.465116279	89	126
Dx17311	0.279069767	113	186

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.279069767	113	186
Dx74332	0.279069767	113	186
Z578	0.279069767	113	186
S79819A	0.186046512	125	126
Dx17311, S14104d	0.186046512	125	126
Dx1740	0.186046512	125	203
Dx7105	0.186046512	125	126
T83510d	0.186046512	125	126
S79819A, T83510D	0.178294574	126	182
Dx17311, Z578	0.178294574	126	182
Dx36842	0.178294574	126	182
T80211d	0.178294574	126	182
Z578	0.178294574	126	182
Dx17311, Z578	0.46875	88	212
Dx1740	0.46875	88	212
s31125a	0.46875	88	212
Z578	0.46875	88	212
Dx17311, Z578	0.375	100	238
Dx1740	0.375	100	238
s14104d	0.375	100	238
Z578	0.375	100	238
Dx17311	0.28125	112	123
Dx1740	0.28125	112	123
Dx74332	0.28125	112	123
T83510d	0.28125	112	123
Z578	0.28125	112	274
S79819A	0.1875	124	125
Dx17311, S14104d	0.1875	124	125
Dx7105	0.1875	124	125
T83510d	0.1875	124	125
Dx17311, Z578	0.661417323	63	182
Dx36842	0.661417323	63	182
T381x5a	0.661417323	63	182
Z578	0.661417323	63	182
Dx17311, Z578	0.472440945	87	97
Dx1740	0.472440945	87	97
s31125a	0.472440945	87	97
Z578	0.472440945	87	97
Dx17311, Z578	0.377952756	99	204
Dx1740	0.377952756	99	204

Covariates	Probability	Minimum RVI	Maximum RVI
s14104d	0.377952756	99	204
Z578	0.377952756	99	204
Dx17311	0.283464567	111	133
Dx1740	0.283464567	111	133
Dx74332	0.283464567	111	133
T83510d	0.283464567	111	133
S79819A	0.188976378	123	139
Dx17311, S14104d	0.188976378	123	139
Dx7105	0.188976378	123	139
T83510d	0.188976378	123	139
S79819A, T83510D	0.173228346	125	126
Dx17311, Z578	0.173228346	125	126
Dx36842	0.173228346	125	126
T80211d	0.173228346	125	126
Z578	0.173228346	125	126
Dx17311, Z578	0.761904762	50	273
Dx36842	0.761904762	50	273
T381x5a	0.761904762	50	273
Z578	0.761904762	50	273
Dx17311, Z578	0.666666667	62	187
Dx36842	0.666666667	62	187
T381x5a	0.666666667	62	187
Z578	0.666666667	62	187
Dx17311, Z578	0.380952381	98	119
Dx1740	0.380952381	98	119
s14104d	0.380952381	98	119
Z578	0.380952381	98	119
Dx17311	0.373015873	99	174
Dx1740	0.373015873	99	174
Dx74332	0.373015873	99	174
Z578	0.373015873	99	174
Dx17311	0.285714286	110	146
Dx1740	0.285714286	110	146
Dx74332	0.285714286	110	146
T83510d	0.285714286	110	146
S79819A	0.19047619	122	123
Dx17311, S14104d	0.19047619	122	123
Dx7105	0.19047619	122	123
T83510d	0.19047619	122	123
Dx17311, Z578	0.768	49	183

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.768	49	183
T381x5a	0.768	49	183
Z578	0.768	49	183
Dx17311, Z578	0.672	61	153
Dx36842	0.672	61	153
T381x5a	0.672	61	153
Z578	0.672	61	153
Dx17311, Z578	0.48	85	113
Dx1740	0.48	85	113
s31125a	0.48	85	113
Z578	0.48	85	113
Dx17311, Z578	0.384	97	99
Dx1740	0.384	97	99
s14104d	0.384	97	99
Z578	0.384	97	99
Dx17311	0.288	109	158
Dx1740	0.288	109	158
Dx74332	0.288	109	158
T83510d	0.288	109	158
S79819A	0.192	121	123
Dx17311, S14104d	0.192	121	123
Dx7105	0.192	121	123
T83510d	0.192	121	123
Z578	0.192	121	123
Dx17311, Z578	0.483870968	84	96
Dx1740	0.483870968	84	96
s31125a	0.483870968	84	96
Z578	0.483870968	84	96
Dx17311, Z578	0.387096774	96	98
Dx1740	0.387096774	96	98
s14104d	0.387096774	96	98
Z578	0.387096774	96	98
Dx17311	0.290322581	108	150
Dx1740	0.290322581	108	150
Dx74332	0.290322581	108	150
T83510d	0.290322581	108	150
S79819A	0.282258065	109	222
Dx1740	0.282258065	109	222
Dx74332	0.282258065	109	222
T83510d	0.282258065	109	222

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.193548387	120	121
Dx17311, S14104d	0.193548387	120	121
Dx7105	0.193548387	120	121
T83510d	0.193548387	120	121
Z578	0.193548387	120	121
S79819A, T83510D	0.185483871	121	134
Dx17311, Z578	0.185483871	121	134
Dx36842	0.185483871	121	134
T80211d	0.185483871	121	134
Z578	0.185483871	121	134
Dx17311, Z578	0.682926829	59	120
Dx36842	0.682926829	59	120
T381x5a	0.682926829	59	120
Z578	0.682926829	59	120
Dx17311, Z578	0.585365854	71	102
Dx1740	0.585365854	71	102
T381x5a	0.585365854	71	102
Z578	0.585365854	71	102
Dx17311, Z578	0.487804878	83	96
Dx1740	0.487804878	83	96
s31125a	0.487804878	83	96
Z578	0.487804878	83	96
Dx17311, Z578	0.479674797	84	125
Dx1740	0.479674797	84	125
s14104d	0.479674797	84	125
Z578	0.479674797	84	125
Dx17311, Z578	0.390243902	95	121
Dx1740	0.390243902	95	121
s14104d	0.390243902	95	121
Z578	0.390243902	95	121
Dx17311	0.292682927	107	126
Dx1740	0.292682927	107	126
Dx74332	0.292682927	107	126
T83510d	0.292682927	107	126
S79819A	0.195121951	119	122
Dx17311, S14104d	0.195121951	119	122
Dx7105	0.195121951	119	122
T83510d	0.195121951	119	122
Z578	0.195121951	119	122
Dx17311, Z578	0.786885246	46	193

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.786885246	46	193
T381x5a	0.786885246	46	193
Z578	0.786885246	46	193
Dx17311, Z578	0.590163934	70	168
Dx1740	0.590163934	70	168
T381x5a	0.590163934	70	168
Z578	0.590163934	70	168
Dx17311, Z578	0.393442623	94	102
Dx1740	0.393442623	94	102
s14104d	0.393442623	94	102
Z578	0.393442623	94	102
Dx17311	0.295081967	106	120
Dx1740	0.295081967	106	120
Dx74332	0.295081967	106	120
T83510d	0.295081967	106	120
S79819A	0.196721311	118	119
Dx17311, S14104d	0.196721311	118	119
Dx7105	0.196721311	118	119
T83510d	0.196721311	118	119
Z578	0.196721311	118	119
Dx17311, Z578	0.694214876	57	141
Dx36842	0.694214876	57	141
T381x5a	0.694214876	57	141
Z578	0.694214876	57	141
Dx17311, Z578	0.595041322	69	193
Dx1740	0.595041322	69	193
T381x5a	0.595041322	69	193
Z578	0.595041322	69	193
Dx17311, Z578	0.495867769	81	84
Dx1740	0.495867769	81	84
s31125a	0.495867769	81	84
Z578	0.495867769	81	84
Dx17311, Z578	0.396694215	93	97
Dx1740	0.396694215	93	97
s14104d	0.396694215	93	97
Z578	0.396694215	93	97
Dx17311	0.297520661	105	115
Dx1740	0.297520661	105	115
Dx74332	0.297520661	105	115
T83510d	0.297520661	105	115

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.198347107	117	119
Dx17311, S14104d	0.198347107	117	119
Dx7105	0.198347107	117	119
T83510d	0.198347107	117	119
Z578	0.198347107	117	119
S79819A, T83510D	0.181818182	119	126
Dx17311, Z578	0.181818182	119	126
Dx36842	0.181818182	119	126
T80211d	0.181818182	119	126
Z578	0.181818182	119	126
Dx17311, Z578	0.7	56	97
Dx36842	0.7	56	97
T381x5a	0.7	56	97
Z578	0.7	56	97
Dx17311, Z578	0.5	80	175
Dx1740	0.5	80	175
s14104d	0.5	80	175
Z578	0.5	80	175
Dx17311, Z578	0.4	92	98
Dx1740	0.4	92	98
m10361	0.4	92	98
s14104d	0.4	92	210
Z578	0.4	92	98
T83510D	0.3	104	118
Dx17311	0.3	104	167
Dx1740	0.3	104	118
Dx74332	0.3	104	118
S79819A, T83510D	0.20	116	209
S79819A	0.2	116	119
Dx17311, S14104d	0.2	116	119
Dx529	0.2	116	119
Dx7105	0.2	116	120
Dx7105	0.2	116	129
T83510d	0.2	116	119
Z578	0.2	116	119
Dx17311, Z578	0.504201681	79	98
Dx1740	0.504201681	79	98
s14104d	0.504201681	79	98
Z578	0.504201681	79	98
Dx17311, Z578	0.403361345	91	92

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.403361345	91	92
Dx74332	0.403361345	91	92
m10361	0.403361345	91	217
m10361	0.403361345	91	224
Z578	0.403361345	91	92
T83510D	0.302521008	103	161
Dx1740	0.302521008	103	161
Dx74332	0.302521008	103	161
S79819A	0.201680672	115	118
Dx17311, S14104d	0.201680672	115	118
Dx529	0.201680672	115	118
T83510d	0.201680672	115	118
Z578	0.201680672	115	118
Dx17311, Z578	0.406779661	90	91
Dx1740	0.406779661	90	91
Dx74332	0.406779661	90	91
Z578	0.406779661	90	91
Dx17311	0.398305085	91	186
Dx1740	0.398305085	91	186
Dx74332	0.398305085	91	186
Z578	0.398305085	91	186
T83510D	0.305084746	102	125
Dx1740	0.305084746	102	125
Dx74332	0.305084746	102	125
S79819A	0.203389831	114	119
Dx17311, S14104d	0.203389831	114	119
Dx529	0.203389831	114	119
T83510d	0.203389831	114	119
Z578	0.203389831	114	119
Dx17311, Z578	0.717948718	53	63
Dx36842	0.717948718	53	63
T381x5a	0.717948718	53	63
Z578	0.717948718	53	63
Dx17311, Z578	0.615384615	65	154
Dx1740	0.615384615	65	154
T381x5a	0.615384615	65	154
Z578	0.615384615	65	154
Dx17311, Z578	0.512820513	77	96
Dx1740	0.512820513	77	96
s14104d	0.512820513	77	96

Covariates	Probability	Minimum RVI	Maximum RVI
Z578	0.512820513	77	96
Dx17311, Z578	0.41025641	89	93
Dx1740	0.41025641	89	93
Dx74332	0.41025641	89	93
Z578	0.41025641	89	93
Dx17311	0.401709402	90	268
Dx1740	0.401709402	90	268
Dx74332	0.401709402	90	268
Z578	0.401709402	90	268
T83510D	0.307692308	101	105
Dx1740	0.307692308	101	105
Dx74332	0.307692308	101	105
S79819A	0.205128205	113	114
Dx17311, S14104d	0.205128205	113	114
Dx529	0.205128205	113	114
T83510d	0.205128205	113	114
Z578	0.205128205	113	114
Dx17311, Z578	0.413793103	88	91
Dx1740	0.413793103	88	91
Dx74332	0.413793103	88	91
Z578	0.413793103	88	91
Dx17311	0.396551724	90	126
Dx1740	0.396551724	90	126
Dx74332	0.396551724	90	126
Z578	0.396551724	90	126
Dx17311	0.379310345	92	162
Dx1740	0.379310345	92	162
Dx74332	0.379310345	92	162
Z578	0.379310345	92	162
T83510D	0.310344828	100	105
Dx1740	0.310344828	100	105
Dx74332	0.310344828	100	105
S79819A	0.206896552	112	113
Dx17311, S14104d	0.206896552	112	113
Dx37231	0.206896552	112	113
Dx529	0.206896552	112	218
T83510d	0.206896552	112	113
Z578	0.206896552	112	113
Dx17311, Z578	0.730434783	51	168
Dx36842	0.730434783	51	168

Covariates	Probability	Minimum RVI	Maximum RVI
T381x5a	0.730434783	51	168
Z578	0.730434783	51	168
Dx17311, Z578	0.626086957	63	70
Dx1740	0.626086957	63	70
T381x5a	0.626086957	63	70
Z578	0.626086957	63	70
Dx17311, Z578	0.52173913	75	102
Dx1740	0.52173913	75	102
s14104d	0.52173913	75	102
Z578	0.52173913	75	102
Dx17311, Z578	0.417391304	87	183
Dx1740	0.417391304	87	183
Dx74332	0.417391304	87	183
Z578	0.417391304	87	183
Dx17311	0.408695652	88	184
Dx1740	0.408695652	88	184
Dx74332	0.408695652	88	184
Z578	0.408695652	88	184
T83510D	0.313043478	99	119
Dx1740	0.313043478	99	119
Dx74332	0.313043478	99	119
S79819A	0.208695652	111	112
Dx17311, S14104d	0.208695652	111	112
Dx37231	0.208695652	111	112
T83510d	0.208695652	111	112
Z578	0.208695652	111	112
S79819A, T83510D	0.008695652	134	189
Dx17311, Z578	0.008695652	134	189
Dx36842	0.008695652	134	189
T80211d	0.008695652	134	189
Z578	0.008695652	134	189
Dx17311, Z578	0.622807018	63	161
Dx1740	0.622807018	63	161
T381x5a	0.622807018	63	161
Z578	0.622807018	63	161
Dx17311, Z578	0.526315789	74	239
Dx1740	0.526315789	74	239
s14104d	0.526315789	74	239
Z578	0.526315789	74	239
Dx17311, Z578	0.421052632	86	167

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.421052632	86	167
Dx74332	0.421052632	86	167
Z578	0.421052632	86	167
T83510D	0.315789474	98	103
Dx1740	0.315789474	98	103
Dx74332	0.315789474	98	103
S79819A	0.307017544	99	154
Dx1740	0.307017544	99	154
Dx74332	0.307017544	99	154
T83510d	0.307017544	99	154
S79819A	0.210526316	110	114
Dx17311, S14104d	0.210526316	110	114
Dx37231	0.210526316	110	114
T83510d	0.210526316	110	114
Z578	0.210526316	110	114
Dx17311, Z578	0.743362832	49	217
Dx36842	0.743362832	49	217
T381x5a	0.743362832	49	217
Z578	0.743362832	49	217
Dx17311, Z578	0.424778761	85	97
Dx1740	0.424778761	85	97
Dx74332	0.424778761	85	97
Z578	0.424778761	85	97
T83510D	0.318584071	97	99
Dx1740	0.318584071	97	99
Dx74332	0.318584071	97	99
S79819A	0.212389381	109	132
Dx17311, S14104d	0.212389381	109	132
Dx37231	0.212389381	109	132
T83510d	0.212389381	109	132
Z578	0.212389381	109	132
S79819A, T83510D	0.017699115	131	183
Dx17311, Z578	0.017699115	131	183
Dx36842	0.017699115	131	183
T80211d	0.017699115	131	183
Z578	0.017699115	131	183
Dx17311, Z578	0.428571429	84	96
Dx1740	0.428571429	84	96
Dx74332	0.428571429	84	96
Z578	0.428571429	84	96

Covariates	Probability	Minimum RVI	Maximum RVI
T83510D	0.321428571	96	101
Dx1740	0.321428571	96	101
Dx74332	0.321428571	96	101
S79819A	0.214285714	108	110
Dx17311, S14104d	0.214285714	108	110
Dx37231	0.214285714	108	110
T83510d	0.214285714	108	110
Z578	0.214285714	108	110
Dx17311, Z578	0.540540541	71	169
Dx1740	0.540540541	71	169
s14104d	0.540540541	71	169
Z578	0.540540541	71	169
Dx17311, Z578	0.432432432	83	91
Dx1740	0.432432432	83	91
Dx74332	0.432432432	83	91
Z578	0.432432432	83	91
Dx17311	0.423423423	84	92
Dx1740	0.423423423	84	92
Dx74332	0.423423423	84	92
Z578	0.423423423	84	92
T83510D	0.324324324	95	113
Dx1740	0.324324324	95	113
Dx74332	0.324324324	95	113
S79819A	0.315315315	96	141
Dx1740	0.315315315	96	141
Dx74332	0.315315315	96	141
T83510d	0.315315315	96	141
S79819A	0.216216216	107	117
Dx17311, S14104d	0.216216216	107	117
Dx37231	0.216216216	107	117
T83510d	0.216216216	107	117
Z578	0.216216216	107	117
Dx17311, Z578	0.545454545	70	92
Dx1740	0.545454545	70	92
s14104d	0.545454545	70	92
Z578	0.545454545	70	92
Dx17311, Z578	0.436363636	82	133
Dx1740	0.436363636	82	133
Dx74332	0.436363636	82	133
Z578	0.436363636	82	133

Covariates	Probability	Minimum RVI	Maximum RVI
T83510D	0.327272727	94	98
Dx1740	0.327272727	94	98
Dx74332	0.327272727	94	98
S79819A	0.218181818	106	108
Dx17311, S14104d	0.218181818	106	108
Dx37231	0.218181818	106	108
T83510d	0.218181818	106	108
Z578	0.218181818	106	108
Dx17311, Z578	0.660550459	57	194
Dx1740	0.660550459	57	194
T381x5a	0.660550459	57	194
Z578	0.660550459	57	194
Dx17311, Z578	0.550458716	69	100
Dx1740	0.550458716	69	100
s14104d	0.550458716	69	100
Z578	0.550458716	69	100
Dx17311, Z578	0.440366972	81	98
Dx1740	0.440366972	81	98
Dx74332	0.440366972	81	98
Z578	0.440366972	81	98
T83510D	0.330275229	93	98
Dx1740	0.330275229	93	98
Dx74332	0.330275229	93	98
S79819A	0.321100917	94	215
Dx1740	0.321100917	94	215
Dx74332	0.321100917	94	215
T83510d	0.321100917	94	215
S79819A	0.220183486	105	107
Dx17311, S14104d	0.220183486	105	107
Dx36847	0.220183486	105	107
Dx37231	0.220183486	105	203
T83510d	0.220183486	105	107
Z578	0.220183486	105	107
S79819A, T83510D	0.201834862	107	183
Dx17311, Z578	0.201834862	107	183
Dx36842	0.201834862	107	183
T80211d	0.201834862	107	183
Z578	0.201834862	107	183
S79819A, T83510D	0.009174312	128	187
Dx17311, Z578	0.009174312	128	187

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.009174312	128	187
T80211d	0.009174312	128	187
Z578	0.009174312	128	187
Dx17311, Z578	0.666666667	56	126
Dx1740	0.666666667	56	126
T381x5a	0.666666667	56	126
Z578	0.666666667	56	126
T83510D	0.333333333	92	94
Dx1740	0.333333333	92	94
Dx74332	0.333333333	92	94
S79819A	0.222222222	104	114
Dx17311, S14104d	0.222222222	104	114
Dx36847	0.222222222	104	114
T83510d	0.222222222	104	114
Z578	0.222222222	104	114
T83510D	0.336448598	91	92
Dx1740	0.336448598	91	92
Dx74332	0.336448598	91	92
S79819A	0.327102804	92	195
Dx1740	0.327102804	92	195
Dx74332	0.327102804	92	195
T83510d	0.327102804	92	195
S79819A, T83510D	0.22	103	106
S79819A	0.224299065	103	106
Dx17311, S14104d	0.224299065	103	106
Dx36847	0.224299065	103	106
T83510d	0.224299065	103	106
Z578	0.224299065	103	106
S79819A, T83510D	0.205607477	105	121
Dx17311, Z578	0.205607477	105	121
Dx36842	0.205607477	105	121
T80211d	0.205607477	105	121
Z578	0.205607477	105	121
S79819A, T83510D	0.018691589	125	126
Dx17311, Z578	0.018691589	125	126
Dx36842	0.018691589	125	126
T80211d	0.018691589	125	126
Z578	0.018691589	125	126
S79819A, T83510D	0.009345794	126	182
Dx17311, Z578	0.009345794	126	182

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.009345794	126	182
T80211d	0.009345794	126	182
Z578	0.009345794	126	182
Dx17311, Z578	0.79245283	42	156
Dx36842	0.79245283	42	156
T381x5a	0.79245283	42	156
Z578	0.79245283	42	156
Dx17311, Z578	0.679245283	54	126
Dx1740	0.679245283	54	126
T381x5a	0.679245283	54	126
Z578	0.679245283	54	126
Dx17311, Z578	0.566037736	66	91
Dx1740	0.566037736	66	91
s14104d	0.566037736	66	91
Z578	0.566037736	66	91
Dx17311, Z578	0.452830189	78	102
Dx1740	0.452830189	78	102
Dx74332	0.452830189	78	102
Z578	0.452830189	78	102
T83510D	0.339622642	90	91
Dx1740	0.339622642	90	91
Dx74332	0.339622642	90	91
S79819A	0.330188679	91	113
Dx1740	0.330188679	91	113
Dx74332	0.330188679	91	113
T83510d	0.330188679	91	113
S79819A, T83510D	0.23	102	172
S79819A	0.226415094	102	103
Dx17311, S14104d	0.226415094	102	103
Dx36847	0.226415094	102	103
T83510d	0.226415094	102	103
Z578	0.226415094	102	103
Dx17311, Z578	0.8	41	220
Dx36842	0.8	41	220
T381x5a	0.8	41	220
Z578	0.8	41	220
Dx17311, Z578	0.571428571	65	96
Dx1740	0.571428571	65	96
s14104d	0.571428571	65	96
Z578	0.571428571	65	96

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.457142857	77	92
Dx1740	0.457142857	77	92
Dx74332	0.457142857	77	92
Z578	0.457142857	77	92
T83510D	0.342857143	89	93
Dx1740	0.342857143	89	93
Dx74332	0.342857143	89	93
S79819A	0.333333333	90	130
Dx1740	0.333333333	90	130
Dx74332	0.333333333	90	130
T83510d	0.333333333	90	130
S79819A	0.323809524	91	99
Dx1740	0.323809524	91	99
Dx74332	0.323809524	91	99
T83510d	0.323809524	91	99
S79819A	0.228571429	101	115
Dx17311, S14104d	0.228571429	101	115
Dx36847	0.228571429	101	115
T83510d	0.228571429	101	115
Z578	0.228571429	101	115
Dx17311, Z578	0.807692308	40	120
Dx36842	0.807692308	40	120
T381x5a	0.807692308	40	120
Z578	0.807692308	40	120
Dx17311, Z578	0.548076923	67	121
Dx1740	0.548076923	67	121
s14104d	0.548076923	67	121
Z578	0.548076923	67	121
Dx17311, Z578	0.461538462	76	143
Dx1740	0.461538462	76	143
Dx74332	0.461538462	76	143
Z578	0.461538462	76	143
T83510D	0.346153846	88	97
Dx1740	0.346153846	88	97
Dx74332	0.346153846	88	97
S79819A	0.336538462	89	142
Dx1740	0.336538462	89	142
Dx74332	0.336538462	89	142
T83510d	0.336538462	89	142
S79819A	0.317307692	91	97

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.317307692	91	97
Dx74332	0.317307692	91	97
T83510d	0.317307692	91	97
S79819A	0.230769231	100	101
Dx17311, S14104d	0.230769231	100	101
Dx36847	0.230769231	100	101
T83510d	0.230769231	100	101
Z578	0.230769231	100	101
S79819A, T83510D	0.221153846	101	169
Dx17311, Z578	0.221153846	101	169
Dx36842	0.221153846	101	169
T80211d	0.221153846	101	169
Z578	0.221153846	101	169
Dx17311, Z578	0.699029126	51	294
Dx1740	0.699029126	51	294
T381x5a	0.699029126	51	294
Z578	0.699029126	51	294
Dx17311, Z578	0.582524272	63	91
Dx1740	0.582524272	63	91
s14104d	0.582524272	63	91
Z578	0.582524272	63	91
Dx17311	0.466019417	75	110
Dx17311, Z578	0.466019417	75	122
Dx1740	0.466019417	75	110
Dx74332	0.466019417	75	110
Z578	0.466019417	75	110
Dx17311	0.45631068	76	146
Dx1740	0.45631068	76	146
Dx74332	0.45631068	76	146
Z578	0.45631068	76	146
T83510D	0.349514563	87	91
Dx1740	0.349514563	87	91
Dx74332	0.349514563	87	91
S79819A	0.233009709	99	100
Dx17311, S14104d	0.233009709	99	100
Dx36847	0.233009709	99	100
T83510d	0.233009709	99	100
Z578	0.233009709	99	100
Dx17311, Z578	0.588235294	62	117
Dx1740	0.588235294	62	117

Covariates	Probability	Minimum RVI	Maximum RVI
s14104d	0.588235294	62	117
Z578	0.588235294	62	117
T83510D	0.352941176	86	91
Dx1740	0.352941176	86	91
Dx74332	0.352941176	86	91
S79819A	0.235294118	98	99
Dx17311	0.235294118	98	99
Dx36847	0.235294118	98	99
T83510d	0.235294118	98	99
Z578	0.235294118	98	99
S79819A, T83510D	0.215686275	100	137
Dx17311, Z578	0.215686275	100	137
Dx36842	0.215686275	100	137
T80211d	0.215686275	100	137
Z578	0.215686275	100	137
S79819A, T83510D	0.009803922	121	134
Dx17311, Z578	0.009803922	121	134
Dx36842	0.009803922	121	134
T80211d	0.009803922	121	134
Z578	0.009803922	121	134
Dx17311, Z578	0.712871287	49	75
Dx1740	0.712871287	49	75
T381x5a	0.712871287	49	75
Z578	0.712871287	49	75
Dx17311, Z578	0.594059406	61	143
Dx1740	0.594059406	61	143
s14104d	0.594059406	61	143
Z578	0.594059406	61	143
Dx17311	0.475247525	73	147
Dx1740	0.475247525	73	147
Dx74332	0.475247525	73	147
Z578	0.475247525	73	147
T83510D	0.356435644	85	102
Dx1740	0.356435644	85	102
Dx74332	0.356435644	85	102
S79819A	0.326732673	88	126
Dx1740	0.326732673	88	126
Dx74332	0.326732673	88	126
T83510d	0.326732673	88	126
S79819A, T83510D	0.24	97	98

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.237623762	97	98
Dx17311	0.237623762	97	98
Dx36847	0.237623762	97	98
T83510d	0.237623762	97	98
Z578	0.237623762	97	98
S79819A, T83510D	0.01980198	119	126
Dx17311, Z578	0.01980198	119	126
Dx36842	0.01980198	119	126
T80211d	0.01980198	119	126
Z578	0.01980198	119	126
Dx17311, Z578	0.6	60	80
Dx1740	0.6	60	80
s14104d	0.6	60	80
Z578	0.6	60	80
Dx17311	0.48	72	111
Dx1740	0.48	72	111
Dx74332	0.48	72	111
Z578	0.48	72	111
S79819A, T83510D	0.36	84	100
T83510D	0.36	84	86
Dx1740	0.36	84	86
Dx74332	0.36	84	86
S79819A	0.24	96	97
Dx17311	0.24	96	97
Dx36847	0.24	96	97
T83510d	0.24	96	97
Z578	0.24	96	97
S79819A, T83510D	0.01	119	182
Dx17311, Z578	0.01	119	182
Dx36842	0.01	119	182
T80211d	0.01	119	182
Z578	0.01	119	182
Dx17311	0.484848485	71	119
Dx1740	0.484848485	71	119
Dx74332	0.484848485	71	119
Z578	0.484848485	71	119
T83510D	0.363636364	83	91
Dx1740	0.363636364	83	91
Dx74332	0.363636364	83	91
S79819A	0.353535354	84	98

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.353535354	84	98
Dx74332	0.353535354	84	98
T83510d	0.353535354	84	98
S79819A	0.242424242	95	98
Dx17311	0.242424242	95	98
Dx36847	0.242424242	95	98
T83510d	0.242424242	95	98
Z578	0.242424242	95	98
Dx17311, Z578	0.734693878	46	119
Dx1740	0.734693878	46	119
T381x5a	0.734693878	46	119
Z578	0.734693878	46	119
Dx17311, Z578	0.612244898	58	126
Dx1740	0.612244898	58	126
s14104d	0.612244898	58	126
Z578	0.612244898	58	126
S79819A, T83510D	0.49	70	123
Dx17311	0.489795918	70	90
Dx1740	0.489795918	70	90
Dx74332	0.489795918	70	90
Z578	0.489795918	70	90
T83510D	0.367346939	82	104
Dx1740	0.367346939	82	104
Dx74332	0.367346939	82	104
S79819A, T83510D	0.24	94	169
S79819A	0.244897959	94	97
Dx17311	0.244897959	94	97
Dx36847	0.244897959	94	97
T83510d	0.244897959	94	97
Z578	0.244897959	94	97
Dx17311, Z578	0.742268041	45	102
Dx1740	0.742268041	45	102
T381x5a	0.742268041	45	102
Z578	0.742268041	45	102
Dx17311, Z578	0.618556701	57	88
Dx1740	0.618556701	57	88
s14104d	0.618556701	57	88
Z578	0.618556701	57	88
Dx17311	0.494845361	69	106
Dx1740	0.494845361	69	106

Covariates	Probability	Minimum RVI	Maximum RVI
Dx74332	0.494845361	69	106
Z578	0.494845361	69	106
Dx17311	0.484536082	70	95
Dx1740	0.484536082	70	95
Dx74332	0.484536082	70	95
Z578	0.484536082	70	95
T83510D	0.371134021	81	91
Dx1740	0.371134021	81	91
Dx74332	0.371134021	81	91
S79819A	0.340206186	84	112
Dx1740	0.340206186	84	112
Dx74332	0.340206186	84	112
T83510d	0.340206186	84	112
S79819A	0.24742268	93	94
Dx17311	0.24742268	93	94
Dx36847	0.24742268	93	94
T83510d	0.24742268	93	94
Z578	0.24742268	93	94
Dx17311, Z578	0.625	56	91
Dx1740	0.625	56	91
s14104d	0.625	56	91
Z578	0.625	56	91
Dx17311	0.5	68	92
Dx1740	0.5	68	92
Dx74332	0.5	68	92
Z578	0.5	68	92
T83510D	0.375	80	89
Dx1740	0.375	80	89
Dx74332	0.375	80	89
S79819A	0.25	92	93
Dx17311, Dx7433	0.25	92	93
Dx17311	0.25	92	99
Dx36847	0.25	92	93
T83510d	0.25	92	93
Z578	0.25	92	93
S79819A, T83510D	0.239583333	93	145
Dx17311, Z578	0.239583333	93	145
Dx36842	0.239583333	93	145
T80211d	0.239583333	93	145
Z578	0.239583333	93	145

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.757894737	43	218
Dx1740	0.757894737	43	218
T381x5a	0.757894737	43	218
Z578	0.757894737	43	218
Dx17311, Z578	0.631578947	55	98
Dx1740	0.631578947	55	98
s14104d	0.631578947	55	98
Z578	0.631578947	55	98
Dx17311	0.505263158	67	113
Dx1740	0.505263158	67	113
Dx74332	0.505263158	67	113
Z578	0.505263158	67	113
T83510D	0.378947368	79	92
Dx1740	0.378947368	79	92
Dx74332	0.378947368	79	92
S79819A, T83510D	0.25	91	153
S79819A	0.252631579	91	92
Dx17311, Dx7433	0.252631579	91	92
Dx36842	0.252631579	91	92
Dx36847	0.252631579	91	95
T83510d	0.252631579	91	92
Z578	0.252631579	91	92
S79819A, T83510D	0.242105263	92	96
Dx17311, Z578	0.242105263	92	96
Dx36842	0.242105263	92	96
T80211d	0.242105263	92	96
Z578	0.242105263	92	96
Dx17311, Z578	0.765957447	42	91
Dx1740	0.765957447	42	91
T381x5a	0.765957447	42	91
Z578	0.765957447	42	91
Dx17311, Z578	0.638297872	54	63
Dx1740	0.638297872	54	63
s14104d	0.638297872	54	63
Z578	0.638297872	54	63
Dx17311	0.510638298	66	126
Dx1740	0.510638298	66	126
Dx74332	0.510638298	66	126
Z578	0.510638298	66	126
T83510D	0.382978723	78	91

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.382978723	78	91
Dx74332	0.382978723	78	91
S79819A	0.255319149	90	91
Dx17311	0.255319149	90	91
Dx17311, Dx7433	0.255319149	90	102
Dx36842	0.255319149	90	91
T83510d	0.255319149	90	91
Z578	0.255319149	90	91
S79819A, T83510D	0.244680851	91	133
Dx17311, Z578	0.244680851	91	133
Dx36842	0.244680851	91	133
T80211d	0.244680851	91	133
Z578	0.244680851	91	133
Dx17311, Z578	0.774193548	41	182
Dx1740	0.774193548	41	182
T381x5a	0.774193548	41	182
Z578	0.774193548	41	182
Dx17311, Z578	0.64516129	53	91
Dx1740	0.64516129	53	91
s14104d	0.64516129	53	91
Z578	0.64516129	53	91
Dx17311	0.516129032	65	98
Dx1740	0.516129032	65	98
Dx74332	0.516129032	65	98
Z578	0.516129032	65	98
T83510D	0.387096774	77	94
Dx1740	0.387096774	77	94
Dx74332	0.387096774	77	94
S79819A	0.258064516	89	90
Dx17311	0.258064516	89	90
Dx36842	0.258064516	89	90
T83510d	0.258064516	89	90
Z578	0.258064516	89	90
S79819A, T83510D	0.247311828	90	122
Dx17311, Z578	0.247311828	90	122
Dx36842	0.247311828	90	122
T80211d	0.247311828	90	122
Z578	0.247311828	90	122
S79819A, T83510D	0.23655914	91	98
Dx17311, Z578	0.23655914	91	98

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.23655914	91	98
T80211d	0.23655914	91	98
Z578	0.23655914	91	98
Dx17311, Z578	0.782608696	40	196
Dx1740	0.782608696	40	196
T381x5a	0.782608696	40	196
Z578	0.782608696	40	196
Dx17311, Z578	0.652173913	52	133
Dx1740	0.652173913	52	133
s14104d	0.652173913	52	133
Z578	0.652173913	52	133
Dx17311	0.52173913	64	105
Dx1740	0.52173913	64	105
Dx74332	0.52173913	64	105
Z578	0.52173913	64	105
T83510D	0.391304348	76	91
Dx1740	0.391304348	76	91
Dx74332	0.391304348	76	91
S79819A	0.369565217	78	196
Dx1740	0.369565217	78	196
Dx74332	0.369565217	78	196
T83510d	0.369565217	78	196
S79819A	0.260869565	88	91
Dx17311	0.260869565	88	91
Dx36842	0.260869565	88	91
T83510d	0.260869565	88	91
Z578	0.260869565	88	91
Dx17311	0.527472527	63	91
Dx1740	0.527472527	63	91
Dx74332	0.527472527	63	91
Z578	0.527472527	63	91
S79819A, T83510D	0.40	75	97
T83510D	0.395604396	75	97
Dx1740	0.395604396	75	97
Dx74332	0.395604396	75	97
S79819A	0.263736264	87	88
Dx17311	0.263736264	87	88
Dx36842	0.263736264	87	88
T83510d	0.263736264	87	88
Z578	0.263736264	87	88

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.666666667	50	134
Dx1740	0.666666667	50	134
s14104d	0.666666667	50	134
Z578	0.666666667	50	134
Dx17311	0.533333333	62	123
Dx1740	0.533333333	62	123
Dx74332	0.533333333	62	123
Z578	0.533333333	62	123
T83510D	0.4	74	96
Dx1740	0.4	74	96
Dx74332	0.4	74	96
S79819A	0.377777778	76	117
Dx1740	0.377777778	76	117
Dx74332	0.377777778	76	117
T83510d	0.377777778	76	117
S79819A	0.266666667	86	89
Dx17311	0.266666667	86	89
Dx36842	0.266666667	86	89
T83510d	0.266666667	86	89
Z578	0.266666667	86	89
S79819A, T83510D	0.255555556	87	224
Dx17311, Z578	0.255555556	87	224
Dx36842	0.255555556	87	224
T80211d	0.255555556	87	224
Z578	0.255555556	87	224
S79819A, T83510D	0.011111111	109	222
Dx17311, Z578	0.011111111	109	222
Dx36842	0.011111111	109	222
T80211d	0.011111111	109	222
Z578	0.011111111	109	222
Dx17311, Z578	0.674157303	49	119
Dx1740	0.674157303	49	119
s14104d	0.674157303	49	119
Z578	0.674157303	49	119
Dx17311	0.539325843	61	119
Dx1740	0.539325843	61	119
Dx74332	0.539325843	61	119
Z578	0.539325843	61	119
Dx17311	0.516853933	63	131
Dx1740	0.516853933	63	131

Covariates	Probability	Minimum RVI	Maximum RVI
Dx74332	0.516853933	63	131
Z578	0.516853933	63	131
T83510D	0.404494382	73	140
Dx1740	0.404494382	73	140
Dx74332	0.404494382	73	140
S79819A, T83510D	0.269662921	85	86
S79819A	0.269662921	85	127
Dx17311	0.269662921	85	86
Dx36842	0.269662921	85	86
T83510d	0.269662921	85	86
Z578	0.269662921	85	86
S79819A, T83510D	0.258426966	86	93
Dx17311, Z578	0.258426966	86	93
Dx36842	0.258426966	86	93
T80211d	0.258426966	86	93
Z578	0.258426966	86	93
S79819A, T83510D	0.02247191	107	183
Dx17311, Z578	0.02247191	107	183
Dx36842	0.02247191	107	183
T80211d	0.02247191	107	183
Z578	0.02247191	107	183
Dx17311, Z578	0.772727273	40	156
Dx1740	0.772727273	40	156
T381x5a	0.772727273	40	156
Z578	0.772727273	40	156
Dx17311, Z578	0.681818182	48	84
Dx1740	0.681818182	48	84
s14104d	0.681818182	48	84
Z578	0.681818182	48	84
Dx17311	0.545454545	60	93
Dx1740	0.545454545	60	93
Dx74332	0.545454545	60	93
Z578	0.545454545	60	93
Dx17311	0.534090909	61	128
Dx1740	0.534090909	61	128
Dx74332	0.534090909	61	128
Z578	0.534090909	61	128
T83510D	0.409090909	72	159
Dx1740	0.409090909	72	159
Dx74332	0.409090909	72	159

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A, T83510D	0.272727273	84	89
Dx17311 , Dx37231	0.272727273	84	89
Dx17311	0.272727273	84	160
Dx36842	0.272727273	84	89
T83510d	0.272727273	84	89
Z578	0.272727273	84	89
Dx17311, Z578	0.689655172	47	116
Dx1740	0.689655172	47	116
s14104d	0.689655172	47	116
Z578	0.689655172	47	116
Dx17311	0.551724138	59	92
Dx1740	0.551724138	59	92
Dx74332	0.551724138	59	92
Z578	0.551724138	59	92
Dx17311	0.540229885	60	151
Dx1740	0.540229885	60	151
Dx74332	0.540229885	60	151
Z578	0.540229885	60	151
T83510D	0.413793103	71	91
Dx1740	0.413793103	71	91
Dx74332	0.413793103	71	91
S79819A, T83510D	0.275862069	83	84
Dx17311 , Dx37231	0.275862069	83	84
Dx36842	0.275862069	83	84
T83510d	0.275862069	83	84
Z578	0.275862069	83	84
S79819A, T83510D	0.264367816	84	100
Dx17311, Z578	0.264367816	84	100
Dx36842	0.264367816	84	100
T80211d	0.264367816	84	100
Z578	0.264367816	84	100
S79819A, T83510D	0.022988506	105	121
Dx17311, Z578	0.022988506	105	121
Dx36842	0.022988506	105	121
T80211d	0.022988506	105	121
Z578	0.022988506	105	121
Dx17311, Z578	0.697674419	46	87
Dx1740	0.697674419	46	87
s14104d	0.697674419	46	87
Z578	0.697674419	46	87

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311	0.558139535	58	77
Dx1740	0.558139535	58	77
Dx74332	0.558139535	58	77
Z578	0.558139535	58	77
T83510D	0.418604651	70	77
Dx1740	0.418604651	70	77
Dx74332	0.418604651	70	77
S79819A, T83510D	0.279069767	82	84
Dx17311 , Dx37231	0.279069767	82	84
Dx36842	0.279069767	82	84
T83510d	0.279069767	82	84
Z578	0.279069767	82	84
S79819A, T83510D	0.26744186	83	84
Dx17311, Z578	0.26744186	83	84
Dx36842	0.26744186	83	84
T80211d	0.26744186	83	84
Z578	0.26744186	83	84
Dx17311, Z578	0.705882353	45	90
Dx1740	0.705882353	45	90
s14104d	0.705882353	45	90
Z578	0.705882353	45	90
Dx17311	0.564705882	57	119
Dx1740	0.564705882	57	119
Dx74332	0.564705882	57	119
Z578	0.564705882	57	119
Dx17311	0.552941176	58	77
Dx1740	0.552941176	58	77
Dx74332	0.552941176	58	77
Z578	0.552941176	58	77
T83510D	0.423529412	69	107
Dx1740	0.423529412	69	107
Dx74332	0.423529412	69	107
S79819A	0.411764706	70	107
Dx1740	0.411764706	70	107
Dx74332	0.411764706	70	107
T83510d	0.411764706	70	107
S79819A, T83510D	0.282352941	81	91
Dx17311 , Dx37231	0.282352941	81	91
Dx36842	0.282352941	81	91
T83510d	0.282352941	81	91

Covariates	Probability	Minimum RVI	Maximum RVI
Z578	0.282352941	81	91
Dx17311, Z578	0.714285714	44	154
Dx1740	0.714285714	44	154
s14104d	0.714285714	44	154
Z578	0.714285714	44	154
Dx17311, Z578	0.678571429	47	112
Dx1740	0.678571429	47	112
s14104d	0.678571429	47	112
Z578	0.678571429	47	112
Dx17311	0.571428571	56	64
Dx1740	0.571428571	56	64
Dx74332	0.571428571	56	64
Z578	0.571428571	56	64
S79819A, T83510D	0.43	68	77
T83510D	0.428571429	68	77
Dx1740	0.428571429	68	77
Dx74332	0.428571429	68	77
S79819A	0.416666667	69	231
Dx1740	0.416666667	69	231
Dx74332	0.416666667	69	231
T83510d	0.416666667	69	231
S79819A	0.404761905	70	159
Dx1740	0.404761905	70	159
Dx74332	0.404761905	70	159
T83510d	0.404761905	70	159
S79819A, T83510D	0.285714286	80	92
Dx17311 , Dx37231	0.285714286	80	92
Dx36842	0.285714286	80	92
T83510d	0.285714286	80	92
Z578	0.285714286	80	92
S79819A, T83510D	0.273809524	81	126
Dx17311, Z578	0.273809524	81	126
Dx36842	0.273809524	81	126
T80211d	0.273809524	81	126
Z578	0.273809524	81	126
Dx17311, Z578	0.722891566	43	120
Dx1740	0.722891566	43	120
s14104d	0.722891566	43	120
Z578	0.722891566	43	120
Dx17311	0.578313253	55	99

Covariates	Probability	Minimum RVI	Maximum RVI
Dx1740	0.578313253	55	99
Dx74332	0.578313253	55	99
Z578	0.578313253	55	99
S79819A, T83510D	0.43	67	71
T83510D	0.43373494	67	71
Dx1740	0.43373494	67	71
Dx74332	0.43373494	67	71
S79819A, T83510D	0.289156627	79	92
Dx17311 , Dx37231	0.289156627	79	92
Dx36842	0.289156627	79	92
T83510d	0.289156627	79	92
Z578	0.289156627	79	92
Dx17311, Z578	0.731707317	42	91
Dx1740	0.731707317	42	91
s14104d	0.731707317	42	91
Z578	0.731707317	42	91
Dx17311	0.573170732	55	197
Dx1740	0.573170732	55	197
Dx74332	0.573170732	55	197
Z578	0.573170732	55	197
T83510D	0.43902439	66	105
Dx1740	0.43902439	66	105
Dx74332	0.43902439	66	105
S79819A, T83510D	0.292682927	78	80
Dx17311 , Dx37231	0.292682927	78	80
Dx36842	0.292682927	78	80
T83510d	0.292682927	78	80
Z578	0.292682927	78	80
S79819A, T83510D	0.024390244	100	137
Dx17311, Z578	0.024390244	100	137
Dx36842	0.024390244	100	137
T80211d	0.024390244	100	137
Z578	0.024390244	100	137
S79819A, T83510D	0.012195122	101	169
Dx17311, Z578	0.012195122	101	169
Dx36842	0.012195122	101	169
T80211d	0.012195122	101	169
Z578	0.012195122	101	169
Dx17311, Z578	0.740740741	41	165
Dx1740	0.740740741	41	165

Covariates	Probability	Minimum RVI	Maximum RVI
s14104d	0.740740741	41	165
Z578	0.740740741	41	165
Dx17311, Z578	0.728395062	42	84
Dx1740	0.728395062	42	84
s14104d	0.728395062	42	84
Z578	0.728395062	42	84
Dx17311	0.592592593	53	62
Dx1740	0.592592593	53	62
Dx74332	0.592592593	53	62
Z578	0.592592593	53	62
T83510D	0.444444444	65	84
Dx1740	0.444444444	65	84
Dx74332	0.444444444	65	84
S79819A, T83510D	0.30	77	133
S79819A, T83510D	0.296296296	77	80
Dx17311 , Dx37231	0.296296296	77	80
Dx36842	0.296296296	77	80
T83510d	0.296296296	77	80
Z578	0.296296296	77	80
Dx17311, Z578	0.75	40	122
Dx1740	0.75	40	122
s14104d	0.75	40	122
Z578	0.75	40	122
Dx17311, Z578	0.7375	41	185
Dx1740	0.7375	41	185
s14104d	0.7375	41	185
Z578	0.7375	41	185
Dx17311, Z578	0.725	42	92
Dx1740	0.725	42	92
s14104d	0.725	42	92
Z578	0.725	42	92
Dx17311	0.6	52	84
Dx1740	0.6	52	84
Dx74332	0.6	52	84
Z578	0.6	52	84
T83510D	0.45	64	70
Dx1740	0.45	64	70
Dx74332	0.45	64	70
S79819A, T83510D	0.3	76	91
Dx17311 , Dx37231	0.3	76	91

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.3	76	91
T83510d	0.3	76	91
Z578	0.3	76	91
S79819A, T83510D	0.0125	99	154
Dx17311, Z578	0.0125	99	154
Dx36842	0.0125	99	154
T80211d	0.0125	99	154
Z578	0.0125	99	154
Dx17311	0.607594937	51	110
Dx1740	0.607594937	51	110
Dx74332	0.607594937	51	110
Z578	0.607594937	51	110
Dx17311	0.594936709	52	283
Dx1740	0.594936709	52	283
Dx74332	0.594936709	52	283
Z578	0.594936709	52	283
S79819A, T83510D	0.46	63	182
T83510D	0.455696203	63	85
Dx1740	0.455696203	63	85
Dx74332	0.455696203	63	85
S79819A, T83510D	0.303797468	75	81
Dx17311 , Dx37231	0.303797468	75	81
Dx36842	0.303797468	75	81
T83510d	0.303797468	75	81
Z578	0.303797468	75	81
Dx17311	0.615384615	50	198
Dx1740	0.615384615	50	198
Dx74332	0.615384615	50	198
Z578	0.615384615	50	198
T83510D	0.461538462	62	72
Dx1740	0.461538462	62	72
Dx74332	0.461538462	62	72
S79819A	0.448717949	63	144
Dx1740	0.448717949	63	144
Dx74332	0.448717949	63	144
T83510d	0.448717949	63	144
S79819A, T83510D	0.307692308	74	91
Dx17311 , Dx37231	0.307692308	74	91
Dx36842	0.307692308	74	91
T83510d	0.307692308	74	91

Covariates	Probability	Minimum RVI	Maximum RVI
Z578	0.307692308	74	91
Dx17311	0.623376623	49	71
Dx1740	0.623376623	49	71
Dx74332	0.623376623	49	71
Z578	0.623376623	49	71
T83510D	0.467532468	61	91
Dx1740	0.467532468	61	91
Dx74332	0.467532468	61	91
S79819A	0.454545455	62	119
Dx1740	0.454545455	62	119
Dx74332	0.454545455	62	119
T83510d	0.454545455	62	119
S79819A, T83510D	0.311688312	73	80
Dx17311 , Dx37231	0.311688312	73	80
Dx36842	0.311688312	73	80
T83510d	0.311688312	73	80
Z578	0.311688312	73	80
S79819A, T83510D	0.012987013	96	141
Dx17311, Z578	0.012987013	96	141
Dx36842	0.012987013	96	141
T80211d	0.012987013	96	141
Z578	0.012987013	96	141
Dx17311	0.631578947	48	126
Dx1740	0.631578947	48	126
Dx74332	0.631578947	48	126
Z578	0.631578947	48	126
Dx17311	0.605263158	50	81
Dx1740	0.605263158	50	81
Dx74332	0.605263158	50	81
Z578	0.605263158	50	81
T83510D	0.473684211	60	94
Dx1740	0.473684211	60	94
Dx74332	0.473684211	60	94
S79819A, T83510D	0.315789474	72	77
Dx17311 , Dx37231	0.315789474	72	77
Dx36842	0.315789474	72	77
T83510d	0.315789474	72	77
Z578	0.315789474	72	77
S79819A, T83510D	0.052631579	92	162
Dx17311, Z578	0.052631579	92	162

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.052631579	92	162
T80211d	0.052631579	92	162
Z578	0.052631579	92	162
Dx17311	0.64	47	69
Dx1740	0.64	47	69
Dx74332	0.64	47	69
Z578	0.64	47	69
T83510D	0.48	59	82
Dx1740	0.48	59	82
Dx74332	0.48	59	82
S79819A, T83510D	0.32	71	85
Dx17311 , Dx37231	0.32	71	85
Dx36842	0.32	71	85
T83510d	0.32	71	85
Z578	0.32	71	85
S79819A, T83510D	0.306666667	72	130
Dx17311, Z578	0.306666667	72	130
Dx36842	0.306666667	72	130
T80211d	0.306666667	72	130
Z578	0.306666667	72	130
S79819A, T83510D	0.013333333	94	215
Dx17311, Z578	0.013333333	94	215
Dx36842	0.013333333	94	215
T80211d	0.013333333	94	215
Z578	0.013333333	94	215
Dx17311	0.648648649	46	121
Dx1740	0.648648649	46	121
Dx74332	0.648648649	46	121
Z578	0.648648649	46	121
T83510D	0.486486486	58	133
Dx1740	0.486486486	58	133
Dx74332	0.486486486	58	133
S79819A	0.459459459	60	98
Dx1740	0.459459459	60	98
Dx74332	0.459459459	60	98
T83510d	0.459459459	60	98
S79819A, T83510D	0.324324324	70	73
Dx17311 , Dx37231	0.324324324	70	73
Dx36842	0.324324324	70	73
T83510d	0.324324324	70	73

Covariates	Probability	Minimum RVI	Maximum RVI
Z578	0.324324324	70	73
S79819A, T83510D	0.310810811	71	91
Dx17311, Z578	0.310810811	71	91
Dx36842	0.310810811	71	91
T80211d	0.310810811	71	91
Z578	0.310810811	71	91
S79819A, T83510D	0.040540541	91	97
Dx17311, Z578	0.040540541	91	97
Dx36842	0.040540541	91	97
T80211d	0.040540541	91	97
Z578	0.040540541	91	97
S79819A, T83510D	0.027027027	92	222
Dx17311, Z578	0.027027027	92	222
Dx36842	0.027027027	92	222
T80211d	0.027027027	92	222
Z578	0.027027027	92	222
S79819A, T83510D	0.013513514	93	145
Dx17311, Z578	0.013513514	93	145
Dx36842	0.013513514	93	145
T80211d	0.013513514	93	145
Z578	0.013513514	93	145
Dx17311	0.657534247	45	105
Dx1740	0.657534247	45	105
Dx74332	0.657534247	45	105
Z578	0.657534247	45	105
T83510D	0.493150685	57	92
Dx1740	0.493150685	57	92
Dx74332	0.493150685	57	92
S79819A, T83510D	0.328767123	69	70
Dx17311 , Dx37231	0.328767123	69	70
Dx36842	0.328767123	69	70
T83510d	0.328767123	69	70
Z578	0.328767123	69	70
S79819A, T83510D	0.02739726	91	98
Dx17311, Z578	0.02739726	91	98
Dx36842	0.02739726	91	98
T80211d	0.02739726	91	98
Z578	0.02739726	91	98
S79819A, T83510D	0.01369863	92	96
Dx17311, Z578	0.01369863	92	96

Covariates	Probability	Minimum RVI	Maximum RVI
Dx36842	0.01369863	92	96
T5694xa	0.01369863	92	96
Z578	0.01369863	92	96
Dx17311	0.666666667	44	99
Dx1740	0.666666667	44	99
Dx74332	0.666666667	44	99
Z578	0.666666667	44	99
T83510D	0.5	56	62
Dx1740	0.5	56	62
Dx74332	0.5	56	62
S79819A, T83510D	0.333333333	68	77
Dx17311 , Dx37231	0.333333333	68	77
Dx36842	0.333333333	68	77
T83510d	0.333333333	68	77
Z578	0.333333333	68	77
S79819A, T83510D	0.027777778	90	91
Dx17311, Z578	0.027777778	90	91
Dx36842	0.027777778	90	91
T80211d	0.027777778	90	91
Z578	0.027777778	90	91
S79819A, T83510D	0.013888889	91	113
Dx17311, Z578	0.013888889	91	113
Dx36842	0.013888889	91	113
T5694xa	0.013888889	91	113
Z578	0.013888889	91	113
Dx17311	0.676056338	43	147
Dx1740	0.676056338	43	147
Dx74332	0.676056338	43	147
Z578	0.676056338	43	147
T83510D	0.507042254	55	85
Dx1740	0.507042254	55	85
Dx74332	0.507042254	55	85
S79819A	0.492957746	56	95
Dx1740	0.492957746	56	95
Dx74332	0.492957746	56	95
T83510d	0.492957746	56	95
S79819A, T83510D	0.338028169	67	90
Dx17311 , Dx37231	0.338028169	67	90
Dx36842	0.338028169	67	90
T83510d	0.338028169	67	90

Covariates	Probability	Minimum RVI	Maximum RVI
Z578	0.338028169	67	90
S79819A, T83510D	0.042253521	88	126
Dx17311, Z578	0.042253521	88	126
Dx36842	0.042253521	88	126
T80211d	0.042253521	88	126
Z578	0.042253521	88	126
S79819A, T83510D	0.014084507	90	122
Dx17311, Z578	0.014084507	90	122
Dx36842	0.014084507	90	122
T5694xa	0.014084507	90	122
Z578	0.014084507	90	122
Dx17311	0.685714286	42	77
Dx1740	0.685714286	42	77
Dx74332	0.685714286	42	77
Z578	0.685714286	42	77
T83510D	0.514285714	54	98
Dx1740	0.514285714	54	98
Dx74332	0.514285714	54	98
S79819A, T83510D	0.342857143	66	72
Dx17311 , Dx37231	0.342857143	66	72
Dx36842	0.342857143	66	72
T83510d	0.342857143	66	72
Z578	0.342857143	66	72
S79819A, T83510D	0.014285714	89	142
Dx17311, Z578	0.014285714	89	142
Dx36842	0.014285714	89	142
T5694xa	0.014285714	89	142
Z578	0.014285714	89	142
Dx17311	0.695652174	41	99
Dx1740	0.695652174	41	99
Dx74332	0.695652174	41	99
Z578	0.695652174	41	99
T83510D	0.52173913	53	67
Dx1740	0.52173913	53	67
Dx74332	0.52173913	53	67
S79819A, T83510D	0.347826087	65	68
Dx17311 , Dx37231	0.347826087	65	68
Dx36842	0.347826087	65	68
T83510d	0.347826087	65	68
Z578	0.347826087	65	68

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A, T83510D	0.014492754	88	184
Dx17311, Z578	0.014492754	88	184
Dx36842	0.014492754	88	184
T5694xa	0.014492754	88	184
Z578	0.014492754	88	184
Dx17311	0.705882353	40	105
Dx1740	0.705882353	40	105
Dx74332	0.705882353	40	105
Z578	0.705882353	40	105
T83510D	0.529411765	52	57
Dx1740	0.529411765	52	57
Dx74332	0.529411765	52	57
S79819A, T83510D	0.35	64	98
S79819A, T83510D	0.352941176	64	65
Dx17311 , Dx37231	0.352941176	64	65
Dx36842	0.352941176	64	65
T83510d	0.352941176	64	65
Z578	0.352941176	64	65
S79819A, T83510D	0.34	65	98
S79819A, T83510D	0.338235294	65	67
Dx17311, Z578	0.338235294	65	67
Dx36842	0.338235294	65	67
T80211d	0.338235294	65	67
Z578	0.338235294	65	67
S79819A, T83510D	0.014705882	87	224
Dx17311, Z578	0.014705882	87	224
Dx36842	0.014705882	87	224
T5694xa	0.014705882	87	224
Z578	0.014705882	87	224
T83510D	0.537313433	51	64
Dx1740	0.537313433	51	64
Dx74332	0.537313433	51	64
S79819A	0.507462687	53	182
Dx1740	0.507462687	53	182
Dx74332	0.507462687	53	182
T83510d	0.507462687	53	182
S79819A, T83510D	0.36	63	70
S79819A, T83510D	0.358208955	63	64
Dx17311 , Dx37231	0.358208955	63	64
Dx36842	0.358208955	63	64

Covariates	Probability	Minimum RVI	Maximum RVI
T83510d	0.358208955	63	64
Z578	0.358208955	63	64
S79819A, T83510D	0.044776119	84	112
Dx17311, Z578	0.044776119	84	112
Dx36842	0.044776119	84	112
T80211d	0.044776119	84	112
Z578	0.044776119	84	112
S79819A, T83510D	0.014925373	86	93
Dx17311, Z578	0.014925373	86	93
Dx36842	0.014925373	86	93
T5694xa	0.014925373	86	93
Z578	0.014925373	86	93
T83510D	0.545454545	50	90
Dx1740	0.545454545	50	90
Dx74332	0.545454545	50	90
S79819A, T83510D	0.363636364	62	63
Dx17311 , Dx37231	0.363636364	62	63
Dx36842	0.363636364	62	63
T83510d	0.363636364	62	63
Z578	0.363636364	62	63
S79819A, T83510D	0.348484848	63	67
Dx17311, Z578	0.348484848	63	67
Dx36842	0.348484848	63	67
T80211d	0.348484848	63	67
Z578	0.348484848	63	67
T83510D	0.553846154	49	63
Dx1740	0.553846154	49	63
Dx74332	0.553846154	49	63
S79819A, T83510D	0.37	61	68
S79819A, T83510D	0.369230769	61	63
Dx17311 , Dx37231	0.369230769	61	63
Dx36842	0.369230769	61	63
T83510d	0.369230769	61	63
Z578	0.369230769	61	63
S79819A, T83510D	0.353846154	62	75
Dx17311, Z578	0.353846154	62	75
Dx36842	0.353846154	62	75
T80211d	0.353846154	62	75
Z578	0.353846154	62	75
S79819A, T83510D	0.030769231	83	111

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.030769231	83	111
Dx36842	0.030769231	83	111
T80211d	0.030769231	83	111
Z578	0.030769231	83	111
S79819A, T83510D	0.015384615	84	92
Dx17311, Z578	0.015384615	84	92
Dx36842	0.015384615	84	92
T5694xa	0.015384615	84	92
Z578	0.015384615	84	92
T83510D	0.5625	48	92
Dx1740	0.5625	48	92
Dx74332	0.5625	48	92
S79819A, T83510D	0.375	60	65
Dx17311 , Dx37231	0.375	60	65
Dx36842	0.375	60	65
T83510d	0.375	60	65
Z578	0.375	60	65
S79819A, T83510D	0.015625	83	84
Dx17311, Z578	0.015625	83	84
Dx36842	0.015625	83	84
T5694xa	0.015625	83	84
Z578	0.015625	83	84
T83510D	0.571428571	47	91
Dx1740	0.571428571	47	91
Dx74332	0.571428571	47	91
S79819A, T83510D	0.38	59	92
S79819A, T83510D	0.380952381	59	70
Dx17311	0.380952381	59	70
Dx17311 , Dx37231	0.380952381	59	92
Dx17311 , Dx37231	0.380952381	59	129
Dx36842	0.380952381	59	70
T83510d	0.380952381	59	70
Z578	0.380952381	59	70
S79819A, T83510D	0.365079365	60	98
Dx17311, Z578	0.365079365	60	98
Dx36842	0.365079365	60	98
T80211d	0.365079365	60	98
Z578	0.365079365	60	98
T83510D	0.580645161	46	57
Dx1740	0.580645161	46	57

Covariates	Probability	Minimum RVI	Maximum RVI
Dx74332	0.580645161	46	57
S79819A	0.548387097	48	183
Dx1740	0.548387097	48	183
Dx74332	0.548387097	48	183
T83510d	0.548387097	48	183
S79819A, T83510D	0.387096774	58	63
Dx17311, Z578	0.387096774	58	63
Dx17311	0.387096774	58	110
Dx36842	0.387096774	58	63
T83510d	0.387096774	58	63
Z578	0.387096774	58	63
S79819A, T83510D	0.370967742	59	91
Dx17311, Z578	0.370967742	59	91
Dx36842	0.370967742	59	91
T80211d	0.370967742	59	91
Z578	0.370967742	59	91
S79819A, T83510D	0.016129032	81	126
Dx17311, Z578	0.016129032	81	126
Dx36842	0.016129032	81	126
T5694xa	0.016129032	81	126
Z578	0.016129032	81	126
T83510D	0.590163934	45	61
Dx1740	0.590163934	45	61
Dx74332	0.590163934	45	61
S79819A, T83510D	0.393442623	57	58
Dx17311, Z578	0.393442623	57	58
Dx36842	0.393442623	57	58
T83510d	0.393442623	57	58
Z578	0.393442623	57	58
S79819A, T83510D	0.37704918	58	126
Dx17311, Z578	0.37704918	58	126
Dx36842	0.37704918	58	126
T80211d	0.37704918	58	126
Z578	0.37704918	58	126
T83510D	0.6	44	62
Dx1740	0.6	44	62
Dx74332	0.6	44	62
S79819A, T83510D	0.4	56	57
Dx17311, Z578	0.4	56	57
Dx36842	0.4	56	57

Covariates	Probability	Minimum RVI	Maximum RVI
T83510d	0.4	56	57
Z578	0.4	56	57
S79819A, T83510D	0.366666667	58	131
Dx17311, Z578	0.366666667	58	131
Dx36842	0.366666667	58	131
T80211d	0.366666667	58	131
Z578	0.366666667	58	131
S79819A, T83510D	0.033333333	78	196
Dx17311, Z578	0.033333333	78	196
Dx36842	0.033333333	78	196
T80211d	0.033333333	78	196
Z578	0.033333333	78	196
T83510D	0.610169492	43	63
Dx1740	0.610169492	43	63
Dx74332	0.610169492	43	63
S79819A	0.593220339	44	103
Dx1740	0.593220339	44	103
Dx74332	0.593220339	44	103
T83510d	0.593220339	44	103
S79819A, T83510D	0.406779661	55	56
Dx17311, Z578	0.406779661	55	56
Dx36842	0.406779661	55	56
T83510d	0.406779661	55	56
Z578	0.406779661	55	56
S79819A, T83510D	0.389830508	56	180
Dx17311, Z578	0.389830508	56	180
Dx36842	0.389830508	56	180
T80211d	0.389830508	56	180
Z578	0.389830508	56	180
S79819A	0.620689655	42	53
T83510D	0.620689655	42	223
Dx1740	0.620689655	42	53
Dx74332	0.620689655	42	53
S79819A, T83510D	0.41	54	94
S79819A, T83510D	0.413793103	54	58
Dx17311, Z578	0.413793103	54	58
Dx36842	0.413793103	54	58
T83510d	0.413793103	54	58
Z578	0.413793103	54	58
S79819A, T83510D	0.396551724	55	114

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.396551724	55	114
Dx36842	0.396551724	55	114
T80211d	0.396551724	55	114
Z578	0.396551724	55	114
S79819A, T83510D	0.034482759	76	117
Dx17311, Z578	0.034482759	76	117
Dx36842	0.034482759	76	117
T80211d	0.034482759	76	117
Z578	0.034482759	76	117
S79819A	0.631578947	41	72
Dx1740	0.631578947	41	72
Dx74332	0.631578947	41	72
T83510d	0.631578947	41	72
S79819A	0.614035088	42	70
Dx1740	0.614035088	42	70
Dx74332	0.614035088	42	70
T83510d	0.614035088	42	70
S79819A, T83510D	0.421052632	53	66
Dx17311, Z578	0.421052632	53	66
Dx36842	0.421052632	53	66
T83510d	0.421052632	53	66
Z578	0.421052632	53	66
S79819A, T83510D	0.403508772	54	96
Dx17311, Z578	0.403508772	54	96
Dx36842	0.403508772	54	96
T80211d	0.403508772	54	96
Z578	0.403508772	54	96
S79819A, T83510D	0.01754386	76	146
Dx17311, Z578	0.01754386	76	146
Dx36842	0.01754386	76	146
T5694xa	0.01754386	76	146
Z578	0.01754386	76	146
S79819A	0.642857143	40	93
Dx1740	0.642857143	40	93
Dx74332	0.642857143	40	93
T83510d	0.642857143	40	93
S79819A	0.625	41	78
Dx1740	0.625	41	78
Dx74332	0.625	41	78
T83510d	0.625	41	78

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A	0.607142857	42	56
Dx1740	0.607142857	42	56
Dx74332	0.607142857	42	56
T83510d	0.607142857	42	56
S79819A, T83510D	0.428571429	52	57
Dx17311, Z578	0.428571429	52	57
Dx36842	0.428571429	52	57
	0.428571429	52	57
T83510d	0.428571429	52	69
Z578	0.428571429	52	57
S79819A, T83510D	0.410714286	53	116
Dx17311, Z578	0.410714286	53	116
Dx36842	0.410714286	53	116
T80211d	0.410714286	53	116
Z578	0.410714286	53	116
S79819A	0.6	42	57
Dx1740	0.6	42	57
Dx74332	0.6	42	57
T83510d	0.6	42	57
S79819A, T83510D	0.436363636	51	64
Dx17311, Z578	0.436363636	51	64
Dx36842	0.436363636	51	64
	0.436363636	51	64
Z578	0.436363636	51	64
S79819A, T83510D	0.444444444	50	55
Dx17311, Z578	0.444444444	50	55
Dx36842	0.444444444	50	55
	0.444444444	50	55
Z578	0.444444444	50	55
S79819A, T83510D	0.425925926	51	59
Dx17311, Z578	0.425925926	51	59
Dx36842	0.425925926	51	59
T80211d	0.425925926	51	59
Z578	0.425925926	51	59
S79819A, T83510D	0.452830189	49	51
Dx17311, Z578	0.452830189	49	51
Dx36842	0.452830189	49	51
	0.452830189	49	51
Z578	0.452830189	49	51
S79819A, T83510D	0.018867925	72	130

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.018867925	72	130
Dx36842	0.018867925	72	130
T5694xa	0.018867925	72	130
Z578	0.018867925	72	130
S79819A, T83510D	0.461538462	48	58
Dx17311, Z578	0.461538462	48	58
Dx36842	0.461538462	48	58
	0.461538462	48	58
Z578	0.461538462	48	58
S79819A, T83510D	0.442307692	49	56
Dx17311, Z578	0.442307692	49	56
Dx36842	0.442307692	49	56
T80211d	0.442307692	49	56
Z578	0.442307692	49	56
S79819A, T83510D	0.038461538	70	159
Dx17311, Z578	0.038461538	70	159
Dx36842	0.038461538	70	159
T80211d	0.038461538	70	159
Z578	0.038461538	70	159
S79819A, T83510D	0.019230769	71	91
Dx17311, Z578	0.019230769	71	91
Dx36842	0.019230769	71	91
T5694xa	0.019230769	71	91
Z578	0.019230769	71	91
S79819A, T83510D	0.470588235	47	52
Dx17311, Z578	0.470588235	47	52
Dx36842	0.470588235	47	52
	0.470588235	47	52
Z578	0.470588235	47	52
S79819A, T83510D	0.019607843	70	95
Dx17311, Z578	0.019607843	70	95
Dx36842	0.019607843	70	95
T5694xa	0.019607843	70	95
Z578	0.019607843	70	95
S79819A, T83510D	0.48	46	62
Dx17311, Z578	0.48	46	62
Dx36842	0.48	46	62
	0.48	46	62
Z578	0.48	46	62
S79819A, T83510D	0.06	67	121

Covariates	Probability	Minimum RVI	Maximum RVI
Dx17311, Z578	0.06	67	121
Dx36842	0.06	67	121
T80211d	0.06	67	121
Z578	0.06	67	121
S79819A, T83510D	0.02	69	231
Dx17311, Z578	0.02	69	231
Dx36842	0.02	69	231
T5694xa	0.02	69	231
Z578	0.02	69	231
S79819A, T83510D	0.489795918	45	46
Dx17311, Z578	0.489795918	45	46
Dx36842	0.489795918	45	46
	0.489795918	45	46
Z578	0.489795918	45	46
S79819A, T83510D	0.06122449	66	221
Dx17311, Z578	0.06122449	66	221
Dx36842	0.06122449	66	221
T80211d	0.06122449	66	221
Z578	0.06122449	66	221
S79819A, T83510D	0.5	44	48
Dx17311, Z578	0.5	44	48
Dx36842	0.5	44	48
	0.5	44	48
Z578	0.5	44	48
S79819A, T83510D	0.479166667	45	224
Dx17311, Z578	0.479166667	45	224
Dx36842	0.479166667	45	224
T80211d	0.479166667	45	224
Z578	0.479166667	45	224
S79819A, T83510D	0.510638298	43	49
Dx17311, Z578	0.510638298	43	49
Dx36842	0.510638298	43	49
	0.510638298	43	49
Z578	0.510638298	43	49
S79819A, T83510D	0.52173913	42	44
Dx17311, Z578	0.52173913	42	44
Dx36842	0.52173913	42	44
T80211d	0.52173913	42	44
	0.52173913	42	92
Z578	0.52173913	42	44

Covariates	Probability	Minimum RVI	Maximum RVI
S79819A, T83510D	0.02	65	98
S79819A, T83510D	0.02173913	65	67
Dx17311, Z578	0.02173913	65	67
Dx36842	0.02173913	65	67
T5694xa	0.02173913	65	67
Z578	0.02173913	65	67
S79819A, T83510D	0.533333333	41	42
Dx17311, Z578	0.533333333	41	42
Dx36842	0.533333333	41	42
T80211d	0.533333333	41	42
Z578	0.533333333	41	42
S79819A, T83510D	0.511111111	42	118
Dx17311, Z578	0.511111111	42	118
Dx36842	0.511111111	42	118
T80211d	0.511111111	42	118
Z578	0.511111111	42	118
S79819A, T83510D	0.044444444	63	131
Dx17311, Z578	0.044444444	63	131
Dx36842	0.044444444	63	131
T80211d	0.044444444	63	131
Z578	0.044444444	63	131
S79819A, T83510D	0.55	40	122
S79819A, T83510D	0.545454545	40	42
Dx17311, Z578	0.545454545	40	42
Dx36842	0.545454545	40	42
T80211d	0.545454545	40	42
Z578	0.545454545	40	42
S79819A, T83510D	0.022727273	63	67
Dx17311, Z578	0.022727273	63	67
Dx36842	0.022727273	63	67
T5694xa	0.022727273	63	67
Z578	0.022727273	63	67
S79819A, T83510D	0.534883721	40	126
Dx17311, Z578	0.534883721	40	126
Dx36842	0.534883721	40	126
T80211d	0.534883721	40	126
Z578	0.534883721	40	126
S79819A, T83510D	0.023255814	62	75
Dx17311, Z578	0.023255814	62	75
Dx36842	0.023255814	62	75

Covariates	Probability	Minimum RVI	Maximum RVI
T5694xa	0.023255814	62	75
Z578	0.023255814	62	75
S79819A, T83510D	0.047619048	60	98
Dx17311, Z578	0.047619048	60	98
Dx36842	0.047619048	60	98
T80211d	0.047619048	60	98
Z578	0.047619048	60	98
S79819A, T83510D	0.023809524	61	128
Dx17311, Z578	0.023809524	61	128
Dx36842	0.023809524	61	128
T5694xa	0.023809524	61	128
Z578	0.023809524	61	128
S79819A, T83510D	0.024390244	60	98
Dx17311, Z578	0.024390244	60	98
Dx36842	0.024390244	60	98
T5694xa	0.024390244	60	98
Z578	0.024390244	60	98
S79819A, T83510D	0.05	58	131
Dx17311, Z578	0.05	58	131
Dx36842	0.05	58	131
T80211d	0.05	58	131
Z578	0.05	58	131
S79819A, T83510D	0.025	59	91
Dx17311, Z578	0.025	59	91
Dx36842	0.025	59	91
T5694xa	0.025	59	91
Z578	0.025	59	91
S79819A, T83510D	0.025641026	58	77
Dx17311, Z578	0.025641026	58	77
Dx36842	0.025641026	58	77
T5694xa	0.025641026	58	77
Z578	0.025641026	58	77
S79819A, T83510D	0.027027027	56	95
Dx17311, Z578	0.027027027	56	95
Dx36842	0.027027027	56	95
T5694xa	0.027027027	56	95
Z578	0.027027027	56	95
S79819A, T83510D	0.027777778	55	114
Dx17311, Z578	0.027777778	55	114
Dx36842	0.027777778	55	114

Covariates	Probability	Minimum RVI	Maximum RVI
T5694xa	0.027777778	55	114
Z578	0.027777778	55	114
S79819A, T83510D	0.057142857	53	182
Dx17311, Z578	0.057142857	53	182
Dx36842	0.057142857	53	182
T80211d	0.057142857	53	182
Z578	0.057142857	53	182
S79819A, T83510D	0.028571429	54	96
Dx17311, Z578	0.028571429	54	96
Dx36842	0.028571429	54	96
T5694xa	0.028571429	54	96
Z578	0.028571429	54	96
S79819A, T83510D	0.029411765	53	116
Dx17311, Z578	0.029411765	53	116
Dx36842	0.029411765	53	116
T5694xa	0.029411765	53	116
Z578	0.029411765	53	116
S79819A, T83510D	0.03030303	52	283
Dx17311, Z578	0.03030303	52	283
Dx36842	0.03030303	52	283
T50992a	0.03030303	52	283
Z578	0.03030303	52	283
S79819A, T83510D	0.0625	50	81
Dx17311, Z578	0.0625	50	81
Dx36842	0.0625	50	81
T80211d	0.0625	50	81
Z578	0.0625	50	81
S79819A, T83510D	0.03125	51	59
Dx17311, Z578	0.03125	51	59
Dx36842	0.03125	51	59
T50992a	0.03125	51	59
Z578	0.03125	51	59
S79819A, T83510D	0.1	47	112
Dx17311, Z578	0.1	47	112
Dx36842	0.1	47	112
T80211d	0.1	47	112
Z578	0.1	47	112
S79819A, T83510D	0.066666667	48	183
Dx17311, Z578	0.066666667	48	183
Dx36842	0.066666667	48	183

Covariates	Probability	Minimum RVI	Maximum RVI
T80211d	0.066666667	48	183
Z578	0.066666667	48	183
S79819A, T83510D	0.033333333	49	56
Dx17311, Z578	0.033333333	49	56
Dx36842	0.033333333	49	56
T50992a	0.033333333	49	56
Z578	0.033333333	49	56
S79819A, T83510D	0.038461538	45	224
Dx17311, Z578	0.038461538	45	224
Dx36842	0.038461538	45	224
T50992a	0.038461538	45	224
Z578	0.038461538	45	224
S79819A, T83510D	0.12	42	57
Dx17311, Z578	0.12	42	57
Dx36842	0.12	42	57
T80211d	0.12	42	57
Z578	0.12	42	57
S79819A, T83510D	0.04	44	103
Dx17311, Z578	0.04	44	103
Dx36842	0.04	44	103
T50992a	0.04	44	103
Z578	0.04	44	103
S79819A, T83510D	0.166666667	40	156
Dx17311, Z578	0.166666667	40	156
Dx36842	0.166666667	40	156
T80211d	0.166666667	40	156
Z578	0.166666667	40	156
S79819A, T83510D	0.083333333	42	56
Dx17311, Z578	0.083333333	42	56
Dx36842	0.083333333	42	56
T80211d	0.083333333	42	56
Z578	0.083333333	42	56
S79819A, T83510D	0.043478261	42	70
Dx17311, Z578	0.043478261	42	70
Dx36842	0.043478261	42	70
T50992a	0.043478261	42	70
Z578	0.043478261	42	70
S79819A, T83510D	0.045454545	41	78
Dx17311, Z578	0.045454545	41	78
Dx36842	0.045454545	41	78

Covariates	Probability	Minimum RVI	Maximum RVI
T50992a	0.045454545	41	78
Z578	0.045454545	41	78
S79819A, T83510D	0.047619048	40	126
Dx17311, Z578	0.047619048	40	126
Dx36842	0.047619048	40	126
T50992a	0.047619048	40	126
Z578	0.047619048	40	126

BIOGRAPHY

Annie Lasway is a Senior Health Systems Analyst at the MITRE Corporation supporting Public Sector Healthcare Research. Prior to joining MITRE, Ms. Lasway worked at Altarum Institute and at the National Institutes of Health. Annie holds a bachelor's degree in Community Health from the University of Maryland College Park, a Master of Public Health with a concentration of Global and Community Health from George Mason University (GMU). She is a certified Project Management Professional (PMP) and a Certified Professional Coder (CPC).

---

1 Matthew Shardlow: An Analysis of Feature Selection Techniques
2 National Institute of Kidney and Digestive and Kidney Diseases. National Institutes of Health. Retrieved on April 9, 2019. From https://www.niddk.nih.gov/health-information/diabetes
3 National Institute of Kidney and Digestive and Kidney Diseases. National Institutes of Health. Retrieved on April 9, 2019. From https://www.niddk.nih.gov/health-information/diabetes
4 M.C.M. de Gunst: Statistical Models. 2013
5 Hui Zou and Trevor Hastie: Regularization and variable selection via the elastic net. Stanford University, USA, 2005.
6 Hui Zou and Trevor Hastie: Regularization and variable selection via the elastic net. Stanford University, USA, 2005.
7 Hui Zou and Trevor Hastie: Regularization and variable selection via the elastic net. Stanford University, USA, 2005.
8 Hui Zou and Trevor Hastie: Regularization and variable selection via the elastic net. Stanford University, USA, 2005.
9 Hui Zou and Trevor Hastie: Regularization and variable selection via the elastic net. Stanford University, USA, 2005.
10 M.C.M. de Gunst: Statistical Models. 2013
11 R. Bergamaschi • A. Romani • S. Tonietti • A. Citterio • C. Berzuini • V. Cosi. Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. Neurol Sci (2000) 21: S819-S823
12 R. Bergamaschi • A. Romani • S. Tonietti • A. Citterio • C. Berzuini • V. Cosi. Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. Neurol Sci (2000) 21: S819-S823
13 R. Bergamaschi • A. Romani • S. Tonietti • A. Citterio • C. Berzuini • V. Cosi. Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. Neurol Sci (2000) 21: S819-S823
14 R. Bergamaschi • A. Romani • S. Tonietti • A. Citterio • C. Berzuini • V. Cosi. Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. Neurol Sci (2000) 21: S819-S823
15 Ali Shojaie, George Michailidis, Discovering graphical Granger causality using the truncating lasso penalty, Bioinformatics, Volume 26, Issue 18, 15 September 2010, Pages i517–i523, https://doi.org/10.1093/bioinformatics/btq377

16 P.Bu¨hlmann, S.van de Geer: Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer 201
17 M.C.M. de Gunst: Statistical Models. 2013
18 M.C.M. de Gunst: Statistical Models. 2013
19 McCoy Rozalina G et al. HbA1c overtesting and overtreatment among US adults with controlled type 2 diabetes, 2001-13
20 McCoy Rozalina G et al. HbA1c overtesting and overtreatment among US adults with controlled type 2 diabetes, 2001-13
21 National Institute of Kidney and Digestive and Kidney Diseases. National Institutes of Health. Retrieved on April 9, 2019. From https://www.niddk.nih.gov/health-information/diabetes
22 National Institute of Kidney and Digestive and Kidney Diseases. National Institutes of Health. Retrieved on April 9, 2019. From https://www.niddk.nih.gov/health-information/diabetes
23 Emilia Javorsky, MPH; Amanda Robinson, MD; and Alexa Boer Kimball, MD, MPH. Evidence-Based Guidelines to Determine Follow-up Intervals: A Call for Action. The American Journal of Managed Care, January 2014 – Published on: January 10, 2014.
24 Emilia Javorsky, MPH; Amanda Robinson, MD; and Alexa Boer Kimball, MD, MPH. Evidence-Based Guidelines to Determine Follow-up Intervals: A Call for Action. The American Journal of Managed Care, January 2014 – Published on: January 10, 2014.
25 Emilia Javorsky, MPH; Amanda Robinson, MD; and Alexa Boer Kimball, MD, MPH. Evidence-Based Guidelines to Determine Follow-up Intervals: A Call for Action. The American Journal of Managed Care, January 2014 – Published on: January 10, 2014.
26 McCoy Rozalina G et al. HbA1c overtesting and overtreatment among US adults with controlled type 2 diabetes, 2001-13
27 McCoy Rozalina G et al. HbA1c overtesting and overtreatment among US adults with controlled type 2 diabetes, 2001-13
28 McCoy Rozalina G et al. HbA1c overtesting and overtreatment among US adults with controlled type 2 diabetes, 2001-13
29 National Institute of Kidney and Digestive and Kidney Diseases. National Institutes of Health. Retrieved on April 9, 2019. From https://www.niddk.nih.gov/health-information/diabetes
30 National Institute of Kidney and Digestive and Kidney Diseases. National Institutes of Health. Retrieved on April 9, 2019. From https://www.niddk.nih.gov/health-information/diabetes
31 Centers for Disease Control and Prevention. National Ambulatory Medical Care Survey: 2009 Summary Tables. Fact Sheet: Physician Office Visits http://www.cdc.gov/nchs/data/ahcd/NAMCS_Factsheet_All_2009.pdf. Accessed June 11, 2012
32 Schoen C, Osborn R; for the Commonwealth Fund. 2010 Commonwealth Fund International Health Policy Survey in Eleven Countries. http://www.commonwealthfund.org/Surveys/2010/Nov/2010-International-Survey.aspx. Published November 2010. Accessed October 30,2012
33 Matthew Shardlow: An Analysis of Feature Selection Techniques
34 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3
35 Robert Tibshirani: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society B, Vol.58, Issue 1, pag. 267-288, 1996
36 Robert Tibshirani: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society B, Vol.58, Issue 1, pag. 267-288, 1996
37 Farrokh Alemi, Sanja Avramovic, and Mark D. Schwartz.Big Data.Sep 2018.ahead of printhttp://doi.org/10.1089/big.2018.0002

38 Centers for Disease Control and Prevention. National Ambulatory Medical Care Survey: 2009 Summary Tables. Fact Sheet: Physician Office Visits http://www.cdc.gov/nchs/data/ahcd/NAMCS_Factsheet_All_2009.pdf. Accessed June 11, 2012
39 Kominski, G. F., Nonzee, N. J., & Sorensen, A. (2016). The Affordable Care Act's Impacts on Access to Insurance and Health Care for Low-Income Populations. Annual review of public health, 38, 489-505
40 Kominski, G. F., Nonzee, N. J., & Sorensen, A. (2016). The Affordable Care Act's Impacts on Access to Insurance and Health Care for Low-Income Populations. Annual review of public health, 38, 489-505
41 Parente, S. T., Feldman, R., Spetz, J., Dowd, B., & Baggett, E. E. (2016). Wage Growth for the Health Care Workforce: Projecting the Affordable Care Act Impact. Health services research, 52(2), 741-762
42 Simoens S, Hurst J. The Supply of Physician Services in OECD Countries. OECD Health Working Papers No. 21. http://www.oecd.org/dataoecd/27/22/35987490.pdf. Published January 2006. Accessed June 11, 2012
43 Yanling Wu,1,2, Yanping Ding,1,2 Yoshimasa Tanaka,3 and Wen Zhang2, Risk Factors Contributing to Type 2 Diabetes and Recent Advances in the Treatment and Prevention. Int J Med Sci. 2014; 11(11): 1185–1200
Published online 2014 Sep 6. doi: 10.7150/ijms.10001
44 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
45 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
46 Judea Pearl, Causality, second edition. Models, Reasoning and Inference. (1999). Page 254.
47 Judea Pearl, Causality, second edition. Models, Reasoning and Inference. (1999). Page 254.
48 Tan Y, Liu Z. Feature selection and prediction with a Markov blanket structure learning algorithm. BMC Bioinformatics. 2013;14(Suppl 17):A3. Published 2013 Oct 22. doi:10.1186/1471-2105-14-S17-A3
49 Robert Tibshirani: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society B, Vol.58, Issue 1, pag. 267-288, 1996
50 http://bayes.cs.ucla.edu/jp_home.html
51 Pearl J. An introduction to causal inference. Int J Biostat. 2010 Feb 26;6(2):Article 7
52 Sonnenberg, Frank A; Beck, J. Robert. Markov Models in Medical Decision Making: A Practical Guide
53 Sonnenberg, Frank A; Beck, J. Robert. Markov Models in Medical Decision Making: A Practical Guide
54 Ali Shojaie, George Michailidis, Discovering graphical Granger causality using the truncating lasso penalty, Bioinformatics, Volume 26, Issue 18, 15 September 2010, Pages i517–i523, https://doi.org/10.1093/bioinformatics/btq377
55 Emilia Javorsky, MPH; Amanda Robinson, MD; and Alexa Boer Kimball, MD, MPH. Evidence-Based Guidelines to Determine Follow-up Intervals: A Call for Action
The American Journal of Managed Care, January 2014 – Published on: January 10, 2014
56 Emilia Javorsky, MPH; Amanda Robinson, MD; and Alexa Boer Kimball, MD, MPH. Evidence-Based Guidelines to Determine Follow-up Intervals: A Call for Action
The American Journal of Managed Care, January 2014 – Published on: January 10, 2014
57 Agency for Healthcare Research and Quality. Table 2: Number of events for selected conditions by type of service: United States, 2010. Medical Expenditure Panel Survey Household

Component Data. Rockville, MD: Agency for Healthcare Research and Quality; 2010. Accessed June 11, 2012

58 Judea Pearl, Causality, second edition. Models, Reasoning and Inference. (1999). Page 254.

59 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

60 Surfing Uncertainty: Prediction, Action, and the Embodied Mind 1st Edition by Andy Clark (Author)

61 Agency for Healthcare Research and Quality. Total expenses and percent distribution for selected conditions by type of service: United States, 2010. Medical Expenditure Panel Survey Household Component Data. Rockville, MD: Agency for Healthcare Research and Quality; 2010. Accessed June 11, 2012

62 Agency for Healthcare Research and Quality. Table 2: Number of events for selected conditions by type of service: United States, 2010. Medical Expenditure Panel Survey Household Component Data. Rockville, MD: Agency for Healthcare Research and Quality; 2010. Accessed June 11, 2012

63 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

64 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

65 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

66 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

67 Schwartz LM, Woloshin S, Wasson JH, Renfrew RA, Welch HG.. Setting the revisit interval in primary care. J Gen Intern Med 1999; 14: 230–235

68 Schwartz LM, Woloshin S, Wasson JH, Renfrew RA, Welch HG.. Setting the revisit interval in primary care. J Gen Intern Med 1999; 14: 230–235.

69 Alemi, Cary, Raya. The Multi-Morbidity Index: A Tool for Assessing the Prognosis of Patients from History of Illness. eGEMs : Generating Evidence and Methods to Improve Patient Outcomes. 2016. 10.13063/2327-9214.1235

70 DeSalvo K, Bowdish B, Alper A, Grossman D, Merrill W.Physician practice variation in assignment of return interval. Arch Intern Med 2000; 160: 205–208.

71 Lurie N, Margolis KL, McGovern PG, Mink PJ, Slater JS. Why do patients of female physicians have higher rates of breast and cervical cancer screening? J Gen Intern Med 1997; 12: 34–43.

72 Borum ML. Cancer screening in women by internal medicine resident physicians. South Med J 1997; 90: 1101–1105.

73 Schwartz LM, Woloshin S, Wasson JH, Renfrew RA, Welch HG. Setting the revisit interval in primary care. J Gen Intern Med 1999; 14: 230–235.
74 Roos NP, Carriere KC, Friesen D. Factors inXuencing the frequency of visits by hypertensive patients to primary care
physicians in Winnipeg. *CMAJ* 1998; **159**: 777–796.
75 Welch HG, Chapko MK, James KE, Schwartz LM, Woloshin S. The role of patients and providers in the timing of follow-up visits. J Gen Intern Med 1999; 14: 223–229.
76 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248
77 Birtwhistle RV1, Godwin MS, Delva MD, Casson RI, Lam M, MacDonald SE, Seguin R, Rühland L. Randomised equivalence trial comparing three month and six month follow up of patients with hypertension by family practitioners. BMJ. 2004 Jan 24;328(7433):204. Epub 2004 Jan 15.
78 KAREN B. DESALVO, JASON P. BLOCK, PAUL MUNTNER AND WILLIAM MERRILL. Predictors of Variation in Office Visit Interval Assignment. International Journal for Quality in Health Care 2003; Volume 15, Number 5: pp. 399–405.
79 Schectman G1, Barnas G, Laud P, Cantwell L, Horton M, Zarling EJ. Prolonging the return visit interval in primary care. Am J Med. 2005 Apr;118(4):393-9.
80 Heckerman D, Geiger D, Chickering D. Learning Bayesian networks: the combination of knowledge and statistical data. Mach Learn. 1995;20:197–243.
81 P.Bu¨hlmann, S.van de Geer: Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer 201
82 P.Bu¨hlmann, S.van de Geer: Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer 201
83 Robert Tibshirani: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society B, Vol.58, Issue 1, pag. 267-288, 1996
84 P.Bu¨hlmann, S.van de Geer: Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer 201
85 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
86 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
87 Rosenbaum, P. R., and D. B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” Biometrika 70: 41–55
88 Rosenbaum, P. R., and D. B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” Biometrika 70: 41–55
89 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
90 P.Bu¨hlmann, S.van de Geer: Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer 201 Ali Shojaie, George Michailidis, Discovering graphical Granger causality using the truncating lasso penalty, Bioinformatics, Volume 26, Issue 18, 15 September 2010, Pages i517–i523, https://doi.org/10.1093/bioinformatics/btq377
91 R. Bergamaschi • A. Romani • S. Tonietti • A. Citterio • C. Berzuini • V. Cosi. Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. Neurol Sci (2000) 21: S819-S823

92 Alemi, F., A. ElRafey, and I. Avramovic. 2018. “Covariate Balancing Through Naturally Occurring Strata.” Health Services Research 53 (1): 273–92
93 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
94 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
95 Introduction to Algorithms for Data Mining and Machine Learning, 2019
96 Javed, Kashif; Maruf, Sameen; Babri, Haroon A. A two-stage Markov blanket-based feature selection algorithm for text classification. 2015
97 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3
98 https://www.wired.com/story/karl-friston-free-energy-principle-artificial-intelligence/
99 http://bayes.cs.ucla.edu/jp_home.html
100 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3
101 Judea Pearl, Causality, second edition. Models, Reasoning and Inference. (1999). Page 254.
102 Rosenbaum, P. R., and D. B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” Biometrika 70: 41–55.
103 Tan, Y., and Z. Liu. 2013. “Feature Selection and Prediction with a Markov Blanket Structure Learning Algorithm.” BMC Bioinformatics 14 (Suppl 17): A3.
104 R. Bergamaschi • A. Romani • S. Tonietti • A. Citterio • C. Berzuini • V. Cosi. Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. Neurol Sci (2000) 21:S819-S823
105 Abdel Majeed, Yazan; Awadalla, Saria S.; Patton, James L. Regression techniques employing feature selection to predict clinical outcomes in stroke.
106 Varol HA, Mani S, Compton DL, Fuchs LS, Fuchs D. Early prediction of reading disability using machine learning. AMIA Annu Symp Proc. 2009;2009:667–671. Published 2009 Nov 14.
107Tan Y, Liu Z. Feature selection and prediction with a Markov blanket structure learning algorithm. *BMC Bioinformatics*. 2013;14(Suppl 17):A3. Published 2013 Oct 22. doi:10.1186/1471-2105-14-S17-A3
108 Pellet, Jean-Philippe; Elisseeff, Andre. Using Markov Blankets for Causal Structure Learning.
109 Sonnenberg, Frank A; Beck, J. Robert. Markov Models in Medical Decision Making: A Practical Guide
110 Tan Y, Liu Z. Feature selection and prediction with a Markov blanket structure learning algorithm. BMC Bioinformatics. 2013;14(Suppl 17):A3. Published 2013 Oct 22. doi:10.1186/1471-2105-14-S17-A3
111 Javed, Kashif; Maruf, Sameen; Babri, Haroon A. A two-stage Markov blanket based feature selection algorithm for text classification. 2015.
112 Aliferis CF, Tsamardinos I, Statnikov A. HITON: a novel Markov Blanket algorithm for optimal variable selection. AMIA Annu Symp Proc. 2003;2003:21–25.
113 Statnikov A, Lytkin NI, Lemeire J, Aliferis CF. Algorithms for Discovery of Multiple Markov Boundaries. J Mach Learn Res. 2013;14:499–566.
114 Farrokh Alemi, Ph.D.; Amr Hassan Elrafey. Stratification of High-Dimensional Data through Blocking Backdoors.
115 Li, Xiaomao; Yin, Jianxin. Sparse Sufficient Dimension Reduction for Markov Blanket Discovery. Communications in Statistics - Simulation and Computation, Volume 45, 2016 - Issue 4

116 Casini, Lorenzo; Illari, Phyllis McKay; Russo, Federica; Williamson, Jon, Models for Prediction, Explanation and Control: Recursive Bayesian Networks
117 K. Yu; X. Wu; W. Ding; Y. Mu; H. Wang, Markov Blanket Feature Selection Using Representative Sets.
118 Javed, Kashif; Maruf, Sameen; Babri, Haroon A. A two-stage Markov blanket-based feature selection algorithm for text classification. 2015
119 Javed, Kashif; Maruf, Sameen; Babri, Haroon A. A two-stage Markov blanket-based feature selection algorithm for text classification. 2015
120 Javed, Kashif; Maruf, Sameen; Babri, Haroon A. A two-stage Markov blanket-based feature selection algorithm for text classification. 2015
121 Javed, Kashif; Maruf, Sameen; Babri, Haroon A. A two-stage Markov blanket-based feature selection algorithm for text classification. 2015
122 Alemi, F., A. ElRafey, and I. Avramovic. 2018. "Covariate Balancing Through Naturally Occurring Strata." Health Services Research 53 (1): 273–92
123 Alemi, F., A. ElRafey, and I. Avramovic. 2018. "Covariate Balancing Through Naturally Occurring Strata." Health Services Research 53 (1): 273–92
124 Alemi, F., A. ElRafey, and I. Avramovic. 2018. "Covariate Balancing Through Naturally Occurring Strata." Health Services Research 53 (1): 273–92
125 Alemi, F., A. ElRafey, and I. Avramovic. 2018. "Covariate Balancing Through Naturally Occurring Strata." Health Services Research 53 (1): 273–92
126 Alemi, F., A. ElRafey, and I. Avramovic. 2018. "Covariate Balancing Through Naturally Occurring Strata." Health Services Research 53 (1): 273–92
127 The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses S. S. Wilks. The Annals of Mathematical Statistics. Vol. 9, No. 1 (Mar., 1938), pp. 60-62. Published by: Institute of Mathematical Statistics https://www.jstor.org/stable/2957648. Page Count: 3
128 The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses S. S. Wilks. The Annals of Mathematical Statistics. Vol. 9, No. 1 (Mar., 1938), pp. 60-62. Published by: Institute of Mathematical Statistics https://www.jstor.org/stable/2957648. Page Count: 3
129 The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses S. S. Wilks. The Annals of Mathematical Statistics. Vol. 9, No. 1 (Mar., 1938), pp. 60-62. Published by: Institute of Mathematical Statistics https://www.jstor.org/stable/2957648. Page Count: 3
130 https://www.cs.ubc.ca/labs/lci/mlrg/slides/doCalc.pdf
131 https://www.cs.ubc.ca/labs/lci/mlrg/slides/doCalc.pdf
132 https://www.cs.ubc.ca/labs/lci/mlrg/slides/doCalc.pdf
133 https://www.cs.ubc.ca/labs/lci/mlrg/slides/doCalc.pdf
134 https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/StandardAnalyticalFiles.html
135 https://www.resdac.org/articles/strengths-and-limitations-cms-administrative-data-research
136 Andrew B. Rosenkrantz MD, MPA a, Danny R. Hughes PhD b, c, Richard Duszak Jr. MD d, Medicare Claims Data Resources: A Primer for Policy-Focused Radiology Health Services Researchers. https://doi.org/10.1016/j.jacr.2017.04.005
137 https://www.resdac.org/articles/strengths-and-limitations-cms-administrative-data-research
138 https://www.resdac.org/articles/strengths-and-limitations-cms-administrative-data-research
139 https://www.resdac.org/articles/strengths-and-limitations-cms-administrative-data-research

140 Nicolai Meinshausena , Alain Hauserb , Joris M. Mooijc , Jonas Petersd , Philip Versteegc , and Peter Bühlmanna,1 Methods for causal inference from gene perturbation experiments and validation. Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 5, 2016 (received for review June 5, 2015)
141 Pearl J. An introduction to causal inference. Int J Biostat. 2010 Feb 26;6(2):Article 7
142 Binbing Yu and Joseph L Gastwirth. 2008. A method of assessing the sensitivity of the Cochran–Mantel–Haenszel test to an unobserved confounder.Phil. Trans. R. Soc. A.366:2377–2388. http://doi.org/10.1098/rsta.2008.0030
143 4th Schizoaffective disorder, chronic with acute exacerbation (295.74), 482 days before SA
144 **3rd Obsessive/compulsive disorder (301.4), 482 days before SA**
145 Standards of medical care in diabetes--2013. American Diabetes Association. Diabetes Care. 2013 Jan; 36 Suppl 1():S11-66.
146 Physician practice variation in assignment of return interval. DeSalvo KB, Bowdish BE, Alper AS, Grossman DM, Merrill WW. Arch Intern Med. 2000 Jan 24; 160(2):205-8.
147 Variation in physician opinion about scheduling of return visits for common ambulatory care conditions.
Tobacman JK, Zeitler RR, Cilursu AM, Mori M
J Gen Intern Med. 1992 May-Jun; 7(3):312-6.
148 Predictors of variation in office visit interval assi gnment. DeSalvo KB, Block JP, Muntner P, Merrill W. Int J Qual Health Care. 2003 Oct; 15(5):399-405.
149 Physician practice variation in assignment of return interval. DeSalvo KB, Bowdish BE, Alper AS, Grossman DM, Merrill WW. Arch Intern Med. 2000 Jan 24; 160(2):205-8.
150 Encounter frequency and serum glucose level, blood pressure, and cholesterol level control in patients with diabetes mellitus. Morrison F, Shubina M, Turchin A. Arch Intern Med. 2011 Sep 26; 171(17):1542-50.
151 Variation in physicians' recommendations about revisit interval for three common conditions. Petitti DB, Grumbach K. J Fam Pract. 1993 Sep; 37(3):235-40.
152 Setting the revisit interval in primary care. Schwartz LM, Woloshin S, Wasson JH, Renfrew RA, Welch HG
J Gen Intern Med. 1999 Apr; 14(4):230-5
153 Variation in physician opinion about scheduling of return visits for common ambulatory care conditions.
Tobacman JK, Zeitler RR, Cilursu AM, Mori M.J Gen Intern Med. 1992 May-Jun; 7(3):312-6.
154 The role of patients and providers in the timing of follow-up visits. Telephone Care Study Group.
Welch HG, Chapko MK, James KE, Schwartz LM, Woloshin S. J Gen Intern Med. 1999 Apr; 14(4):223-9.
155 Variation in physicians' recommendations about revisit interval for three common conditions. Petitti DB, Grumbach K.J Fam Pract. 1993 Sep; 37(3):235-40.
156 Predictors of variation in office visit interval assignment. DeSalvo KB, Block JP, Muntner P, Merrill W.Int J Qual Health Care. 2003 Oct; 15(5):399-405.
157 Physician practice variation in assignment of return interval. DeSalvo KB, Bowdish BE, Alper AS, Grossman DM, Merrill WW. Arch Intern Med. 2000 Jan 24; 160(2):205-8.
158 Variation in physicians' recommendations about revisit interval for three common conditions. Petitti DB, Grumbach K.J Fam Pract. 1993 Sep; 37(3):235-40.
159 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of

hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248
160 Thomas L. Nickolas 1, Mary B. Leonard 2, Elizabeth Shane 3. Chronic kidney disease and bone fracture: a growing concern. https://doi.org/10.1038/ki.2008.264
161 Thomas L. Nickolas 1, Mary B. Leonard 2, Elizabeth Shane 3. Chronic kidney disease and bone fracture: a growing concern. https://doi.org/10.1038/ki.2008.264
162 Thomas L. Nickolas 1, Mary B. Leonard 2, Elizabeth Shane 3. Chronic kidney disease and bone fracture: a growing concern. https://doi.org/10.1038/ki.2008.264
163 Simon J. Bennet a, Olivia M.B. Berry b, Jane Goddard c, John F. Keating d. Acute renal dysfunction following hip fracture. https://doi.org/10.1016/j.injury.2009.07.009
164 B. Novis, et al. Association of pre-operative risk factors with postoperative acute renal failure. Anaesth Analg, 74 (1994), pp. 143-148
165 Simon J. Bennet a, Olivia M.B. Berry b, Jane Goddard c, John F. Keating d. Acute renal dysfunction following hip fracture. https://doi.org/10.1016/j.injury.2009.07.009
166 Simon J. Bennet a, Olivia M.B. Berry b, Jane Goddard c, John F. Keating d. Acute renal dysfunction following hip fracture. https://doi.org/10.1016/j.injury.2009.07.009
167 Simon J. Bennet a, Olivia M.B. Berry b, Jane Goddard c, John F. Keating d. Acute renal dysfunction following hip fracture. https://doi.org/10.1016/j.injury.2009.07.009
168 M.L. Cittanova, et al. The chronic inhibition of ACE impairs postoperative renal function. Anaesth Analg, 93 (2001), pp. 111-115
169 S.R. Cummings, L.J. Melton. Epidemiology and outcomes of osteoporotic fractures. Lancet, 359 (2002), pp. 1761-1767.
170 Greenwell MW, Mangold TM, Tolley EA, Wall BM: Kidney disease as a predictor of mortality in chronic spinal cord injury. Am J Kidney Dis 2007;49:383–393.
171 St. Andre JR, Smith BM, Stroupe KT, Burns SP, Evans CT, Cowper Ripley C, Li K, Huo Z, Hogan TP, Weaver FM: A comparison of costs and health care utilization for veterans with traumatic and nontraumatic spinal cord injury. Top Spinal Cord Inj Rehabil 2011;1:27–42.
172 McKinely WO, Jackson AB, Cardenas DD, DeVivo MJ: Long-term medical complications after traumatic spinal cord injury: a regional model systems analysis. Arch Phys Med Rhabil 1999;80:140–210.
173 Pettersson-Hammerstad K, Jonsson O, Svennung IB, Karlsson AK: Impaired renal function in newly spinal cord injured patients improves in the chronic state – effect of clean intermittent catheterization? J Urol 2008;180:187–191.
174 Weld KJ, Wall BM, Mangold TA, Steere EL, Dmochowski RR: Influences on renal function in chronic spinal cord injured patients. J Urol 2000;164:1490–1493.
175 Wall BM, Huch KM, Mangold TA, Steere EL, Cooke CR: Risk factors for development of proteinuria in chronic spinal cord injury. Am J Kidney Dis 1999;33:899–903.
176 Giannantoni A, Scivoletto G, Di Stasi SM, Silecchia A, Finazzi-Agro E, Micali I, Castellano V: Clean intermittent catherization and prevention of renal disease in spinal cord injury patients. Spinal Cord 1998;36:29–32.
177 Roddy E, Doherty M. Epidemiology of gout. Arthritis Res Ther. 2010;12:223.
178 Matthew J Roughley, John Belcher, Christian D Mallen & Edward Roddy. Gout and risk of chronic kidney disease and nephrolithiasis: meta-analysis of observational studies. *Arthritis Research & Therapy* **volume 17**, Article number: 90 (2015)
179 Matthew J Roughley, John Belcher, Christian D Mallen & Edward Roddy. Gout and risk of chronic kidney disease and nephrolithiasis: meta-analysis of observational studies. *Arthritis Research & Therapy* **volume 17**, Article number: 90 (2015)
180 Muntner P, He J, Vupputuri S, Coresh J, Batuman V. Blood lead and chronic kidney disease in the general United States population: results from NHANES III. Kidney Int 2003;63:1044-50.
181 Batuman V. Lead nephropathy, gout, and hypertension. Am J Med Sci 1993; 305:241-7
182 Brewster UC, Perazella MA. A review of chronic lead intoxication: an unrecognized cause of chronic kidney disease. Am J Med Sci 2004;327:341-7

183 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

184 Bénédicte Stengel. Chronic kidney disease and cancer: a troubling connection. J Nephrol. 2010; 23(3): 253–262. PMCID: PMC4823382. PMID: 20349418

185 Bénédicte Stengel. Chronic kidney disease and cancer: a troubling connection. J Nephrol. 2010; 23(3): 253–262. PMCID: PMC4823382. PMID: 20349418

186 Bénédicte Stengel. Chronic kidney disease and cancer: a troubling connection. J Nephrol. 2010; 23(3): 253–262. PMCID: PMC4823382. PMID: 20349418

187 Cledes J, Allain P. Chronic lead nephropathy. Epidemiology and diagnosis. Presse Med 1992;21:759-62

188 Meryem Benjelloun, Faissal Tarrass, Khadija Hachim, Ghislaine Medkouri, Mohamed Gharbi Benghanem, Benyounes Ramdani. Department of Nephrology and Dialysis, Ibn-Rochd University Hospital Center, Casablanca, Morocco. Year : 2007 | Volume : 18 | Issue : 1 | Page : 83-86

189 Brewster UC, Perazella MA. A review of chronic lead intoxication: an unrecognized cause of chronic kidney disease. Am J Med Sci 2004;327:341-7

190 Cledes J, Allain P. Chronic lead nephropathy. Epidemiology and diagnosis. Presse Med 1992;21:759-62

191 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

192 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

193 Quinn RR, Hemmelgarn BR, Padwal RS, et al; Canadian Hypertension Education Program. The 2010 Canadian Hypertension Education Program recommendations for the management of hypertension: part I—blood pressure measurement, diagnosis and assessment of risk. Can J Cardiol. 2010;26(5):241-248

194 Rosenbaum, P. R., and D. B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” Biometrika 70: 41–55

195 David Heckerman, Christopher Meek, and Gregory Cooper. 2006. A Bayesian approach to causal discovery. In Innovations in Machine Learning. Springer, 1–28.

196 David Heckerman, Christopher Meek, and Gregory Cooper. 2006. A Bayesian approach to causal discovery. In Innovations in Machine Learning. Springer, 1–28.

197 Schwartz LM, Woloshin S, Wasson JH, Renfrew RA, Welch HG.. Setting the revisit interval in primary care. J Gen Intern Med 1999; 14: 230–235.