RISK PREDICTION MODELS AND VISUALIZATIONS FOR HEPATORENAL SYNDROME

By

Jejo David Koola

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Informatics

December 15, 2018

Nashville, Tennessee

Approved:

Michael E. Matheny, M.D., M.S., M.P.H.

Bradley Malin, Ph.D.

Daniel Fabbri, Ph.D.

Dedicated to my mother, for her
support and her unwavering faith in me.

ACKNOWLEDGEMENTS

I would like to sincerely thank my thesis committee for their support in completing this project. I would like to especially thank Dr. Michael Matheny for his amazing mentoring, help, and guidance in helping me take my first steps in Biomedical Informatics and a physician scientist career. I would also like to thank Dr. Brad Malin and Dr. Daniel Fabbri who supported me with guidance and insightful comments regarding this work. Thank you for giving me the chance to make this an unforgettable experience.

My work would not have been possible without the generous financial support of the Department of Veterans Affairs and the Advanced Fellowship Program in Medical Informatics through the Office of Academic Affiliations. Additionally, I whole heartedly thank the Department of Biomedical Informatics, Vanderbilt University for their generous financial in-kind support in allowing me to obtain this Master's.

My work wouldn't have been possible without the help of so many other people who were involved in this project. Thank you to the faculty, staff, and fellow students at the Department of Veterans Affairs. Most specifically, I would like to thank Dr. Ruth Reeves, Dr. Glenn Gobbel, Jason Denton, Dax Westerman, Vincent Messina, and Dr. Freneka Minter. Thank you also to the faculty, staff, and students of the Department of Biomedical Informatics at Vanderbilt University. Specifically, I would like to thank Dr. Kevin Johnson, Dr. Cynthia Gadd, and Ms. Rischelle Jenkins.

Finally, I would like to thank my parents and my sister for supporting me and encouraging me throughout this experience and work.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER I: INTRODUCTION

In 2012 the Institute of Medicine (now the National Academy of Medicine) released a desiderata for a learning healthcare system, where evidence informs practice and practice informs evidence.[1] Though the randomized clinical trial (RCT) serves as the gold standard for informing clinical decisions, flaws exist in terms of achieving recruitment, overly stringent inclusion/exclusion criteria, and lack of patient-centered decision making.[2,3] Moreover, the majority of medical decisions may not be informed by randomized controlled trials; a recent review found that only 19% of Class I cardiology guidelines have Level A evidence.[4]

Clinicians often have questions when caring for patients but do not pursue answers to many of them.[5] Observational cohort studies have grown as an important complement to RCTs allowing comparative effectiveness research and patient-centered trials.[6] This thesis centers on leveraging observational cohort data to create and interpret models improving healthcare for cirrhosis patients.

CLINICAL BACKGROUND

Cirrhosis, a late stage of chronic liver damage where scarring replaces hepatic tissue, carries significant morbidity and mortality due to decreased mental, physical, and biochemical function. The prevalence is estimated between 400,000 and 3,000,000 persons in the United States, and the disease causes 44,000 deaths annually.[7–11] Over fifteen etiologies exist, including hemochromatosis, Wilson's disease, autoimmune hepatitis, and primary biliary cirrhosis.[12] In the United States, the most common causes are alcohol abuse, viral hepatitis, and nonalcoholic fatty liver disease (NAFLD).[12–14] Due to increasing obesity and diabetes in the US, the prevalence of NAFLD has been growing (from 5.5% in the original National Health and Nutrition Examination Survey up to 11% in 2008, refer to Figure 1).[8] Because cirrhosis may remain latent for years, only autopsy or late sequelae implicate the diagnosis in 20-30% percent of cases.[15–19]

The Department of Veterans Affairs (VA) faces an increasing burden of chronic liver disease due to substance use disorders, chronic viral hepatitis, and increasing numbers of patients with NAFLD. The VA's patients face a higher burden of substance abuse, particularly among Iraq and Afghanistan veterans where the prevalence has been estimated at 11%.[20] Overall, 90-100% of alcoholics develop liver steatosis, 10-35% develop alcoholic fibrosis and/or hepatitis, and 10% develop cirrhosis.[21] The VA is the largest single provider of Hepatitis C Virus (HCV) care in the US, and has approximately 186,000 patients have chronic active HCV.[22] Refer to Figure 2 for a description of the changes in VA cirrhosis etiology prevalence. These dramatic increases in overall prevalence of cirrhosis at the VA will impact the VHA system with greatly increased costs due to complications occurring from end-stage liver disease.

Figure 1: Change in prevalence of chronic liver disease in the United States from 1988 to 2008.

Data obtained from Younossi et al.[8]

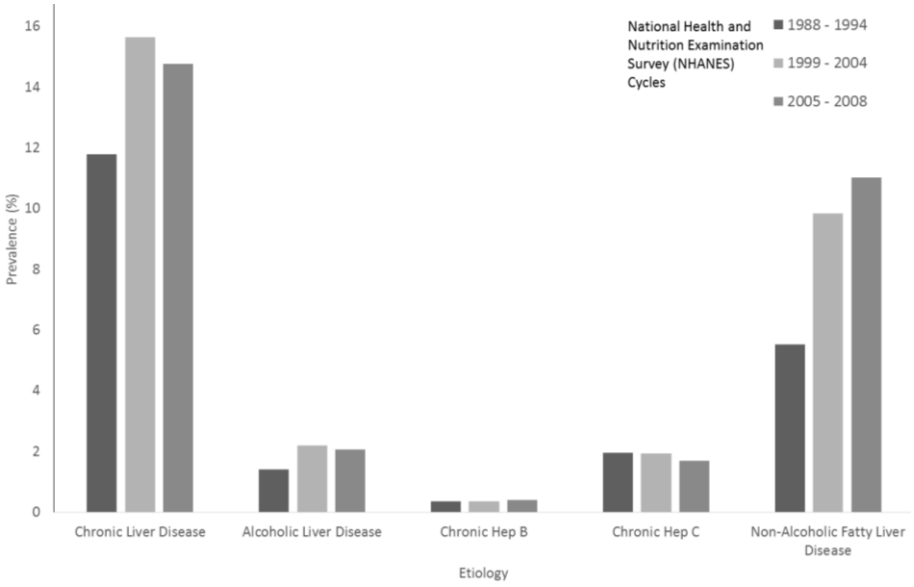
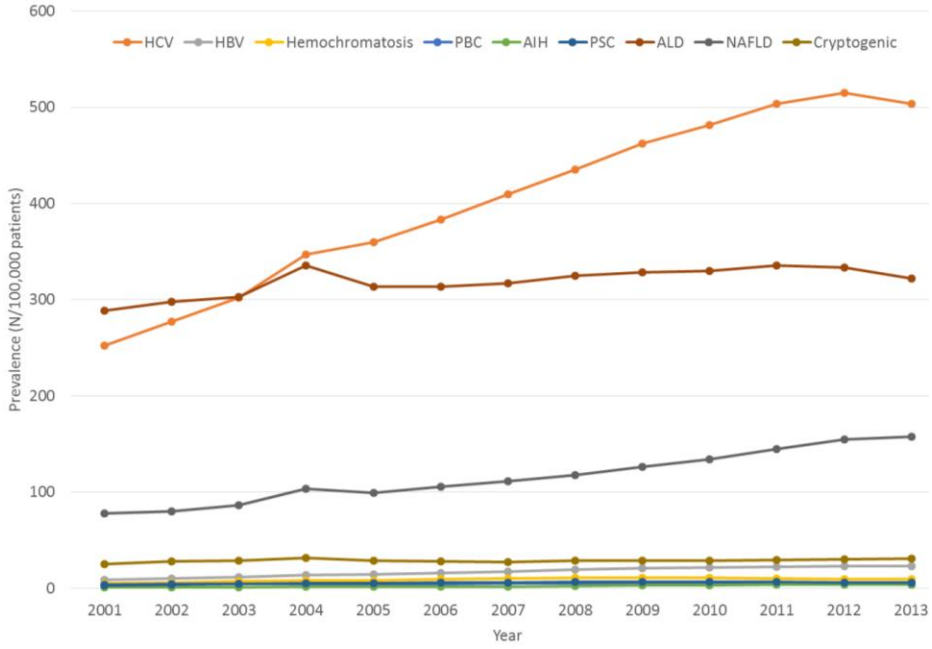


Figure 2: Change in prevalence of cirrhosis etiology for Veterans Affairs patients from 2001 to 2013.

Data obtained from Beste et al.[22] Note: HCV: Hepatitis C Virus; HBV: Hepatitis B Virus: PBC: Primary Biliary Cirrhosis; AIH: Autoimmune Hepatitis; PSC: Primary Sclerosing Cholangitis; ALD: Alcoholic Liver Disease; NAFLD: Non-Alcoholic Fatty Liver Disease.



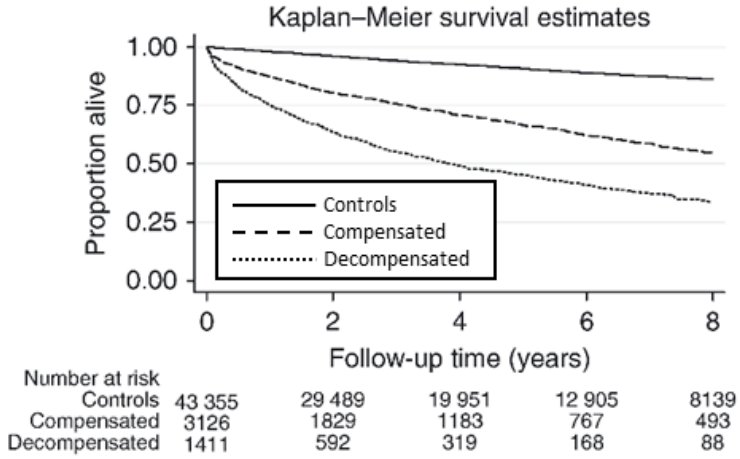
Cirrhosis Complications

Cirrhosis impacts the health care system broadly because of the breadth and severity of end-stage liver disease complications. Complications stem from liver synthetic dysfunction and portal hypertension. Synthetic dysfunction causes coagulation disorders, low serum albumin, low

platelet counts, and hepatic encephalopathy.[23] Portal hypertension, increased pressure in the veins that carry blood from the gastrointestinal tract to the liver, leads to ascites, varices, renal failure, gastrointestinal bleeding, and spontaneous bacterial peritonitis.[24] Because many medications are metabolized by the liver, cirrhosis can cause heightened medication sensitivity and toxicity.[25] Once a cirrhotic patient develops a complication they follow a significantly different disease trajectory. The median survival for a cirrhotic patient without a complication is 10 years, however, once a patient becomes decompensated, i.e. develops a complication, their median survival drops to four years.[26] Refer to Figure 3 for a description of the survival curve.

Figure 3: Survival estimates for controls, subjects with compensated cirrhosis, and subjects with decompensated cirrhosis.

Data taken from Fleming et al.[26]



The only cure for cirrhosis is liver transplant. However because livers are a scarce resource, relatively few are transplanted annually. In 2016, 7,841 livers were transplanted with 13,725 patients still on the waiting list at the end of the year.[27] Patients are candidates for liver transplant once their Model for End-stage Liver Disease (MELD) score, a measure of overall liver dysfunction, is ≥ 15. Patients may qualify for earlier liver transplant if they have certain exceptional conditions, such as hepatocellular carcinoma. Liver transplantation is contraindicated in select cases such as severe extra-hepatic disease, acquired immunodeficiency syndrome, and persistent non-adherence with medical care. Because transplantation is expensive and livers are scarce, patients must undergo an extensive pre-transplant evaluation that includes exhaustive laboratory testing, cardiopulmonary assessment, cancer screening, infectious diseases evaluation, and psychosocial appraisal. Guidelines recommend all patients with a MELD > 10 be referred for liver transplant evaluation to initiate pre-transplant evaluation and substance-abuse counseling (if needed). One particular complication, renal failure, heralds poor survival and warrants urgent transplant evaluation.

Patients with cirrhosis are particularly prone to renal failure from multiple etiologies. Causes include hypovolemia from diuretics, medication toxicity, parenchymal renal disease such viral hepatitis induced cryoglobulinemia, and changes in the circulatory system due to cirrhosis. Cirrhotics with renal failure have a higher mortality and increased frequency of complications compared to cirrhotics without renal failure.[28] Acute renal failure is also a common occurrence in patients hospitalized with cirrhosis, occurring 20% of the time.[29] The deadliest form of renal failure in cirrhosis is termed Hepatorenal Syndrome.

HEPATORENAL SYNDROME (HRS)

Hepatorenal syndrome (HRS), is a particularly challenging complication of end-stage cirrhosis, and represents an archetype of multi-organ failure.[30–32] HRS represents functional kidney dysfunction due to intense renal vasoconstriction with concomitant splanchnic vasodilation. Diagnosing HRS requires exclusion of other kidney disorders, absence of shock, no concurrent or recent treatment with nephrotoxic drugs, and no improvement in serum creatinine after at least two days of diuretic withdrawal and volume expansion with albumin.[30]

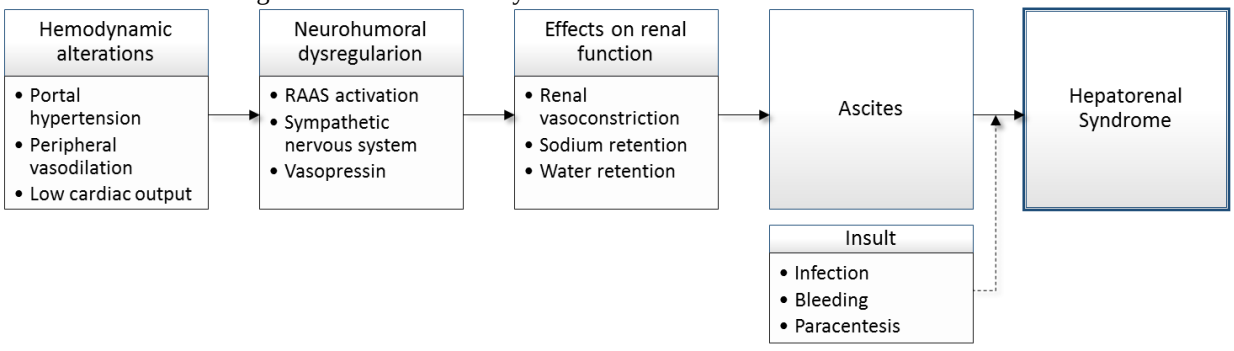
Table 1: Diagnostic criteria for Hepatorenal Syndrome.

Cirrhosis with ascites
Serum creatinine > 1.5 mg/dl (old guidelines) OR diagnosis of AKI using established guideline agency AKI criteria (new guidelines)
No improvement of serum creatinine after at least 2 days with diuretic withdrawal and volume expansion with albumin. The recommended dose of albumin is 1 g/kg of body weight per day up to a maximum of 100 g/day.
Absence of shock
No current or recent treatment with nephrotoxic drugs
Absence of parenchymal kidney disease as indicated by proteinuria 500 mg/day, micro-hematuria (50 red blood cells per high power field) and/or abnormal renal ultrasonography.

Renal biopsy often does not show intrinsic disease nor sufficient tissue damage to explain the amount of dysfunction. Hepatorenal Syndrome's underlying etiology is understood to be a cumulative effect of multiple clinical conditions starting with progressive portal hypertension causing splanchnic vasodilation resulting in lowered intravascular perfusion pressures to the kidneys. Subsequent neurohormonal compensation leads to sodium and water retention resulting in ascites. Further degradation in effective blood flow to the kidneys, potentially brought on by an acute insult such as an infection, leads to Hepatorenal Syndrome. Refer to Figure 4.

Figure 4: Steps leading to Hepatorenal Syndrome.

Note: RAAS: Renin-Angiotensin Aldosterone System.



The disorder is broadly divided into two types, I and II, based mainly on rate of progression and a few clinical indicators. The median survival for Type I HRS is two weeks, and is six months for Type II HRS.[31,33,34] Over a five year span, 39% of cirrhotic patients will experience HRS.[35]

The definitive treatment for HRS is liver transplant,[30,31,34] but several case series suggest that only 4.5 – 35% of patients (median 18.5%) receive a liver[36–41] due to HRS' high mortality and organ scarcity.[42,43] The standard of care includes several temporizing and palliative measures

including vasopressors,[30,31,34] dialysis,[44] Molecular Adsorbent Recirculating System (MARS) therapy,[45,46] Transjugular Intrahepatic Portosystemic Shunt (TIPS),[47,48] and hospice.[49,50] However, there is large variability in survival among these patients, particularly those with Type II HRS. Some patients may benefit from early identification and initiation of some medical treatments.

Temporizing measures such as dialysis do not change overall survival, and as such there is increasing focus on palliative treatment, which may still result in net improvements in patient satisfaction and quality of life in this time period. Earlier diagnosis would be beneficial to initiate timely triaging or specific treatments for HRS, including the use of vasopressor and somatostatin agonists, albumin expansion, dialysis, expedite evaluation for transplantation, or referral for palliative care. Several studies have shown that renal function prior to the initiation of vasopressor therapy is predictive of response,[51–53] particularly a creatinine less than 3.0 mg/dL.[54] The International Ascites Club has stressed earlier diagnosis and treatment of HRS in its updated 2015 guidelines.[55] Moreover, increasing changes in serum creatinine from baseline have shown linear increases in hospitalization costs amongst patients admitted with acute kidney injury (AKI) from $4,886 for an increase ≥ 0.3 mg/dl up to $22,023 for an increase ≥ 2.0 mg/dl.[56]

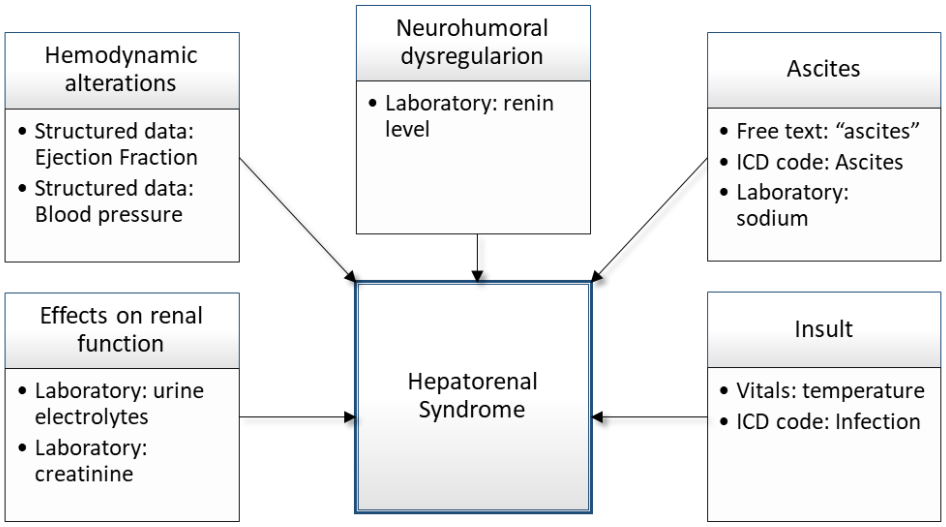
INFORMATICS BACKGROUND

DATA DRIVEN PHENOTYPING

Electronic health record (EHR) phenotyping helps identify sufficiently large cohorts to perform observational studies that inform clinical care in a wide variety of domains; refer to Shivade *et al. and* Xu *et al.* for a review.[57,58] Phenotyping is especially important as larger observational cohort datasets have been generated due to collaboration from multiple institutions.[59,60] EHR phenotyping has been applied to various conditions including cancer,[61–63] diabetes,[64] heart failure,[65] rheumatoid arthritis,[66] cataracts,[67] drug side effects,[68] pneumonia,[69] asthma,[70] and hypertension.[71]

Phenotyping has not been applied to acute kidney injury (AKI), a common acute complication sometimes necessitating hospitalization and a challenging problem because of the close overlap between multiple causes of kidney injury. There are more than ten causes of AKI;[72] and in observational cohort studies, though laboratory markers can be used for some etiologies, the majority of etiologies are represented by the International Classification of Diseases (ICD) code. Using ICD-9 codes alone is well known to have limited sensitivity and sub-optimal specificity.[73] Hepatorenal syndrome is a serious form of AKI that can occur among patients with cirrhosis, and stands as an archetype of multi-organ failure.[30–32] Refer to Figure 5 for some EHR elements that may be used to identify patients with HRS.

Figure 5: Hypothetical phenotyping elements for Hepatorenal Syndrome.
Note: ICD: International Classification of Diseases.



Rule based systems serve as the most basic phenotyping model. The simplest rule based systems employ a single ICD code to identify a cohort, e.g. ICD-9 code 571.2 for alcoholic cirrhosis. Typical rule-based systems apply other constraints (such as age, biological sex, laboratory value thresholds, etc.) in a sequence of steps. Rule-based systems can be *a priori* defined, either by medical experts or based on healthcare guidelines. Conversely, they can be automatically derived using, for example, a decision tree algorithm. The pros of a rule-based system include their ease in interpretation, implementation speed, and they tend to yield good results on limited datasets. Because rule-based systems often rely on structured data elements that are universal in EHRs, they are also more portable. Examples of rule-based systems include rules for reporting quality metrics to guideline agencies, e.g. National Quality Forum[74] measure #0018[75] which identifies individuals with hypertension based on age, blood pressure readings, and ICD codes.

Unlike rule-based systems, systems utilizing machine learning and statistical analysis attempt a data-driven approach to identifying a phenotype. Popular machine learning models employed in these systems include support vector machines,[66] random forest,[76] and Bayesian algorithms.[77] Statistical methods for classification, such as logistic regression, are not infrequently used for phenotyping when the overall number of variables is small.[78] As EHR datasets have grown, particularly in the number of variables available for model construction, machine learning algorithms have overtaken rule-based and statistical algorithms.

Algorithm development has given way to phenotyping system design. Phenotyping systems attempt to generate reproducible phenotyping algorithms that can be shared and validated across systems. Examples include the Electronic Medical Records and Genomics (eMERGE) network,[59] Strategic Health IT Advanced Research Projects (SHARP),[79] and the Cross Institutional Translational Research (CICTR).[80] Rule-based algorithms are naturally easier to share across systems and can be posted online, e.g. on PheKB.org.[60] Important questions remain about using free text from clinical notes and variable selection in phenotyping.

USE OF NATURAL LANGUAGE PROCESSING PRODUCTS IN PHENOTYPING & RISK MODELING

Phenotyping applications leverage multiple information sources from the EHR, with clinical text playing an increasingly important role. Clinical text, as recorded by healthcare providers usually

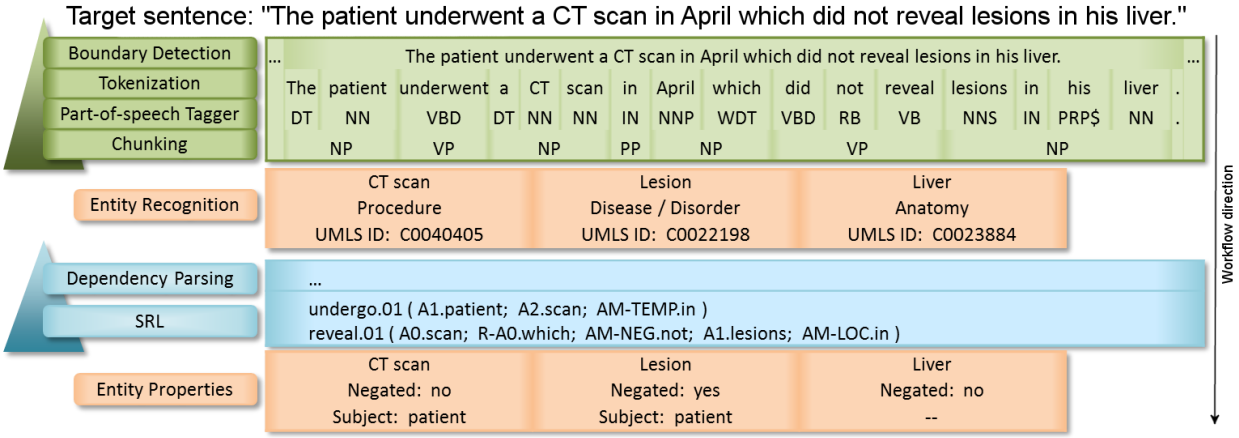
during routine care of patients, can contain data inexpressible in a structured format, tedious to express in structured formats, or more nuanced then that which is contained in structured data.[81] Natural Language Processing (NLP) of clinical text plays a crucial role in converting free text into computable structured data.[66,82,83] NLP systems may proceed along one of two workflows. One choice is to build a custom NLP pipeline and train the system to identify specific phrases/concepts within the documents. This process frequently requires painstaking annotation of a document corpus at the phrase level. Because the concepts of interest are *a priori* defined using domain knowledge, this method will produce a more parsimonious feature set and the results likely will have higher specificity.

Conversely, one may utilize an "out-of-the-box" NLP system that attempts to identify all medical concepts within the documents. Natural Language Processing has increasingly turned to replacing raw text with standardized concepts from ontologies such as the Unified Medical Language System (UMLS).[84] Natural Language Processing pipelines from clinical Text Analysis Knowledge Extraction System (cTAKES),[85] MedLEE,[86] and MetaMap[87] allow for replacing free text concepts with UMLS concept unique identifiers (CUIs). Refer to Figure 6 for an example of the steps taken by the cTAKES pipeline to convert free text into computable CUIs. The workflow for using these systems for model development is often different, with annotation more likely to be performed at the document or at the patient level. In either case, some sort of manual chart review is required; however, depending on resources and level of domain knowledge one workflow may be superior to the other. Because the "out-of-the-box" approach is often non-specific, i.e. it attempts to identify all medical terms in the text, some sort of post-processing is required. This post-processing often requires dimension-reduction and/or machine learning to match the identified concepts with the document- or patient-level annotation.

Phenotyping has traditionally been a time intensive process, often requiring the assistance of domain experts. As a result, increasing emphasis has recently been placed on automated methods, termed high throughput phenotyping, requiring less domain knowledge.[88–90] These high throughput methods have focused on using NLP to augment the phenotyping process.[91–93] To date, however, they have only been validated on chronic medical conditions. Performance may be biased due to the much higher data density for chronic conditions, particularly in terms of clinical text.

Figure 6: Description of workflow and modules within the clinical Text Analysis Knowledge Extraction System (cTAKES) NLP system.

Note: Taken from apache.org.[94] DT: determiner; NN: noun; VBD: verb, past tense; IN: preposition; NNP: proper noun; WDT: Wh-determiner; VBD: RB: adverb; VB: verb, base form; NNS: noun, plural; PRP$: possessive pronoun; NP: noun phrase; VB: verb phrase; Unified Medical Language System (UMLS).



Risk Prediction Models

Risk prediction models increasingly complement clinical reasoning and decision making in modern medicine. Models have been developed to predict a wide array of outcomes including ICU mortality[95], various types of cancer[96–98], quality control[99], post-acute coronary syndrome outcomes[100], and other forms of acute kidney injury.[101–105]

Traditional views of medicine incorporated the ideal of the master clinician who, through years of experience and inductive reasoning, could appropriately diagnose and treat patients. This view of physician healthcare delivery changed in the 20th century, most notably with the birth of Gordon Guyatt's "Evidence Based Medicine."[106] However, its seeds had taken root much earlier. Possibly the first use of rigorous modeling in healthcare had to do with population tracking and prediction. Verhulst developed the logistic equation to describe population growth in 1845.[107] The logistic function had subsequent uses from the U.S. Food Administration to model food shortages during World War I.[108] Eventually, Cox would publish his seminal work on logistic regression.[109] Though rigorous modeling was nipping at the heels of medicine in the early 20th century, the majority of scientific thinking, communication, and training were still grounded in inductive knowledge.

Perhaps one of the earliest changes to this dogma occurred with the Framingham Heart Study started in 1948 in the town of Framingham, Massachusetts.[110] A cohort of 5,209 subjects were followed prospectively to quantify various heart disease risk factors. However, the subsequent multivariable survival model was too complex for everyday use and it was not until a simpler, point-based formula allowed the Framingham Risk Score to be employed in routine clinical care.[111] Other early forays into risk models include the Child-Turcotte-Pugh score[112,113] to predict surgical mortality in cirrhotic patients, Maddrey's discriminant function for alcoholic hepatitis,[114] and Ranson's criteria for pancreatitis mortality.[115] Though these were early forays into risk prediction models in medicine, it was not until the 1980s when the field exploded with several new models.[116–119]

Nevertheless, modeling was invariably expert driven, partially as a commentary on the climate in medicine at the time, but also as a matter of necessity. These models were usually based on carefully conducted prospective studies where the variables had to be prespecified and reasonable to collect. The concept of "data mining," and its promise, was coined in the 1990s with the increasing reliance on large database systems in finance, transportation, and communication.[120] Data-driven medicine did not really become a concept until two important events in the 21st century, sequencing of the human genome and the HITECH Act, which introduced a tremendous amount of data into healthcare. With the proliferation of EHRs, risk modeling became more complex as it consumed vastly more information,[105,121–124] and even allowed EHR incorporation of predictions at the point of care.[125–128] These EHR driven, point-of-care risk models have led to probabilistic clinical decision support (CDS), as opposed to reporting a binary outcome, making calibration all the more important.

Assessment of Performance (Discrimination / Calibration)

For a risk prediction model to be clinically useful, one must consider both its discrimination as well as its calibration.[129] Discrimination refers to the model's ability to distinguish individuals who experienced the outcome from those who remained event free. Calibration refers to agreement between the probability of developing the outcome as estimated by the model and the observed outcome frequencies. Although clinical decision rules have often focused on a model's discriminative ability (e.g., instituting a statin medication in high cardiovascular risk patients)[130], proper calibration is required when multiple decision options are available at differing levels of risk (e.g., management of a solitary pulmonary nodule found on computed tomography).[131] Moreover, model performance degrades when used in a cohort outside of its development, making careful validation of discrimination and calibration essential.[132]

Several statistics are available to summarize discrimination for binary classification models, which tend to be the most common in healthcare, including the c-index[133] and the area under the receiver operator characteristic curve (AUC). One may best assess calibration graphically by plotting observed outcome frequencies against mean predicted probabilities within subgroups of the observations, usually split by deciles of predicted probabilities.[134] The plot can be supplemented with formal statistical testing for goodness of fit, frequently done using the Hosmer-Lemeshow (HL) test. However, because the HL test applies the chi-square distribution, whose power scales with the sample size,[135] the null hypothesis for the HL-test may be accepted under a small sample size, but rejected under a large sample size.[136]

Newer measures of discrimination and calibration have been developed. Discrimination has been reframed as a reclassification task, i.e. how well does a new model correctly classify observations that the old model misclassified. Two metrics for measuring reclassification include the Net Reclassification Index (NRI) and the Integrated Discrimination Index.[129] The NRI may be interpreted as the net improvement in the true positive rate plus the net improvement in the false positive rate under the new model. The NRI may be measured at specific thresholds of interest of the underlying model, which may be more informative than a global measure such as the AUC. Healthcare delivery is often interested in model performance at certain cutpoints.

Feature Generation & Selection

Feature selection and generation are common steps in data pre-processing. Feature selection hopes to identify a relevant subset of the original features; whereas, feature generation creates new features (optionally replacing some original features) to enhance model performance.

Feature selection has always been essential to produce parsimonious models, prevent loss of statistical power, and prevent overfitting. However, as the amount of healthcare data has exploded feature selection has gained in importance. Traditional healthcare models were often created with 5 – 20 *a priori* selected variables; however, newer data-driven models may have hundreds of potential variables. Popular clinical NLP systems can generate hundreds to thousands of features from reviewed documents.[85–87,137] Often, dimensionality reduction is necessary to either make the classification task more tractable or improve performance.[138,139]

Perhaps the most common traditional selection method has been forward selection and backward elimination (FBS). Forward/backward selection has drawbacks, however; particularly in large datasets with many variables or collinearity among its predictors.[140] Newer methods include using random forest classifier based variable importance and [141] penalized logistic regression.[142,143] For very high dimensional data, which is more common in biomedical datasets, penalized logistic regression methods such as LASSO and elastic net appear to work well.[144]

Though one may produce an adequate model via the original feature set, often times one must consider feature generation as both a way to perform dimension reduction and identify novel relationships. Feature generation methods such as *a priori* specifying interaction terms, Principle Component Analysis (PCA),[145] and Latent Discriminant Analysis (LDA)[146] serve an important role in discovering underlying structure. Methods such as PCA and LDA have the added benefit of reducing the feature space.

A relatively recent dimensionality reduction technique involves a distributed vector representation of words, or word embeddings, which has shown good performance in many NLP tasks.[147,148] Google's word2vec, an increasingly popular embedding algorithm,[149] has been generalized to vector representations of an entire document (termed doc2vec).[150] Although word embeddings have been used to improve classification in healthcare tasks,[151–153] it is still relatively new to assess improvement in phenotyping. Zhang *et al.* assessed word embedding's benefit in identifying phrases suggestive of psychiatric illness,[154] and Turner *et al* used word embedding to identify an overall phenotype for Systemic Lupus Erythematosus.[155] However, they applied it to chronic conditions and used raw text.

Regardless of whether one is interested in feature selection or feature generation, the primary goal is to improve model accuracy. However, and this is particularly true in data-driven methods, a secondary measure of importance is understandability.[156] Validation of the model not only requires statistical measures such as discrimination and calibration, but also biological/physiological plausibility. Feature generation methods such as PCA, latent discriminant analysis, and neural networks may obscure understandability of the model.[157] In this regard, knowledge driven variable selection tends to be superior.

INFORMATION VISUALIZATION

The surge of Electronic Health Records, and its resulting zettabyte of data,[158] allows us to realize the vision of the learning healthcare system. Despite the growth of observational cohort studies, challenges still remain bringing the knowledge from the bench-to-the-bedside. Observational cohort studies employ data for secondary use, i.e. data collected for other purposes. The most common secondary use scenario is data collected as part of routine clinical care.[159] However, due to cognitive and perceptual limitations, healthcare providers need increasing help to digest the vast amounts of information generated during clinical care. Information visualization can be

defined as "the process of transforming data, information, and knowledge into visual form making use of humans' natural visual capabilities."[160] Bertin, a pioneer in information visualization, defined it as, "… finding the artificial memory that best supports our natural means of perception."[161] Ultimately it is the depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system. It is important to differentiate scientific visualization from information visualization. Scientific visualization's goals are to depict scientific data, often physically based, over a compact domain. Information visualization often deals with abstract data types mapped onto abstract domains.[162]

Cognitive Principles of Visualization

There are two basic cognitive principles in visualization, pre-attentive properties and gestalt properties.[161] Pre-attentive processing occurs without need for focusing attention. Tasks which are completed in less than 250 milliseconds are considered pre-attentive. As reference, a non-exhaustive list of the various pre-attentive properties that have been studied in Information Visualization include length, width, size, curvature, number, terminators, intersection, closure, color, intensity, flicker, motion, binocular luster, stereoscopic depth, 3D depth curves, and lighting direction (See Figure 7 for examples).[163] Accurate application of pre-attentive properties can communicating information rapidly, as a pre-attentive task takes the same amount of time irrespective of the number of distractors.[164] The gestalt properties are forms or patterns that transcend the stimuli used to create them. They include proximity, similarity, enclosure, closure, continuity, and connection. In Information Visualization, proximity, similarity, and enclosure tend to play larger roles.

Taxonomy of Visualization

Visualization techniques can be subdivided into a few different types. Certain methods may be more appropriate for particular types of data.[165,166] Table 2 describes some of the patterns of information visualization and the data sizes and dimensionality for which they are suited. Geometric visualization is one of the most common and has the closest resemblance to Scientific Visualization. A classic example is the use of hierarchical parallel coordinates, which allows the plotting of multiple variables within one figure.[167] Iconographic visualization techniques have gained more popularity as a means of communicating complex information to lay people.[168] Pixel based visualizations allow display of very large datasets by mapping multi-dimensional datapoints onto a single pixel within a larger diagram.[169] Hierarchical display visualizations allow the user to see underlying relationships within the data; the Treemap is a popular tool to demonstrate hierarchical relationships and semantic information.[170] Network visualizations are common in displaying non-hierarhical relationships within a wide area of research, including social relationships and bioinformatics.[171,172]

Figure 7: Examples of pre-attentive/gestalt properties in visual processing.
Each panel demonstrates a different pre-attentive property. A: orientation; B: length; C: closure; D: size; E: curvature; F: color; G: density; H: shape. Panels C, F, and G also demonstrate gestalt properties of closure, similarity, and proximity, respectively. Taken from Healey.[173]

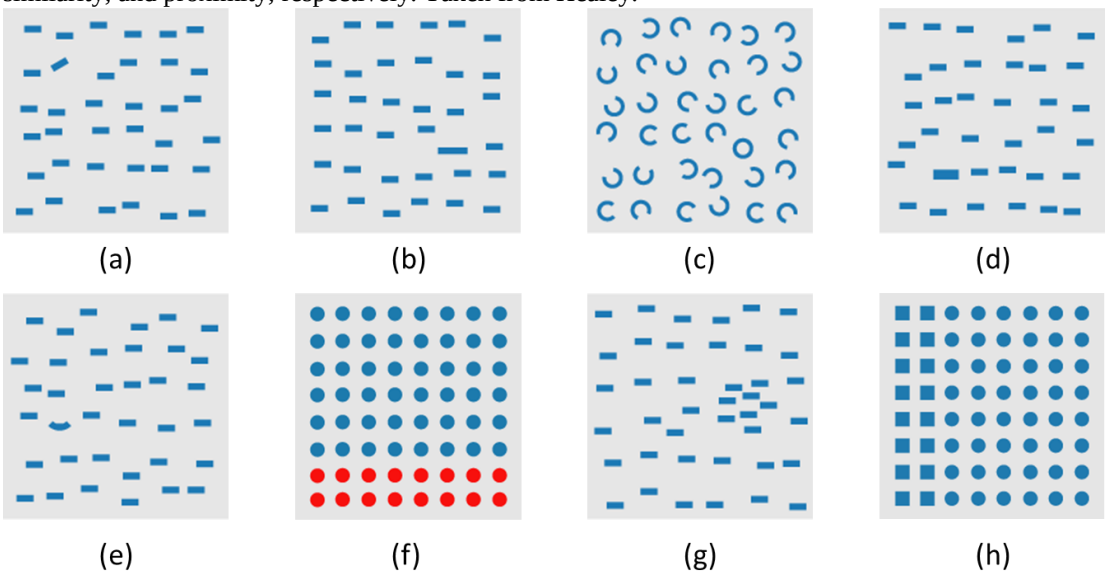


Table 2: Patterns of information visualization and their relationships to data size and dimensionality.

	Data Size	Dimensions	Examples
Geometric	Large to very large	Medium to high	Scatter matrix, PCA, factor analysis, MDS, GridViz, Hierarchical Parallel Coordinates
Icon	Small to medium	Medium to high	Chernoff faces, color icons, driftweed, shape coding, sound icons
Pixel	Large to very large	Medium	Circle Segments, Recursive Pattern, Space Filling Curves
Hierarchical	Small to medium	Low to medium	TreeMap, Dimensional Stacking, Worlds-within-Worlds
Graph	Medium	Low	NetMap, NetViz, SocialNet
Hybrid	Variable	Variable	Variable

General Paradigm of Visualization

Ben Shneiderman constructed a paradigm of visualization, divided into tasks and data types. Tasks includes: (1) provide an overview; (2) zoom into relevant detail; (3) filter; (4) provide details-on-demand; (5) relate what is being viewed to the bigger picture; (6) be able to provide a history of actions; and (7) extract relevant details.[174] Shneiderman also suggested (possibly orthogonal) data-types that in some ways mirror the taxonomy presented in Table 2. In particular, he considered data to be: 1-dimensional (1-D), 2-D, 3-D, Multi-dimensional, Temporal, Hierarchical, and Networked. Naturally, certain visualization techniques work well

with certain data types in Shneiderman's hierarchy, e.g. Hierarchical Parallel Coordinates for multi-dimensional data and a TreeMap for hierarchical data. Table 3 outlines Shneiderman's visualization "tasks" and methods to effectuate these tasks.

Table 3: Tasks in Shneiderman's Paradigm of Data Visualization.

Tasks	Example Methods
Overview	Fisheye distortion
Zoom	Brushing
Filter	Dynamic querying
Details-on-Demand	Pop-up windows
Relate	LifeLines[175]
History	Undo/Redo controls
Extract	Image export

Visualization is increasingly interactive, especially when dealing with large datasets with multidimensional data. Central paradigms in dealing with complex data sets, as is often encountered in healthcare, are multiple coordinated views and dynamic queries.[176,177] Dynamic queries allow the user to interact with the visualization in a pragmatic manner: (1) visual presentation of the query's components; (2) visual presentation of results; (3) rapid, incremental, and reversible control; (4) selection by pointing; and (5) immediate, continuous feedback.

Temporal Visualizations

Temporal visualizations seek to identify patterns in time-dependent data changes. Several general purpose visualization tools have been developed to explore temporal changes in data.[178–181] Healthcare temporal visualization has tackled time series[182–184] and event series data.[175,185–187] Time series data often have issues with scale and data density, particularly in healthcare. Berry *et al.*[179] used "brushing," a technique that allows the user to examine a data segment in greater detail by serially highlight and zooming in on areas of interest. Another technique used to zoom in on detail is "distortion visualization." Kincaid[180] applies distortion allowing the user to employ a "fish-eye lens" type effect. Doing so allows the user to view detail about a specific part of the time-line, while still maintaining global perspective.

Different layouts can help alleviate some of the challenges of temporal data. Multi-dimensionality is a common hurdle with healthcare data. Multi-dimensionality makes visualizing temporal data all the more difficult. Some basic solutions to the problem include utilizing interactivity to allow the user to select subsets of variables to display, such as in Rind *et al.*[182] Another design principle combines "multiple coordinated views," i.e. different perspectives of the same data, to handle multi-dimensionality. Zhao *et al.*[181] uses a radial layout, but combines it with alternate linear views offering different perspectives of the data. Disease specific views, a common design in healthcare visualizations of patient data, attempt to align the patient's disease course, usually summarized by a numeric measure, compared to treatment interventions.[183] Brodbeck *et al.*,[184] for example, utilizes a simple linear view but juxtaposes estimates of lung function with breathing treatments in patients with obstructive pulmonary disease.

Though frequently healthcare visualizations attempt to plot data regarding a single patient, another common theme is plotting data of multiple patients together on the same visualization in an attempt to identify patterns. One of the earliest multi-patient visualizations is the seminal

work by Plaisant *et al*. with *LifeLines*.[175] Plaisant expanded on this work by tackling a common problem with temporal visualization of healthcare data: often healthcare data records discrete events, e.g. medication refills; however, these discrete events are semantically part of a continuous temporal period, i.e. the time from when the healthcare provider starts prescribing the medication to when it stops. Wang *et al*.[187] operationalized an iterative method to combine events to visualize patterns in patients with heparin-induced thrombocytopenia. Meyer[186] updated the prior work to assess patterns in medication use for patients with obstructive lung disease.

Challenge of Big Data

Big Data offers unique challenges for Information Visualization. Both the scale of data and the dimensionality provide challenges for constructing effective visual analytic tools. Two major approaches have emerged to handle these challenges: distortion and non-distortion based approaches. Non-distortion based approaches primarily focus on presenting a part of the data at any one time using scrolling or paging access, providing hierarchical access, and structure specific presentation. Distortion based approaches try to preserve the global data presentation on the macro scale, while zooming into the relevant area of interest on the micro scale. One common distortion based visualization is the Fisheye Lens, which zooms in on a local area of the visualization, compressing the areas not of interest, though they are still displayed on screen.[188] The prior discussed paradigms of dynamic queries and multiple coordinated views can also help the user explore and interpret very large datasets.

Application in Clinical Informatics

Complex informatics solutions have had difficulty gaining traction in routine clinical practice because of esoteric analytic techniques and outputs. Information visualization, a field devoted to conveying complex data, can address these shortcomings.[165,177] Visual Analytics in healthcare has been used for improving radiology interpretation,[169,189] investigating temporality,[185,187,190–193] explaining social networks,[171] analyzing spatial patterns,[194,195] documenting workflows,[177] identify latent structure,[176] and analyzing high dimensional data.[196,197] Research on uncertainty visualization has predominated within geographic information science, geographic visualization, and scientific visualization fields.[198–203]

Though visualizing risk prediction and uncertainty has received some attention for patient-facing tools,[168,204] physician-facing clinical decision support (CDS) at the point of care has received sparse investigation.[205] Most visualizations in medical practice are designed to be used "off-line." Because medical decision making rarely has an obvious correct answer, further research in uncertainty visualization for CDS would be beneficial.

Visualization has also been utilized to better understand patient cohorts. For example, Mane *et al.*[205] employed visualization to better understand results in comparative effectiveness trials to make decisions in psychiatry. Early efforts have also used visualization to show clinicians "Patients-like-me" cohorts.[206] An essential component of these kinds of studies is ensuring that one has an appropriate cohort of patients. Research studies are often done under the assumption that a cohort of patients based on fairly rudimentary principles are relatively uniform. As discussed in simple rule-based phenotyping algorithms, a common assumption is that if one selects a group of patients with Hepatorenal Syndrome via simple structured data (e.g., ICD-9 code) the cohort will be rather homogenous.

However, this assumption often does not hold. The case mix (i.e. the heterogeneity of subjects in the population) used to develop the model affects both discrimination and calibration.[207,208] More importantly, population heterogeneity can hide subgroups for which the risk model underperforms (e.g. pharmacogenomic model based dosing of warfarin[209] and a geriatric mortality model[210] did worse in African Americans and the EuroSCORE cardiac surgery mortality model performed worse for high-risk subgroups).[211]

The process of finding a cohort of similar patients goes by many names, including phenotyping, matching, cohort discovery, etc. Matching is an increasingly popular method to improve causal inference in observational studies.[212] Matching attempts to reduce the bias inherently involved in observational studies because of measured and unmeasured confounders.[213] The end result is a matched group of study subjects that could theoretically have been produced by a randomized control trial. However, traditional matching methods may cause spurious results based on the methodologies.[214,215]

Information Visualization, when combined with clustering, may help identify true pockets of heterogeneity. Clustering has been applied to various medical problems such as identifying disease subtypes[70,216,217] and risk stratifying patients.[218,219] At its heart, Information Visualization aims at making the user an active participant in identifying patterns in the data, which may be undiagnosed by a computational algorithm. For example Gotz *et al.*[220] clustered patients into similar groups, then attempted to visualize common clinical trajectories. The goal being to use the group's trajectory to predict a new patient's outcomes.

PRIOR HRS INFORMATICS RESEARCH

Current HRS mortality risk models suffer from one or more flaws: they were developed prior to the current standard of care (particularly the use of vasopressors);[35] they were developed with modest sample sizes;[37,39,221–224] used specialized, non-routine laboratory tests;[221] did not include Type I and Type II HRS;[39,222,225] or they were limited to a small, *a priori* set of variables.[39,222] Refer to Table 4 for a summary of risk prediction studies involving HRS. Because Hepatorenal Syndrome exhibits different phenotypes, risk model performance may vary widely among sub groups.[95,226,227]

Several of the studies in Table 4 use an outdated definition of HRS, either the 2007[30] or the 1996 International Ascites Club criteria.[228] Most of the studies do not perform any validation of their risk model, whether internal or external. The listed studies had a median subject size of 64 (IQR: 41 – 105) and they evaluated a median of 7.5 variables (IQR: 1 – 16.5) in univariate testing. Studies that performed multivariable model building invariably only included 3 – 5 variables. Nevertheless, several smaller studies show promising avenues for investigation. Several studies have investigated the beneficial use of Transjugular Intrahepatic Portosystemic Shunts (TIPS), a common procedure performed in cirrhotics. The use of beta-blockers, a common medication used in cirrhosis, has also come under scrutiny.

For example Guevara et al.[229] assessed vascular hemodynamics in 7 patients before and after TIPS placement, resulting in (expected) reduced portal pressures and favorable metabolite profiles. Testino et al.[230] assessed the role of TIPS in 9 patients with severe alcoholic hepatitis and HRS, and identified favorable outcomes in this small sample study. Brensing et al.[231] prospectively followed 41 non-transplanted cirrhotics with HRS who received TIPS, and also recorded favorable short-term hemodynamic and long-term outcomes. Regardless all three

studies show the limitation of studying TIPS in HRS patients as no one study or TIPS specialized center has enough patients for sufficient inference due to the usually restrictive nature of TIPS selection. Sersté et al.[232] performed a single-center study involving 151 patients with refractory ascites (often a precursor to HRS) and found poorer survival. Mandorfer et al.[233] retrospectively analyzed 607 patients with spontaneous bacterial peritonitis and identified increased risk of HRS. Because beta-blockers are the standard of care for many patients with cirrhosis (to prevent variceal bleeding), there is significant confounding and retrospective studies require larger numbers to tease out any potential causal effect. To the best of our knowledge, no study has evaluated the necessity of stopping a beta-blocker when a patient develops HRS.

In both of these cases improved phenotyping efforts could advance the field by building larger observational cohort datasets. Additionally, phenotyping could be deployed for real-time identification of HRS patients as current practice guidelines often leave the diagnosis in doubt early in the illness's trajectory. Biomedical informatics tools and techniques may help advance risk prediction for HRS by allowing a significantly larger candidate predictor pool, improving accuracy, and allowing for other data types such as NLP augmented clinical text. Medical informatics solutions have been sought to improve cirrhosis care; however, these methods invariably rely on structured data and ignore social determinants of illness.

THESIS OBJECTIVE

Because of the prior small sample sizes of patient cohorts and complexity of the disease, there has been little work in attempting to use clinical informatics and data science to identify and characterize individual phenotypes within the overall syndrome of HRS. In addition, there are opportunities to attempt to create early detection algorithms in order to identify and potentially triage and manage patients at high risk for developing HRS in the near term. Lastly, interpretation of complex, high density clinical data is a challenge in health care, particularly as data collection and data become harder and harder to interpret and efficiently manage, and there are needs for interactive visualizations to help present actionable clinical information in a way that is believable, accepted, and actionable for clinicians. Within the context of the use case for HRS, we aim to explore some of these challenges.

Table 4: Summary of Hepatorenal Syndrome Risk Prediction Papers.

Note: PC: Prospective cohort; RC: Retrospective cohort; RCT: Randomized Controlled Trial; IAC: International Ascites Club.

Author (Year)	Study Type	N	Inclusion Criteria	Exclusion Criteria	# of Var	Outcome Definition	Event Rate
Ginès (1993)[35]	PC	234	Cirrhosis with ascites	GI hemorrhage within 1 month of admission, hepatic encephalopathy or bacterial infection at the time of study, hepatocellular carcinoma, respiratory, cardiac, or renal diseases.	39	Development of HRS by criteria similar to 1996 IAC criteria[228]	56 (24%)
Gungor (2014)[221]	PC	64	Group 1: Healthy control Group 2: serum creatinine value <1.5 mg/dl. Group 3: Type-2 HRS (serum creatinine >1.5 mg/dl) Group 4: Type-1 HRS (serum creatinine >2.5 mg/dl)	Group 2: renal dysfunction, active infection or malignancy Groups 3/4: shock, fluid losses, active infection, and patients who did not fulfil HRS diagnostic	6	6-month Mortality	28 (43.7%)
Maddukuri (2014)[39]	RC	59	2007 IAC criteria for HRS,[30] received vasoconstrictors plus albumin	No baseline creatinine, received dialysis within 3 days of therapy, received liver transplant within 3 days of therapy, died within 3 days of therapy	1	Treatment response, defined as decrease in serum creatinine gen-erally to a level < 1.5 mg/dl	15 (25.4%)
Martinez (2012)[222]	RC	68	2007 IAC criteria for HRS Type 1[30]	None provided	9	Death or discharge from hospital	Median survival: 13 days
Salerno (2011)[224]	PC	253 (76 w/ HRS)	Cirrhosis patients admitted with renal failure (creatinine > 1.5 mg/dl)	Age < 18 years, prior kidney or liver transplant	21	3-month mortality	58 (76%)
Alessandria (2005)[37]	RC	105	1996 IAC criteria for HRS[228]	None specified	15	Mortality	78 (74%)
Barreto (2014)[225]	PC	70	2007 IAC criteria for HRS[30] and infection	Admitted for elective diagnostic or therapeutic procedures, history of liver and/or kidney transplantation, hemodialysis before admission.	27	Treatment response, mortality	Treatment response: 23 (33%), 3-month mor-tality: 53 (75%)

Author (Year)	Study Type	N	Inclusion Criteria	Exclusion Criteria	# of Var	Outcome Definition	Event Rate
Krag (2010)[234]	PC	24	Alcoholic cirrhosis with ascites without HRS type 1 by 2007 IAC criteria[30]	GI hemorrhage within the week before the study, spontaneous bacterial peritonitis, insulin-dependent diabetes, acute or chronic intrinsic renal or cardiovascular diseases, arterial hypertension, abnormal electrocardiogram or any acute medical conditions such as infections or acute heart or lung diseases, alcohol abstinence for 6 weeks was required	1	Development of type 1 HRS within 3 months, 12-month mortality	Type 1 hrs: 4 (16.7%), death: 9 (37.5%)
Mandorfer (2014)[233]	RC	165	Cirrhosis with ascites who received their first paracentesis	Non-liver causes of ascites	1	Development of HRS within 90 days of SBP development	29 (18%)
Sersté (2010)[232]	RC	151	Cirrhotic patients with refractory ascites	Not specified	15	Mortality	97 (64.2%), median survival: 8 months
Ghosh (2013)[52]	RCT	46	Type 2 HRS based on IAC 2007 criteria[30]	Severe coronary artery disease, sepsis in, HCC, diabetic nephropathy	17	HRS reversal	34 (74%); 17 in intervention group A and 17 in group B
Guevara (1998)[229]	PC	7	Type 1 HRS patients by 1996 IAC criteria[228]	None described.	1	Improvement in renal function	7 (100%)
Brensing (2000)[235]	PC	41	HRS patients by 1996 IAC criteria[228] ineligible for transplant	Transplant eligibility	5	Multiple outcomes. 3-month survival	63%
Testino (2012)[230]	PC	9	Severe Alcoholic hepatitis with HRS	Ongoing infections, malignancy, symptomatic cardiac or respiratory diseases, GI hemorrhage in the last week	1	Improvement in renal function	Unclear, presumably all

CHAPTER II: PHENOTYPING HEPATORENAL SYNDROME

CHAPTER OBJECTIVE

In this study, we sought first to assess the performance characteristics of ICD-9-CM codes for determining HRS occurring during a patient hospitalization. We then evaluated commonly used machine learning methods and dimensionality reduction techniques among a large number of variables derived from EHR structured data and NLP processed outputs in order to develop probabilistic predictions for phenotyping HRS during hospitalization of patients that have both cirrhosis and acute kidney injury. We report on the performance of these methods by comparing each of the HRS predictors to a reference standard of clinical patient chart reviews.

MATERIALS AND METHODS

STUDY POPULATION

We analyzed a retrospective cohort of patients hospitalized from among 124 medical centers in the Department of Veterans Affairs (VA) between January 1, 2005 and December 31, 2013. The VA is an integrated care network that includes acute inpatient hospitals, outpatient primary care and sub-specialist clinics, outpatient pharmacies, rehabilitation facilities, long-term care facilities and domiciliaries. All VA personnel use the same EHR, Veterans Information Systems and Technology Architecture/Computerized Patient Record System (ViSTa/CPRS), for documentation and administration of clinical care.[236] The institutional review board and research and development committees of the Tennessee Valley Health Care System VA Medical Center, Nashville, TN, approved this study.

DATA COLLECTION

All data were collected from the EHR and accessed via the national Corporate Data Warehouse. The clinical data included vital signs, laboratory data, inpatient and outpatient medication data, narrative text notes, ICD-9 codes for diagnoses, and Current Procedural Terminology (CPT) codes for procedures.

COHORT SELECTION

We examined a cohort of patients hospitalized at a VA facility during the study years. We included all hospitalizations for patients who had a cirrhosis diagnosis (based on a history of two outpatient or one inpatient) ICD-9 code (571.2 or 571.5) and had AKI during their hospitalization with a maximum inpatient creatinine of at least 1.5 mg/dl. The maximum inpatient creatinine cutoff was used to be compliant with International Ascites Club criteria for HRS (Refer to Table 5).[30] We excluded hospitalizations where the patient was on dialysis prior to admission, did not have at least one serum creatinine value within the year prior to admission or during the inpatient stay, who had a diagnosis of HRS prior to the hospitalization, who had a prior hospitalization with AKI, or who were discharged in less than forty eight hours.

We performed stratified sampling based on presence/absence of an ICD-9 code for HRS, level of kidney injury, and level of liver disease. Acute Kidney Injury was defined by the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines: with Stage I being a defined as a rise in creatinine of $\geq$0.3 mg/dl from baseline; Stage II being defined as a doubling of serum creatinine from baseline; and Stage III being defined as a tripling of serum creatinine or initiation

of dialysis. Severity of liver disease was defined by the Model for End Stage Liver Disease (MELD) score: a combination of three laboratory values: the serum creatinine, international normalized ratio, and platelet count. We sampled in blocks of twelve: six patients were selected if they had an ICD-9 code for HRS (572.4) anytime during their hospitalization; six patients (without an HRS ICD-9 code) were selected based on two levels of kidney injury (KDIGO Stage I versus KDIGO Stages II and III) and three levels of MELD (< 20, >= 20, and unable to calculate). We randomly selected a total of 42 blocks (504 inpatient admissions) to serve as the gold standard cohort.

Table 5: Diagnostic criteria for Hepatorenal Syndrome from the International Ascites Club.

Criteria
Cirrhosis with ascites
Serum creatinine > 1.5 mg/dl
No improvement of serum creatinine after at least 2 days with diuretic withdrawal and volume expansion
Absence of shock
No current treatment with nephrotoxic drugs
Absence of parenchymal kidney disease

Outcome

Two physician annotators reviewed the 504 hospitalizations reviewing all clinical notes, relevant laboratory values, medications, and radiology reports to assign each hospitalization into one of five categories: HRS Type I, HRS Type II, HRS Type Indeterminate, Maybe HRS, and Not HRS. Reviewers were instructed to differentiate Type I, Type II, and Not HRS based on International Ascites Club criteria.[30] Type Indeterminate was reserved for cases where the reviewer felt the patient had enough evidence for HRS, but could not differentiate between Type I and II; whereas, Maybe HRS was reserved for cases of clinical uncertainty. We employed a practice phase where the two annotators worked in blocks of twelve patients until the inter-annotator agreement was ≥ 0.8. Disagreements on the 504 patient set were adjudicated by a board certified nephrologist. We report the inter-annotator agreement for the 504 charts that were reviewed. To reduce the problem to a two-class classification measure, we combined HRS Type I, Type II, Type Indeterminate, and Maybe HRS into a "Yes HRS" category. We performed a sensitivity analysis to examine classification performance after excluding "Maybe HRS" from model building and validation.

Predictor Variables

We included 649 variables from the structured data in the EHR, including demographics (3), laboratory values (92), vital signs (21), home medications (99), inpatient medications (116), medical history (129), inpatient diagnoses and procedures (176), and four other miscellaneous variables. To the structured data we added nine engineered variables comprised of the patient's creatinine response to various events during hospitalization. Variable engineering was performed

using the training set and validated on the test set prior to inclusion. A detailed summary of these variables and associated definitions are included in Online Appendix A.1 and A.2. To the structured variables, we added variables from natural language processing of the clinical notes as outlined in the next section.

With the exception of cirrhosis-related or nephrotoxic medications (e.g., lactulose, rifaximin, albumin, norepinephrine, cyclosporine), which were coded as separate variables, all medications were represented by their corresponding VA drug class code (e.g., "cephalosporin 3rd generation"). The VA drug class codes are available publicly through the VA National Drug File.[237] With the exception of three prehospitalization laboratory variables, the inpatient laboratory values and vital signs were summarized by their maximum, minimum, and mean or median. Missing values for laboratory test results were filled in using Markov Chain Monte Carlo multiple imputation using a subset of co-morbid conditions, medications, and procedures (See Online Appendix A.3).[238]

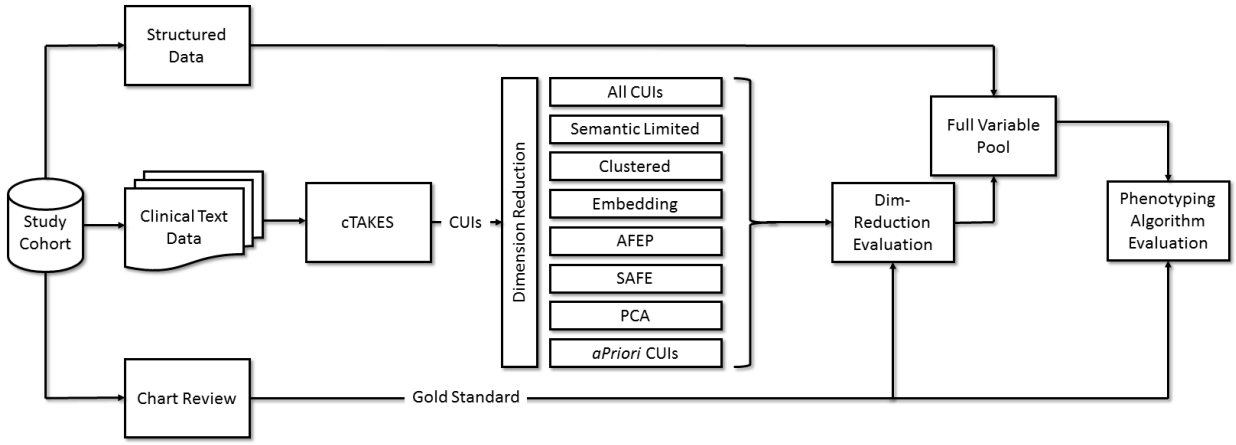
Natural Language Processing

We filtered all available clinical notes based on authorship by first including only physicians and advanced practice providers, and then excluding specialties unlikely to address hepatic pathology (podiatry, ophthalmology, and dentistry). We converted the documents into a string of CUIs mapped to the UMLS (version 2013AB)[84] using the clinical Text Analysis Knowledge Extraction System (cTAKES) version 3.2.[85] To manage the large number of unique CUIs and data sparsity, based on inspection and evaluation of instability of modeling within the training data, we first filtered the output by removing CUIs with a less than 2% or greater than 90% prevalence among documents. All CUI counts were log transformed. From this data, we evaluated nine different dimensionality reduction techniques: (1) using the full set of CUIs; (2) CUIs limited by semantic type; (3) CUIs aggregated by semantic similarity; (4) document embedding using the raw text; (5) document embedding using CUIs; (6) an *a priori* selection of CUIs based on domain knowledge; (7) Yu's Automated Feature Extraction for Phenotyping (AFEP);[93] (8) Yu's Surrogate-Assisted Feature Extraction (SAFE); [92] and (9) principal component analysis (PCA). We refer the reader to Online Appendix Tables A.4, A.5, and A.8 for the list of semantic type filters, *a priori* selected CUIs, and AFEP/SAFE selected CUIs, respectively.

To aggregate CUIs by semantic similarity we first limited by semantic type and then constructed a pairwise similarity matrix using the Information Content based on the Leacock and Chodorow distance measure, which has been shown to exhibit good performance when compared against other semantic similarity measures.[239] We subsequently performed *k*-medoids clustering to find groups of similar CUIs. Seventy clusters were chosen using the gap statistic and the "1-standard-error" rule.[240] For models (4) and (5) we used the Distributed Memory Model of Paragraph Vectors (doc2vec)[150] as implemented by the python gensim package.[241] We utilize the term "document embedding," as opposed to "word embedding," signifying doc2vec's ability to consume variable length text, and therefore obviate the need to combine word vectors. Similar to Turner *et al.*[155] we pre-processed raw text by removing non-alpha numeric characters and eliminating stopwords before using the doc2vec algorithm to generate vectors. No processing of the CUIs was performed other than the default parameters within cTAKES. For PCA, we kept sufficient components (395) to explain 95% of the variance. Refer to Figure 8 for the workflow.

Figure 8: Workflow describing Natural Language Processing pipeline.

Note: cTAKES: clinical Text Analysis Knowledge Extraction System; CUI: Concept Unique Identifier; AFEP: Automated Feature Extraction for Phenotyping; SAFE: Surrogate-Assisted Feature Extraction; PCA: Principal Component Analysis



Final Phenotyping Model Development

We tested five different classification models: logistic regression (LR), support vector machines (SVM), gradient boosting (GBM), random forest, and naïve Bayes. For LR and naïve Bayes we first performed variable selection using penalized LR, using the L_1 penalty (Least Absolute Shrinkage and Selection Operator—LASSO), to select a subset of the predictor variables.[42] For the remainder of the models we used the full set of predictor variables. The hyperparameters for SVM, GBM, and random forest were optimized using five-fold cross validation on the training set. A Gaussian distribution was assumed for naïve Bayes.

NLP Dimensionality Reduction and Phenotyping Model Assessment

We assessed the NLP dimensionality reduction techniques by constructing an SVM model using only the NLP variables with HRS as the outcome measure. The Radial Basis Function served as the SVM kernel and hyperparameters, C and γ, were optimized using grid search and 5-fold cross validation. While it is possible that the dimensionality reduction techniques may perform differently using an alternative model assessment method, we elected to test NLP variables with an SVM model because we wanted to utilize a method that had a low bias and few assumptions about the model parameter development, to allow for complex interactions to be discoverable in the CUI data. While this can result in high variance, we limited the values of C in the grid search to prevent very small C values that would increase the variance and over-fitting to observed data. In addition, this machine learning framework has been shown to work well with NLP variables.[57,242,243]

Performance of the NLP dimension reduction technique and the final phenotyping algorithm were calculated using bootstrapping (100 bootstrap samples) to estimate discrimination (area under the receiver operating characteristic [ROC] curve [AUC], F1-measure, precision, recall) and calibration (slope and intercept of the best fit line through the observed to predicted probability plot and Brier score) metrics.[129,244] We defined statistical significance as non-crossing of the 95% bootstrapped confidence intervals. We compared the discriminatory performance of the machine learning algorithms to the ICD-9 code.

We conducted an error analysis using the best machine learning method and studied the false positives and false negatives. We looked at false positives and false negatives at three cut-points for the probabilistic phenotype: the optimal sensitivity and specificity based on Youden's index, sensitivity of 0.95, and specificity of 0.95. For each of these scenarios, we examined the annotators' notes on the gold standard to understand why the errors occurred.

RESULTS

Based on manual annotation there were 87 cases with Type I HRS, 19 with Type II HRS, 16 with Type Indeterminate, 88 with Maybe HRS, and 294 without HRS. Table 6 shows a summary of the cohort after the case annotations were dichotomized as noted in the methods, resulting in a total of 210 (41.7%) hospitalizations with HRS. Eighty cases were adjudicated, yielding a weighted Cohen's kappa of 0.83. Males represented 98.2% of the total admissions, with a median age of 61. White patients accounted for the majority of hospital admissions (71.1%). The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of a discharge ICD-9 code for HRS were 57.6%, 88.8%, 78.6%, and 74.6%, respectively. The sensitivity, specificity, PPV, and NPV of an HRS ICD-9 code at any time during hospitalization were 87.1%, 76.5%, 72.6%, and 89.3%, respectively.

Table 6: Characteristics of the cohort of cirrhotic patients with and without HRS as determined by chart review.

[1] Note: A patient may have more than one etiology of cirrhosis, hence percentages add up to greater than 100%.

Characteristic	**HRS Diagnosis (n = 210)**	**No HRS Diagnosis (n = 294)**
DEMOGRAPHICS		
Age, mean (SD)	60 (7.9)	62 (10.2)
Gender (male), n (%)	208 (99.0%)	292 (99.3%)
Race, n (%)		
White	154 (73.3%)	201 (68.4%)
Black	28 (13.3%)	65 (22.1%)
Other	28 (13.3%)	28 (9.5%)
PRE-ADMISSION CHARACTERISTICS		
Cirrhosis Etiology, n (%)[1]		
Alcoholic	130 (61.9%)	151 (51.4%)
Viral (Hepatitis B and C)	112 (53.3%)	130 (44.2%)
NAFLD	31 (14.7%)	41 (13.9%)
Congestive Heart Failure, n (%)	19 (9.0%)	61 (20.7%)
Diabetes Mellitus, n (%)	59 (28.1%)	122 (41.5%)

Characteristic	**HRS Diagnosis (n = 210)**	**No HRS Diagnosis (n = 294)**
Chronic Kidney Disease, n (%)	20 (9.5%)	54 (18.4%)
Prior Cirrhosis Complications, n (%)		
Hepatic Encephalopathy	75 (35.7%)	61 (20.7%)
Varices	58 (27.6%)	66 (22.4%)
SBP	30 (14.3%)	19 (6.5%)
Ascites	122 (58.1%)	132 (44.9%)
Hepatocellular Carcinoma	28 (13.3%)	22 (7.5%)
Baseline Creatinine, mean (SD)	1.04 (0.42)	1.15 (0.49)
INDEX HOSPITALIZATION CHARACTERISTICS		
Maximum Creatinine, mean (SD)	4.16 (2.10)	2.75 (1.50)
Maximum Blood Urea Nitrogen, mean (SD)	78.2 (49.0)	49.9 (26.0)
Average Sodium, mean (SD)	132.6 (5.7)	135.8 (5.0)
Average Bilirubin, mean (SD)	12.7 (11.5)	4.4 (6.2)
Average Albumin, mean (SD)	2.4 (0.6)	2.6 (0.7)
Average INR, mean (SD)	2.0 (0.7)	1.7 (0.7)
Admission MELD, mean (SD)	26.3 (8.4)	20.5 (7.1)
Discharge HRS ICD-9 Code, n (%)	170 (81.0%)	63 (21.4%)

There were a total of 23,415 distinct CUIs within the entire document corpus, and a total of 6,985 distinct CUIs after initial frequency filtering. Limiting based on semantic type reduced the total number of distinct CUIs to 2082. The median number of CUIs per cluster was 12 (IQR: 5 – 18). AFEP and SAFE selected thirty-six and three CUIs, respectively. Table 7 presents the total number of variables and evaluation results for each of the nine NLP strategies. Document embedding using CUIs (AUC of 0.79, 95% CI: 0.79 – 0.80) significantly improved performance compared to embedding using raw text (AUC of 0.66, 95% CI 0.66 – 0.67). The *a priori* CUI selection, semantically informed clustering, and the high-throughput phenotyping methods (SAFE and AFEP) had statistically similar performance (AUC of 0.81 – 0.82). The *a priori* CUI set was selected for further analysis due to their clinical relevance and ease of interpretation.

Table 7: Evaluation of dimension reduction techniques for handling Natural Language Processing outputs for phenotyping.

Note: ¥: the counts for these models are doubled because they include both the positive assertion and the negative assertion; AFEP and SAFE include an extra variable for note count. AFEP: Automated Feature Extraction for Phenotyping; SAFE: Surrogate-Assisted Feature Extraction; CUI: Concept Unique Identifier; AUC: Area Under the Curve.

Model	No. of Variables	Precision	Recall	F-measure	AUC
Full CUI Set	13,970¥	0.56 (0.55, 0.57)	0.84 (0.83, 0.84)	0.68 (0.67, 0.70)	0.74 (0.74, 0.75)
Semantic Type Limited CUI Set	4,164¥	0.63 (0.62, 0.64)	0.80 (0.79, 0.81)	0.70 (0.68, 0.71)	0.73 (0.72, 0.73)
AFEP	37	0.66 (0.65, 0.67)	0.84 (0.83, 0.86)	0.74 (0.73, 0.74)	0.82 (0.81, 0.82)
SAFE	4	0.73 (0.72, 0.74)	0.79 (0.78, 0.80)	0.76 (0.75, 0.76)	0.82 (0.81, 0.82)
Principal Component Analysis	395	0.53 (0.52, 0.54)	0.77 (0.74, 0.80)	0.61 (0.60, 0.63)	0.57 (0.56, 0.57)
Document Embedding with Raw Text	500	0.58 (0.57, 0.59)	0.65 (0.62, 0.67)	0.60 (0.59, 0.61)	0.66 (0.66, 0.67)
Document Embedding with CUIs	500	0.66 (0.65, 0.67)	0.79 (0.78, 0.81)	0.72 (0.71, 0.72)	0.79 (0.79, 0.80)
Clustered CUIs	140¥	0.72 (0.71, 0.73)	0.78 (0.77, 0.79)	0.73 (0.72, 0.73)	0.82 (0.81, 0.82)
A priori CUIs	52¥	0.66 (0.65, 0.67)	0.84 (0.83, 0.85)	0.74 (0.73, 0.74)	0.81 (0.80, 0.81)

Combining the structured and NLP variables, there were a total of 701 candidate predictors. LASSO selected 21 variables. The results of the model comparisons are shown in Table 8. Logistic regression had the best performance in terms of AUC, though modest performance in terms of calibration. Figure 9 (Panel A) shows the ROC curves with 95% confidence intervals for the 5 methods. The sensitivity and specificity are also plotted for the HRS ICD-9 code (both for a discharge ICD-9 code and any ICD-9 code during the inpatient stay). Logistic regression dominated the other methods and was superior to using just the ICD-9 code. Figure 9 (Panel B) shows the smoothed calibration curves for the different methods based on Van Hoorde et al.[244] Though calibration appears relatively uniform for regression, GBM, SVM, and random forest based on the Brier score; the calibration curve shows GBM and SVM had superior performance. As part of our sensitivity analysis, appendix Table A.6 shows the classifier performance after building the five classifiers after excluding "Maybe HRS" from the model building and validation. We note slight improvement for regression (AUC of 0.94); however, we elected to maintain "Maybe HRS" within the model to account for edge cases. By varying the probability threshold, the user may include/exclude clinically uncertain cases. Appendix Table A.7 shows model performance using the SAFE CUIs for comparison. Overall model performance for

logistic regression is largely unchanged, though the individual variables selected by LASSO identify more structured variables to make up for the fewer NLP variables.

Table 8: Discrimination and calibration performance of the five models to phenotype Hepatorenal Syndrome.

Note: Slope and Intercept refer to the parameters of the best-fit line through the observed-to-predicted probability plot; AUC: Area Under the Curve

Model	AUC (95% CI)	Slope (95% CI)	Intercept (95% CI)	Brier Score (95% CI)
Logistic Regression	0.93 (0.92, 0.93)	0.68 (0.65, 0.71)	0.18 (0.13, 0.24)	0.11 (0.11, 0.11)
Gradient Boosting	0.88 (0.88,0.88)	1.26 (1.21, 1.31)	0.15 (0.10, 0.20)	0.14 (0.13, 0.14)
Naïve Bayes	0.73 (0.72, 0.74)	0.04 (0.03, 0.04)	-0.41 (-0.53, -0.29)	0.32 (0.30, 0.33)
Random Forest	0.91 (0.91, 0.91)	2.01 (1.95, 2.06)	0.29 (0.24, 0.35)	0.13 (0.13, 0.13)
Support Vector Machine	0.90 (0.90, 0.91)	0.74 (0.71, 0.77)	-0.12 (-0.17, -0.07)	0.13 (0.12, 0.13)

Figure 9: Discrimination (Panel A), via the ROC curve, and calibration (Panel B), via smoothed observed-to-expected probability plots, for the five different various models for phenotyping Hepatorenal Syndrome phenotyping models.

Note: The grey square represents performance for a Hepatorenal Syndrome ICD-9 code anytime during the admission. The grey circle represents a Hepatorenal Syndrome ICD-9 code as a discharge diagnosis. LR: Logistic Regression; SVM: Support Vector Machine; GBM: Gradient Boosting Machine; NB: Naïve Bayes; RF: Random Forest

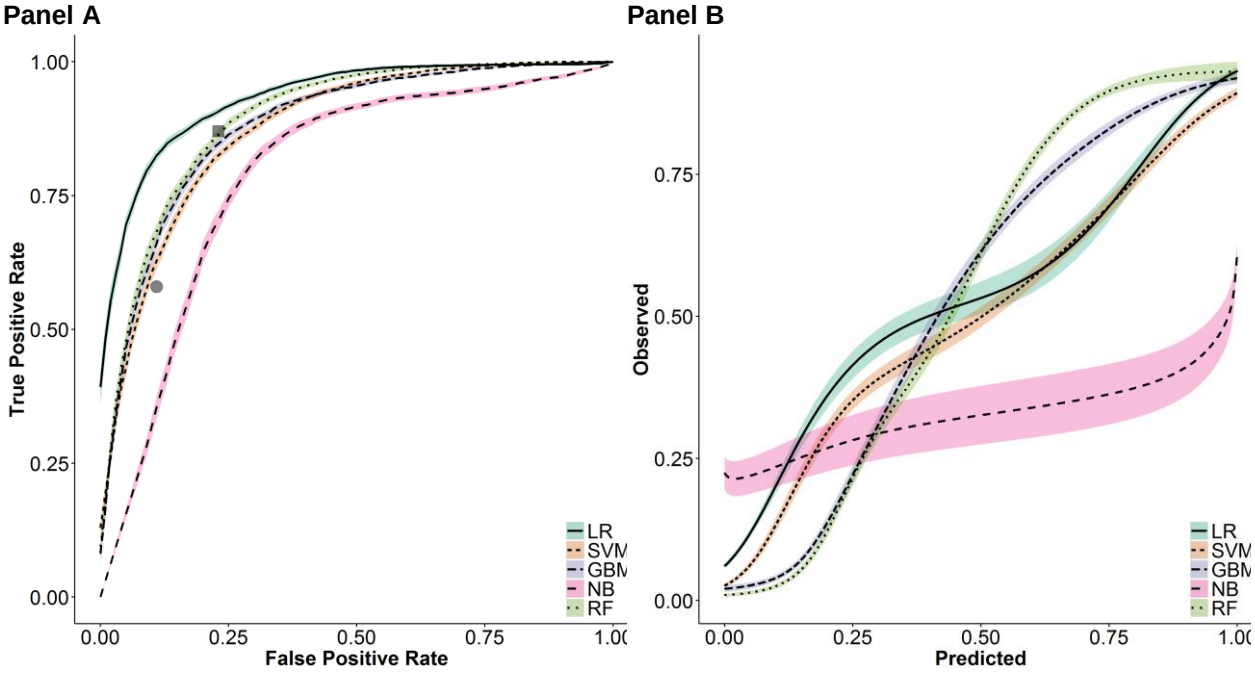


Table 9 reports the odds ratios for the variables used in the LR model. Significant variables predictive of HRS include an ICD-9 code for HRS, NLP mention of HRS, inpatient use of midodrine, the peak serum creatinine after the first 48 hours of admission, and the average mean corpuscular hemoglobin concentration (MCHC). Variables predictive of other causes of renal failure include an ICD-9 code for acute tubular necrosis (ATN), NLP mention of shock, high urine sodium, a significant difference between the maximum inpatient serum creatinine versus at discharge, and higher serum sodium.

Table 9: Odds ratios and confidence intervals for the logistic regression model based on 100 bootstrap samples.

Note: INR: International Normalized Ratio; MCHC: Mean Corpuscular Hemoglobin Concentration; NLP: Natural Language Processing; HRS: Hepatorenal Syndrome; ATN: Acute Tubular Necrosis; NAFLD: Non-alcoholic Fatty Liver Disease

Variable	Odds Ratio (95% CI)	Variable	Odds Ratio (95% CI)
Inpatient Labs	-	**Temporal**	-
Average Serum Sodium	0.67 (0.64, 0.70)	Creatinine Diff. (max inpt. to discharge)	0.21 (0.20, 0.23)
Average Urine Sodium	0.73 (0.70, 0.77)	Creatinine Diff. (1st 48 hours vs. rest of stay)	0.55 (0.51, 0.60)
Average Bicarbonate	0.79 (0.76, 0.83)	Peak Creatinine After First 48h	1.78 (1.66, 1.91)
Minimum Albumin	0.84 (0.81, 0.88)		
Average Glucose	0.94 (0.90, 0.97)	**ICD 9 Codes**	-
Average Total Bilirubin	1.15 (1.09, 1.20)	Inpatient ATN	0.40 (0.36, 0.45)
Minimum INR	1.16 (1.11, 1.21)	Inpatient NAFLD	1.07 (1.03, 1.11)
Average Blood Urea Nitrogen	1.16 (1.07, 1.26)	Inpatient Ascites	1.59 (1.51, 1.67)
Minimum Blood Urea Nitrogen	1.77 (1.63, 1.93)	Inpatient HRS	9.98 (9.12, 10.93)
Average MCHC	1.96 (1.87, 2.05)		
		NLP	-
Inpatient Medications	-	(+) Shock	0.21 (0.20, 0.23)
Midodrine	3.24 (2.89, 3.62)	(+) Paracentesis	1.37 (1.30, 1.43)
		(+) HRS	1.78 (1.67, 1.90)

Table 10 reports our error analysis at three levels of cut-offs: optimal using Youden's index, high sensitivity, and high specificity. As expected, false positives versus false negatives dominate at higher sensitivity and higher specificity, respectively. False positives at high sensitivity are primarily caused by the algorithm's inability to detect improvement with fluid administration, separating chronic kidney disease from HRS, and other causes of AKI in cirrhotics. At higher specificity, false negatives are caused by high urine sodium, chronic kidney disease, and competing diagnoses. At an optimal threshold, the majority of errors stemmed from an inability to identify improvement with fluid administration. Insufficient information caused errors at all cut-points, though a relatively small percentage of errors.

Table 10: Error analysis of false positive and false negatives using the logistic regression model on the test set at three different thresholds.

Note: FP: False Positive; FN: False Negative; CKD: Chronic Kidney Disease; ATN: Acute Tubular Necrosis; GI: Gastrointestinal; HIVAN: Human Immunodeficiency Virus Associated Nephropathy

	Sensitivity = 0.95		**Specificity = 0.95**		**Optimal Threshold (Youden's Index)**	
	FP (n=21)	FN (n=3)	FP (n=3)	FN (n=15)	FP (n=9)	FN (n=5)
High Urine Sodium		2		3		2
Improved with fluids	9		1	1	5	
CKD	4			2		2
Competing Diagnosis (sepsis)				2		1
Competing Diagnosis (contrast)	1				1	
Competing Diagnosis (ATN)	1		1		1	
Competing Diagnosis (hypotension or shock)	1			2	1	
Competing Diagnosis (multiple)	1			1		
Competing Diagnosis (GI Bleed)	1			1		
Competing Diagnosis (HIVAN)	1					
Insufficient Information	1	1	1	3	1	
Error in Underlying Data	1					

DISCUSSION

This research demonstrates that it is possible to create a high performance probabilistic phenotyping algorithm to detect cases of HRS. This is one of the first efforts to phenotype AKI etiology, a condition that effects up to 2% of hospitalized patients.[245] Penalized LR achieved the best performance with an AUC of 0.93 (95% CI: 0.92-0.93). NLP significantly boosted the performance of the model from an AUC of 0.82 (95% CI: 0.81-0.83). The sensitivity and specificity of an ICD-9 code anytime during the hospitalization were 87.1% and 76.5%, respectively; whereas, a discharge ICD-9 code had a sensitivity and specificity of 57.6% and 88.8%, respectively. At Youden's index, the LR algorithm would have a sensitivity of 85.4% and a specificity of 84.0%. The probabilistic phenotyping algorithm allows one to alter the thresholds for varying levels of sensitivity and specificity depending on the needs of the user.

Optimizing the algorithm required handling the large number of NLP variables. Automated dimensionality reduction in NLP based classification has been shown to improve performance in multiple studies.[246–248] Increasing effort has been placed on high-throughput phenotyping to perform automated feature selection/dimension reduction, though to date they have been primarily tested in chronic conditions where the data density is much higher. In our study, manual NLP variable selection using domain knowledge performed similarly to dimensionality reduction using SAFE, AFEP, and semantic similarity informed clustering. Manual variable selection has been shown to perform favorably in other studies.[61,67] For instance, Chen et. al. showed that a feature set selected by domain experts outperformed a data driven approach in phenotyping algorithms for Rheumatoid Arthritis, Colorectal Cancer, and Venous Thromboembolism.[249]

Although embeddings have been used for phenotyping tasks, we demonstrate its performance in acute illness and using CUIs instead of raw text.[154,155] Turner *et al.* showed their word embedding task using raw text outperformed bag-of-words models but did not outperform machine learning models using CUIs. We show that CUI based models (including embedding) outperform embedding models using free text. Increasing effort is being applied to mapping free text to a domain ontology for purposes of improving a wide variety of NLP tasks[246,250] and constructing shareable, computable clinical data warehouses.[251]

Though machine learning algorithms are increasingly popular for cohort identification,[57] our study showed superior performance with penalized LR. Regression has been used for phenotyping efforts[78] and, in at least one risk prediction study comparing regression to machine learning models, regression performed better.[105] Machine learning methods such as support vector machines and random forests tend to perform well on classification tasks where multiple interactions exist between the predictor variables, which suggests that complex interactions may not have been highly prevalent in these data. Additionally, despite the better discriminatory power of the logistic regression model, calibration was better with gradient boosting and support vector machines, which suggests that for some cut-points performance may still favor the machine learning methods.

The most important variable based on odds ratio was the HRS ICD-9 code. Inpatient codes for ascites also significantly increased the probability of HRS. This makes pathophysiologic sense because development of ascites and HRS are tightly related, particularly in HRS Type II. Inpatient administration of midodrine, a medication that increases the blood pressure, was significantly predictive of HRS. This is also a logical finding because midodrine is used in only a

few contexts in medicine and one of them is treatment of HRS. NLP variables that were predictive of HRS include mention of HRS and mention of paracentesis (removal of accumulated fluid in the abdomen), which is indicative of the presence of clinically significant ascites. Predictors with good negative predictive value for HRS include variables that indicate less severe portal hypertension (increased blood pressure in the abdominal blood vessels), other causes of acute kidney injury (ATN and shock), and significant improvement in creatinine levels at time of discharge.

To better understand failure points and edge cases, we performed an error analysis, revealing three common themes. First, errors were made in the system assessing response to fluid administration. In essence, this is a temporal pattern recognition problem. Though some temporal type variables were included in the model, they were insufficient to capture the full variation of response waveforms. Second, there were challenges differentiating HRS from other causes of kidney failure in cirrhotics. HRS is commonly one of several competing diagnoses in clinical practice when diagnosing the etiology of AKI in cirrhosis. The phenotyping system performed well in most cases. Finally, insufficient information caused a low level of persistent error across all cut-points. While this is unavoidable when using retrospective data, it may be mitigated when using the system prospectively. Importantly, our probabilistic phenotyping model allows the user to tailor the cutoff to the intended use: higher sensitivity for clinical decision support and higher specificity for defining cohorts in secondary data use analyses.

LIMITATIONS

There are some limitations to this research that are worth highlighting for refinement and extension of this investigation. First, this is a retrospective observational cohort and there were gaps in documentation that likely lowered ascertainment from chart review for the phenotype. Second, the VA data may not be representative of other clinical environments due to the slightly older average age and predominance of men. The other clinical variables, however, are not significantly different than other studies published regarding HRS.[37,52,53] We only performed internal validation; however, we aimed to increase generalizability by sampling across a broad range of kidney injury and liver disease. Moreover, all variables are common to other electronic health records, and the selected variables make pathophysiologic sense. Third, several significant predictors were ICD-9 codes, but with the transition to ICD-10 in the US, the algorithm's performance cannot be assured. At the same time, it is worth noting that there are one-to-one mappings for two of the important ICD-9 codes (ATN and HRS) based on the General Equivalent Maps (GEMs) framework.[252] The code sets defining non-alcoholic fatty liver disease and ascites would require additional validation.

NLP dimension reduction was assessed with SVM, and it is possible that an alternate method may have ranked the methods in a different order. We did not test expectation-maximization methods of clustering, such as Gaussian mixture modeling, for dimension reduction as we do not know the inherent probability distribution of the data. Lastly, a more thoughtful exploration of mapping temporal changes using established methods may have improved performance.[253–255]

CONCLUSION

This study demonstrated the utility of a probabilistic phenotype that used machine learning based methods to retrospectively classify patients with HRS. Though we focused on one form of AKI due to its high mortality, lessons learned could be applied to phenotyping other forms of kidney injury. Domain knowledge and several automated dimension reduction methods demonstrated

similar performance for identifying acute illness. Penalized logistic regression identified a parsimonious set of features with excellent performance. Unlike the fixed sensitivity and specificity of the HRS ICD-9 code, this probabilistic model can be used at multiple set points depending on the use case (e.g., a bias towards specificity or sensitivity). Future directions include external validation and identifying HRS cohorts for predictive analytics, clinical decision making, and population management.

CHAPTER III: RISK PREDICTION MODELS FOR HEPATORENAL SYNDROME

CHAPTER OBJECTIVE

Risk prediction models increasingly complement clinical reasoning and decision making in modern medicine. Within the domain of HRS risk prediction, prior models focused on predicting the long-term risk of developing HRS (6 months to 2 years) with none predicting short-term risk during hospitalization, which could be useful to support immediate decision making regarding treatment. In this study, we developed and internally validated an HRS risk prediction algorithm using data available in the peri-admission window of patient hospitalization among a large nationwide veteran cohort of patients. We sought to develop an algorithm to support clinical decision making and initiate treatment earlier, thus improving anticipated outcomes.

PATIENTS AND METHODS

STUDY POPULATION

We analyzed a retrospective cohort of patients hospitalized from among 122 medical centers in the U.S. Department of Veterans Affairs (VA) between January 1, 2005 and December 31, 2013. The VA is an integrated care network that includes acute inpatient hospitals, outpatient care services, and long-term care facilities. All VA personnel use the same EHR, Veterans Information Systems and Technology Architecture/Computerized Patient Record System (ViSTa/CPRS), for documentation and administration of clinical care.[20] The institutional review board and research and development committees of the Tennessee Valley Health Care System VA Medical Center, Nashville, TN, approved this study.

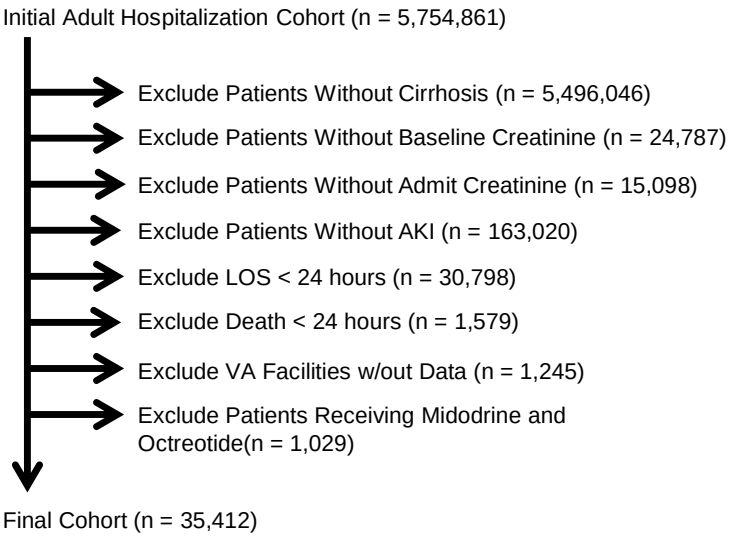
DATA COLLECTION

All data were collected from the EHR and accessed via the national corporate data warehouse. The clinical data included International Classification of Diseases - version 9 (ICD-9) codes for diagnoses, Current Procedural Terminology (CPT) codes for procedures, vital signs, laboratory data, and inpatient and outpatient medication data.

COHORT SELECTION

We included patients who had a cirrhosis diagnosis (based on a history of two outpatient or one inpatient) ICD-9 code (571.2 or 571.5) and had AKI on admission. AKI on admission was defined according to the Kidney Disease Improving Global Outcomes (KDIGO) guidelines to be an absolute rise in serum creatinine of at least 0.3 mg/, relative rise of 50%, or the new initiation of dialysis in a patient.[256] Baseline creatinine was defined as the average outpatient creatinine from values collected from 365 days to 7 days prior to admission[257] and peak creatinine was defined as the maximum creatinine between -24 and +24 hours of admission.

Figure 10: Cohort selection process from an initial sample of all inpatient admissions after applying exclusion criteria.



We excluded patients who were: (1) on dialysis prior to admission; (2) did not have at least one serum creatinine value within the year prior to admission; (3) did not have an admission creatinine; or (4) died within 24 hours of hospitalization. We also excluded patients from two VA facilities that had no laboratory measurements in the available source data. Furthermore, we excluded patients who received both octreotide and midodrine within 24 hours of hospitalization as this combination was pathognomonic for the treatment of HRS—indicating that a presumptive diagnosis was established at time of admission. We did not exclude other treatments for HRS (e.g., norepinephrine) because they could be used to treat alternate diseases (e.g., sepsis). We refer the reader to Figure 10 for a breakdown of the cohort selection process.

OUTCOME

The main outcome of interest was the presence of an ICD-9 code for HRS (572.4) during hospitalization or at the time of discharge, and the patient must have had a documented history of ascites or presented with ascites. See Appendix 1 for details regarding validation of the HRS ICD-9 code and ascites status ascertainment. If patients did not meet the ascites requirement, then they were assigned to the non-HRS AKI group (even if their EHR contained an ICD-9 code for HRS).

PREDICTOR VARIABLES

We started with a total of 404 variables during the pre-admission (all data up to -24 hours of admit) and the admission timeframe (-24 to +24 hours of admit). We included data -24 hours prior to admit as part of the admission timeframe to incorporate emergency room data. Refer to Table 11 for a breakdown of the variables. Since by definition 100% of HRS patients have ascites, the complication was excluded from the candidate predictor pool. A detailed summary of these variables and associated definitions are included in Appendix 2.

Table 11: Breakdown of the candidate predictor variables used in the Hepatorenal Syndrome risk prediction model.

(Note: KDIGO: Kidney Diseases Improving Global Outcomes; MELD: Model for End Stage Liver Disease; SBP: Spontaneous Bacterial Peritonitis)

Variable Group	Number of Variables in Group
Pre-Admission Timeframe (all data up to -24 hours of admit)	
Demographics	3
Comorbid conditions	64
Cirrhosis etiologies	3
h/o Cirrhosis complications	5
Home medications	142
Paracentesis within 3 days of admission	1
# of paracenteses within past 90 days	1
Admission Timeframe (-24 to +24 hours of admit)	
Inpatient medications	137
Laboratory values	24
Vital signs	12
Procedures (including paracentesis)	6
KDGIO renal failure stage	1
MELD score	1
SBP diagnosed at admission	1
Total IV Fluids	1

With the exception of cirrhosis-related medications (e.g., lactulose, rifaximin, albumin, norepinephrine, and vasopressin), which were coded as separate variables, all medications were represented by their corresponding VA drug class code (e.g., "cephalosporin 3rd generation"). The VA drug class codes are available publicly through the VA National Drug File.[237] Except for the baseline creatinine value, the remaining laboratory values corresponded to the 23 most commonly collected test results on inpatient admission. We summarized the inpatient laboratory values and vital signs as the average value during the admission timeframe, SBP was defined either by administrative code or > 250 neutrophils/mm^3 in ascites fluid. We eliminated any categorical variables that were present for less than 0.2% of admissions or showed perfect collinearity. The remaining 287 variables are outlined in Appendix 2. Missing values for race were replaced with "Unknown." Missing values for laboratory values were filled in using multiple imputation using a subset of co-morbid conditions, medications, and procedures (as

outlined in Appendix 3).[238,258] Missing vital signs and ages were imputed with the median admission values from the entire cohort.

MODEL DEVELOPMENT

We performed a penalized logistic regression, using the L_1 penalty (Least Absolute Shrinkage and Selection Operator — LASSO) , to select a subset of the predictor variables.[42] Refer to Appendix 1 for details on the variable selection procedure. We subsequently used the variables identified by the LASSO procedure in a generalized estimating equations (GEE) model clustered by patient using an exchangeable covariance structure, which adjusted for correlation due to the multiple admissions per patient.[259] Finally, we produced a traditional point-based scoring model, similar to the Framingham risk study,[260] based on the statistically significant variables from the GEE model.

MODEL ASSESSMENT

We reported the AUC of the GEE model with a 95% confidence interval (CI) calculated from the bootstrap samples and variable odds ratios.[129] We assessed model calibration using the Brier score (range from 0 to 1, where 0 implies perfect calibration), slope and intercept of the regression line between O/E probabilities, and an O/E probability plot.[261] We performed two sensitivity analyses: first, excluding hospitalizations where patients received vasopressin or norepinephrine on admission; second, excluding patients who could possibly have cardiorenal syndrome. Finally, we compared our model to a baseline that included only the Model for End-Stage Liver Disease (MELD) score as a predictor. All statistical analysis was performed using the R statistical programming suite, version 3.2.2.

RESULTS

After applying the inclusion and exclusion criteria, we identified 19,146 patients comprising 35,412 inpatient admissions. Hospitalization characteristics of patients with and without HRS diagnosis are summarized in Table 12. There was a median of one admission per patient with an interquartile range of one to two admissions, and a maximum of 23 admissions for one patient. The distribution of admissions is shown in Appendix 5. Males represented 98.2% of the total admissions, with a median age of 61. White patients accounted for the majority of hospital admissions (71.1%). The event rate for hospitalized HRS was 2,258 (6.4%).

Table 12: Characteristics of patients with and without HRS.

Characteristic	**HRS Diagnosis (n=2435)**	**No HRS Diagnosis (n=32977)**
Age, mean (SD)	61 (7.8)	63 (9.2)
Gender (male), n (%)	2392 (98.2%)	32393 (98.2%)
Race, n (%)		
White	1796 (73.8%)	23372 (70.9%)
Black	388 (15.9%)	6803 (20.6%)
Other	251 (10.3%)	2802 (8.5%)
Etiology, n (%)		
Alcoholic	1488 (61.1%)	17513 (53.1%)
Viral (Hep B and C)	1337 (54.9%)	17211 (52.2%)
NAFLD	497 (20.4%)	4893 (14.8%)

KDIGO Renal Failure Stage, n (%)		
Stage I	1092 (44.8%)	22828 (69.2%)
Stage II	670 (27.5%)	5243 (15.9%)
Stage III	673 (27.6%)	4906 (14.9%)
Congestive Heart Failure, n (%)	398 (16.3%)	9774 (29.6%)
Diabetes Mellitus, n (%)	950 (39.0%)	17030 (51.6%)
Chronic Kidney Disease, n (%)	872 (35.8%)	12713 (38.5%)
h/o Cirrhosis Complications, n (%)		
Hepatic Encephalopathy	1289 (52.9%)	10357 (31.4%)
Varices	859 (35.3%)	8270 (25.1%)
SBP	592 (24.3%)	4099 (12.4%)
Ascites	1936 (79.5%)	17959 (54.5%)
Hepatocellular Carcinoma	352 (14.5%)	3782 (11.5%)
Baseline Creatinine, mean (SD)	1.30 (0.60)	1.57 (1.42)
Admit Creatinine, mean (SD)	2.96 (1.74)	2.41 (1.82)
Admit Blood Urea Nitrogen, mean (SD)	51.0 (27.1)	39.6 (25.5)
Admit Sodium, mean (SD)	131.8 (6.4)	134.4 (5.7)
Admit Bilirubin, mean (SD)	7.5 (9.1)	3.0 (4.8)
Admit Albumin, mean (SD)	2.3 (0.7)	2.5 (0.9)
Admit INR, mean (SD)	1.8 (0.7)	1.6 (0.7)
Admit MELD, mean (SD)	26.5 (7.4)	19.5 (6.5)
Admit Mean Arterial Pressure, mean (SD)	94.8 (14.1)	101.2 (16.3)

Table 13 presents the 26 out of 287 variables selected by LASSO and the corresponding odds ratios from the GEE model. Not unexpectedly, the MELD score increased the probability of HRS (per 1-point increase, OR 1.16, 95% CI: 1.14 - 1.17). Other strong predictors included KDIGO Stage II renal failure (OR 1.23, 95% CI: 1.08 - 1.39), hepatic encephalopathy (OR 1.58, 95% CI: 1.43 - 1.75), diagnosis of SBP at time of admission (OR 1.57, 95% 1.37 – 1.81), and a paracentesis on the day of admission (OR 1.50, 95% CI: 1.35 - 1.68). Medication exposure to drugs the VA classifies as 'Non Opioid Analgesics' (primarily aspirin) significantly reduced the likelihood of HRS (OR 0.73, 95% CI: 0.64 - 0.83). Of the 'Non Opioid Analgesic' group, 167 out of 9,986 admissions had acute tubular necrosis (ATN), by ICD-9 code, versus 476 out of 25,426 admissions (p=0.22), which suggested that the difference was not borne out of non-steroidal anti-inflammatory drug induced AKI. The risk model had an AUC of 0.84 (95% CI: 0.83 – 0.85), a Brier score of 0.053 (95% CI: 0.050—0.055), slope of 0.98 (95% CI: 0.92—1.04), intercept of -0.04 (95% CI: -0.20—-0.12) and the O/E calibration curve is shown in Figure 11. When converting the model to a traditional point-based scoring model, points ranged from 0 to 55 (median 19, IQR: 14 – 25). Using the point-based model preserved discrimination (AUC of 0.82), though calibration was significantly worse (slope and intercept of the calibration line and Brier score of -3.50, 1.00, and 0.30, respectively). Refer to Appendix 6 for details.

The two sensitivity analyses as outlined in the Methods did not change the results of the LASSO variable selection or significantly affect the odds ratios from the GEE model. See Appendix 7 for

details. The MELD score-only model had an AUC of 0.76 (95% CI: 0.75—0.78), Brier score of 0.056 (95% CI: 0.053—0.059), slope of 1.00 (95% CI: 0.91—1.09), and intercept of 0.00 (95% CI: -0.23—0.24). We also investigated performance for two patient populations for whom clinical suspicion of HRS is typically low. First, in patients with an admission MELD score < 15 (roughly corresponding to Child-Pugh Classification A), the model had an AUC of 0.82. Second, in patients with a history of CKD and who did not present with SBP, the model had an AUC of 0.85.

Figure 11: Observed-to-expected probability plot from the GEE model. Each point represents a decile within the predicted probability.

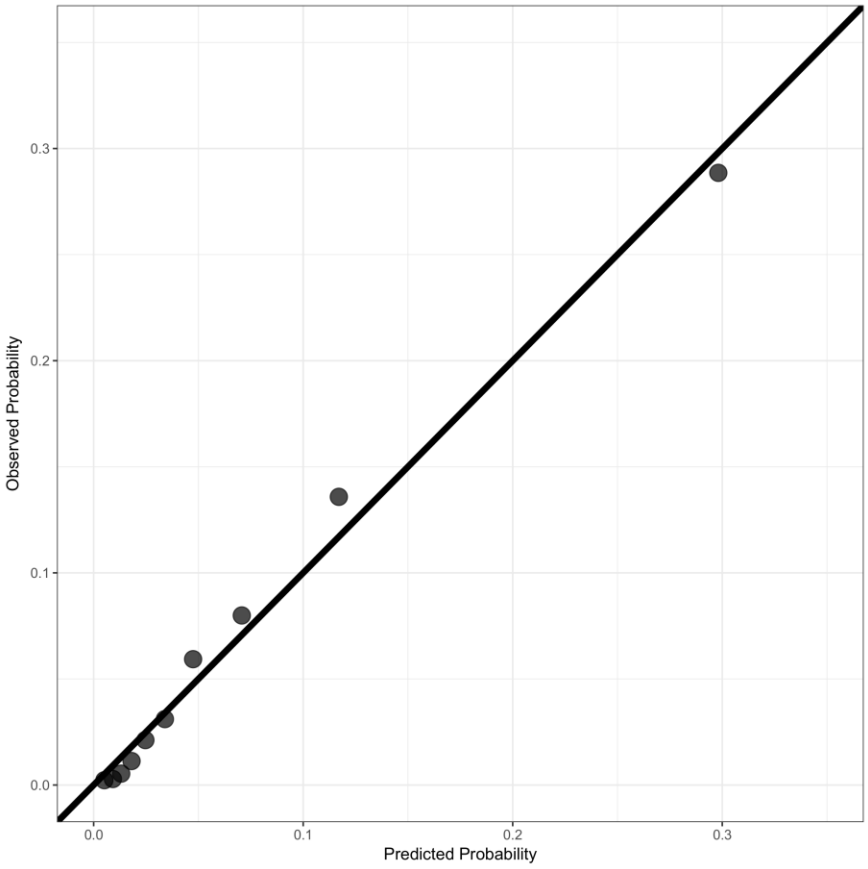


Table 13: Odds ratios for the general estimating equations model predicting HRS for variables selected by penalized logistic regression.

Risk Factor	GEE Odds Ratio (95% CI)
Admit Intravenous Fluids / 1000 mL	0.93 (0.90 - 0.97)
Admit MELD	1.15 (1.14 - 1.17)
Baseline Creatinine	0.79 (0.75 - 0.84)
Admit Sodium	0.99 (0.98 - 0.99)
Admit Bicarbonate	0.98 (0.97 - 0.99)
Admit Blood Urea Nitrogen	1.01 (1.00 - 1.01)
Admit Glucose	1.00 (1.00 - 1.00)
Admit Mean Corpuscular Hemoglobin Conc.	1.06 (1.01 - 1.11)
Admit Mean Corpuscular Hemoglobin	1.00 (0.99 - 1.02)
Admit Alkaline Phosphatase	1.00 (1.00 - 1.00)
Admit Partial Thromboplastin Time	1.00 (0.99 - 1.00)
Admit International Normalized Ratio	0.56 (0.50 - 0.64)
Admit Systolic Blood Pressure	1.00 (0.99 - 1.00)
Admit Temperature	0.98 (0.89 - 1.07)
Admit Weight	1.00 (1.00 - 1.00)
Admit Maximum Temperature	0.98 (0.92 - 1.05)
# Paracentesis in 90 days Pre-Admit	1.09 (1.07 - 1.12)
KDIGO Stage II (vs. KDIGO Stage I as baseline)	1.26 (1.12 - 1.41)
KDIGO Stage III (vs. KDIGO Stage I as baseline)	1.02 (0.88 - 1.18)
Ascites	1.65 (1.46 - 1.86)
Coronary Artery Disease	0.83 (0.74 - 0.94)
Hepatic Encephalopathy	1.42 (1.28 - 1.57)
Home Medication Analgesics	0.85 (0.78 - 0.94)
Admit Medication Albumin Infusion	1.40 (1.25 - 1.55)
Admit Medication Non Opioid Analgesics	0.79 (0.70 - 0.89)
Admit Procedure Paracentesis	1.36 (1.23 - 1.51)
Intercept	0.43

DISCUSSION

In this study, we developed a risk model to diagnose HRS at the point of hospital admission to support clinical diagnosis and decision making during the critical 48 hour surveillance period. Our model achieved good discrimination (AUC of 0.84), with excellent calibration (Brier score of 0.053 and good O/E ratio), and consists of basic medical history, common laboratory values, and initial medical management that can be obtained quickly during the admission window. This model includes modifiable risk factors, such as avoiding large volume paracentesis and use of non-opioid analgesics, and may support earlier initiation of treatment. Earlier treatment before kidney function worsens may improve survival.[51–54] Furthermore, earlier decision making may

reduce costs, reduce length of stay, initiate transplant evaluation,[262] and motivate referral to hospice.

Using the risk model at hospital admission could impact the standard of care in diagnosing HRS, which involves withdrawal of diuretics, plasma expansion with albumin for 48 hours, and ruling out other causes of renal failure. Other studies have attempted to use a variety of tests to diagnose HRS earlier. Investigations have either looked at imaging modalities or novel biomarkers. Promising imaging modalities include magnetic resonance imaging[263] and ultrasound;[264,265] whereas, biomarkers have looked at the predictive ability of arginine metabolism,[266,267] neutrophil gelatinase-associated lipocalin,[268] and cystatin C.[269] Other research looking at long term risk of HRS include low serum sodium concentration, high plasma renin activity, absence of hepatomegaly, low cardiac output, and the MELD score.[35,234,270,271] Though these studies are enticing, they were limited by small sample sizes, overly stringent exclusion criteria, or the need for specialized equipment or laboratory tests. In contrast, our study uses common clinical variables obtainable at the point of care during the admission window.

Our study corroborates known risk factors for HRS, but also highlights new ones. Of all the other cirrhosis complications, a history of hepatic encephalopathy was the only one to achieve a statistically significant relationship with HRS. Interestingly, back in 1972 Fischer and James postulated a connection between hepatic encephalopathy and HRS due to amino acid precursors of false neurochemical transmitters, such as phenylalanine and tyrosine, and their derivatives produced by gut bacterial decarboxylases.[272] Performing a paracentesis on admission, total number of paracenteses in the past 90 days, and SBP on admission were all highly predictive of HRS. Though we do not know if the paracenteses were diagnostic or therapeutic, since a diagnostic paracentesis' association to HRS would likely be mediated through a diagnosis of SBP, it is likely that the additional risk of paracentesis is borne from large volume withdrawal. Paracentesis is known to cause significant hemodynamic changes and activation of the renin angiotensin aldosterone system,[273,274] has been shown to precipitate AKI even with concomitant albumin expansion,[275] and precipitate HRS if not accompanied by albumin infusion.[276] A higher number of paracenteses in the past 90 days likely reflects patients with refractory ascites and is identifying patients with Type II HRS. Unsurprisingly, higher admission MELD score (indicating more advanced liver disease) and hyponatremia (which has already been shown to be predictive of future HRS)[270] were both correlated with HRS.

Counterintuitively, KDIGO Stage II renal failure was a significant predictor of HRS, but KDIGO Stage III was not. A greater percentage of patients with KDIGO Stage III went on to receive dialysis during their inpatient stay (2,642 out of 5,579 admissions, as opposed to 252 out of 5,911 admissions for Stage III and Stage II, respectively, $p < 0.0001$). Dialysis is routinely reserved for HRS patients who are headed for transplant (only 57 of the KDIGO Stage III patients received one) suggesting that their renal failure was due to other causes. Furthermore, patients with KDIGO Stage III had higher levels of CKD Stage IV and IV (2,045 out of 5,579 Stage III admissions versus 168 out of 5,911 Stage II admissions, $p < 0.0001$). For these KDIGO Stage III patients, a small amount of renal injury could have resulted in dialysis initiation. Looking at variables inversely correlated with HRS, patients with higher baseline creatinine were predisposed to other etiologies of renal failure include ATN, pre-renal causes, and medication induced kidney injury likely as a function of reduced renal reserve. Patients with existing kidney disease are more prone to community acquired AKI.[277]

There were several risk factors detected that have not previously been reported. It is unclear why an elevated MCHC is correlated with a diagnosis of HRS. Patients with advanced liver disease often have abnormal red blood cells including acanthocytes, spur cells, echinocytes, and target cells.[278] The reduced cell volume will lead to a higher MCHC calculation. The increased MCHC may be an indicator of advanced liver disease; however, these cells are also prone to hemolysis further impairing oxygen transport that may further predispose to HRS. Interestingly, higher International Normalized Ratio (INR) values decreased the odds of HRS. Since the INR is part of the MELD score, it is possible that the MELD score already captured the patients with the higher INR values. Patients with higher INRs were also more likely to have a GI bleed in our cohort. When looking at admissions with INR > 3.0, 235 out of 1,280 (18.3%) had a GI Bleed versus 4361 out of 29417 (14.8%) ($p = 0.0006$). Although a GI bleed may precipitate HRS, it is more likely to cause prerenal renal failure.

There are several limitations of this study that should be highlighted for further investigation. First, our cohort is largely male and may not generalize to female populations. However, biological sex has not played a role in prior research on HRS or AKI. Second, the administrative definition of the outcome showed limited sensitivity and good specificity. This could impact the study by coding bias that neglects patients less likely to receive a HRS ICD-9 code, which are likely some of the borderline patients with a broader differential for their renal failure. Furthermore, we acknowledge that some patients in the "No HRS" group are likely false negatives; however, this would bias our study towards the null. Our validation of the ICD-9 code was based on older criteria for HRS (particularly a hard cutoff for creatinine) because chart review was conducted for patients treated prior to the 2015 criteria. We note, however, that the inclusion/exclusion criteria for this study did not use a minimum creatinine threshold, and used the KDIGO values for AKI consistent with the 2015 International Ascites Club guidelines. Third, we could not differentiate Type I from Type II HRS. The average maximum creatinine during the inpatient stay for the HRS cohort was 4.22 +/- 2.18 (SD) mg/dL and 1767 out of the 2258 patients had a maximum creatinine > 2.5 mg/dL. Therefore, our cohort likely overrepresented HRS Type I and may not work as well at detecting HRS Type II. Fourth, ICD-9 codes have limitations in identifying comorbidities.[73] Fifth, we used an ICD-9 code for HRS at any point during hospitalization as the outcome measure whereas we limited our cohort to patients who presented with AKI. It is possible that some patients presented with non-HRS AKI and subsequently developed HRS during the hospitalization.

In conclusion, this study constructed a probabilistic risk prediction model to diagnose HRS within 24 hours of hospital admission using routine clinical variables in the largest ever published HRS cohort. Separating HRS from other causes of kidney injury can be challenging, and our model showed good performance even for groups generally thought less likely to have HRS. This would provide clinical utility by allowing earlier treatment initiation.[51–54]

CHAPTER IV: INFORMATION VISUALIZATION FOR MODEL ANALYSIS

CHAPTER OBJECTIVE

Information Visualization can help make sense of complicated, high-dimensional data and complement insights obtained from traditional, numerical analyses. In this study, we sought first to identify the most salient clinical features of cirrhosis and HRS that hepatologists find important. We then constructed two visualizations to help clinical decision making. The first visualization was a per-patient visualization giving a temporal view/summary of a patient's cirrhosis disease course, particularly as a means of helping to diagnose etiology of acute kidney injury. The second visualization uses a multiple coordinated views approach to assess the risk model performance presented in Chapter III. We report on a qualitative analysis of the visualizations.

METHODS

WORKFLOW OBSERVATION

Forty consenting physicians were observed in inpatient and outpatient settings across two institutions as part of a larger study; refer to Miller *et al.*[279] for details. In brief, observations were conducted using a human factors engineering approach including contextual interviews for data collection, interpretive debriefing sessions for data collation, data consolidation using thematic analysis, and additional data analysis using three work modeling approaches. The observations focused on "clinical decision making."

We paid special attention to the interaction between the clinician and the EHR. In particular, what information sources does a clinician use from the EHR when preparing for a patient encounter? Before most patient encounters, clinicians make an effort to obtain a quick overview of the patient's clinical course. Although, the primary aim of the parent study was looking at clinicians interactions with cirrhosis patients in general, several key aspects were relevant to understanding cirrhotics who present with kidney injury. We asked clinicians what parts of the medical record should be abstracted and visualized for quickly summarizing the patient's disease course.

DESIGN RATIONALE AND TOOLS

We constructed two visualizations. Broadly, both visualizations sought to improve early identification of HRS. The first visualization (TIMELINE) was a patient summary tool to help provide an overview of the patient's medical history. The second (CLUSTERVIEW) was an interactive visualization aid to help the user utilize the risk prediction model presented in Chapter III. The second visualization's goal was to use established paradigms for complex data and display a global view of this very complex model, yet allow the user to drill down and examine the relevant components. Doing so users are able to evaluate risk model performance in sub-groups. In a secondary workflow, the user would be able to see the relationship between a hypothetical, new patient and the other patients within the training cohort.

For TIMELINE we used D3.js, an interactive HTML based interface that allows for vector graphics and interactivity.[280] The visualization layer communicated through standard hypertext protocols with Microsoft Internet Information Services. All data meant for visualization was

converted into JavaScript Object Notation (JSON) format. Radial coordinates were calculated with the standard d3 radial transform. Drawing was accomplished using the d3.js "arc" tool.

We used the R programming interface and used Shiny for CLUSTERVIEW. Shiny is an interactive protocol that allows a web-like front end with the backend powered by R.[281] Shiny Dashboards allow the user to streamline the process of data input, statistical analysis, and graphical exploration. The GUI is designed to streamline the process of model performance analysis. Distinct panels are used for various stages of the analysis, including data input and filtering, and outlier detection.

DATA SETS

For TIMELINE, we utilized the patient pool presented in Chapter II. In brief, it consisted of 504 cirrhotic patients who were hospitalized with AKI from a variety of causes. Manual chart review adjudicated the cause of AKI. Though the dataset in Chapter II was primarily interested in the data surrounding the hospitalization of interest, for the visualization we expanded this dataset with time-series data of laboratory measurements, diagnoses, and procedure codes.

For CLUSTERVIEW, we utilized the dataset presented in Chapter III. In brief, it consisted of 35,412 patients with cirrhosis hospitalized with acute kidney injury from a variety of causes. The primary outcome of the model was development of HRS. The model was constructed with 287 predictor variables. Refer to Chapter III for full details.

METHODS ON USING CLUSTERING TO LOOK AT MODEL PERFORMANCE

We also assessed sub-population calibration and discrimination using unsupervised clustering, which collates similar observations together without forehand knowledge of any group membership. We performed unsupervised clustering using Kohonen's Self-Organizing Map (SOM),[282] implemented by the *kohonen* R package,[283] to divide our parent cohort into forty-nine subgroups based on the 286 variables. We used a SOM to perform the clustering because it also allows for abstraction and visualization. We utilized the Gower distance as the similarity measure, which handles both continuous and discrete variables.[284] We used the cluster instability metric to choose a map size (i.e., the number of clusters).[285] We tested the cluster instability metric for map sizes between 4x4 and 13x13 units and selected the size at which the instability metric displayed an "elbow point," the point past which we have diminishing marginal return in cluster stability.

The user can select to plot similar patients; each patient is represented by an asterixis on the plot. Patient similarity was defined in one of two ways: first, patients who all fall within a percentile range for HRS risk based on the model, e.g. top risk decile; second, all patients similar to a hypothetical, new patient. Unlike the clustering, however, when plotting similar patients to a new patient only the variables selected for plotting are used for computing similarity. The Gower distance was used for identifying patient similarity to a new patient.

Figure 12: Example clinical course visualization for a patient with alcoholic cirrhosis.

Note: CREAT: creatinine; MELD: Model for End-Stage Liver Disease; ALB: Albumin; BILI: Total Bilirubin; PLT: Platelet; PT: Prothrombin Time

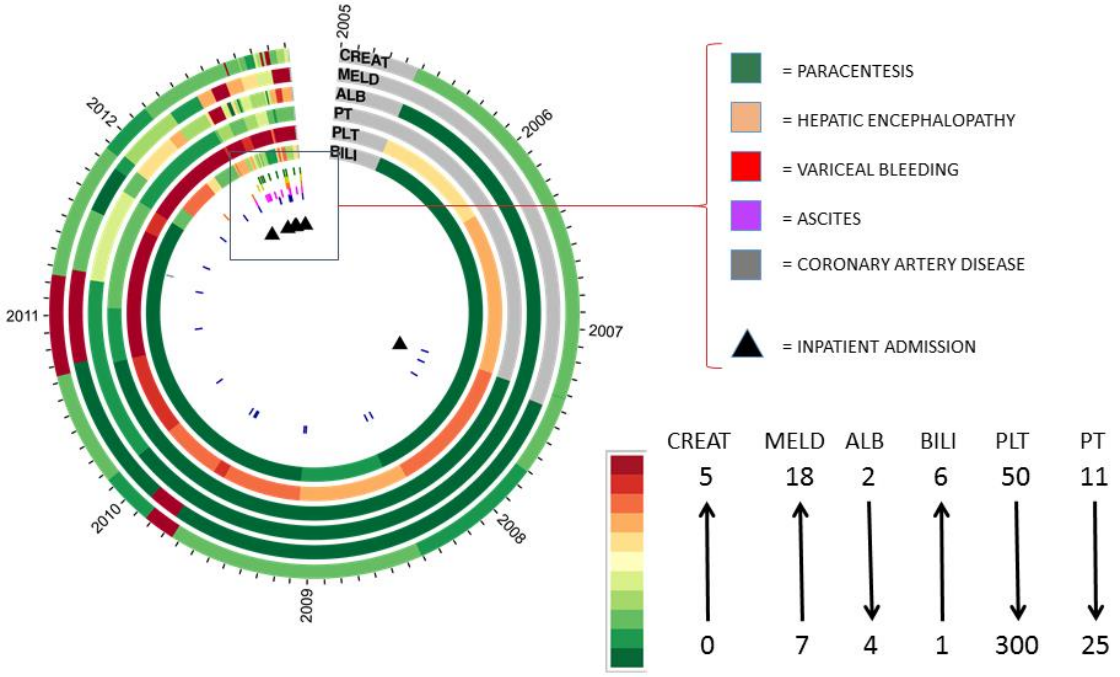
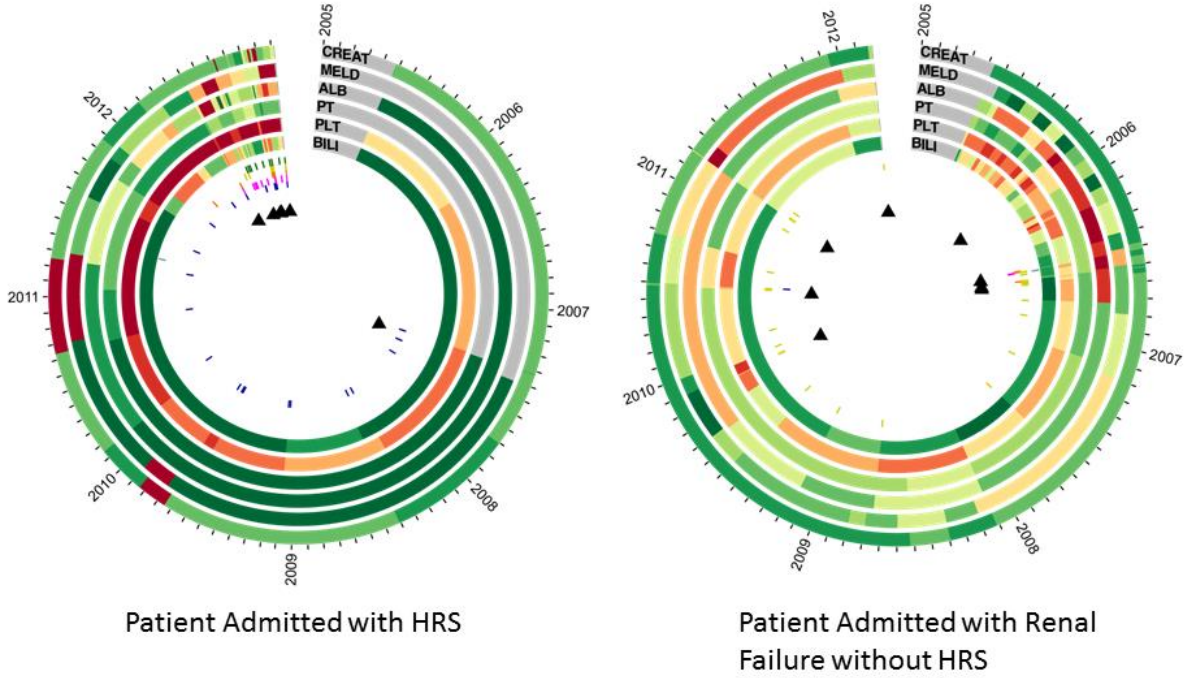


Figure 13: Example comparison of two patients with different clinical trajectories as depicted by the clinical course visualization tool.

The temporal axis is circular and moves clockwise, starting in the year 2005 (labeled on the outermost ring). The end of the ring represents the hospital encounter of interest; for example in patient A, the final hospitalization resulted in death. CREAT: creatinine; MELD: Model for End-Stage Liver Disease; ALB: Albumin; BILI: Total Bilirubin; PLT: Platelet; PT: Prothrombin Time



RESULTS

TIMELINE VISUALIZATION

Figure 12 demonstrates an example patient for the clinical course visualization. Six numerical parameters (five laboratory values and one risk score) are plotted within the rings using color coding for severity. The innermost ring-space identifies decompensating events such as complications of cirrhosis. Finally black triangles indicate inpatient admission. The example patient had a history of alcohol abuse and had a late diagnosis of cirrhosis around 2011. The patient had a rapidly decompensating course, particularly with the development of hepatocellular carcinoma in 2012 marked by multiple hospital admissions prior to his death.

Figure 13 demonstrates a side-by-side comparison of two patients: Panel A demonstrates a patient admitted with HRS and Panel B demonstrates a patient admitted with AKI but not diagnosed with HRS. In order to facilitate comparison, both patient timelines are anchored on the last date visualized – the first day of index admission. Two trends are readily notable. The patient without HRS, has a relatively stable clinical course. His indices suggest moderate cirrhosis with recurring problems due to hepatic encephalopathy. He would be classified as Baveno stage I. The HRS patient, however, demonstrates a rapidly declining course. The patient's timeline exhibits multiple paracenteses before admission—indicative of refractory ascites, a common precursor to HRS.

CLUSTERING TO ANALYZE MODEL PERFORMANCE

To look at population heterogeneity and assess model performance amongst subgroups of similar admissions, the total cohort was subdivided into 49 clusters using a 7x7 SOM. There was a median of 622 observations per cluster with an interquartile range of 321 to 973 and a maximum of 2026 observations in one cluster. Excluding three clusters that had ≤ 2 observations, the minimum number of observations per cluster was 192.

Table 14 reports the data as a table with point estimates and 95% confidence intervals around the estimates.

Table 14: Discrimination and calibration statistics for each cluster along with 95% confidence intervals obtained by bootstrap sampling.

Clusters with "N/A" had ≤ 2 observations. When comparing the cluster number with the cluster map in Figure 3, cluster #1 starts in the bottom left corner and proceeds row-wise until cluster #49 in the top right corner.

Cl.	N	Intercept	Slope	Brier	AUC
1	423	-0.374 (-0.707,-0.04)	0.612 (0.387,0.837)	0.17 (0.15,0.191)	0.665 (0.606,0.725)
2	927	1.647 (-3.471,6.765)	1.555 (0.289,2.82)	0.008 (0.003,0.013)	0.803 (0.554,1.051)
3	1535	0.686 (-0.552,1.923)	1.217 (0.883,1.551)	0.016 (0.01,0.022)	0.859 (0.784,0.934)
4	2026	0.785 (-0.002,1.572)	1.214 (0.966,1.461)	0.027 (0.021,0.033)	0.821 (0.767,0.875)
5	973	0.458 (-0.637,1.553)	1.075 (0.731,1.42)	0.033 (0.022,0.044)	0.762 (0.678,0.846)
6	236	-0.166 (-0.644,0.312)	0.82 (0.507,1.132)	0.146 (0.116,0.176)	0.749 (0.672,0.826)
7	266	-1.001 (-1.747,-0.255)	0.381 (-0.017,0.779)	0.139 (0.11,0.168)	0.617 (0.52,0.714)
8	321	-0.615 (-1.144,-0.087)	0.682 (0.426,0.937)	0.124 (0.1,0.149)	0.757 (0.702,0.812)
9	860	0.366 (-0.196,0.928)	1.037 (0.807,1.266)	0.08 (0.066,0.094)	0.792 (0.745,0.839)
10	1332	2.172 (0.58,3.764)	1.665 (1.2,2.13)	0.02 (0.013,0.028)	0.865 (0.794,0.937)
11	2006	0.104 (-0.523,0.73)	1.004 (0.81,1.197)	0.045 (0.036,0.053)	0.775 (0.733,0.817)
12	1	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)
13	327	-0.262 (-1.273,0.749)	0.69 (0.281,1.1)	0.101 (0.071,0.13)	0.649 (0.555,0.743)
14	0	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)
15	66	4.122 (-47.616,55.859)	4.391 (-14.986,23.769)	0.029 (-0.001,0.059)	0.515 (-0.17,1.2)
16	425	-0.427 (-1.51,0.656)	0.865 (0.482,1.247)	0.05 (0.033,0.067)	0.819 (0.722,0.916)
17	1117	0.757 (0.207,1.306)	1.315 (1.092,1.539)	0.053 (0.043,0.063)	0.868 (0.832,0.904)
18	1223	0.053 (-0.327,0.433)	0.975 (0.807,1.144)	0.092 (0.08,0.104)	0.78 (0.743,0.817)
19	378	0.176 (-0.421,0.773)	0.906 (0.656,1.157)	0.109 (0.083,0.134)	0.782 (0.727,0.838)
20	457	-0.548 (-1.528,0.431)	0.963 (0.544,1.382)	0.035 (0.023,0.048)	0.801 (0.712,0.89)
21	811	-1.029 (-1.621,-0.437)	0.541 (0.313,0.769)	0.077 (0.061,0.092)	0.665 (0.603,0.728)
22	508	1.535 (-0.659,3.729)	1.55 (0.883,2.218)	0.02 (0.009,0.032)	0.835 (0.75,0.92)
23	892	-0.312 (-2.116,1.492)	1.022 (0.544,1.5)	0.013 (0.007,0.02)	0.768 (0.665,0.871)
24	788	-0.435 (-0.77,-0.1)	0.699 (0.496,0.901)	0.137 (0.122,0.152)	0.694 (0.649,0.739)
25	831	-0.518 (-1.105,0.07)	0.888 (0.647,1.13)	0.055 (0.041,0.068)	0.783 (0.728,0.837)

Cl.	N	Intercept	Slope	Brier	AUC
26	819	0.148 (-1.106,1.402)	1.126 (0.642,1.61)	0.037 (0.027,0.047)	0.785 (0.687,0.882)
27	121	7.581 (-13.038,28.201)	4.699 (-6.656,16.053)	0.034 (0.011,0.058)	0.969 (0.917,1.02)
28	1092	0.956 (0.24,1.672)	1.209 (0.955,1.463)	0.054 (0.043,0.066)	0.807 (0.761,0.854)
29	1256	1.191 (-1.188,3.571)	1.573 (0.826,2.321)	0.011 (0.006,0.016)	0.785 (0.661,0.908)
30	907	0.137 (-0.709,0.983)	0.947 (0.636,1.259)	0.061 (0.048,0.075)	0.719 (0.652,0.785)
31	261	-0.081 (-0.465,0.302)	1.02 (0.695,1.344)	0.16 (0.135,0.185)	0.761 (0.699,0.824)
32	506	-0.316 (-0.544,-0.089)	0.584 (0.377,0.791)	0.206 (0.185,0.226)	0.667 (0.612,0.722)
33	622	-0.043 (-0.981,0.894)	1.011 (0.669,1.353)	0.043 (0.03,0.056)	0.85 (0.778,0.923)
34	1142	1.51 (-0.911,3.931)	1.631 (0.868,2.394)	0.012 (0.008,0.017)	0.871 (0.76,0.981)
35	1493	-0.226 (-1.705,1.253)	1.128 (0.703,1.554)	0.012 (0.007,0.017)	0.83 (0.742,0.918)
36	815	-0.211 (-1.09,0.668)	0.87 (0.557,1.183)	0.057 (0.043,0.07)	0.714 (0.644,0.785)
37	449	-0.219 (-1.143,0.705)	1.029 (0.644,1.414)	0.039 (0.024,0.055)	0.829 (0.745,0.913)
38	813	-0.019 (-0.633,0.595)	0.889 (0.631,1.146)	0.083 (0.068,0.098)	0.749 (0.685,0.813)
39	253	-0.914 (-1.583,-0.246)	0.411 (0.162,0.66)	0.124 (0.091,0.156)	0.661 (0.575,0.748)
40	192	0.566 (-1.037,2.169)	1.535 (0.664,2.407)	0.058 (0.036,0.08)	0.936 (0.886,0.986)
41	562	0.275 (-3.947,4.496)	1.457 (0.262,2.653)	0.008 (0.001,0.014)	0.85 (0.636,1.064)
42	604	-0.296 (-0.707,0.116)	0.821 (0.607,1.036)	0.115 (0.098,0.132)	0.782 (0.736,0.827)
43	1456	0.756 (0.034,1.479)	1.23 (0.994,1.466)	0.036 (0.028,0.044)	0.819 (0.774,0.864)
44	308	1.862 (-0.798,4.523)	1.7 (0.767,2.633)	0.023 (0.01,0.036)	0.874 (0.786,0.962)
45	488	-0.619 (-1.257,0.018)	0.696 (0.378,1.014)	0.109 (0.086,0.133)	0.676 (0.602,0.749)
46	2	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)
47	1537	0.181 (-0.325,0.687)	1.063 (0.841,1.285)	0.072 (0.062,0.082)	0.747 (0.706,0.788)
48	131	-1.849 (-9.556,5.859)	1.985 (-5.986,9.957)	0.031 (0.012,0.05)	0.91 (0.805,1.015)
49	854	-1.399 (-4.175,1.376)	1.168 (0.312,2.024)	0.006 (0.002,0.01)	0.908 (0.802,1.013)

There were twenty clusters with an AUC ≥ 0.8, seventeen with an AUC between 0.7 and 0.8, and nine with an AUC < 0.7. Clusters with AUC ≥ 0.8 demonstrated phenotypes of CKD, cardiac comorbidities with coronary artery disease, congestive heart failure (CHF), and atrial fibrillation, and higher serum sodium on admission. In addition, the group of well performing clusters in the bottom left quadrant showed a phenotype of alcohol abuse and higher serum sodium. Poorly performing clusters displayed a phenotype of patients with a history of decompensated cirrhosis (hepatic encephalopathy, spontaneous bacterial peritonitis, varices, and ascites) along with KDIGO Stage I kidney injury. When comparing the top versus bottom 5 performing clusters, in addition to the aforementioned phenotypic differences, there was a significant difference in degree of kidney injury (average admission creatinine of 6.2 ± 2.9 (SD) versus 2.1 ± 1.1 ($p<0.0001$), respectively). There was KDIGO Stage 3 kidney injury 1860 out of 1860 hospitalizations in the top performing clusters versus 1 out of 1958 in the bottom performing clusters (1691 out of 1958 were KDIGO Stage 1). Finally, clusters where patients received a paracentesis on the day of admission also performed poorly. The clustering results were displayed using CLUSTERVIEW's interface (Refer to Figure 16).

CLUSTERVIEW VISUALIZATION

Layout

The tool consists of three main areas (

Figure 15): Panel A, control panel; Panel B, cluster visualization view; and Panel C, detail view. The control panel in the left area has multi-checkbox interface to choose options for the visualization: (1) selecting variables to display; (2) plotting observed outcomes; (3) plotting similar patients to an index patient; (4) plotting similar patients to the index patient in terms of clustering; and (5) exporting data. The cluster visualization view in the middle area visualizes multiple patient clusters using a glyph visualization. The detail view on the right allows a drill down of all the variables that make up the cluster. The user can retrieve summary statistics on a single cluster, or compare two clusters by selecting them in the middle panel.

Each cluster represents a group of similar patients (Figure 14). The relative importance of features within each cluster is visualized using pie-piece glyphs. The larger the pie piece, the greater the feature's importance to the respective cluster. The features of interest can be chosen from the left control panel.

Figure 14: Visual representation of each patient cluster.

Each cluster has an (optional) halo which color codes a user-selected outcome, e.g. probability of HRS (Arrow A). The user may (optionally) display up to fourteen variables within the visualization, symbolized by pie-pieces (Arrow B). The size of the pie-piece is proportional to the importance of that variable within the cluster.

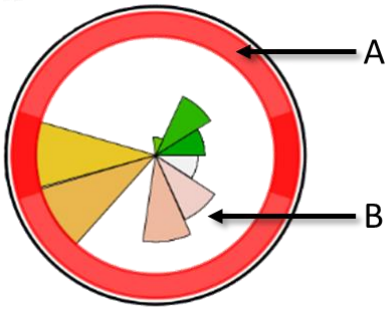


Figure 15: Layout of statistics visualization tool.

The layout includes three components: Panel A displays the control options including selection of variables and measures to display; Panel B contains the visualization; Panel C has drill-down information that dynamically responds to the visualized measure and chosen cluster.

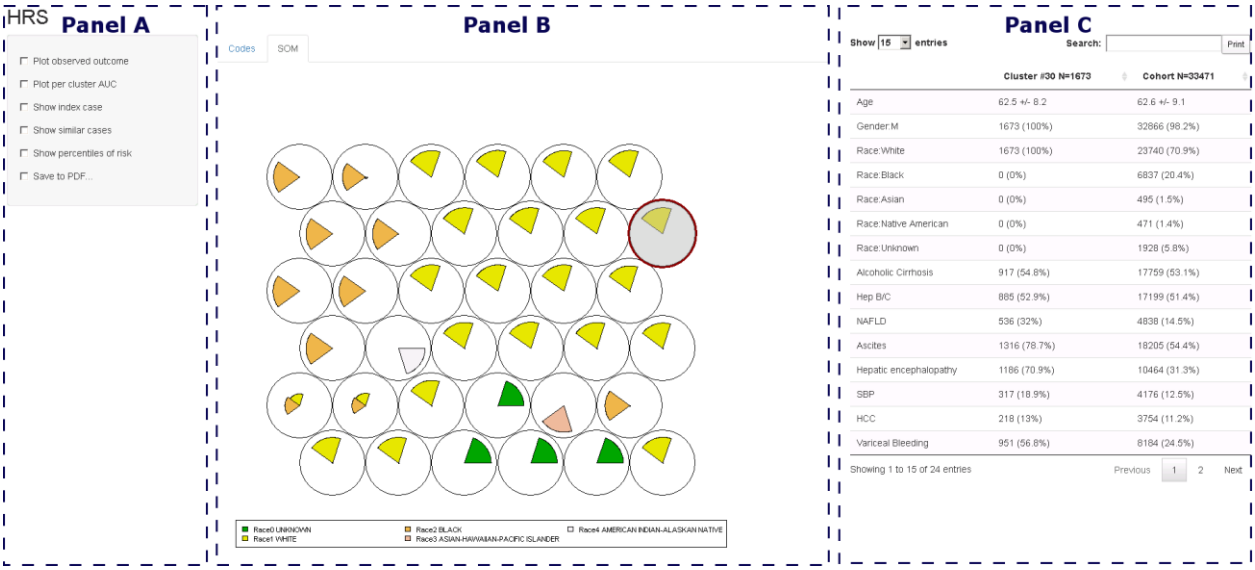


Figure 16: 7x7 Self Organizing Map with the 49 clusters of observations. Each cluster has an outer ring color coded to the risk model's performance (AUC) within that cluster.

The slope and intercept of the risk model are color coded in the bottom semicircle and top semicircle, respectively, for each cluster. Each cluster's affinity for 14 variables is represented by a color coded pie piece, with a larger pie piece showing greater affinity. (CR = Creatinine; NA = Sodium; AFIB = Atrial Fibrillation; CAD = Coronary Artery Disease; CHF = Congestive Heart Failure; CKD = Chronic Kidney Disease; HE = Hepatic Encephalopathy; SBP = Spontaneous Bacterial Peritonitis)

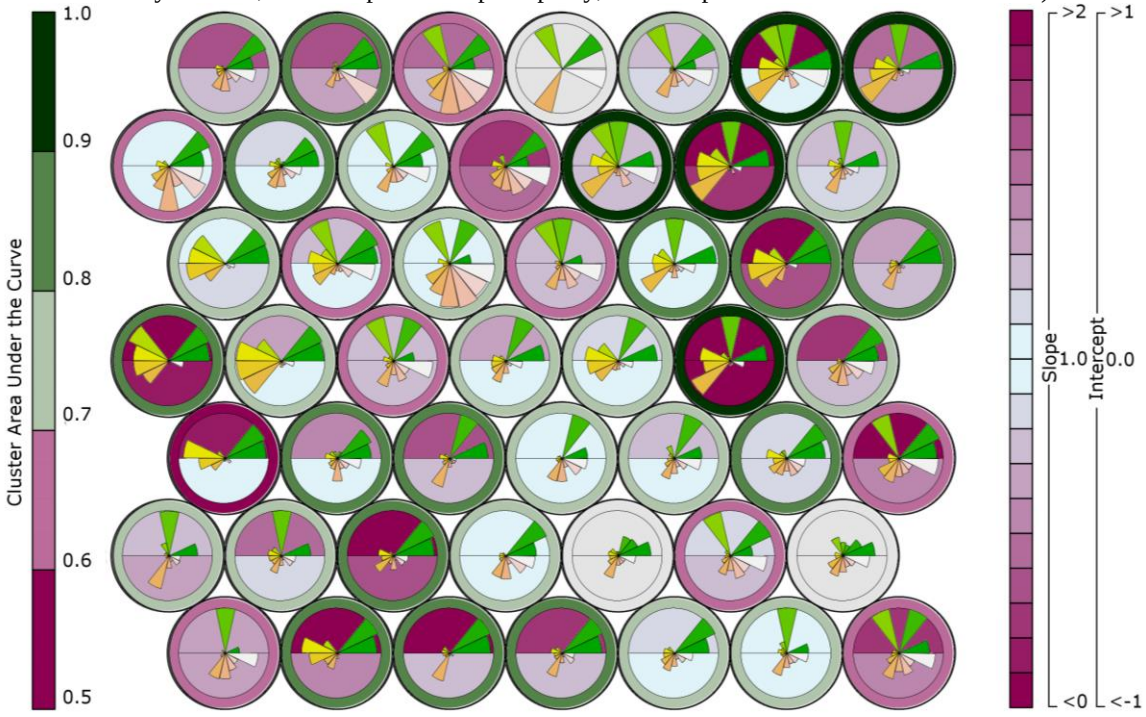
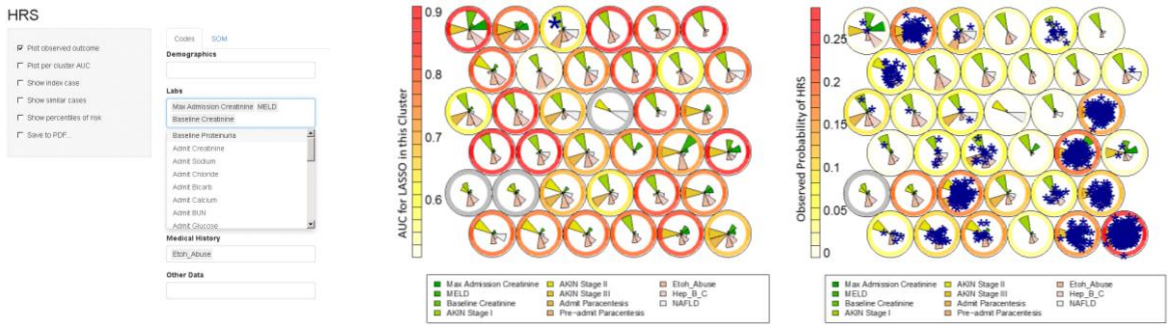


Figure 16 shows how the user can use CLUSTERVIEW to analyze overall model performance. The clustering along with the model's AUC and model calibration using the slope and intercept of the regression line through O/E probability plot; additionally the visualization also shows each cluster's affinity for fourteen key clinical variables.

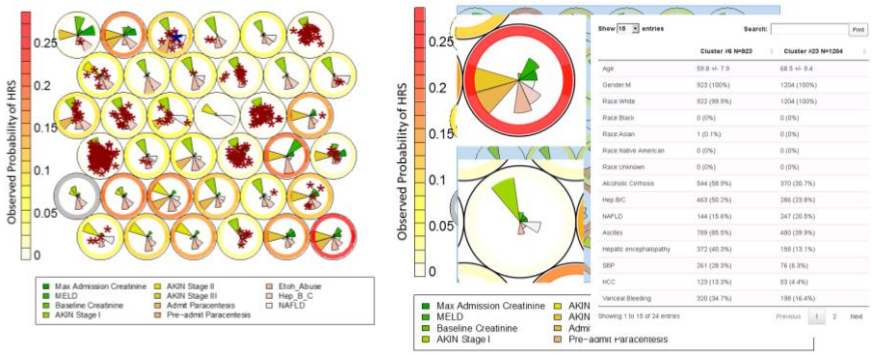
Figure 17: Statistics visualization user interface workflow.

(A) The user utilizes the control panel to select desired outcome and predictors for plotting; (B) Plotting area displays clusters with the cluster measure displayed as a colored halo; (C) Patients similar in terms of risk percentile, chosen from the control panel, are plotted; (D) Patients with similar predictors are plotted; (E) Two clusters can be selected and compared using a drill-down table.

Panel A **Panel B** **Panel C**



Panel D **Panel E**



User Interaction

Refer to Figure 17 for the workflow. Using the left control panel users can select the predictors of interest to display within the cluster visualization from five domains: demographics, laboratory values, home medications, inpatient medications, and comorbidities. The user can also plot either the percentage of patients who have HRS within the cluster or the prediction model's AUC for the respective cluster. This information is visualized as a color-coded halo. Using the selected features from the first step, users can visualize which features or combinations of features predispose patients to HRS. Figure 17 panels C and D demonstrate plotting similar patients via the two protocols discussed in the methods: similar patients based on predicted probability and patients similar to a new, hypothetical patient. Finally, the user may select a cluster (or select two clusters to compare) to obtain more information about the cluster. Cluster information is provided in tabular format in the right hand side of the interface. For example, one may see that patients in the first cluster, with a much higher prevalence of HRS, have a significantly higher history of cirrhosis complications.

DISCUSSION

We constructed visualizations to better understand high dimensional data and represent complex information to help clinicians make healthcare decisions. Human cognition uses pre-attentive systems for significant parts of graphical data visual processing. To leverage the pre-attentive qualities of visual cognition, we used color and proximity in TIMELINE. For CLUSTERVIEW, we employed color, size, proximity, and similarity. We leveraged design paradigms for high-dimensional, temporal data including brushing within the timeline visualization, color mapping, and coordinated views within the clustering visualization.

Radial graphs aren't necessarily best for temporal tasks and can increase processing time.[286] Radial graphs are most useful for finding extreme values, particularly in relation to other dimensions.[287,288] Radar graphs do have some shortcomings other than possibly increased processing time (depending on the task). They can be error prone because the individual needs to follow one of several concentric rings if the goal is to look up an individual value. However, for TIMELINE the goal is less to identify a particular value (e.g., what was the MELD score in January 1st, 2014) and more important to see the evolving pattern of values across multiple dimensions.

Our clustering and visualization technique can be used at two stages of clinical care. During model formulation, it can be invoked to identify clusters with a different case mix and poor model performance. These outlier clusters may indicate the need to refit the model for these sub-groups. Possible etiologies include not accounting for important risk variables or necessary interaction terms. Clustering to perform sub-group analysis also helps us perform assessment of *strong calibration*.[289] Unlike measures such as Hosmer-Lemeshow or observed-to-expected probability plots which assess moderate calibration, strong calibration requires predicted risks to correspond to observed event rates for every covariate pattern. It is both computationally and cognitively infeasible to specify all possible covariate patterns. Clustering, however, provides an intuitive, parsimonious grouping of covariate patterns.

Secondly, we can use CLUSTERVIEW at the point-of-care: when a new patient enters the emergency room they can be assigned to a cluster. Using this alternate clinical workflow, we obtain the model's risk estimate and its performance for patients like the index case. Unlike the standard-of-care, if the algorithm indicates sufficient risk of HRS and the patient falls within a cluster where model performance is acceptable (i.e., AUC $\geq$ 0.8) we can initiate vasopressor therapy immediately. Though visualizing risk prediction and uncertainty has received some attention for patient-facing tools,[168,204] physician-facing CDS has received sparse investigation.[205]

In summary, we developed two cohesive information visualizations that combines model performance with identifying a patient's phenotype. Next steps would be to have a formative and summative evaluation of these visualizations. Techniques could include semi-structured interviews and the use of eye-tracking to monitor areas of visualizations that users are spending more time on.[290] This could help find pre-attentive errors that people are making because the visualization is poor.

CHAPTER V: CONCLUSIONS

The widespread adoption of the electronic health record has created a new source of "Big Data." As the cost of healthcare in the United States has significantly outpaced other countries,[291] we have to turn to new technologies and exploit novel methods of analysis and decision making. Cirrhosis is one of the most expensive diseases in America, and is responsible for over $10 billion dollars of spending annually.[292] The majority of these costs accrue from frequent hospitalizations from cirrhosis-related complications.[293] Hepatorenal Syndrome serves as one of the deadliest cirrhosis complications, with a median survival of weeks to months. Additionally, HRS doubles hospital length of stay and associated costs.[294]

In order to improve the care of these high cost diseases, there is increasing emphasis on observational cohort trials and pragmatic clinical trials. These study designs require highly accurate methods to identify patient cases and controls. Traditional means of identifying cases, including chart review, are infeasible to provide sufficient cases for accurate model building. We have presented a highly accurate, and tunable algorithm to identify cases and controls using EHR phenotyping with an AUC of 0.93. Additionally, our phenotyping algorithm demonstrated the benefit of utilizing fixed vector representations of UMLS CUIs instead of free text.

Predictive analytics plays an increasingly important part in delivering state of the art care and clinical decision support. Particularly in the era of big data we have the opportunity to improve the care of our sickest patients. The Veterans Health Administration has over three decades of experience collecting routine clinical information about the patients it serves. With the development of the VINCI infrastructure, the VA serves as one of the best resources for longitudinal observational cohort research. The VA's decades of investment in its health IT infrastructure is starting to pay dividends.[295]

This study showed that improved risk prediction modeling surrounding HRS patients can identify HRS patients at time of hospitalization. HRS is often difficult to diagnose as it mimics many other causes of kidney injury on presentation; moreover, cirrhotics are prone to kidney damage from many causes. The model helps identify those patients who are more likely to go on to develop acute kidney injury. We compiled a cohort of 2,435 inpatient hospitalizations with HRS, which is to our best knowledge the largest observational cohort for HRS ever assembled. Using this cohort we developed a model with an AUC of 0.84 for identifying patients at high risk for developing HRS with excellent calibration based on data in the peri-admission window.

Clustering allowed us to analyze model performance within patient sub-cohorts. For example, patients with existing cardiac comorbidities or CKD did well. Given the recent evidence showing that insufficient cardiac output induces renal hypoperfusion in HRS,[30,234] patients with heart failure comprise a demographic in need of more accurate prediction. Patients with CKD represent another subpopulation where diagnosing HRS can be difficult because of the increased variability in kidney function in patients with CKD.[296] Subgroups where model performance is subpar show a higher proportion of cirrhosis complications and a lower level of kidney injury on presentation. However, in patients with Baveno stage III or IV decompensated cirrhosis[297] and KDIGO stage III renal failure (3223 admissions), the model had an AUC of 0.81, suggesting that it was a combination of cirrhosis complications and low level kidney injury that comprised a challenging phenotype.

Despite aggressive promises of increasing health technology use and big data to affect healthcare, real world implementations of real-time predictive analytics still appear to be few and far between. One of the challenges faced in utilizing some of these new technologies lies in creating interpretable models.[157] Information Visualization techniques have been employed to make previously ineluctable models such as neural networks more accessible.[298,299] Information Visualization works best when it augments the human brain's ability to find patterns in data. Our study created an interactive information visualization system to improve the care for patients with HRS. We first built a patient timeline viewer to help clinicians quickly understand the relevant history of a patient. During direct workflow observation we identified key attributes of a cirrhotic patient's disease course that could help specialists (and non-specialists alike) understand how sick a patient was.

Leveraged with this data we also constructed CLUSTERVIEW, our risk prediction model exploring and clustering visualization. Our clustering visualization allows rapid visualization of model performance within sub-groups of the cohort to assess for how case-mix may model performance. Additionally, we theorize an additional workflow where such a visualization will help providers provide tailored care for patients by identifying "Patients-like-me" cohorts.

LIMITATIONS

This work has some limitations that are worth noting. Most importantly, this work is based on retrospective observational cohort data from the Department of Veterans Affairs. The VA data may not be representative of other clinical environments due to the slightly older average age and predominance of men. We only performed internal validation; however, we aimed to increase generalizability by sampling across a broad range of kidney injury and liver disease. Moreover, all variables are common to other electronic health records, and the selected variables make pathophysiologic sense. Third, several significant predictors were ICD-9 codes, but with the transition to ICD-10 in the US, the algorithm's performance cannot be assured. At the same time, it is worth noting that there are one-to-one mappings for two of the important ICD-9 codes (ATN and HRS) based on the General Equivalent Maps (GEMs) framework.[252] The code sets defining other conditions would require validation. Though we included etiology of cirrhosis as a covariate in all of our models, it is possible that important interaction effects may go unnoticed.

Our work showed that logistic regression was comparable, and at times superior, to certain machine learning methods. We did leverage penalized regression models to handle overfitting, however. Though this held true in this dataset, it is possible that in a dataset with significantly more observations, a machine learning model would offer superior performance. We did not perform formative or summative evaluation of our visualization for this study; however, we did discuss the evaluations with select domain experts for informal feedback.

FUTURE DIRECTIONS

The studies described in this thesis demonstrate that machine learning assisted models can improve and accelerate care of patients with Hepatorenal Syndrome. The validity of the phenotyping model will need to be assessed in an outside cohort. The phenotyping model may be improved by including temporal feature generation. There is exciting work using recurrent neural networks to capture temporal features within EHR models.[300] Much of healthcare data is inherently temporal as the basic unit of measurement, the patient, is often reassessed multiple

times during their clinical course, e.g. repeat lab measurements for each day of their hospitalization.

The risk prediction model only utilized structured data, and as discussed in the phenotyping model, the natural extension would be to utilize clinical text data within the algorithm in two phases. The first phase would be to implement the phenotyping model developed in Chapter II to identify HRS patients, thereby improving upon the sensitivity and specificity of the ICD code utilized as the gold standard. Secondly, elements in the free text could serve as important predictors of patients who would go on to develop, or have HRS.

The question remains of the clinical utility, which can only be assessed in prospective clinical trial. The next step would be to implement a real-time system that would monitor cirrhotic patients admitted to the hospital with some level of renal failure. The system could not only highlight cases that are apparent based on the phenotyping model, but also cases that are likely to go onto develop HRS. The system would need to be integrated into routine clinical care. Future work would revolve around integrating these models into routine clinical care using technologies such as Fast Healthcare Interoperability Resources and CDS-Hooks[301,302]

REFERENCES

1. Medicine, I. of & America, C. on the L. H. C. S. in. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. (National Academies Press, 2013).
2. Park, S.-J. *et al.* Trial of Everolimus-Eluting Stents or Bypass Surgery for Coronary Disease. *N. Engl. J. Med.* **372**, 1204–1212 (2015).
3. Greenhalgh, T., Howick, J. & Maskrey, N. Evidence based medicine: a movement in crisis? *The BMJ* **348**, (2014).
4. Tricoci, P., Allen, J. M., Kramer, J. M., Califf, R. M. & Smith, S. C. Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines. *JAMA* **301**, 831–841 (2009).
5. Del Fiol, G., Workman, T. E. & Gorman, P. N. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern. Med.* **174**, 710–718 (2014).
6. Fleurence, R. L. *et al.* Launching PCORnet, a national patient-centered clinical research network. *J. Am. Med. Inform. Assoc.* **21**, 578–582 (2014).
7. Kim, W. R., Brown, R. S., Terrault, N. A. & El-Serag, H. Burden of liver disease in the United States: Summary of a workshop. *Hepatology* **36**, 227–242 (2002).
8. Younossi, Z. M. *et al.* Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **9**, 524-530.e1; quiz e60 (2011).
9. ASRANI, S. K., LARSON, J. J., YAWN, B., THERNEAU, T. M. & KIM, W. R. Underestimation of Liver-Related Mortality in the United States. *Gastroenterology* **145**, 375-82.e1–2 (2013).
10. Schuppan, D. & Afdhal, N. H. Liver Cirrhosis. *Lancet* **371**, 838–851 (2008).
11. Everhart, J. E. & Ruhl, C. E. Burden of Digestive Diseases in the United States Part III: Liver, Biliary Tract, and Pancreas. *Gastroenterology* **136**, 1134–1144 (2009).
12. Heidelbaugh, J. J. & Bruderly, M. Cirrhosis and chronic liver failure: part I. Diagnosis and evaluation. *Am. Fam. Physician* **74**, 756–762 (2006).
13. *Sherlock's Diseases of the Liver and Biliary System*. (Wiley-Blackwell, 2011).
14. Brown, J. J., Naylor, M. J. & Yagan, N. Imaging of hepatic cirrhosis. *Radiology* **202**, 1–16 (1997).
15. Fujimoto, K., Sawabe, M., Sasaki, M., Kino, K. & Arai, T. Undiagnosed cirrhosis occurs frequently in the elderly and requires periodic follow ups and medical treatments. *Geriatr. Gerontol. Int.* **8**, 198–203 (2008).
16. Graudal, N., Leth, P., Mårbjerg, L. & Galløe, A. M. Characteristics of cirrhosis undiagnosed during life: a comparative analysis of 73 undiagnosed cases and 149 diagnosed cases of cirrhosis, detected in 4929 consecutive autopsies. *J. Intern. Med.* **230**, 165–171 (1991).
17. Kobayashi, K. *et al.* Studies of clinically latent cirrhosis. *Hepatogastroenterology.* **37 Suppl 2**, 77–80 (1990).
18. Kanwal, F. *et al.* The Quality of Care Provided to Patients With Cirrhosis and Ascites in the Department of Veterans Affairs. *Gastroenterology* **143**, 70–77 (2012).
19. Singal, A. G. *et al.* Failure Rates in the Hepatocellular Carcinoma Surveillance Process. *Cancer Prev. Res. Phila. Pa* **5**, 1124–1130 (2012).

20. Seal, K. H. *et al.* Substance use disorders in Iraq and Afghanistan veterans in VA healthcare, 2001–2010: Implications for screening, diagnosis and treatment. *Drug Alcohol Depend.* **116**, 93–101 (2011).
21. Bellentani, S. *et al.* Drinking habits as cofactors of risk for alcohol induced liver damage. *Gut* **41**, 845–850 (1997).
22. Beste, L. A. *et al.* Trends in Burden of Cirrhosis and Hepatocellular Carcinoma by Underlying Liver Disease in US Veterans, 2001–2013. *Gastroenterology* **149**, 1471-1482.e5 (2015).
23. Tsochatzis, E. A., Bosch, J. & Burroughs, A. K. Liver cirrhosis. *The Lancet* **383**, 1749–1761 (2014).
24. Franchis, R. de. Revising consensus in portal hypertension: Report of the Baveno V consensus workshop on methodology of diagnosis and therapy in portal hypertension. *J. Hepatol.* **53**, 762–768 (2010).
25. Lammert, C., Bjornsson, E., Niklasson, A. & Chalasani, N. Oral medications with significant hepatic metabolism at higher risk for hepatic adverse events. *Hepatology* **51**, 615–620 (2010).
26. Fleming Kate M., Aithal Guruprasad P., Card Tim R. & West Joe. All-cause mortality in people with cirrhosis compared with the general population: a population-based cohort study. *Liver Int.* **32**, 79–84 (2011).
27. Kim, W. R. *et al.* OPTN/SRTR 2016 Annual Data Report: Liver. *Am. J. Transplant.* **18**, 172–253
28. Ginès, P. & Schrier, R. W. Renal Failure in Cirrhosis. *N. Engl. J. Med.* **361**, 1279–1290 (2009).
29. Fede, G. *et al.* Renal failure and cirrhosis: A systematic review of mortality and prognosis. *J. Hepatol.* **56**, 810–818 (2012).
30. Salerno, F., Gerbes, A., Gines, P., Wong, F. & Arroyo, V. Diagnosis, prevention and treatment of hepatorenal syndrome in cirrhosis. *Gut* **56**, 1310–1318 (2007).
31. EASL clinical practice guidelines on the management of ascites, spontaneous bacterial peritonitis, and hepatorenal syndrome in cirrhosis. *J. Hepatol.* **53**, 397–417 (2010).
32. Wadei, H. M. Hepatorenal syndrome: a critical update. *Semin. Respir. Crit. Care Med.* **33**, 55–69 (2012).
33. Wong, F. Recent advances in our understanding of hepatorenal syndrome. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 382–391 (2012).
34. Management of Adult Patients with Ascites Due to Cirrhosis: Update 2012.
35. Ginès, A. *et al.* Incidence, predictive factors, and prognosis of the hepatorenal syndrome in cirrhosis with ascites. *Gastroenterology* **105**, 229–236 (1993).
36. Restuccia, T. *et al.* Effects of treatment of hepatorenal syndrome before transplantation on posttransplantation outcome. A case-control study. *J. Hepatol.* **40**, 140–146 (2004).
37. Alessandria, C. *et al.* MELD score and clinical type predict prognosis in hepatorenal syndrome: Relevance to liver transplantation. *Hepatology* **41**, 1282–1289 (2005).
38. Skagen, C., Einstein, M., Lucey, M. R. & Said, A. Combination Treatment With Octreotide, Midodrine, and Albumin Improves Survival in Patients With Type 1 and Type 2 Hepatorenal Syndrome: *J. Clin. Gastroenterol.* **43**, 680–685 (2009).
39. Maddukuri, G., Cai, C. X., Munigala, S., Mohammadi, F. & Zhang, Z. Targeting an Early and Substantial Increase in Mean Arterial Pressure Is Critical in the Management of Type 1

Hepatorenal Syndrome: A Combined Retrospective and Pilot Study. *Dig. Dis. Sci.* **59**, 471–481 (2014).
40. Nazar, A. *et al.* Predictors of response to therapy with terlipressin and albumin in patients with cirrhosis and type 1 hepatorenal syndrome. *Hepatology* **51**, 219–226 (2010).
41. Martín–Llahí, M. *et al.* Prognostic Importance of the Cause of Renal Failure in Patients With Cirrhosis. *Gastroenterology* **140**, 488-496.e4 (2011).
42. Alqahtani, S. A. & Larson, A. M. Adult liver transplantation in the USA: *Curr. Opin. Gastroenterol.* **27**, 240–247 (2011).
43. Angeli, P. & Gines, P. Hepatorenal syndrome, MELD score and liver transplantation: An evolving issue with relevant implications for clinical practice. *J. Hepatol.* **57**, 1135–1140 (2012).
44. Nadim, M. K. *et al.* Hepatorenal syndrome: the 8th international consensus conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit. Care* **16**, R23 (2012).
45. Mitzner, S. R. *et al.* Extracorporeal Detoxification Using the Molecular Adsorbent Recirculating System for Critically Ill Patients with Liver Failure. *J. Am. Soc. Nephrol.* **12**, S75–S82 (2001).
46. Fabrizi, F., Aghemo, A. & Messa, P. Hepatorenal syndrome and novel advances in its management. *Kidney Blood Press. Res.* **37**, 588–601 (2013).
47. Wong, F., Pantea, L. & Sniderman, K. Midodrine, octreotide, albumin, and TIPS in selected patients with cirrhosis and type 1 hepatorenal syndrome. *Hepatology* **40**, 55–64 (2004).
48. Senzolo, M., Cholongitas, E., Tibballs, J., Burroughs, A. & Patch, D. Transjugular intrahepatic portosystemic shunt in the management of ascites and hepatorenal syndrome. *J. Gastroenterol.* **18**, 1143–1150 (2006).
49. American Academy of Hospice and Palliative Medicine, Center to Advance Palliative Care, Hospice and Palliative Nurses Association, Last Acts Partnership & National Hospice and Palliative Care Organization. National Consensus Project for Quality Palliative Care: Clinical Practice Guidelines for quality palliative care, executive summary. *J. Palliat. Med.* **7**, 611–627 (2004).
50. Salpeter, S. R., Luo, E. J., Malter, D. S. & Stuart, B. Systematic Review of Noncancer Presentations with a Median Survival of 6 Months or Less. *Am. J. Med.* **125**, 512.e1-512.e16 (2012).
51. Sharma, P., Kumar, A., Shrama, B. C. & Sarin, S. K. An Open Label, Pilot, Randomized Controlled Trial of Noradrenaline Versus Terlipressin in the Treatment of Type 1 Hepatorenal Syndrome and Predictors of Response. *Am. J. Gastroenterol.* **103**, 1689–1697 (2008).
52. Ghosh, S. *et al.* Noradrenaline vs terlipressin in the treatment of type 2 hepatorenal syndrome: a randomized pilot study. *Liver Int.* **33**, 1187–1193 (2013).
53. Singh, V. *et al.* Noradrenaline vs. terlipressin in the treatment of hepatorenal syndrome: A randomized study. *J. Hepatol.* **56**, 1293–1298 (2012).
54. Boyer, T. D. *et al.* Predictors of response to terlipressin plus albumin in hepatorenal syndrome (HRS) type 1: Relationship of serum creatinine to hemodynamics. *J. Hepatol.* **55**, 315–321 (2011).
55. Angeli, P. *et al.* Diagnosis and management of acute kidney injury in patients with cirrhosis: Revised consensus recommendations of the International Club of Ascites. *J. Hepatol.* **62**, 968–974 (2015).

56. Chertow, G. M., Burdick, E., Honour, M., Bonventre, J. V. & Bates, D. W. Acute Kidney Injury, Mortality, Length of Stay, and Costs in Hospitalized Patients. *J. Am. Soc. Nephrol.* **16**, 3365–3370 (2005).
57. Shivade, C. *et al.* A review of approaches to identifying patient phenotype cohorts using electronic health records. *J. Am. Med. Inform. Assoc.* **21**, 221–230 (2014).
58. Xu, J. *et al.* Review and evaluation of electronic health records-driven phenotype algorithm authoring tools for clinical and translational research. *J. Am. Med. Inform. Assoc.* **22**, 1251–1260 (2015).
59. Gottesman, O. *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet. Med.* **15**, 761–771 (2013).
60. Kirby, J. C. *et al.* PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *J. Am. Med. Inform. Assoc.* **23**, 1046–1052 (2016).
61. Friedlin, J. *et al.* Comparing Methods for Identifying Pancreatic Cancer Patients Using Electronic Data Sources. *AMIA. Annu. Symp. Proc.* **2010**, 237–241 (2010).
62. McCowan, I. A. *et al.* Collection of Cancer Stage Data by Classifying Free-text Medical Reports. *J. Am. Med. Inform. Assoc.* **14**, 736–745 (2007).
63. Xu, H. *et al.* Extracting and Integrating Data from Entire Electronic Health Records for Detecting Colorectal Cancer Cases. *AMIA. Annu. Symp. Proc.* **2011**, 1564–1572 (2011).
64. Pathak, J., Kiefer, R. C., Bielinski, S. J. & Chute, C. G. Mining the Human Phenome using Semantic Web Technologies: A Case Study for Type 2 Diabetes. *AMIA. Annu. Symp. Proc.* **2012**, 699–708 (2012).
65. Garvin, J. H. *et al.* Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure. *J. Am. Med. Inform. Assoc.* **19**, 859–866 (2012).
66. Carroll, R. J., Eyler, A. E. & Denny, J. C. Naïve Electronic Health Record Phenotype Identification for Rheumatoid Arthritis. *AMIA. Annu. Symp. Proc.* **2011**, 189–196 (2011).
67. Peissig, P. L. *et al.* Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. *J. Am. Med. Inform. Assoc.* **19**, 225–234 (2012).
68. Forster, A. J., Jennings, A., Chow, C., Leeder, C. & Walraven, C. van. A systematic review to evaluate the accuracy of electronic adverse drug event detection. *J. Am. Med. Inform. Assoc.* **19**, 31–38 (2012).
69. Bejan, C. A., Vanderwende, L., Wurfel, M. M. & Yetisgen-Yildiz, M. Assessing Pneumonia Identification from Time-Ordered Narrative Reports. *AMIA. Annu. Symp. Proc.* **2012**, 1119–1128 (2012).
70. Wu, W. *et al.* Unsupervised phenotyping of Severe Asthma Research Program participants using expanded lung data. *J. Allergy Clin. Immunol.* **133**, 1280–1288 (2014).
71. Thompson, W. K. *et al.* An Evaluation of the NQF Quality Data Model for Representing Electronic Health Record Driven Phenotyping Algorithms. *AMIA. Annu. Symp. Proc.* **2012**, 911–920 (2012).
72. Findlay, M. & Isles, C. Causes of Acute Kidney Injury. in *Clinical Companion in Nephrology* 45–52 (Springer International Publishing, 2015). doi:10.1007/978-3-319-14868-7_10
73. Quan, H., Parsons, G. A. & Ghali, W. A. Validity of information on comorbidity derived rom ICD-9-CCM administrative data. *Med. Care* **40**, 675–685 (2002).
74. Kizer, K. W. The National Quality Forum Seeks to Improve Health Care. *Acad. Med.* **75**, 320 (2000).

75. Controlling High Blood Pressure | eCQI Resource Center. Available at: https://ecqi.healthit.gov/ecqm/measures/cms165v4. (Accessed: 9th October 2018)
76. Zheng, T. *et al.* A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int. J. Med. Inf.* **97**, 120–127 (2017).
77. Sesen, M. B., Kadir, T., Alcantara, R.-B., Fox, J. & Brady, M. Survival Prediction and Treatment Recommendation with Bayesian Techniques in Lung Cancer. *AMIA. Annu. Symp. Proc.* **2012**, 838–847 (2012).
78. Fine, A. M. *et al.* Use of population health data to refine diagnostic decision-making for pertussis. *J. Am. Med. Inform. Assoc. JAMIA* **17**, 85–90 (2010).
79. Rea, S. *et al.* Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project. *J. Biomed. Inform.* **45**, 763–771 (2012).
80. Anderson, N. *et al.* Implementation of a deidentified federated data network for population-based cohort discovery. *J. Am. Med. Inform. Assoc.* **19**, e60–e67 (2012).
81. Voorham, J. & Denig, P. Computerized Extraction of Information on the Quality of Diabetes Care from Free Text in Electronic Patient Records of General Practitioners. *J. Am. Med. Inform. Assoc.* **14**, 349–354 (2007).
82. DeLisle, S. *et al.* Combining free text and structured electronic medical record entries to detect acute respiratory infections. *PloS One* **5**, e13377 (2010).
83. Wei, W.-Q. *et al.* Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J. Am. Med. Inform. Assoc.* **23**, e20–e27 (2016).
84. Unified Medical Language System (UMLS). Available at: https://www.nlm.nih.gov/research/umls/. (Accessed: 10th September 2016)
85. Savova, G. K. *et al.* Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J. Am. Med. Inform. Assoc. JAMIA* **17**, 507–513 (2010).
86. Friedman, C., Shagina, L., Lussier, Y. & Hripcsak, G. Automated Encoding of Clinical Documents Based on Natural Language Processing. *J. Am. Med. Inform. Assoc.* **11**, 392–402 (2004).
87. Aronson, A. R. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc. AMIA Symp.* 17–21 (2001).
88. Hripcsak, G. & Albers, D. J. Next-generation phenotyping of electronic health records. *J. Am. Med. Inform. Assoc.* **20**, 117–121 (2013).
89. Richesson, R. L., Sun, J., Pathak, J., Kho, A. N. & Denny, J. C. Clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods. *Artif. Intell. Med.* **71**, 57–61 (2016).
90. Pathak, J., Kho, A. N. & Denny, J. C. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. *J. Am. Med. Inform. Assoc.* e206–e211 doi:10.1136/amiajnl-2013-002428
91. Halpern, Y., Horng, S., Choi, Y. & Sontag, D. Electronic medical record phenotyping using the anchor and learn framework. *J. Am. Med. Inform. Assoc.* **23**, 731–740 (2016).
92. Yu, S. *et al.* Surrogate-assisted feature extraction for high-throughput phenotyping. *J. Am. Med. Inform. Assoc.* **24**, e143–e149 (2017).
93. Yu, S. *et al.* Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. *J. Am. Med. Inform. Assoc.* **22**, 993–1000 (2015).

94. Default Clinical Pipeline - Apache cTAKES - Apache Software Foundation. Available at: https://cwiki.apache.org/confluence/display/CTAKES/Default+Clinical+Pipeline. (Accessed: 10th October 2018)
95. Knaus, W. A. *et al.* THe apache iii prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* **100**, 1619–1636 (1991).
96. Barlow, W. E. *et al.* Prospective Breast Cancer Risk Prediction Model for Women Undergoing Screening Mammography. *J. Natl. Cancer Inst.* **98**, 1204–1214 (2006).
97. Park, Y. *et al.* Validation of a Colorectal Cancer Risk Prediction Model Among White Patients Age 50 Years and Older. *J. Clin. Oncol.* **27**, 694–698 (2009).
98. Cassidy, A. *et al.* The LLP risk model: an individual risk prediction model for lung cancer. *Br. J. Cancer* **98**, 270–276 (2007).
99. Matheny, M. E., Ohno-Machado, L. & Resnic, F. S. Risk-adjusted sequential probability ratio test control chart methods for monitoring operator and institutional mortality rates in interventional cardiology. *Am. Heart J.* **155**, 114–120 (2008).
100. Eagle KA, Lim MJ, Dabbous OH & et al. A validated prediction model for all forms of acute coronary syndrome: Estimating the risk of 6-month postdischarge death in an international registry. *JAMA* **291**, 2727–2733 (2004).
101. Rodríguez, E. *et al.* Risk Factors for Acute Kidney Injury in Severe Rhabdomyolysis. *PLOS ONE* **8**, e82992 (2013).
102. Kim, W. H. *et al.* Simplified Clinical Risk Score to Predict Acute Kidney Injury After Aortic Surgery. *J. Cardiothorac. Vasc. Anesth.* **27**, 1158–1166 (2013).
103. Gurm, H. S., Seth, M., Kooiman, J. & Share, D. A Novel Tool for Reliable and Accurate Prediction of Renal Complications in Patients Undergoing Percutaneous Coronary Intervention. *J. Am. Coll. Cardiol.* **61**, 2242–2248 (2013).
104. Breidthardt, T. *et al.* A Combined Cardiorenal Assessment for the Prediction of Acute Kidney Injury in Lower Respiratory Tract Infections. *Am. J. Med.* **125**, 168–175 (2012).
105. Cronin, R. M. *et al.* National veterans health administration inpatient risk stratification models for hospital-acquired acute kidney injury. *J. Am. Med. Inform. Assoc.* ocv051 (2015). doi:10.1093/jamia/ocv051
106. Guyatt, G. H. *et al.* What is "quality of evidence" and why is it important to clinicians? *BMJ* **336**, 995–998 (2008).
107. *NOUVEAUX MÉMOIRES DE L'ACADÉMIE ROYALE DES SCIENCES ET BELLES-LETTRES DE BRUXELLES*. (L'Académie Royale de Bruxelles et de l'Université Louvain, 1845).
108. Wilson, J. R. & Lorenz, K. A. Short History of the Logistic Regression Model. in *Modeling Binary Correlated Responses using SAS, SPSS and R* **9**, 17–23 (Springer International Publishing, 2015).
109. Cox, D. R. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Ser. B Methodol.* **20**, 215–242 (1958).
110. Truett, J., Cornfield, J. & Kannel, W. A multivariate analysis of the risk of coronary heart disease in Framingham. *J. Chronic Dis.* **20**, 511–524 (1967).
111. Wilson, P. W. F. *et al.* Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation* **97**, 1837–1847 (1998).
112. Pugh, R. N. H., Murray-Lyon, I. M., Dawson, J. L., Pietroni, M. C. & Williams, R. Transection of the oesophagus for bleeding oesophageal varices. *BJS* **60**, 646–649 (1973).

113. Child, C. G. & Turcotte, J. G. Surgery and portal hypertension. *Major Probl. Clin. Surg.* **1**, 1–85 (1964).
114. Maddrey, W. C. *et al.* Corticosteroid therapy of alcoholic hepatitis. *Gastroenterology* **75**, 193–199 (1978).
115. Ranson, J. H. *et al.* Prognostic signs and the role of operative management in acute pancreatitis. *Surg. Gynecol. Obstet.* **139**, 69–81 (1974).
116. Alvarado, A. A practical score for the early diagnosis of acute appendicitis. *Ann. Emerg. Med.* **15**, 557–564 (1986).
117. Knaus, W. A., Zimmerman, J. E., Wagner, D. P., Draper, E. A. & Lawrence, D. E. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit. Care Med.* **9**, 591–597 (1981).
118. Gage, B. F. *et al.* Validation of Clinical Classification Schemes for Predicting Stroke: Results From the National Registry of Atrial Fibrillation. *JAMA* **285**, 2864–2870 (2001).
119. Mallampati, S. R. *et al.* A clinical sign to predict difficult tracheal intubation; a prospective study. *Can. Anaesth. Soc. J.* **32**, 429–434 (1985).
120. Silberschatz, A., Stonebraker, M. & Ullman, J. D. Database Systems: Achievements and Opportunities. *SIGMOD Rec* **19**, 6–22 (1990).
121. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A. & Escobar, G. Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. *Health Aff. (Millwood)* **33**, 1123–1131 (2014).
122. Tabak, Y. P., Sun, X., Nunez, C. M. & Johannes, R. S. Using electronic health record data to develop inpatient mortality predictive model: Acute Laboratory Risk of Mortality Score (ALaRMS). *J. Am. Med. Inform. Assoc.* **21**, 455–463 (2014).
123. McNamara, R. L. *et al.* Development of a Hospital Outcome Measure Intended for Use With Electronic Health Records: 30-Day Risk-standardized Mortality After Acute Myocardial Infarction. *Med. Care* **53**, 818–826 (2015).
124. Bannay, A. *et al.* The Best Use of the Charlson Comorbidity Index With Electronic Health Care Database to Predict Mortality: *Med. Care* **54**, 188–194 (2016).
125. Escobar, G. J. *et al.* Stratification of Risk of Early-Onset Sepsis in Newborns ≥34 Weeks' Gestation. *Pediatrics* **133**, 30–36 (2014).
126. Escobar, G. J., Gardner, M. N., Greene, J. D., Draper, D. & Kipnis, P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med. Care* **51**, 446–453 (2013).
127. Liu, V., Turk, B. J., Ragins, A. I., Kipnis, P. & Escobar, G. J. An Electronic Simplified Acute Physiology Score-Based Risk Adjustment Score for Critical Illness in an Integrated Healthcare System. *Crit. Care Med.* **41**, 41–48 (2013).
128. Hu, Z. *et al.* Online Prediction of Health Care Utilization in the Next Six Months Based on Electronic Health Record Information: A Cohort and Validation Study. *J. Med. Internet Res.* **17**, (2015).
129. Steyerberg, E. W. *et al.* Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiol. Camb. Mass* **21**, 128–138 (2010).
130. Stone, N. J. *et al.* 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in AdultsA Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* **63**, 2889–2934 (2014).

131. MacMahon, H. *et al.* Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society. *Radiology* **237**, 395–400 (2005).
132. Moons, K. G. M. *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* **98**, 691–698 (2012).
133. Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the Yield of Medical Tests. *JAMA* **247**, 2543–2546 (1982).
134. Steyerberg, E. W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. (Springer Science & Business Media, 2008).
135. Dahiya, R. C. & Gurland, J. How Many Classes in the Pearson Chi—Square Test? *J. Am. Stat. Assoc.* **68**, 707–712 (1973).
136. Yu, W., Xu, W. & Zhu, L. A modified Hosmer–Lemeshow test for large data sets. *Commun. Stat. - Theory Methods* **46**, 11813–11825 (2017).
137. Denny, J. C., Irani, P. R., Wehbe, F. H., Smithers, J. D. & Spickard, A. The KnowledgeMap Project: Development of a Concept-Based Medical School Curriculum Database. *AMIA. Annu. Symp. Proc.* **2003**, 195–199 (2003).
138. Koller, D. & Sahami, M. *Toward Optimal Feature Selection*. (Stanford InfoLab, 1996).
139. Yang, Y. & Pedersen, J. O. A Comparative Study on Feature Selection in Text Categorization. in 412–420 (Morgan Kaufmann Publishers, 1997).
140. Derksen, S. & Keselman, H. J. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *Br. J. Math. Stat. Psychol.* **45**, 265–282 (1992).
141. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **31**, 2225–2236 (2010).
142. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1994).
143. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
144. Lu, F. & Petkova, E. A comparative study of variable selection methods in the context of developing psychiatric screening instruments. *Stat. Med.* **33**, 401–421 (2014).
145. Jolliffe, I. Principal Component Analysis. in *International Encyclopedia of Statistical Science* (ed. Lovric, M.) 1094–1096 (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-04898-2_455
146. Fraley, C. & Raftery, A. E. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.* **97**, 611–631 (2002).
147. Levy, O. & Goldberg, Y. Neural Word Embedding as Implicit Matrix Factorization. in *Advances in Neural Information Processing Systems 27* (eds. Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q.) 2177–2185 (Curran Associates, Inc., 2014).
148. Turian, J., Ratinov, L. & Bengio, Y. Word Representations: A Simple and General Method for Semi-supervised Learning. in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* 384–394 (Association for Computational Linguistics, 2010).
149. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient Estimation of Word Representations in Vector Space. *ArXiv13013781 Cs* (2013).
150. Le, Q. V. & Mikolov, T. Distributed Representations of Sentences and Documents. in *ICML* **14**, 1188–1196 (2014).

151. Yang, C. C. & Zhao, M. Determining Associations with Word Embedding in Heterogeneous Network for Detecting Off-Label Drug Uses. in *2017 IEEE International Conference on Healthcare Informatics (ICHI)* 496–501 (2017). doi:10.1109/ICHI.2017.78
152. Cocos, A., Fiks, A. G. & Masino, A. J. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *J. Am. Med. Inform. Assoc.* **24**, 813–821 (2017).
153. Baćac, A. Classification of Large-Scale Biological Annotations Using Word Embeddings Derived from Corpora of Biomedical Research Literature. (Fakultet Elektrotehnike i Računarstva, Sveučilište u Zagrebu, 2017).
154. Zhang, Y. *et al.* Psychiatric symptom recognition without labeled data using distributional representations of phrases and on-line knowledge. *J. Biomed. Inform.* (2017). doi:10.1016/j.jbi.2017.06.014
155. Turner, C. A. *et al.* Word2Vec inversion and traditional text classifiers for phenotyping lupus. *BMC Med. Inform. Decis. Mak.* **17**, 126 (2017).
156. Liu, H. & Motoda, H. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. (Springer Science & Business Media, 1998).
157. Dreiseitl, S. & Ohno-Machado, L. Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed. Inform.* **35**, 352–359 (2002).
158. Raghupathi, W. & Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* **2**, 3 (2014).
159. Safran, C. *et al.* Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *J. Am. Med. Inform. Assoc.* **14**, 1–9 (2007).
160. Gershon, N., Eick, S. G. & Card, S. Information Visualization. *interactions* **5**, 9–15 (1998).
161. Bertin, J. *Semiology of Graphics: Diagrams, Networks, Maps*. (Esri Press, 2010).
162. Card, S. K., Mackinlay, J. D. & Shneiderman, B. *Readings in information visualization: using vision to think*. (Morgan Kaufmann, 1999).
163. Healey, C. G., Booth, K. S. & Enns, J. T. Harnessing preattentive processes for multivariate data visualization. in *Graphics Interface* 107–107 (Citeseer, 1993).
164. Joseph, J. S., Chun, M. M. & Nakayama, K. Attentional requirements in a 'preattentive' feature search task. *Nature* **387**, 805–807 (1997).
165. de Oliveira, M. C. F. & Levkowitz, H. From visual data exploration to visual data mining: a survey. *IEEE Trans. Vis. Comput. Graph.* **9**, 378–394 (2003).
166. Keim, D. . Information visualization and visual data mining. *IEEE Trans. Vis. Comput. Graph.* **8**, 1–8 (2002).
167. Fua, Y.-H., Ward, M. O. & Rundensteiner, E. A. Hierarchical Parallel Coordinates for Exploration of Large Datasets. in *Proceedings of the Conference on Visualization '99: Celebrating Ten Years* 43–50 (IEEE Computer Society Press, 1999).
168. Spiegelhalter, D., Pearson, M. & Short, I. Visualizing Uncertainty About the Future. *Science* **333**, 1393–1400 (2011).
169. Pickett, R. M. & Levkowitz, H. Integrated displays of multiparameter and multimodality images. in *, Proceedings of the First Conference on Visualization in Biomedical Computing, 1990* 58–65 (1990). doi:10.1109/VBC.1990.109302
170. Fekete, J. D. & Plaisant, C. Interactive information visualization of a million items. in *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.* 117–124 (2002). doi:10.1109/INFVIS.2002.1173156

171. Correa, C., Crnovrsanin, T. & Ma, K.-L. Visual Reasoning about Social Networks Using Centrality Sensitivity. *IEEE Trans. Vis. Comput. Graph.* **18**, 106–120 (2012).
172. Pavlopoulos, G. A. *et al.* Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future. *GigaScience* **4**, 38 (2015).
173. Healey, Christopher. Perception in Visualization. *Perception in Visualization* Available at: https://www.csc2.ncsu.edu/faculty/healey/PP/index.html. (Accessed: 23rd October 2018)
174. Bederson, B. & Shneiderman, B. *The Craft of Information Visualization: Readings and Reflections*. (Morgan Kaufmann, 2003).
175. Plaisant, C., Milash, B., Rose, A., Widoff, S. & Shneiderman, B. LifeLines: Visualizing Personal Histories. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 221–227 (ACM, 1996). doi:10.1145/238386.238493
176. Lex, A., Schulz, H., Streit, M., Partl, C. & Schmalstieg, D. VisBricks: Multiform Visualization of Large, Inhomogeneous Data. *IEEE Trans. Vis. Comput. Graph.* **17**, 2291–2300 (2011).
177. Zhang, Z. *et al.* The Five Ws for Information Visualization with Application to Healthcare Informatics. *IEEE Trans. Vis. Comput. Graph.* **19**, 1895–1910 (2013).
178. Reijner, H. & Software, P. *The Development of the Horizon Graph*.
179. Berry, L. & Munzner, T. BinX: Dynamic Exploration of Time Series Datasets Across Aggregation Levels. in *IEEE Symposium on Information Visualization* p2–p2 (2004). doi:10.1109/INFVIS.2004.11
180. Kincaid, R. SignalLens: Focus+Context Applied to Electronic Time Series. *IEEE Trans. Vis. Comput. Graph.* **16**, 900–907 (2010).
181. Zhao, J., Chevalier, F. & Balakrishnan, R. KronoMiner: Using Multi-foci Navigation for the Visual Exploration of Time-series Data. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 1737–1746 (ACM, 2011). doi:10.1145/1978942.1979195
182. Rind, A. *et al.* Visual Exploration of Time-Oriented Patient Data for Chronic Diseases: Design Study and Evaluation. in *Information Quality in e-Health* 301–320 (Springer, Berlin, Heidelberg, 2011). doi:10.1007/978-3-642-25364-5_22
183. Gschwandtner, T., Aigner, W., Kaiser, K., Miksch, S. & Seyfang, A. CareCruiser: Exploring and visualizing plans, events, and effects interactively. in *2011 IEEE Pacific Visualization Symposium* 43–50 (2011). doi:10.1109/PACIFICVIS.2011.5742371
184. Brodbeck, D., Gasser, R., Degen, M., Reichlin, S. & Luthiger, J. Enabling large-scale telemedical disease management through interactive visualization. *Eur. Notes Med. Inform.* **1**, 1172–1177 (2005).
185. Vrotsou, K., Johansson, J. & Cooper, M. ActiviTree: Interactive Visual Exploration of Sequences in Event-Based Data Using Graph Similarity. *IEEE Trans. Vis. Comput. Graph.* **15**, 945–952 (2009).
186. Meyer, T. E. *et al. Visualizing Patterns of Drug Prescriptions with EventFlow: A Pilot Study of Asthma Medications in the Military Health System*. (2013).
187. Wang, T. D. *et al.* Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison. *IEEE Trans. Vis. Comput. Graph.* **15**, 1049–1056 (2009).
188. Sarkar, M. & Brown, M. H. Graphical Fisheye Views of Graphs. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 83–91 (ACM, 1992). doi:10.1145/142750.142763

189. Lundstrom, C., Ljung, P., Persson, A. & Ynnerman, A. Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation. *IEEE Trans. Vis. Comput. Graph.* **13**, 1648–1655 (2007).
190. Glatter, M., Huang, J., Ahern, S., Daniel, J. & Lu, A. Visualizing Temporal Patterns in Large Multivariate Data using Modified Globbing. *IEEE Trans. Vis. Comput. Graph.* **14**, 1467–1474 (2008).
191. Bui, A., Aberle, D. R. & Kangarloo, H. TimeLine: Visualizing Integrated Patient Records. *IEEE Trans. Inf. Technol. Biomed.* **11**, 462–473 (2007).
192. Suntinger, M., Obweger, H., Schiefer, J. & Groller, E. The Event Tunnel: Interactive Visualization of Complex Event Streams for Business Process Pattern Analysis. in *Visualization Symposium, 2008. PacificVIS '08. IEEE Pacific* 111–118 (2008). doi:10.1109/PACIFICVIS.2008.4475466
193. Roque, F. S., Slaughter, L. & Tkatšenko, A. A Comparison of Several Key Information Visualization Systems for Secondary Use of Electronic Health Record Content. in *Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents* 76–83 (Association for Computational Linguistics, 2010).
194. Maries, A. *et al.* GRACE: A Visual Comparison Framework for Integrated Spatial and Non-Spatial Geriatric Data. *IEEE Trans. Vis. Comput. Graph.* **19**, 2916–2925 (2013).
195. Maciejewski, R. *et al.* A Visual Analytics Approach to Understanding Spatiotemporal Hotspots. *IEEE Trans. Vis. Comput. Graph.* **16**, 205–220 (2010).
196. Bae, S.-H., Choi, J. Y., Qiu, J. & Fox, G. C. Dimension Reduction and Visualization of Large High-dimensional Data via Interpolation. in *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing* 203–214 (ACM, 2010). doi:10.1145/1851476.1851501
197. Turkay, C., Lundervold, A., Lundervold, A. & Hauser, H. Representative Factor Generation for the Interactive Visual Analysis of High-Dimensional Data. *IEEE Trans. Vis. Comput. Graph.* **18**, 2621–2630 (2012).
198. Gosink, L. *et al.* Characterizing and Visualizing Predictive Uncertainty in Numerical Ensembles Through Bayesian Model Averaging. *IEEE Trans. Vis. Comput. Graph.* **19**, 2703–2712 (2013).
199. Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P. & Moorhead, R. J. A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets. *IEEE Trans. Vis. Comput. Graph.* **15**, 1209–1218 (2009).
200. MacEachren, A. *et al.* Visual Semiotics amp;amp; Uncertainty Visualization: An Empirical Study. *IEEE Trans. Vis. Comput. Graph.* **18**, 2496–2505 (2012).
201. Zuk, T. & Carpendale, S. Theoretical analysis of uncertainty visualizations. in **6060**, 606007-606007–14 (2006).
202. Grigoryan, G. & Rheingans, P. Point-based probabilistic surfaces to show surface uncertainty. *IEEE Trans. Vis. Comput. Graph.* **10**, 564–573 (2004).
203. Lee, B., Robertson, G. G., Czerwinski, M. & Sims Parr, C. CandidTree: visualizing structural uncertainty in similar hierarchies. *Inf. Vis.* **6**, 233 (2007).
204. Bhaskaran, P. *et al.* ForeTell: Facilitating doctor-patient conversation through interactive information visualization of risk prediction index. in *2013 IEEE International Conference on Bioinformatics and Biomedicine* **0**, 1–4 (IEEE Computer Society, 2012).

205. Mane, K. K. *et al.* VisualDecisionLinc: A visual analytics approach for comparative effectiveness-based clinical decision support in psychiatry. *J. Biomed. Inform.* **45**, 101–106 (2012).
206. Falkman, G. Information visualisation in clinical Odontology: multidimensional analysis and interactive data exploration. *Artif. Intell. Med.* **22**, 133–158 (2001).
207. Vergouwe, Y., Moons, K. G. M. & Steyerberg, E. W. External Validity of Risk Models: Use of Benchmark Values to Disentangle a Case-Mix Effect From Incorrect Coefficients. *Am. J. Epidemiol.* **172**, 971–980 (2010).
208. Nieboer, D., Ploeg, T. van der & Steyerberg, E. W. Assessing Discriminative Performance at External Validation of Clinical Prediction Models. *PLOS ONE* **11**, e0148820 (2016).
209. Kimmel, S. E. *et al.* A Pharmacogenetic versus a Clinical Algorithm for Warfarin Dosing. *N. Engl. J. Med.* **369**, 2283–2293 (2013).
210. Lee SJ, Lindquist K, Segal MR & Covinsky KE. DEvelopment and validation of a prognostic index for 4-year mortality in older adults. *JAMA* **295**, 801–808 (2006).
211. Shanmugam, G., West, M. & Berg, G. Additive and logistic EuroSCORE performance in high risk patients. *Interact. Cardiovasc. Thorac. Surg.* **4**, 299–303 (2005).
212. Dehejia, R. H. & Wahba, S. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Rev. Econ. Stat.* **84**, 151–161 (2002).
213. Osborne, J. W. *Best Practices in Quantitative Methods*. (SAGE, 2008).
214. King, G. & Nielsen, R. Why propensity scores should not be used for matching. *Copy Httpj Mp1sexgVw Download Cit. BibTex Tagged XML Download Pap.* **378**, (2016).
215. Stürmer, T. *et al.* A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J. Clin. Epidemiol.* **59**, 437.e1-437.e24 (2006).
216. Cooper, L. A. D. *et al.* Integrated morphologic analysis for the identification and characterization of disease subtypes. *J. Am. Med. Inform. Assoc.* **19**, 317–323 (2012).
217. Wardlaw, A. J., Silverman, M., Siva, R., Pavord, I. D. & Green, R. Multi-dimensional phenotyping: towards a new taxonomy for airway disease. *Clin. Exp. Allergy* **35**, 1254–1262 (2005).
218. Pontremoli, R. *et al.* Prevalence and Clinical Correlates of Microalbuminuria in Essential Hypertension The MAGIC Study. *Hypertension* **30**, 1135–1143 (1997).
219. Chen, R. *et al.* Patient Stratification Using Electronic Health Records from a Chronic Disease Management Program. *IEEE J. Biomed. Health Inform.* **PP**, 1–1 (2016).
220. Gotz, D., Sun, J., Cao, N. & Ebadollahi, S. Visual cluster analysis in support of clinical decision intelligence. in *AMIA Annual Symposium Proceedings* **2011**, 481 (American Medical Informatics Association, 2011).
221. Gungor, G. *et al.* Neutrophil gelatinase-associated lipocalin in prediction of mortality in patients with hepatorenal syndrome: a prospective observational study. *Liver Int.* **34**, 49–57 (2014).
222. Martinez, M. O., Sayles, H., Vivekanandan, R., Souza, S. D. & Florescu, M. C. Hepatorenal Syndrome: Are We Missing Some Prognostic Factors? *Dig. Dis. Sci.* **57**, 210–214 (2012).
223. Martín-Llahí, M., Guevara, M. & Ginès, P. Hyponatremia in cirrhosis: clinical features and management. *Gastroentérologie Clin. Biol.* **30**, 1144–1151 (2006).
224. Salerno, F. *et al.* Diagnosis, treatment and survival of patients with hepatorenal syndrome: A survey on daily medical practice. *J. Hepatol.* **55**, 1241–1248 (2011).

225. Barreto, R. *et al.* Type-1 hepatorenal syndrome associated with infections in cirrhosis: Natural history, outcome of kidney function, and survival. *Hepatology* **59**, 1505–1513 (2014).
226. Hickey, G. L. *et al.* Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur. J. Cardio-Thorac. Surg. Off. J. Eur. Assoc. Cardio-Thorac. Surg.* **43**, 1146–1152 (2013).
227. Keegan, M. T., Gajic, O. & Afessa, B. Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and Influence of Resuscitation Status on Model Performance. *Chest* **142**, 851–858 (2012).
228. Arroyo, V. *et al.* Definition and diagnostic criteria of refractory ascites and hepatorenal syndrome in cirrhosis. *Hepatology* **23**, 164–176 (1996).
229. Guevara, M. *et al.* Transjugular intrahepatic portosystemic shunt in hepatorenal syndrome: Effects on renal function and vasoactive systems. *Hepatology* **28**, 416–422 (1998).
230. Testino, G., Leone, S., Ferro, C. & Borro, P. Severe acute alcoholic hepatitis and hepatorenal syndrome: role of transjugular intrahepatic portosystemic stent shunt. *J. Med. Life* **5**, 203–205 (2012).
231. Brensing, K. A. *et al.* Long term outcome after transjugular intrahepatic portosystemic stent-shunt in non-transplant cirrhotics with hepatorenal syndrome: a phase II study. *Gut* **47**, 288–295 (2000).
232. Sersté, T. *et al.* Deleterious effects of beta-blockers on survival in patients with cirrhosis and refractory ascites. *Hepatology* **52**, 1017–1022 (2010).
233. Mandorfer, M. *et al.* Nonselective β Blockers Increase Risk for Hepatorenal Syndrome and Death in Patients With Cirrhosis and Spontaneous Bacterial Peritonitis. *Gastroenterology* **146**, 1680-1690.e1 (2014).
234. Krag, A., Bendtsen, F., Henriksen, J. H. & Møller, S. Low cardiac output predicts development of hepatorenal syndrome and survival in patients with cirrhosis and ascites. *Gut* **59**, 105–110 (2010).
235. Brensing, K. *et al.* Transjugular intrahepatic portosystemic stent-shunt for hepatorenal syndrome. *The Lancet* **349**, 697–698 (1997).
236. Brown, S. H., Lincoln, M. J., Groen, P. J. & Kolodner, R. M. VistA—U.S. Department of Veterans Affairs national-scale HIS. *Int. J. Med. Inf.* **69**, 135–156 (2003).
237. VA National Drug File - Data.gov. Available at: https://catalog.data.gov/dataset/va-national-drug-file-may-2015. (Accessed: 13th June 2017)
238. Su, Y.-S., Yajima, M., Gelman, A. E. & Hill, J. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. *J. Stat. Softw.* **45**, 1–31 (2011).
239. Sánchez, D. & Batet, M. Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. *J. Biomed. Inform.* **44**, 749–759 (2011).
240. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63**, 411–423 (2001).
241. Řehůřek, R. & Sojka, P. Software Framework for Topic Modelling with Large Corpora. in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* 45–50 (ELRA, 2010).
242. Bekkerman, R., El-Yaniv, R., Tishby, N. & Winter, Y. Distributional Word Clusters vs. Words for Text Categorization. *J. Mach. Learn. Res.* **3**, 1183–1208 (2003).

243. Dumais, S., Platt, J., Heckerman, D. & Sahami, M. Inductive Learning Algorithms and Representations for Text Categorization. in *Proceedings of the Seventh International Conference on Information and Knowledge Management* 148–155 (ACM, 1998). doi:10.1145/288627.288651
244. Van Hoorde, K., Van Huffel, S., Timmerman, D., Bourne, T. & Van Calster, B. A spline-based tool to assess and visualize the calibration of multiclass risk predictions. *J. Biomed. Inform.* **54**, 283–293 (2015).
245. Bellomo, R., Kellum, J. A. & Ronco, C. Acute kidney injury. *The Lancet* **380**, 756–766 (2012).
246. Garla, V. N. & Brandt, C. Ontology-guided feature engineering for clinical text classification. *J. Biomed. Inform.* **45**, 992–998 (2012).
247. Miller, S., Guinness, J. & Zamanian, A. Name tagging with word clusters and discriminative training. in *Proceedings of HLT* 337–342 (2004).
248. Hofmann, T. Probabilistic Latent Semantic Analysis. in *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* 289–296 (Morgan Kaufmann Publishers Inc., 1999).
249. Chen, Y. *et al.* Applying active learning to high-throughput phenotyping algorithms for electronic health records data. *J. Am. Med. Inform. Assoc.* e253–e259 doi:10.1136/amiajnl-2013-001945
250. Shivade, C., Malewadkar, P., Fosler-Lussier, E. & Lai, A. M. Comparison of UMLS Terminologies to Identify Risk of Heart Disease in Clinical Notes. *J. Biomed. Inform.* **58**, S103–S110 (2015).
251. *CommonDataModel: Definition and DDLs for the OMOP Common Data Model (CDM)*. (Observational Health Data Sciences and Informatics, 2018).
252. Medicare, C. for, Baltimore, M. S. 7500 S. B. & Usa, M. 2017-ICD-10-CM-and-GEMs. (2016). Available at: https://www.cms.gov/medicare/coding/icd10/2017-icd-10-cm-and-gems.html. (Accessed: 22nd March 2017)
253. Moskovitch, R. & Shahar, Y. Medical Temporal-Knowledge Discovery via Temporal Abstraction. *AMIA. Annu. Symp. Proc.* **2009**, 452–456 (2009).
254. Sacchi, L., Capozzi, D., Bellazzi, R. & Larizza, C. JTSA: An open source framework for time series abstractions. *Comput. Methods Programs Biomed.* **121**, 175–188 (2015).
255. Post, A. R. *et al.* Temporal Abstraction-based Clinical Phenotyping with Eureka! *AMIA. Annu. Symp. Proc.* **2013**, 1160–1169 (2013).
256. KDIGO Work Group. Section 2: AKI Definition. *Kidney Int. Suppl.* **2**, 19–36 (2012).
257. Siew, E. D. *et al.* Estimating Baseline Kidney Function in Hospitalized Patients with Impaired Kidney Function. *Clin. J. Am. Soc. Nephrol.* **7**, 712–719 (2012).
258. Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J. & Solenberger, P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* **27**, 85–96 (2001).
259. Yan, J. & Fine, J. P. Estimating Equations for Association Structures. *Stat. Med.* **23**, 859–880 (2004).
260. Sullivan, L. M., Massaro, J. M. & D'Agostino, R. B. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat. Med.* **23**, 1631–1660 (2004).
261. Altman, D. G., Vergouwe, Y., Royston, P. & Moons, K. G. M. Prognosis and prognostic research: validating a prognostic model. *BMJ* **338**, b605 (2009).

262. Pericleous, M., Sarnowski, A., Moore, A., Fijten, R. & Zaman, M. The clinical management of abdominal ascites, spontaneous bacterial peritonitis and hepatorenal syndrome: a review of current guidelines and recommendations. *Eur. J. Gastroenterol. Hepatol.* **28**, e10–e18 (2016).
263. Low, G. *et al.* Magnetic resonance elastography in the detection of hepatorenal syndrome in patients with cirrhosis and ascites. *Eur. Radiol.* **25**, 2851–2858 (2015).
264. Goyal, S. *et al.* Intrarenal resistance index (RI) as a predictor of early renal impairment in patients with liver cirrhosis. *Trop. Gastroenterol. Off. J. Dig. Dis. Found.* **34**, 235–239 (2013).
265. Kastelan, S., Ljubicic, N., Kastelan, Z., Ostojic, R. & Uravic, M. The role of duplex-doppler ultrasonography in the diagnosis of renal dysfunction and hepatorenal syndrome in patients with liver cirrhosis. *Hepatogastroenterology.* **51**, 1408–1412 (2004).
266. Nicković, V. *et al.* Diagnostical significance of dimethylarginine in the development of hepatorenal syndrome in patients with alcoholic liver cirrhosis. *Vojnosanit. Pregl.* **69**, 686–691 (2012).
267. Lluch, P. *et al.* Accumulation of Symmetric Dimethylarginine in Hepatorenal Syndrome. *Exp. Biol. Med.* **231**, 70–75 (2006).
268. Fagundes, C. *et al.* Urinary neutrophil gelatinase-associated lipocalin as biomarker in the differential diagnosis of impairment of kidney function in cirrhosis. *J. Hepatol.* **57**, 267–273 (2012).
269. Ahn, H. S. *et al.* Cystatin C is a good predictor of hepatorenal syndrome and survival in patients with cirrhosis who have normal serum creatinine levels. *Hepatogastroenterology.* **59**, 1168–1173 (2012).
270. Janičko, M., Veselíny, E., Abraldes, J. G. & Jarčuška, P. Serum sodium identifies patients with cirrhosis at high risk of hepatorenal syndrome. *Z. Für Gastroenterol.* **51**, 628–634 (2013).
271. Ruiz-del-Arbol, L. *et al.* Circulatory function and hepatorenal syndrome in cirrhosis. *Hepatology* **42**, 439–447 (2005).
272. Fischer, J. E. & James, J. H. Treatment of hepatic coma and hepatorenal syndrome. *Am. J. Surg.* **123**, 222–230 (1972).
273. Ruiz-del-Arbol, L. *et al.* Paracentesis-induced circulatory dysfunction: Mechanism and effect on hepatic hemodynamics in cirrhosis. *Gastroenterology* **113**, 579–586 (1997).
274. Vila, M. C. *et al.* Hemodynamic changes in patients developing effective hypovolemia after total paracentesis. *J. Hepatol.* **28**, 639–645 (1998).
275. Tsien, C. D., Rabie, R. & Wong, F. Acute kidney injury in decompensated cirrhosis. *Gut* **62**, 131–137 (2013).
276. Salerno, F., Gerbes, A., Ginès, P., Wong, F. & Arroyo, V. Diagnosis, prevention and treatment of hepatorenal syndrome in cirrhosis. *Postgrad. Med. J.* **84**, 662–670 (2008).
277. Wonnacott, A., Meran, S., Amphlett, B., Talabani, B. & Phillips, A. Epidemiology and Outcomes in Community-Acquired Versus Hospital-Acquired AKI. *Clin. J. Am. Soc. Nephrol. CJASN* **9**, 1007–1014 (2014).
278. Grahn, E. P., Dietz, A. A., Stefani, S. S. & Donnelly, W. J. Burr cells, hemolytic anemia and cirrhosis. *Am. J. Med.* **45**, 78–87 (1968).
279. Miller, A. *et al.* Application of contextual design methods to inform targeted clinical decision support interventions in sub-specialty care environments. *Int. J. Med. Inf.* **117**, 55–65 (2018).

280. Bostock, M., Ogievetsky, V. & Heer, J. D3 Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.* **17**, 2301–2309 (2011).
281. R Studio, Inc. Shiny: Easy web applications in R. (2013). Available at: http://shiny.rstudio.com/. (Accessed: 30th October 2018)
282. Kohonen, T. Essentials of the self-organizing map. *Neural Netw.* **37**, 52–65 (2013).
283. Wehrens, R. & Buydens, L. M. C. Self- and Super-organising Maps in R: the kohonen package. *J Stat Softw* **21**, (2007).
284. Gower, J. C. Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra Its Appl.* **67**, 81–97 (1985).
285. Fang, Y. & Wang, J. Selection of the number of clusters via the bootstrap method. *Comput. Stat. Data Anal.* **56**, 468–477 (2012).
286. Goldberg, J. & Helfman, J. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Inf. Vis.* **10**, 182–195 (2011).
287. Few, S. Keep Radar Graphs Below the Radar - Far Below. *https://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf*
288. Harris, R. L. *Information Graphics: A Comprehensive Illustrated Reference*. (Oxford University Press, 2000).
289. Van Calster, B. *et al.* A calibration hierarchy for risk models was defined: from utopia to empirical data. *J. Clin. Epidemiol.* **74**, 167–176 (2016).
290. Poole, A. & Ball, L. J. Eye tracking in HCI and usability research. *Encycl. Hum.-Comput. Interact. C Ghaoui Ed* (2006).
291. Davis, K. *2012 Annual Report President's Message—Health Care Reform: A Journey | Commonwealth Fund*. (The Commonwealth Fund, 2012).
292. Allen, A. M. *et al.* Time trends in the health care burden and mortality of acute on chronic liver failure in the United States. *Hepatology* **64**, 2165–2172 (2016).
293. Bajaj, J. S. *et al.* The 3-month readmission rate remains unacceptably high in a large North American cohort of patients with cirrhosis. *Hepatology* **64**, 200–208 (2016).
294. Parasa, S., Sridhar, A. R. M. & Olden, K. A Study of Trends in Diagnosis, Outcomes and Costs of Hepatorenal Syndrome in End-Stage Liver Disease Patients - Results From Nationwide Inpatient Database 2000-2009. *Gastroenterology* **142**, S-948 (2012).
295. Fihn, S. D. *et al.* Insights from advanced analytics at the Veterans Health Administration. *Health Aff. Proj. Hope* **33**, 1203–1211 (2014).
296. Modification of Diet in Renal Disease Study Group *et al.* Predictors of the progression of renal disease in the Modification of Diet in Renal Disease Study. *Kidney Int.* **51**, 1908–1919 (1997).
297. de Franchis, R. Evolving Consensus in Portal Hypertension Report of the Baveno IV Consensus Workshop on methodology of diagnosis and therapy in portal hypertension. *J. Hepatol.* **43**, 167–176 (2005).
298. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. & Lipson, H. Understanding Neural Networks Through Deep Visualization. 12
299. Hohman, F., Kahng, M., Pienta, R. & Chau, D. H. Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers. *ArXiv180106889 Cs Stat* (2018).
300. Nemati, S. *et al.* An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit. Care Med.* **46**, 547 (2018).

301. Bender, D. & Sartipi, K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* 326–331 (2013). doi:10.1109/CBMS.2013.6627810
302. Spineth, M., Rappelsberger, A. & Adlassnig, K. P. Implementing CDS Hooks Communication in an Arden-Syntax-Based Clinical Decision Support Platform. *Stud. Health Technol. Inform.* **255**, 165–169 (2018).
303. RxNorm API. Available at: https://rxnav.nlm.nih.gov/RxNormAPIs.html#. (Accessed: 2nd November 2017)
304. Belziti, C. A. *et al.* Worsening Renal Function in Patients Admitted With Acute Decompensated Heart Failure: Incidence, Risk Factors and Prognostic Implications. *Rev. Esp. Cardiol.* **63**, 294–302 (2010).
305. Cruz, D. N., Gheorghiade, M., Palazuolli, A., Ronco, C. & Bagshaw, S. M. Epidemiology and outcome of the cardio-renal syndrome. *Heart Fail. Rev.* **16**, 531–542 (2011).
306. Nohria, A. *et al.* Cardiorenal Interactions: Insights From the ESCAPE Trial. *J. Am. Coll. Cardiol.* **51**, 1268–1274 (2008).

APPENDIX A: DETAILS ON PHENOTYPING ALGORITHMS

Appendix Table A.1: List of candidate predictor variables with their data domain (Class) and their timeframe (Preadmission or Home vs. Inpatient) along with summary measures, percent missing, and number of patients with a value > 0.

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Cond AAA	Home	Condition	12.00 (2.38)	0	12
Home Cond AbdomSurg	Home	Condition	22.00 (4.37)	0	22
Home Cond ACS	Home	Condition	39.00 (7.74)	0	39
Home Cond AFIB	Home	Condition	65.00 (12.90)	0	65
Home Cond ALD	Home	Condition	456.00 (90.48)	0	456
Home Cond AnalFisFist	Home	Condition	2.00 (0.40)	0	2
Home Cond Angina	Home	Condition	25.00 (4.96)	0	25
Home Cond ANMA	Home	Condition	260.00 (51.59)	0	260
Home Cond ARF	Home	Condition	232.00 (46.03)	0	232
Home Cond ARRH	Home	Condition	141.00 (27.98)	0	141
Home Cond Ascites	Home	Condition	254.00 (50.40)	0	254
Home Cond ASP	Home	Condition	2.00 (0.40)	0	2
Home Cond Asthma	Home	Condition	26.00 (5.16)	0	26
Home Cond ATN	Home	Condition	18.00 (3.57)	0	18
Home Cond AutoNeuropathy	Home	Condition	3.00 (0.60)	0	3
Home Cond AZ Cancer	Home	Condition	126.00 (25.00)	0	126
Home Cond BilCirrhosis	Home	Condition	3.00 (0.60)	0	3
Home Cond BowelPerf	Home	Condition	3.00 (0.60)	0	3
Home Cond BURN	Home	Condition	4.00 (0.79)	0	4
Home Cond CABG	Home	Condition	25.00 (4.96)	0	25
Home Cond CAD	Home	Condition	120.00 (23.81)	0	120
Home Cond Cancer	Home	Condition	109.00 (21.63)	0	109
Home Cond CANDI	Home	Condition	27.00 (5.36)	0	27
Home Cond CardiacArrest	Home	Condition	5.00 (0.99)	0	5
Home Cond CardSurg	Home	Condition	3.00 (0.60)	0	3
Home Cond CarotidDis	Home	Condition	4.00 (0.79)	0	4
Home Cond Cath	Home	Condition	23.00 (4.56)	0	23
Home Cond CathPCIALL	Home	Condition	24.00 (4.76)	0	24
Home Cond CathPCICont	Home	Condition	23.00 (4.56)	0	23
Home Cond CathPCInoCont	Home	Condition	3.00 (0.60)	0	3
Home Cond CathPCIwCont	Home	Condition	23.00 (4.56)	0	23
Home Cond CDVD	Home	Condition	88.00 (17.46)	0	88

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Cond CHEMO	Home	Condition	17.00 (3.37)	0	17
Home Cond CHF	Home	Condition	80.00 (15.87)	0	80
Home Cond Cirrhosis	Home	Condition	444.00 (88.10)	0	444
Home Cond Cirrhosis Risk 1	Home	Condition	392.00 (77.78)	0	392
Home Cond CKD	Home	Condition	74.00 (14.68)	0	74
Home Cond Colitis	Home	Condition	4.00 (0.79)	0	4
Home Cond Cons Condition	Home	Condition	98.00 (19.44)	0	98
Home Cond COPDAsthma	Home	Condition	160.00 (31.75)	0	160
Home Cond CS	Home	Condition	2.00 (0.40)	0	2
Home Cond CVA	Home	Condition	48.00 (9.52)	0	48
Home Cond DecALD	Home	Condition	467.00 (92.66)	0	467
Home Cond Dementia	Home	Condition	5.00 (0.99)	0	5
Home Cond DIAL	Home	Condition	12.00 (2.38)	0	12
Home Cond DIAR	Home	Condition	73.00 (14.48)	0	73
Home Cond DKA	Home	Condition	2.00 (0.40)	0	2
Home Cond DM	Home	Condition	181.00 (35.91)	0	181
Home Cond DMNeuropathy	Home	Condition	26.00 (5.16)	0	26
Home Cond DMOsm	Home	Condition	4.00 (0.79)	0	4
Home Cond DYS	Home	Condition	170.00 (33.73)	0	170
Home Cond ETOH	Home	Condition	347.00 (68.85)	0	347
Home Cond Etoh Abuse	Home	Condition	281.00 (55.75)	0	281
Home Cond Fatigue	Home	Condition	1.00 (0.20)	0	1
Home Cond Fibromyalgia	Home	Condition	11.00 (2.18)	0	11
Home Cond Gastroparesis	Home	Condition	5.00 (0.99)	0	5
Home Cond GI	Home	Condition	155.00 (30.75)	0	155
Home Cond GIPerf	Home	Condition	3.00 (0.60)	0	3
Home Cond GLOM	Home	Condition	2.00 (0.40)	0	2
Home Cond GLOMNephEx	Home	Condition	2.00 (0.40)	0	2
Home Cond HCC	Home	Condition	50.00 (9.92)	0	50
Home Cond HE	Home	Condition	136.00 (26.98)	0	136
Home Cond Headache	Home	Condition	36.00 (7.14)	0	36
Home Cond Hemorrhoid	Home	Condition	95.00 (18.85)	0	95
Home Cond HEP	Home	Condition	299.00 (59.33)	0	299
Home Cond Hep B C	Home	Condition	242.00 (48.02)	0	242
Home Cond HIV	Home	Condition	12.00 (2.38)	0	12
Home Cond HOTN	Home	Condition	110.00 (21.83)	0	110
Home Cond HRS	Home	Condition	113.00 (22.42)	0	113
Home Cond HTN	Home	Condition	357.00 (70.83)	0	357
Home Cond HTNEmer	Home	Condition	4.00 (0.79)	0	4

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Cond HYPC	Home	Condition	5.00 (0.99)	0	5
Home Cond Hyperkalemia	Home	Condition	62.00 (12.30)	0	62
Home Cond Hyperparathyroidism	Home	Condition	3.00 (0.60)	0	3
Home Cond IBS	Home	Condition	5.00 (0.99)	0	5
Home Cond IVD	Home	Condition	81.00 (16.07)	0	81
Home Cond JAUD	Home	Condition	72.00 (14.29)	0	72
Home Cond LIV	Home	Condition	462.00 (91.67)	0	462
Home Cond LKM	Home	Condition	3.00 (0.60)	0	3
Home Cond LUP	Home	Condition	7.00 (1.39)	0	7
Home Cond MECHVENT	Home	Condition	20.00 (3.97)	0	20
Home Cond Megacolon	Home	Condition	1.00 (0.20)	0	1
Home Cond MetabolicSyndrome	Home	Condition	127.00 (25.20)	0	127
Home Cond MI	Home	Condition	18.00 (3.57)	0	18
Home Cond MVR	Home	Condition	14.00 (2.78)	0	14
Home Cond Myopathies	Home	Condition	20.00 (3.97)	0	20
Home Cond NAFLD	Home	Condition	72.00 (14.29)	0	72
Home Cond NephGLOM	Home	Condition	9.00 (1.79)	0	9
Home Cond Nephrtmy	Home	Condition	2.00 (0.40)	0	2
Home Cond NFXss	Home	Condition	403.00 (79.96)	0	403
Home Cond NSTEMI	Home	Condition	12.00 (2.38)	0	12
Home Cond OA	Home	Condition	125.00 (24.80)	0	125
Home Cond Obesity	Home	Condition	108.00 (21.43)	0	108
Home Cond OFss	Home	Condition	367.00 (72.82)	0	367
Home Cond OLG	Home	Condition	1.00 (0.20)	0	1
Home Cond PALL	Home	Condition	47.00 (9.33)	0	47
Home Cond Parkinsons	Home	Condition	3.00 (0.60)	0	3
Home Cond PCI	Home	Condition	17.00 (3.37)	0	17
Home Cond PCR	Home	Condition	50.00 (9.92)	0	50
Home Cond PFT	Home	Condition	40.00 (7.94)	0	40
Home Cond Plegia CC	Home	Condition	9.00 (1.79)	0	9
Home Cond Porphyria	Home	Condition	2.00 (0.40)	0	2
Home Cond PUD CC	Home	Condition	42.00 (8.33)	0	42
Home Cond PVD	Home	Condition	28.00 (5.56)	0	28
Home Cond RA	Home	Condition	4.00 (0.79)	0	4
Home Cond RectalProlapse	Home	Condition	1.00 (0.20)	0	1
Home Cond RHBD	Home	Condition	8.00 (1.59)	0	8
Home Cond Rheum CC	Home	Condition	6.00 (1.19)	0	6

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Cond SBP	Home	Condition	49.00 (9.72)	0	49
Home Cond Sickle	Home	Condition	1.00 (0.20)	0	1
Home Cond SIRS	Home	Condition	51.00 (10.12)	0	51
Home Cond SpinalCord	Home	Condition	4.00 (0.79)	0	4
Home Cond SPS	Home	Condition	33.00 (6.55)	0	33
Home Cond StableAngina	Home	Condition	21.00 (4.17)	0	21
Home Cond STEMI	Home	Condition	13.00 (2.58)	0	13
Home Cond STROKE	Home	Condition	30.00 (5.95)	0	30
Home Cond TB	Home	Condition	3.00 (0.60)	0	3
Home Cond TIA	Home	Condition	13.00 (2.58)	0	13
Home Cond TIPS	Home	Condition	13.00 (2.58)	0	13
Home Cond TOBC	Home	Condition	234.00 (46.43)	0	234
Home Cond TRAU	Home	Condition	30.00 (5.95)	0	30
Home Cond UNAN	Home	Condition	34.00 (6.75)	0	34
Home Cond UrinaryObst	Home	Condition	46.00 (9.13)	0	46
Home Cond Varices	Home	Condition	124.00 (24.60)	0	124
Home Cond VascProc	Home	Condition	73.00 (14.48)	0	73
Home Cond VascSurg	Home	Condition	58.00 (11.51)	0	58
Home Cond VHD	Home	Condition	25.00 (4.96)	0	25
Home Cond VLP	Home	Condition	58.00 (11.51)	0	58
Home Cond VMT	Home	Condition	22.00 (4.37)	0	22
Inpt Cond AA, Elixhauser	Inpatient	Condition	1.16 (2.01)	0	276
Inpt Cond AAA	Inpatient	Condition	0.04 (0.73)	0	5
Inpt Cond AbdomSurg	Inpatient	Condition	0.08 (0.44)	0	18
Inpt Cond ACS	Inpatient	Condition	0.05 (0.33)	0	14
Inpt Cond aDIAL	Inpatient	Condition	0.33 (1.68)	0	39
Inpt Cond AFIB	Inpatient	Condition	0.20 (0.90)	0	50
Inpt Cond AFL	Inpatient	Condition	0.27 (1.16)	0	68
Inpt Cond AIDS/HIV, Elixhauser	Inpatient	Condition	0.04 (0.41)	0	8
Inpt Cond ALD	Inpatient	Condition	2.89 (3.49)	0	445

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond Angina	Inpatient	Condition	0.02 (0.19)	0	7
Inpt Cond ANMA	Inpatient	Condition	0.67 (1.62)	0	185
Inpt Cond ARF	Inpatient	Condition	1.54 (2.77)	0	346
Inpt Cond ARRH	Inpatient	Condition	0.38 (1.04)	0	102
Inpt Cond Arrhythmias, Elixhauser	Inpatient	Condition	0.09 (0.37)	0	30
Inpt Cond Ascites	Inpatient	Condition	1.06 (1.69)	0	256
Inpt Cond Asthma	Inpatient	Condition	0.06 (0.54)	0	16
Inpt Cond ATN	Inpatient	Condition	0.11 (0.74)	0	35
Inpt Cond AutoNeuropathy	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond AZ Cancer	Inpatient	Condition	0.51 (1.84)	0	94
Inpt Cond BilCirrhosis	Inpatient	Condition	0.00 (0.09)	0	1
Inpt Cond BLA, Elixhauser	Inpatient	Condition	6.00 (1.19)	0	6
Inpt Cond BowelPerf	Inpatient	Condition	0.01 (0.10)	0	2
Inpt Cond CABG	Inpatient	Condition	0.04 (0.44)	0	10
Inpt Cond CAD	Inpatient	Condition	0.23 (0.73)	0	75
Inpt Cond Cancer	Inpatient	Condition	0.48 (1.82)	0	87
Inpt Cond CANDI	Inpatient	Condition	0.05 (0.24)	0	21
Inpt Cond CardiacArrest	Inpatient	Condition	8.00 (1.59)	0	8
Inpt Cond CardSurg	Inpatient	Condition	0.01 (0.14)	0	4
Inpt Cond Cath	Inpatient	Condition	0.02 (0.16)	0	7
Inpt Cond CathPCIALL	Inpatient	Condition	0.02 (0.16)	0	7

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond CathPCICont	Inpatient	Condition	0.02 (0.15)	0	6
Inpt Cond CathPCInoCont	Inpatient	Condition	4.00 (0.79)	0	4
Inpt Cond CathPCIwCont	Inpatient	Condition	0.02 (0.15)	0	6
Inpt Cond CDVD	Inpatient	Condition	0.15 (0.85)	0	39
Inpt Cond CGP, Elixhauser	Inpatient	Condition	0.38 (1.06)	0	134
Inpt Cond CHF	Inpatient	Condition	0.41 (1.78)	0	79
Inpt Cond CHF bansal	Inpatient	Condition	0.41 (1.78)	0	79
Inpt Cond CHF, Elixhauser	Inpatient	Condition	0.40 (1.78)	0	78
Inpt Cond Cirrhosis	Inpatient	Condition	2.01 (2.53)	0	420
Inpt Cond Cirrhosis Risk 1	Inpatient	Condition	1.73 (3.09)	0	344
Inpt Cond CKD	Inpatient	Condition	0.36 (1.22)	0	98
Inpt Cond Cons Condition	Inpatient	Condition	0.09 (0.60)	0	24
Inpt Cond Cons Proc	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond COPD	Inpatient	Condition	0.57 (1.98)	0	114
Inpt Cond COPDAsthma	Inpatient	Condition	0.61 (2.04)	0	121
Inpt Cond CPD, Elixhauser	Inpatient	Condition	0.63 (2.04)	0	130
Inpt Cond CS	Inpatient	Condition	0.01 (0.11)	0	3
Inpt Cond CVA	Inpatient	Condition	0.10 (0.86)	0	23
Inpt Cond DA, Elixhauser	Inpatient	Condition	0.08 (0.32)	0	34
Inpt Cond DecALD	Inpatient	Condition	3.09 (3.60)	0	451
Inpt Cond Deyo's CHF	Inpatient	Condition	0.39 (1.77)	0	73

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond Deyo's CPD	Inpatient	Condition	0.57 (1.98)	0	115
Inpt Cond Deyo's CVD	Inpatient	Condition	0.11 (0.87)	0	26
Inpt Cond Deyo's DM w/ chronic comp	Inpatient	Condition	0.06 (0.38)	0	20
Inpt Cond Deyo's DM w/o chronic comp	Inpatient	Condition	0.47 (1.16)	0	135
Inpt Cond Deyo's Hemiplegia or Paraplegia	Inpatient	Condition	0.01 (0.10)	0	2
Inpt Cond Deyo's HIV	Inpatient	Condition	0.04 (0.41)	0	8
Inpt Cond Deyo's Malignancy	Inpatient	Condition	0.48 (1.82)	0	87
Inpt Cond Deyo's MI	Inpatient	Condition	0.04 (0.29)	0	15
Inpt Cond Deyo's Mild Liver Disease	Inpatient	Condition	2.03 (2.54)	0	420
Inpt Cond Deyo's Moderate to Severe Liver Disease	Inpatient	Condition	1.44 (2.12)	0	337
Inpt Cond Deyo's PUD	Inpatient	Condition	0.06 (0.32)	0	18
Inpt Cond Deyo's PVD	Inpatient	Condition	0.14 (0.91)	0	25
Inpt Cond Deyo's Renal	Inpatient	Condition	0.45 (1.33)	0	120
Inpt Cond Deyo's Rheumatic	Inpatient	Condition	0.01 (0.11)	0	3
Inpt Cond Deyo's Tumor	Inpatient	Condition	0.05 (0.25)	0	20
Inpt Cond DiabetesC, Elixhauser	Inpatient	Condition	0.14 (0.86)	0	31
Inpt Cond DiabetesU, Elixhauser	Inpatient	Condition	0.44 (1.07)	0	134
Inpt Cond DIAL	Inpatient	Condition	0.48 (2.49)	0	44
Inpt Cond DIAR	Inpatient	Condition	0.07 (0.33)	0	26
Inpt Cond DKA	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond DM	Inpatient	Condition	0.57 (1.49)	0	149

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond DMNeuropathy	Inpatient	Condition	8.00 (1.59)	0	8
Inpt Cond DMOsm	Inpatient	Condition	0.03 (0.42)	0	4
Inpt Cond DP, Elixhauser	Inpatient	Condition	0.13 (0.56)	0	42
Inpt Cond Drug Abuse, Elixhauser	Inpatient	Condition	0.11 (0.89)	0	24
Inpt Cond DYS	Inpatient	Condition	0.13 (0.41)	0	53
Inpt Cond ETOH	Inpatient	Condition	1.64 (3.13)	0	301
Inpt Cond Etoh Abuse	Inpatient	Condition	0.73 (2.23)	0	178
Inpt Cond FED, Elixhauser	Inpatient	Condition	0.88 (1.95)	0	249
Inpt Cond Gastroparesis	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond GI	Inpatient	Condition	0.31 (1.14)	0	69
Inpt Cond GIPerf	Inpatient	Condition	0.01 (0.10)	0	2
Inpt Cond GLOM	Inpatient	Condition	3.00 (0.60)	0	3
Inpt Cond GLOMNephEx	Inpatient	Condition	3.00 (0.60)	0	3
Inpt Cond HBC	Inpatient	Condition	0.91 (1.99)	0	215
Inpt Cond HCC	Inpatient	Condition	0.23 (1.19)	0	44
Inpt Cond HE	Inpatient	Condition	0.50 (1.46)	0	137
Inpt Cond Headache	Inpatient	Condition	0.01 (0.11)	0	3
Inpt Cond HEMOCH	Inpatient	Condition	0.03 (0.63)	0	2
Inpt Cond Hemorrhoid	Inpatient	Condition	0.03 (0.20)	0	12
Inpt Cond HEP	Inpatient	Condition	1.15 (2.20)	0	261
Inpt Cond Hep B C	Inpatient	Condition	0.91 (1.99)	0	215

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond HF, Elixhauser	Inpatient	Condition	0.53 (1.02)	0	178
Inpt Cond HFC, Elixhauser	Inpatient	Condition	0.14 (0.56)	0	56
Inpt Cond HIV	Inpatient	Condition	0.06 (0.48)	0	11
Inpt Cond HOSP	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond HOTN	Inpatient	Condition	0.22 (0.62)	0	77
Inpt Cond HRS	Inpatient	Condition	0.69 (1.37)	0	233
Inpt Cond HTD, Elixhauser	Inpatient	Condition	0.08 (0.31)	0	37
Inpt Cond HTN	Inpatient	Condition	0.67 (1.14)	0	221
Inpt Cond Hydronephrosis	Inpatient	Condition	4.00 (0.79)	0	4
Inpt Cond HYPC	Inpatient	Condition	0.02 (0.24)	0	6
Inpt Cond Hyperkalemia	Inpatient	Condition	0.17 (0.79)	0	59
Inpt Cond Hyperparathyroidism	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond IABP	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond IBS	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond iDIAL	Inpatient	Condition	0.15 (0.72)	0	38
Inpt Cond IVD	Inpatient	Condition	0.14 (0.43)	0	59
Inpt Cond JAUD	Inpatient	Condition	0.19 (0.77)	0	56
Inpt Cond LD, Elixhauser	Inpatient	Condition	3.40 (4.03)	0	458
Inpt Cond LIV	Inpatient	Condition	2.60 (3.08)	0	446
Inpt Cond LKM	Inpatient	Condition	0.01 (0.17)	0	3
Inpt Cond LM, Elixhauser	Inpatient	Condition	0.01 (0.20)	0	2

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond LUP	Inpatient	Condition	0.01 (0.15)	0	5
Inpt Cond LvrTx	Inpatient	Condition	0.02 (0.40)	0	2
Inpt Cond MC, Elixhauser	Inpatient	Condition	0.05 (0.25)	0	20
Inpt Cond MECHVENT	Inpatient	Condition	0.17 (0.49)	0	65
Inpt Cond MI	Inpatient	Condition	0.03 (0.25)	0	9
Inpt Cond MVR	Inpatient	Condition	0.01 (0.13)	0	5
Inpt Cond Myopathies	Inpatient	Condition	9.00 (1.79)	0	9
Inpt Cond NAFLD	Inpatient	Condition	0.10 (0.42)	0	37
Inpt Cond NAS	Inpatient	Condition	0.06 (0.40)	0	20
Inpt Cond NephGLOM	Inpatient	Condition	0.01 (0.14)	0	2
Inpt Cond Nephrtmy	Inpatient	Condition	0.02 (0.17)	0	5
Inpt Cond NFXss	Inpatient	Condition	1.46 (2.91)	0	298
Inpt Cond NSTEMI	Inpatient	Condition	0.01 (0.13)	0	6
Inpt Cond OA	Inpatient	Condition	0.03 (0.28)	0	10
Inpt Cond Obesity	Inpatient	Condition	0.04 (0.26)	0	13
Inpt Cond OFss	Inpatient	Condition	2.33 (3.62)	0	416
Inpt Cond OLG	Inpatient	Condition	0.01 (0.19)	0	4
Inpt Cond OND, Elixhauser	Inpatient	Condition	0.04 (0.46)	0	10
Inpt Cond OrganTrans	Inpatient	Condition	0.02 (0.36)	0	2
Inpt Cond PALL	Inpatient	Condition	0.59 (1.88)	0	117
Inpt Cond Paracentesis	Inpatient	Condition	0.76 (1.17)	0	215

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond Paralysis, Elixhauser	Inpatient	Condition	0.01 (0.10)	0	2
Inpt Cond Parkinsons	Inpatient	Condition	0.02 (0.40)	0	1
Inpt Cond PCD, Elixhauser	Inpatient	Condition	0.05 (0.27)	0	19
Inpt Cond PCI	Inpatient	Condition	3.00 (0.60)	0	3
Inpt Cond PCR	Inpatient	Condition	0.08 (0.55)	0	22
Inpt Cond PFT	Inpatient	Condition	0.04 (0.22)	0	15
Inpt Cond Plegia CC	Inpatient	Condition	0.01 (0.10)	0	2
Inpt Cond PUD CC	Inpatient	Condition	0.06 (0.32)	0	18
Inpt Cond PVD	Inpatient	Condition	0.10 (0.86)	0	18
Inpt Cond PVD, Elixhauser	Inpatient	Condition	0.16 (1.06)	0	31
Inpt Cond PY, Elixhauser	Inpatient	Condition	0.24 (2.79)	0	25
Inpt Cond RA	Inpatient	Condition	0.01 (0.11)	0	3
Inpt Cond RA, Elixhauser	Inpatient	Condition	0.01 (0.17)	0	3
Inpt Cond RenalTrans	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond RF, Elixhauser	Inpatient	Condition	0.52 (1.53)	0	129
Inpt Cond RHBD	Inpatient	Condition	0.01 (0.11)	0	3
Inpt Cond Rheum CC	Inpatient	Condition	0.01 (0.17)	0	3
Inpt Cond RUD, Elixhauser	Inpatient	Condition	0.04 (0.26)	0	12
Inpt Cond SBP	Inpatient	Condition	0.17 (0.63)	0	56
Inpt Cond SIRS	Inpatient	Condition	0.17 (0.61)	0	77
Inpt Cond SPS	Inpatient	Condition	0.19 (1.26)	0	53

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Cond ST, Elixhauser	Inpatient	Condition	0.45 (1.81)	0	82
Inpt Cond StableAngina	Inpatient	Condition	0.01 (0.17)	0	3
Inpt Cond STEMI	Inpatient	Condition	0.02 (0.19)	0	5
Inpt Cond STROKE	Inpatient	Condition	0.07 (0.75)	0	13
Inpt Cond TB	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond TIA	Inpatient	Condition	1.00 (0.20)	0	1
Inpt Cond TIPS	Inpatient	Condition	0.02 (0.15)	0	6
Inpt Cond TOBC	Inpatient	Condition	0.23 (0.87)	0	81
Inpt Cond TRAU	Inpatient	Condition	0.02 (0.24)	0	4
Inpt Cond UNAN	Inpatient	Condition	0.02 (0.19)	0	7
Inpt Cond UrinaryObst	Inpatient	Condition	0.02 (0.16)	0	7
Inpt Cond Valvular, Elixhauser	Inpatient	Condition	0.10 (0.46)	0	31
Inpt Cond Varices	Inpatient	Condition	0.23 (0.71)	0	73
Inpt Cond VascProc	Inpatient	Condition	0.37 (0.71)	0	136
Inpt Cond VascSurg	Inpatient	Condition	0.34 (0.67)	0	129
Inpt Cond VHD	Inpatient	Condition	0.05 (0.29)	0	16
Inpt Cond VLP	Inpatient	Condition	0.13 (0.52)	0	39
Inpt Cond VMT	Inpatient	Condition	0.04 (0.23)	0	19
Inpt Cond WL, Elixhauser	Inpatient	Condition	0.09 (0.30)	0	45
X3DPreAdmitProc_Paracentesis	Inpatient	Condition	8.00 (1.59)	0	8
Race: 0 UNKNOWN	Inpatient	Demographic	38.00 (7.54)	0	NA

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Race: 1 WHITE	Inpatient	Demographic	355.00 (70.44)	0	NA
Race: 2 BLACK	Inpatient	Demographic	93.00 (18.45)	0	NA
Race: 3 ASIAN-HAWAIIAN-PACIFIC ISLANDER	Inpatient	Demographic	9.00 (1.79)	0	NA
Race: 4 AMERICAN INDIAN-ALASKAN NATIVE	Inpatient	Demographic	9.00 (1.79)	0	NA
Age	Inpatient	Demographic	61.55 (9.35)	0	504
Gender	Inpatient	Demographic	500.00 (99.21)	0	NA
AdmissionToPostAdmDifference	Inpatient	Laboratory	-0.73 (1.79)	7.94	187
Inpt Lab Avg Alb	Inpatient	Laboratory	2.55 (0.68)	2.98	504
Inpt Lab Avg AlkPhos	Inpatient	Laboratory	147.90 (107.48)	3.97	504
Inpt Lab Avg ALT	Inpatient	Laboratory	80.10 (153.46)	3.77	504
Inpt Lab Avg AST	Inpatient	Laboratory	153.80 (304.61)	5.56	504
Inpt Lab Avg BilirubinD	Inpatient	Laboratory	3.56 (6.26)	38.1	411
Inpt Lab Avg BilirubinT	Inpatient	Laboratory	7.86 (9.67)	3.77	504
Inpt Lab Avg BNP	Inpatient	Laboratory	6040000000.00 (96949181037.19)	68.06	504
Inpt Lab Avg BS	Inpatient	Laboratory	132.10 (47.11)	0.4	504
Inpt Lab Avg BUN	Inpatient	Laboratory	41.20 (20.96)	6.15	504
Inpt Lab Avg CA	Inpatient	Laboratory	8.27 (0.94)	1.59	504
Inpt Lab Avg CK	Inpatient	Laboratory	27560.00 (581533.94)	58.33	504
Inpt Lab Avg CL	Inpatient	Laboratory	102.60 (6.49)	0	504
Inpt Lab Avg FeNa	Inpatient	Laboratory	2.80 (26.09)	40.08	504

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Lab Avg HCO3	Inpatient	Laboratory	22.51 (4.20)	0	504
Inpt Lab Avg HCT	Inpatient	Laboratory	31.55 (5.47)	0	504
Inpt Lab Avg HGB	Inpatient	Laboratory	10.80 (2.16)	3.57	504
Inpt Lab Avg INR	Inpatient	Laboratory	1.83 (0.69)	8.33	504
Inpt Lab Avg MCH	Inpatient	Laboratory	32.85 (3.39)	0.4	504
Inpt Lab Avg MCHC	Inpatient	Laboratory	34.08 (1.16)	0.2	504
Inpt Lab Avg MCV	Inpatient	Laboratory	96.31 (8.45)	0.2	504
Inpt Lab Avg NA	Inpatient	Laboratory	134.50 (5.58)	0.6	504
Inpt Lab Avg PLT	Inpatient	Laboratory	121.00 (75.28)	0.2	504
Inpt Lab Avg PT	Inpatient	Laboratory	20.53 (7.56)	11.31	504
Inpt Lab Avg PTT	Inpatient	Laboratory	42.18 (13.46)	20.04	504
Inpt Lab Avg TropI	Inpatient	Laboratory	0.44 (3.18)	57.54	400
Inpt Lab Avg UrineNa	Inpatient	Laboratory	28.86 (26.68)	32.94	504
Inpt Lab Avg WBC	Inpatient	Laboratory	10.14 (7.04)	0.2	504
Inpt Lab Max Alb	Inpatient	Laboratory	3.01 (0.85)	2.98	504
Inpt Lab Max AlkPhos	Inpatient	Laboratory	188.20 (139.48)	3.97	504
Inpt Lab Max ALT	Inpatient	Laboratory	146.70 (389.40)	3.77	504
Inpt Lab Max AST	Inpatient	Laboratory	334.90 (1078.05)	5.56	504
Inpt Lab Max BilirubinD	Inpatient	Laboratory	4.30 (7.48)	38.1	410
Inpt Lab Max BilirubinT	Inpatient	Laboratory	10.21 (12.10)	3.77	504

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Lab Max BNP	Inpatient	Laboratory	5017000000.00 (84550541645.45)	68.06	504
Inpt Lab Max BS	Inpatient	Laboratory	225.30 (144.24)	0.4	504
Inpt Lab Max BUN	Inpatient	Laboratory	61.69 (39.84)	6.15	504
Inpt Lab Max CA	Inpatient	Laboratory	8.96 (0.91)	1.59	504
Inpt Lab Max CK	Inpatient	Laboratory	50190.00 (515634.62)	58.33	504
Inpt Lab Max CL	Inpatient	Laboratory	108.00 (7.94)	0	504
Inpt Lab Max FeNa	Inpatient	Laboratory	3.45 (24.64)	40.08	504
Inpt Lab Max HCO3	Inpatient	Laboratory	26.45 (4.98)	0	504
Inpt Lab Max HCT	Inpatient	Laboratory	36.29 (5.70)	0	504
Inpt Lab Max HGB	Inpatient	Laboratory	12.53 (4.19)	3.57	504
Inpt Lab Max INR	Inpatient	Laboratory	2.36 (1.70)	8.33	504
Inpt Lab Max MCH	Inpatient	Laboratory	33.74 (3.59)	0.4	504
Inpt Lab Max MCHC	Inpatient	Laboratory	34.90 (1.23)	0.2	504
Inpt Lab Max MCV	Inpatient	Laboratory	98.71 (9.02)	0.2	504
Inpt Lab Max NA	Inpatient	Laboratory	139.40 (6.53)	0.6	504
Inpt Lab Max PLT	Inpatient	Laboratory	171.00 (97.60)	0.2	504
Inpt Lab Max PT	Inpatient	Laboratory	25.91 (18.83)	11.31	504
Inpt Lab Max PTT	Inpatient	Laboratory	54.83 (37.02)	20.04	504
Inpt Lab Max TropI	Inpatient	Laboratory	0.82 (6.75)	57.54	396
Inpt Lab Max UrineNa	Inpatient	Laboratory	38.25 (40.16)	32.94	504

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Lab Max WBC	Inpatient	Laboratory	15.08 (10.52)	0.2	504
Inpt Lab Min Alb	Inpatient	Laboratory	2.17 (0.69)	2.98	504
Inpt Lab Min AlkPhos	Inpatient	Laboratory	120.40 (90.32)	3.97	504
Inpt Lab Min ALT	Inpatient	Laboratory	47.70 (76.25)	3.77	504
Inpt Lab Min AST	Inpatient	Laboratory	83.30 (114.25)	5.56	504
Inpt Lab Min BilirubinD	Inpatient	Laboratory	2.90 (5.52)	38.1	409
Inpt Lab Min BilirubinT	Inpatient	Laboratory	5.89 (7.93)	3.77	504
Inpt Lab Min BNP	Inpatient	Laboratory	399900000000.00 (8803822474195.57)	68.06	504
Inpt Lab Min BS	Inpatient	Laboratory	79.34 (30.20)	0.4	504
Inpt Lab Min BUN	Inpatient	Laboratory	26.37 (18.74)	6.15	504
Inpt Lab Min CA	Inpatient	Laboratory	7.64 (1.28)	1.59	504
Inpt Lab Min CK	Inpatient	Laboratory	2578000.00 (52985939.99)	58.33	504
Inpt Lab Min CL	Inpatient	Laboratory	96.58 (7.04)	0	504
Inpt Lab Min FeNa	Inpatient	Laboratory	1.47 (21.43)	40.08	452
Inpt Lab Min HCO3	Inpatient	Laboratory	18.45 (5.10)	0	504
Inpt Lab Min HCT	Inpatient	Laboratory	27.31 (6.72)	0	504
Inpt Lab Min HGB	Inpatient	Laboratory	9.31 (2.30)	3.57	504
Inpt Lab Min INR	Inpatient	Laboratory	1.54 (0.44)	8.33	504
Inpt Lab Min MCH	Inpatient	Laboratory	32.05 (3.39)	0.4	504
Inpt Lab Min MCHC	Inpatient	Laboratory	33.24 (1.32)	0.2	504

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Lab Min MCV	Inpatient	Laboratory	94.25 (8.32)	0.2	504
Inpt Lab Min NA	Inpatient	Laboratory	129.40 (6.36)	0.6	504
Inpt Lab Min PLT	Inpatient	Laboratory	93.29 (67.59)	0.2	502
Inpt Lab Min PT	Inpatient	Laboratory	17.54 (4.63)	11.31	504
Inpt Lab Min PTT	Inpatient	Laboratory	36.05 (8.77)	20.04	504
Inpt Lab Min TropI	Inpatient	Laboratory	1667.00 (37272.08)	57.54	405
Inpt Lab Min UrineNa	Inpatient	Laboratory	26.40 (49.23)	32.94	504
Inpt Lab Min WBC	Inpatient	Laboratory	6.82 (5.12)	0.2	504
Inpt Lab ProteinStick High	Inpatient	Laboratory	1.91 (7.21)	13.69	366
Inpt Lab ProteinStick Median	Inpatient	Laboratory	1.18 (2.04)	13.69	350
MaxAdmissionCreatinine	Inpatient	Laboratory	2.36 (1.55)	0.2	504
MaxCreatinineChange	Inpatient	Laboratory	2.23 (1.92)	0	504
MaxInpatientCreatinine	Inpatient	Laboratory	3.34 (1.90)	0	504
MeanInpatientCreatinine	Inpatient	Laboratory	2.24 (1.37)	0	504
PeakPostAdmToDisCreatinine	Inpatient	Laboratory	3.07 (2.06)	7.74	504
PreAdmitLab_ProteinStick_High	Inpatient	Laboratory	1.20 (1.11)	21.83	351
PreAdmitLab_ProteinStick_Median	Inpatient	Laboratory	0.68 (1.07)	21.83	196
PreAdmMeanCreatinine	Inpatient	Laboratory	1.10 (0.47)	0	504
Home Med Lactulose	Home	Medication	91.00 (18.06)	0	91
Home Med Rifaximin	Home	Medication	21.00 (4.17)	0	21
Inpt Med Cyclosporine	Inpatient	Medication	1.00 (0.20)	0	1
Inpt Med Dobutamine	Inpatient	Medication	8.00 (1.59)	0	8

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Med Dopamine	Inpatient	Medication	33.00 (6.55)	0	33
Inpt Med Human Albumin	Inpatient	Medication	246.00 (48.81)	0	246
Inpt Med Lactulose	Inpatient	Medication	299.00 (59.33)	0	299
Inpt Med Midodrine	Inpatient	Medication	127.00 (25.20)	0	127
Inpt Med Nacetylcysteine	Inpatient	Medication	40.00 (7.94)	0	40
Inpt Med Norepinephrine	Inpatient	Medication	66.00 (13.10)	0	66
Inpt Med Octreotide	Inpatient	Medication	164.00 (32.54)	0	164
Inpt Med Phenylephrine	Inpatient	Medication	20.00 (3.97)	0	20
Inpt Med Rifaximin	Inpatient	Medication	71.00 (14.09)	0	71
Inpt Med Septra	Inpatient	Medication	20.00 (3.97)	0	20
Inpt Med Trimethoprim	Inpatient	Medication	16.00 (3.17)	0	16
Inpt Med Vancomycin	Inpatient	Medication	205.00 (40.67)	0	205
Inpt Med Vasopressin	Inpatient	Medication	33.00 (6.55)	0	33
Octreotide_and_Midodrine	Inpatient	Medication	119.00 (23.61)	0	119
Octreotide_and_Norepinephrine	Inpatient	Medication	31.00 (6.15)	0	31
Octreotide_and_Vasopressin	Inpatient	Medication	15.00 (2.98)	0	15
Any_IV_Vasopressor	Inpatient	Medication Class	86.00 (17.06)	0	86
Home Med Ace Inhibitors	Home	Medication Class	96.00 (19.05)	0	96
Home Med Alcohol Deterrents	Home	Medication Class	1.00 (0.20)	0	1
Home Med Alpha Blockers Related	Home	Medication Class	45.00 (8.93)	0	45
Home Med Aminoglycosides	Home	Medication Class	4.00 (0.79)	0	4

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Med Amphetamine Like Stimulants	Home	Medication Class	1.00 (0.20)	0	1
Home Med Analgesics	Home	Medication Class	155.00 (30.75)	0	155
Home Med Analgesics Topical	Home	Medication Class	9.00 (1.79)	0	9
Home Med Angiotensin Ii Inhibitor	Home	Medication Class	21.00 (4.17)	0	21
Home Med Anti Infective Topical	Home	Medication Class	27.00 (5.36)	0	27
Home Med Anti Infective Topical Other	Home	Medication Class	2.00 (0.40)	0	2
Home Med Anti Infectives Other	Home	Medication Class	30.00 (5.95)	0	30
Home Med Anti Inflammatory Topical	Home	Medication Class	23.00 (4.56)	0	23
Home Med Antiacne Agents	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antiacne Agents Topical	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antianginals	Home	Medication Class	17.00 (3.37)	0	17
Home Med Antiarrhythmics	Home	Medication Class	4.00 (0.79)	0	4
Home Med Antibacterial Topical	Home	Medication Class	9.00 (1.79)	0	9
Home Med Anticoagulants	Home	Medication Class	16.00 (3.17)	0	16
Home Med Anticonvulsants	Home	Medication Class	44.00 (8.73)	0	44
Home Med Antidepressants	Home	Medication Class	119.00 (23.61)	0	119
Home Med Antidepressants Other	Home	Medication Class	114.00 (22.62)	0	114
Home Med Antidotes Deterrents And Poison Control	Home	Medication Class	11.00 (2.18)	0	11
Home Med Antidotes Deterrents And Poison Control Exchange Resins	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antidotes Deterrents Other	Home	Medication Class	7.00 (1.39)	0	7

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Med Antifungal Topical	Home	Medication Class	18.00 (3.57)	0	18
Home Med Antifungals	Home	Medication Class	4.00 (0.79)	0	4
Home Med Antihistamines	Home	Medication Class	51.00 (10.12)	0	51
Home Med Antihistamines Alkylamine	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antihistamines Ethanolamine	Home	Medication Class	7.00 (1.39)	0	7
Home Med Antihistamines Other	Home	Medication Class	21.00 (4.17)	0	21
Home Med Antihistamines Phenothiazine	Home	Medication Class	9.00 (1.79)	0	9
Home Med Antihistamines Piperazine	Home	Medication Class	14.00 (2.78)	0	14
Home Med Antihistamines Piperidine	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antihypertensive Combinations	Home	Medication Class	13.00 (2.58)	0	13
Home Med Antihypertensives Other	Home	Medication Class	11.00 (2.18)	0	11
Home Med Antilipemic Agents	Home	Medication Class	58.00 (11.51)	0	58
Home Med Antimalarials	Home	Medication Class	4.00 (0.79)	0	4
Home Med Antimicrobials	Home	Medication Class	97.00 (19.25)	0	97
Home Med Antimigraine Agents	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antineoplastic Hormones	Home	Medication Class	3.00 (0.60)	0	3
Home Med Antineoplastic Other	Home	Medication Class	7.00 (1.39)	0	7
Home Med Antineoplastics	Home	Medication Class	11.00 (2.18)	0	11
Home Med Antineoplastics Antimetabolites	Home	Medication Class	1.00 (0.20)	0	1
Home Med Antiparasitics	Home	Medication Class	4.00 (0.79)	0	4

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Med Antiparkinson Agents	Home	Medication Class	3.00 (0.60)	0	3
Home Med Antiprotozoals	Home	Medication Class	4.00 (0.79)	0	4
Home Med Antipsoriatic	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antipsoriatics Topical	Home	Medication Class	2.00 (0.40)	0	2
Home Med Antipsychotics	Home	Medication Class	28.00 (5.56)	0	28
Home Med Antipsychotics Other	Home	Medication Class	27.00 (5.36)	0	27
Home Med Antituberculars	Home	Medication Class	1.00 (0.20)	0	1
Home Med Antivirals	Home	Medication Class	16.00 (3.17)	0	16
Home Med Autonomic Medications	Home	Medication Class	7.00 (1.39)	0	7
Home Med Benzodiazepine Derivative Sedatives Hypnotics	Home	Medication Class	40.00 (7.94)	0	40
Home Med Beta Blockers Related	Home	Medication Class	189.00 (37.50)	0	189
Home Med Blood Formation Products	Home	Medication Class	2.00 (0.40)	0	2
Home Med Blood Products Modifiers Volume Expanders	Home	Medication Class	30.00 (5.95)	0	30
Home Med Calcium Channel Blockers	Home	Medication Class	59.00 (11.71)	0	59
Home Med Carbonic Anhydrase Inhibitor Diuretics	Home	Medication Class	1.00 (0.20)	0	1
Home Med Cardiovascular Agents Other	Home	Medication Class	7.00 (1.39)	0	7
Home Med Cardiovascular Medications	Home	Medication Class	352.00 (69.84)	0	352
Home Med Central Nervous System Medications	Home	Medication Class	232.00 (46.03)	0	232
Home Med Cephalosporin 1st Generation	Home	Medication Class	1.00 (0.20)	0	1

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Med Cephalosporin 3rd Generation	Home	Medication Class	1.00 (0.20)	0	1
Home Med Cns Medications Other	Home	Medication Class	6.00 (1.19)	0	6
Home Med Cns Stimulants	Home	Medication Class	1.00 (0.20)	0	1
Home Med Dermatological Agents	Home	Medication Class	67.00 (13.29)	0	67
Home Med Dermatologicals Topical Other	Home	Medication Class	7.00 (1.39)	0	7
Home Med Digitalis Glycosides	Home	Medication Class	10.00 (1.98)	0	10
Home Med Diuretics	Home	Medication Class	257.00 (50.99)	0	257
Home Med Emollients	Home	Medication Class	16.00 (3.17)	0	16
Home Med Erythromycins Macrolides	Home	Medication Class	6.00 (1.19)	0	6
Home Med Heavy Metal Antagonists	Home	Medication Class	1.00 (0.20)	0	1
Home Med Keratolytics Caustics Topical	Home	Medication Class	2.00 (0.40)	0	2
Home Med Lithium Salts	Home	Medication Class	1.00 (0.20)	0	1
Home Med Local Anesthetics Topical	Home	Medication Class	3.00 (0.60)	0	3
Home Med Loop Diuretics	Home	Medication Class	202.00 (40.08)	0	202
Home Med Nitrofurans Antimicrobials	Home	Medication Class	1.00 (0.20)	0	1
Home Med Non Opioid Analgesics	Home	Medication Class	66.00 (13.10)	0	66
Home Med Opioid Analgesics	Home	Medication Class	112.00 (22.22)	0	112
Home Med Opioid Antagonist Analgesics	Home	Medication Class	1.00 (0.20)	0	1
Home Med Parasympatholytics	Home	Medication Class	4.00 (0.79)	0	4
Home Med Parasympathomimetics Cholinergics	Home	Medication Class	3.00 (0.60)	0	3

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Home Med Penicillinase Resistant Penicillins	Home	Medication Class	1.00 (0.20)	0	1
Home Med Penicillins Amino Derivatives	Home	Medication Class	11.00 (2.18)	0	11
Home Med Penicillins And Beta Lactam Antimicrobials	Home	Medication Class	13.00 (2.58)	0	13
Home Med Phenothiazine Related Antipsychotics	Home	Medication Class	1.00 (0.20)	0	1
Home Med Platelet Aggregation Inhibitors	Home	Medication Class	13.00 (2.58)	0	13
Home Med Potassium Sparing Combinations Diuretics	Home	Medication Class	184.00 (36.51)	0	184
Home Med Quinolones	Home	Medication Class	46.00 (9.13)	0	46
Home Med Sedatives Hypnotics Other	Home	Medication Class	21.00 (4.17)	0	21
Home Med Sedatives Hypontics	Home	Medication Class	58.00 (11.51)	0	58
Home Med Sulfonamide Related Antimicrobials	Home	Medication Class	7.00 (1.39)	0	7
Home Med Sun Protectants Screens Topical	Home	Medication Class	2.00 (0.40)	0	2
Home Med Tetracyclines	Home	Medication Class	4.00 (0.79)	0	4
Home Med Thiazides Related Diuretics	Home	Medication Class	33.00 (6.55)	0	33
Home Med Tricyclic Antidepressants	Home	Medication Class	9.00 (1.79)	0	9
Inpt Med Class Ace	Inpatient	Medication Class	81.00 (16.07)	0	81
Inpt Med Class Aminoglycosides	Inpatient	Medication Class	29.00 (5.75)	0	29
Inpt Med Class Arb	Inpatient	Medication Class	10.00 (1.98)	0	10
Inpt Med Class Benzodiazepines	Inpatient	Medication Class	205.00 (40.67)	0	205
Inpt Med Class Betablockers	Inpatient	Medication Class	301.00 (59.72)	0	301
Inpt Med Class Fluoroquinolones	Inpatient	Medication Class	141.00 (27.98)	0	141

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Med Class Glucocorticoids	Inpatient	Medication Class	100.00 (19.84)	0	100
Inpt Med Class Insulin	Inpatient	Medication Class	216.00 (42.86)	0	216
Inpt Med Class Ksparingdiuretic	Inpatient	Medication Class	191.00 (37.90)	0	191
Inpt Med Class Nsaids	Inpatient	Medication Class	43.00 (8.53)	0	43
Inpt Med Class Opioids	Inpatient	Medication Class	418.00 (82.94)	0	418
Inpt Med Class Statins	Inpatient	Medication Class	57.00 (11.31)	0	57
Inpt Med Va Class Ace Inhibitors	Inpatient	Medication Class	81.00 (16.07)	0	81
Inpt Med Va Class Alcohol Deterrents	Inpatient	Medication Class	3.00 (0.60)	0	3
Inpt Med Va Class Alpha Blockers Related	Inpatient	Medication Class	66.00 (13.10)	0	66
Inpt Med Va Class Aminoglycosides	Inpatient	Medication Class	26.00 (5.16)	0	26
Inpt Med Va Class Amphetamine Like Stimulants	Inpatient	Medication Class	2.00 (0.40)	0	2
Inpt Med Va Class Analgesics Topical	Inpatient	Medication Class	9.00 (1.79)	0	9
Inpt Med Va Class Angiotensin Ii Inhibitor	Inpatient	Medication Class	10.00 (1.98)	0	10
Inpt Med Va Class Anti Infective Anti Inflammatory Combinations Topical	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Anti Infective Topical Other	Inpatient	Medication Class	8.00 (1.59)	0	8
Inpt Med Va Class Anti Infectives Other	Inpatient	Medication Class	274.00 (54.37)	0	274
Inpt Med Va Class Anti Inflammatory Topical	Inpatient	Medication Class	28.00 (5.56)	0	28
Inpt Med Va Class Antiacne Agents Topical	Inpatient	Medication Class	2.00 (0.40)	0	2
Inpt Med Va Class Antianginals	Inpatient	Medication Class	26.00 (5.16)	0	26
Inpt Med Va Class Antiarrhythmics	Inpatient	Medication Class	19.00 (3.77)	0	19

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Med Va Class Antibacterial Topical	Inpatient	Medication Class	28.00 (5.56)	0	28
Inpt Med Va Class Anticoagulants	Inpatient	Medication Class	233.00 (46.23)	0	233
Inpt Med Va Class Anticonvulsants	Inpatient	Medication Class	76.00 (15.08)	0	76
Inpt Med Va Class Antidepressants Other	Inpatient	Medication Class	151.00 (29.96)	0	151
Inpt Med Va Class Antidotes Deterrents And Poison Control Exchange Resins	Inpatient	Medication Class	110.00 (21.83)	0	110
Inpt Med Va Class Antidotes Deterrents Other	Inpatient	Medication Class	17.00 (3.37)	0	17
Inpt Med Va Class Antifungal Topical	Inpatient	Medication Class	64.00 (12.70)	0	64
Inpt Med Va Class Antifungals	Inpatient	Medication Class	69.00 (13.69)	0	69
Inpt Med Va Class Antihemorrhagics	Inpatient	Medication Class	4.00 (0.79)	0	4
Inpt Med Va Class Antihistamines Ethanolamine	Inpatient	Medication Class	86.00 (17.06)	0	86
Inpt Med Va Class Antihistamines Other	Inpatient	Medication Class	12.00 (2.38)	0	12
Inpt Med Va Class Antihistamines Phenothiazine	Inpatient	Medication Class	43.00 (8.53)	0	43
Inpt Med Va Class Antihistamines Piperazine	Inpatient	Medication Class	28.00 (5.56)	0	28
Inpt Med Va Class Antihistamines Piperidine	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Antihypertensive Combinations	Inpatient	Medication Class	8.00 (1.59)	0	8
Inpt Med Va Class Antihypertensives Other	Inpatient	Medication Class	45.00 (8.93)	0	45
Inpt Med Va Class Antilipemic Agents	Inpatient	Medication Class	67.00 (13.29)	0	67
Inpt Med Va Class Antimalarials	Inpatient	Medication Class	4.00 (0.79)	0	4
Inpt Med Va Class Antineoplastic Other	Inpatient	Medication Class	4.00 (0.79)	0	4

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Med Va Class Antiparkinson Agents	Inpatient	Medication Class	3.00 (0.60)	0	3
Inpt Med Va Class Antiprotozoals Other	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Antipsoriatics Topical	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Antipsychotics Other	Inpatient	Medication Class	84.00 (16.67)	0	84
Inpt Med Va Class Antituberculars	Inpatient	Medication Class	2.00 (0.40)	0	2
Inpt Med Va Class Antivertigo Agents	Inpatient	Medication Class	9.00 (1.79)	0	9
Inpt Med Va Class Antivirals	Inpatient	Medication Class	15.00 (2.98)	0	15
Inpt Med Va Class Barbituric Acid Derivative Sedatives Hypnotics	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Benzodiazepine Derivative Sedatives Hypnotics	Inpatient	Medication Class	205.00 (40.67)	0	205
Inpt Med Va Class Beta Blockers Related	Inpatient	Medication Class	301.00 (59.72)	0	301
Inpt Med Va Class Blood Derivatives	Inpatient	Medication Class	247.00 (49.01)	0	247
Inpt Med Va Class Blood Formation Products	Inpatient	Medication Class	19.00 (3.77)	0	19
Inpt Med Va Class Calcium Channel Blockers	Inpatient	Medication Class	81.00 (16.07)	0	81
Inpt Med Va Class Carbonic Anhydrase Inhibitor Diuretics	Inpatient	Medication Class	4.00 (0.79)	0	4
Inpt Med Va Class Cardiovascular Agents Other	Inpatient	Medication Class	139.00 (27.58)	0	139
Inpt Med Va Class Cephalosporin 1st Generation	Inpatient	Medication Class	31.00 (6.15)	0	31
Inpt Med Va Class Cephalosporin 2nd Generation	Inpatient	Medication Class	6.00 (1.19)	0	6

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Med Va Class Cephalosporin 3rd Generation	Inpatient	Medication Class	171.00 (33.93)	0	171
Inpt Med Va Class Cns Medications Other	Inpatient	Medication Class	7.00 (1.39)	0	7
Inpt Med Va Class Dermatologicals Topical Other	Inpatient	Medication Class	46.00 (9.13)	0	46
Inpt Med Va Class Digitalis Glycosides	Inpatient	Medication Class	25.00 (4.96)	0	25
Inpt Med Va Class Emollients	Inpatient	Medication Class	37.00 (7.34)	0	37
Inpt Med Va Class Erythromycins Macrolides	Inpatient	Medication Class	45.00 (8.93)	0	45
Inpt Med Va Class Extended Spectrum Penicillins	Inpatient	Medication Class	171.00 (33.93)	0	171
Inpt Med Va Class Heavy Metal Antagonists	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Lincomycins	Inpatient	Medication Class	21.00 (4.17)	0	21
Inpt Med Va Class Lithium Salts	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Local Anesthetics Injection	Inpatient	Medication Class	22.00 (4.37)	0	22
Inpt Med Va Class Local Anesthetics Topical	Inpatient	Medication Class	13.00 (2.58)	0	13
Inpt Med Va Class Loop Diuretics	Inpatient	Medication Class	305.00 (60.52)	0	305
Inpt Med Va Class Nitrofurans Antimicrobials	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Non Opioid Analgesics	Inpatient	Medication Class	218.00 (43.25)	0	218
Inpt Med Va Class Opioid Analgesics	Inpatient	Medication Class	370.00 (73.41)	0	370
Inpt Med Va Class Opioid Antagonist Analgesics	Inpatient	Medication Class	13.00 (2.58)	0	13
Inpt Med Va Class Parasympatholytics	Inpatient	Medication Class	16.00 (3.17)	0	16
Inpt Med Va Class Parasympathomimetics Cholinergics	Inpatient	Medication Class	49.00 (9.72)	0	49

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Inpt Med Va Class Penicillin G Related Penicillins	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Penicillinase Resistant Penicillins	Inpatient	Medication Class	11.00 (2.18)	0	11
Inpt Med Va Class Penicillins Amino Derivatives	Inpatient	Medication Class	47.00 (9.33)	0	47
Inpt Med Va Class Phenothiazine Related Antipsychotics	Inpatient	Medication Class	6.00 (1.19)	0	6
Inpt Med Va Class Platelet Aggregation Inhibitors	Inpatient	Medication Class	16.00 (3.17)	0	16
Inpt Med Va Class Potassium Sparing Combinations Diuretics	Inpatient	Medication Class	191.00 (37.90)	0	191
Inpt Med Va Class Quinolones	Inpatient	Medication Class	201.00 (39.88)	0	201
Inpt Med Va Class Sedatives Hypnotics Other	Inpatient	Medication Class	42.00 (8.33)	0	42
Inpt Med Va Class Sedatives Hypontics	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Soaps Shampoos Soap Free Cleansers	Inpatient	Medication Class	6.00 (1.19)	0	6
Inpt Med Va Class Sulfonamide Related Antimicrobials	Inpatient	Medication Class	20.00 (3.97)	0	20
Inpt Med Va Class Sun Protectants Screens Topical	Inpatient	Medication Class	1.00 (0.20)	0	1
Inpt Med Va Class Tetracyclines	Inpatient	Medication Class	7.00 (1.39)	0	7
Inpt Med Va Class Thiazides Related Diuretics	Inpatient	Medication Class	38.00 (7.54)	0	38
Inpt Med Va Class Tricyclic Antidepressants	Inpatient	Medication Class	7.00 (1.39)	0	7
Inpt Med Va Class Volume Expanders	Inpatient	Medication Class	9.00 (1.79)	0	9
LOSHours	Inpatient	Misc	311.10 (736.92)	0	504
MELD	Inpatient	Misc	22.98 (8.16)	19.05	504

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
Num_paracentesis_90D	Inpatient	Misc	1.17 (0.76)	0	504
PalliativeConsult	Inpatient	Misc	181.00 (35.91)	0	181
n_Anuria	Inpatient	NLP	0.02 (0.19)	2.38	7
n_ARDS	Inpatient	NLP	0.04 (0.38)	2.38	7
n_Ascites	Inpatient	NLP	3.03 (5.00)	2.38	335
n_ATN	Inpatient	NLP	0.02 (0.45)	2.38	1
n_Casts	Inpatient	NLP	0.05 (0.41)	2.38	13
n_Dehydration	Inpatient	NLP	0.04 (0.30)	2.38	16
n_Edema	Inpatient	NLP	7.05 (6.90)	2.38	460
n_Glomerulonephritis	Inpatient	NLP	1.00 (0.20)	2.38	1
n_HE	Inpatient	NLP	0.35 (1.28)	2.38	78
n_Hematemesis	Inpatient	NLP	0.69 (1.69)	2.38	169
n_HRS	Inpatient	NLP	1.19 (2.61)	2.38	203
n_Hydronephrosis	Inpatient	NLP	0.98 (2.61)	2.38	144
n_Hypotension	Inpatient	NLP	0.31 (1.03)	2.38	76
n_Nephritis	Inpatient	NLP	0.11 (0.69)	2.38	18
n_Nephrotoxic	Inpatient	NLP	0.25 (1.24)	2.38	45
n_NSAIDS	Inpatient	NLP	0.44 (1.51)	2.38	99
n_NVD	Inpatient	NLP	6.02 (6.85)	2.38	417
n_Paracentesis	Inpatient	NLP	1.95 (3.87)	2.38	237
n_Peritonitis	Inpatient	NLP	0.26 (1.19)	2.38	50

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
n_Prerenal	Inpatient	NLP	0.00 (0.00)	2.38	0
n_RBCs	Inpatient	NLP	1.59 (4.21)	2.38	211
n_sepsis	Inpatient	NLP	0.95 (2.96)	2.38	119
n_Shock	Inpatient	NLP	0.30 (1.29)	2.38	50
n_SIRS	Inpatient	NLP	0.06 (0.38)	2.38	20
n_TubularCells	Inpatient	NLP	0.02 (0.26)	2.38	3
n_UrineSediment	Inpatient	NLP	0.05 (0.43)	2.38	9
p_Anuria	Inpatient	NLP	0.15 (0.77)	1.98	32
p_ARDS	Inpatient	NLP	0.57 (3.65)	1.98	26
p_Ascites	Inpatient	NLP	26.25 (36.76)	1.98	406
p_ATN	Inpatient	NLP	0.07 (0.53)	1.98	14
p_Casts	Inpatient	NLP	0.25 (1.07)	1.98	54
p_Dehydration	Inpatient	NLP	1.16 (3.91)	1.98	149
p_Edema	Inpatient	NLP	16.42 (20.32)	1.98	438
p_Glomerulonephritis	Inpatient	NLP	0.15 (1.38)	1.98	17
p_HE	Inpatient	NLP	5.34 (11.36)	1.98	245
p_Hematemesis	Inpatient	NLP	0.70 (3.20)	1.98	68
p_HRS	Inpatient	NLP	10.71 (16.78)	1.98	368
p_Hydronephrosis	Inpatient	NLP	0.19 (1.75)	1.98	30
p_Hypotension	Inpatient	NLP	5.43 (11.63)	1.98	304
p_Nephritis	Inpatient	NLP	1.01 (6.72)	1.98	55

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
p_Nephrotoxic	Inpatient	NLP	0.51 (1.81)	1.98	79
p_NSAIDS	Inpatient	NLP	0.66 (2.27)	1.98	97
p_NVD	Inpatient	NLP	15.56 (28.71)	1.98	398
p_Paracentesis	Inpatient	NLP	14.93 (22.78)	1.98	353
p_Peritonitis	Inpatient	NLP	1.24 (6.23)	1.98	104
p_Prerenal	Inpatient	NLP	0.00 (0.00)	1.98	0
p_RBCs	Inpatient	NLP	11.18 (14.12)	1.98	453
p_sepsis	Inpatient	NLP	3.68 (10.76)	1.98	172
p_Shock	Inpatient	NLP	3.18 (13.91)	1.98	91
p_SIRS	Inpatient	NLP	0.26 (1.33)	1.98	37
p_TubularCells	Inpatient	NLP	0.12 (0.75)	1.98	23
p_UrineSediment	Inpatient	NLP	0.31 (1.93)	1.98	43
FluidResponsive	Inpatient	Temporal	91.00 (18.06)	0	91
MaxInptToDisCreatDifference	Inpatient	Temporal	0.71 (1.66)	0	331
PostAlbuminSlope	Inpatient	Temporal	0.10 (0.24)	66.07	434
PostContrastSlope	Inpatient	Temporal	0.05 (0.19)	81.35	463
PostAnyFluidSlope	Inpatient	Temporal	0.01 (0.33)	0	129
PostHypotensionSlope	Inpatient	Temporal	0.01 (0.14)	89.48	18
PostInsultSlope	Inpatient	Temporal	0.08 (0.25)	0	441
PostIVFSlope	Inpatient	Temporal	0.01 (0.33)	48.21	102
PostVasopressorSlope	Inpatient	Temporal	0.16 (0.09)	93.65	497

Varname	Time	Class	Summary, n(%) or mean (SD)	% missing	N > 0
AvgInptDiastolic	Inpatient	Vitals	65.69 (9.52)	0.6	504
AvgInptMAP	Inpatient	Vitals	82.12 (10.65)	0.6	504
AvgInptPulse	Inpatient	Vitals	81.23 (13.04)	0.4	504
AvgInptResp	Inpatient	Vitals	19.37 (2.09)	0.6	504
AvgInptSystolic	Inpatient	Vitals	116.30 (16.69)	0.6	504
AvgInptTemp	Inpatient	Vitals	97.71 (0.77)	0.6	504
AvgInptWeight	Inpatient	Vitals	203.00 (49.40)	7.54	504
MaxInptDiastolic	Inpatient	Vitals	86.10 (14.14)	0.6	504
MaxInptMAP	Inpatient	Vitals	94.79 (13.35)	0.6	504
MaxInptPulse	Inpatient	Vitals	104.70 (21.75)	0.4	504
MaxInptResp	Inpatient	Vitals	26.80 (12.32)	0.6	504
MaxInptSystolic	Inpatient	Vitals	148.00 (25.10)	0.6	504
MaxInptTemp	Inpatient	Vitals	99.46 (1.31)	0.6	504
MaxInptWeight	Inpatient	Vitals	211.80 (51.43)	7.54	504
MinInptDiastolic	Inpatient	Vitals	47.10 (12.14)	0.6	501
MinInptMAP	Inpatient	Vitals	69.89 (12.02)	0.6	504
MinInptPulse	Inpatient	Vitals	60.81 (16.12)	0.4	499
MinInptResp	Inpatient	Vitals	14.73 (3.14)	0.6	502
MinInptSystolic	Inpatient	Vitals	89.17 (18.77)	0.6	501
MinInptTemp	Inpatient	Vitals	95.95 (1.49)	0.6	504
MinInptWeight	Inpatient	Vitals	192.60 (53.09)	7.54	503

Appendix Table A.2: Code definitions for co-morbid conditions and procedures used in the model based on International Classification of Diseases-Version 9, Current Procedural Terminology, and ICD Procedure Code.

Condition Abbreviation	Description	Codes
AA, Elixhauser	Alcohol Abuse	265.2, 291.[12356789], 303.[09], 305.0, 357.5, 425.5, 535.3, 571.[0123], 980.*, V11.3
AAA	Abdominal Aortic Aneurysm, Procedure	3480[02-5], 3482[56], 3483[012], 350[89][12], 3510[23], 33877, 441.4, 39.71
AbdomSurg	Abdominal Surgery, Procedure	3810[012], 38115, 3812[09], 4310[78], 4311[23678], 4312[1-4], 4330[05], 4331[02], 43320, 4334[01], 4335[012], 4336[01], 4350[012], 43510, 4362[012], 4363[1-489], 43659, 4384[023678], 4385[05], 4386[05], 43870, 43880, 43999, 44010, 4402[015], 4405[05], 4411[01], 4412[015], 44130, 4414[013-7], 4415[0-356], 44160, 4420[1-8], 4421[012], 4423[89], 44300, 4431[0246], 4432[02], 4434[056], 4460[2-5], 4462[056], 44640, 44650, 4466[01], 44680, 44799, 44820, 44850, 44900, 4495[05], 44960, 44899, 4511[0-469], 4512[0136], 45805, 45825, 4712[025], 47130, 47350, 4736[012], 47399, 4760[05], 4761[02], 47620, 47701, 4772[01], 4774[01], 4776[05], 4778[05], 47800, 48005, 4814[056], 4815[02-5], 48180, 48520, 48540,

		4854[57], 48556, 48999, 4900[02], 49010, 49020, 49062, 49040, 49060, 49085, 4920[01], 49220, 49999, 41.4[123], 41.5, 41.9[59], 42.1[0129], 42.4[012], 42.5[1-689], 42.6[1-689], 43.0, 43.5, 43.6, 43.7, 43.8[129], 43.9[19], 44.3[189], 44.4[012], 44.5, 44.6[123589], 44.99, 45.0[123], 45.6[123], 45.7[1-69], 45.8[123], 45.9[0-5], 46.0[1-4], 46.1[0134], 46.2[0-4], 46.3[19], 46.4[0-3], 46.5[012], 46.6[0-4], 46.7[1-69], 46.8[012], 46.9[349], 48.4[239], 48.5[0129], 48.6[1-59], 50.0, 50.2[2-69], 50.3, 50.4, 50.6[19], 51.0[34], 51.2[1-4], 52.22, 52.5[1239], 52.6, 52.7, 52.9[59], 54.1[129], 54.99
ACS	Acute Coronary Syndrome,Condition	410*, 411*
aDIAL	Acute Dialysis,Procedure	90935, 90937, 90945, 90947, 90999, V45.1, V56.0, V56.1, 39.95
AFIB	Atrial Fibrillation, Condition	427.3[12]
AFL	Alcoholic Fatty Liver	571.[013]
AIDS/HIV, Elixhauser	AIDS/HIV	04[234].*
ALD	Advanced Liver Disease,Condition	070.22, 070.23, 070.44, 456.0, 456.1, 456.20, 456.21, 571.2, 571.3, 571.5, 571.6, 572.[2348]
Amyloidosis	Amyloidosis, Diagnosis	277.3, 277.3[019]
AnalFisFist	Anal Fissures or Fistula, Diagnosis	565.1
Angina	Angina, Condition	413*, 411.1
ANMA	Anemia,Condition	280*, 281*, 282.01, 282.2*, 282.3*, 282.4*, 282.71, 282.8, 282.9,

		283.[019]*, 284*, 285*, 648.2*, 776.5*
ARF	Acute Renal Failure, Condition	584.[5-9], 669.3[0124]
ARRH	Arrhythmia,Condition	427*, 785.0, 785.1, 779.81, 426*, V45.0*, V53.3*, 746.86
Arrhythmias, Elixhauser	Cardiac Arrhythmias, Diagnosis	426.[079], 426.[1][023], 427.[012346789], 785.0, 996.0[14], V45.0, V53.3
Ascites	Ascites, Condition or Procedure (paracentesis)	4908[0-3], 789.5*, 54.91
ASP	Aspergillosis,Condition	117.3, 484.6, 518.6
Asthma	Asthma, Condition	493.*
ATN	Acute Tubular Necrosis, Condition	584.5
AUTOHEP	Autoimmune Hepatitis	571.42
AutoNeuropathy	Autonomic Neuropathy, Diagnosis	337.9
AZ_Cancer	Cancer,Condition	1[4-9][0-9]*, 20[0-8]*, 209.[0-3]*, 23[0-3]*
BilCirrhosis	Biliary Cirrhosis	571.6
BLA, Elixhauser	Blood Loss Anemia	280
BmTx	Bone Marrow Transplant, Procedure	3824[012], 996.8[58], V42.8[12], 41.0*
BowelPerf	BowelPerforation, Diagnosis	569.83
BURN	Burns,Condition	906.5, 906.6, 906.7, 906.8, 906.9, 906.9, 940*, 941*, 942*, 943*, 944*, 945*, 946*, 947*, 948.1*, 948.2*, 948.3*, 948.4*, 948.5*, 948.6*, 948.7*, 948.8*, 948.9*, 949*
CABG	CABG,Procedure	3351[012346789], 3352[123], 3353[3-6], V45.81, 414.04, 36.1*, 36.2*
CAD	Coronary Artery Disease,Condition	410.*, 411.*, 412.*, 413.*, 414.[02-9]*, V45.81, V45.82
Cancer	Cancer,Condition	1[4568]_.*, 17[012456789].*, 19[0124].*, 195.[012345678]*, 20[012345678].*, 238.6*

CANDI	Candidiasis,Condition	112*
CardiacArrest	Cardiac Arrest, Diagnosis	427.4[12], 427.5
CardSurg	Cardiac Surgery, Procedure	33020, 33120, 33130, 3314[01], 3323[678], 33243, 3325[13], 33261, 3330[05], 3331[05], 3332[012], 3333[025], 33404, 3341[4-7], 3347[68], 33496, 3340[01356], 3341[0-3], 3342[02567], 33430, 3346[03458], 3347[012579], 3350[0-6], 3351[0-46-9], 3352[1-3], 3353[03-6], 3354[25], 33572, 3360[0268], 3361[012579], 3364[157], 3368[148], 3369[0247], 3386[013], 3387[057], 3391[0567], 3366[05], 33670, 33702, 33710, 3372[02], 3373[02567], 3375[05], 3376[2467], 3377[014-9], 3378[0168], 3380[023], 3381[34], 33824, 3384[05], 3385[123], 3391[89], 3392[02], 33999, 35.*, 35.1[0-4], 35.2[0-8], 35.3[1-59], 35.5[134], 35.[67][0-3], 35.8[1-4], 35.9*, 36.03, 36.1[0-79], 36.2, 36.3[129], 36.9[19], 37.1[01], 37.3[12356], 37.49, 39.6[1-46]
CarotidDis	Carotid Disease, Condition	433.1, 38.12
Cath	Cardiac Catheterization, Procedure	9350[138], 9351[014], 9352[46789], 9353[01239], 9354[0235], 9355[56], 9356[12], 37.2[1-3], 88.5[2-7]
CathPCIALL	Cath/PCI ALL, Procedure	9350[138], 9351[014], 9352[46789],

		9353[01239], 9354[0235], 9355[56], 9356[12], 9297[34], 9298[0124], 9299[56], G029[01], 37.2[123], 88.5[2-7], 00.66, 36.0[125679]
CathPCICont	Cath/PCI w Contrast,Procedure	93508, 9351[014], 9352[46789], 9353[1239], 9354[035], 9355[56], 9297[34], 9298[0124], 9299[56], G029[01], 37.2[23], 88.5[2-7], 00.66, 36.0[125679]
CathPCInoCont	Cath/PCI NO Contrast, Procedure	9350[13], 93530, 93542, 9356[12], 37.21
CDVD	Cardiovascular Disease,Condition	3353[03456]*, 3351[0123456789]*, 3352[012358]*, 3353[03456]*, 9298[0124]*, 9299[5678]*, 3480[023456]*, 3525*, 3528*, 3530*, 3535[15]*, 3537[12]*, 35381, 354[5789]*, 355[468]*, 3555[168]*, 3557[01]*, 356[4567]*, 410.*, 411.*, 412.*, 413.*, 429.7*, 430.*, 431.*, 433._1, 435.*, 436.*, 434.0, 434.01, 434.1, 434.11, 434.9, 434.91
CGP, Elixhauser	Coagulopathy	286.*, 287.[1345]
CHEMO	Chemotherapy, Procedure	4180F, 9640[01289], 9641[0-7], 9650[012589], 9651[012], 99555, C895[345], G0292, G035[59], 99.25, 00.10
CHF	Congestive Heart Failure,Condition	398.91, 402.11, 404.01, 404.11, 404.91, 428*, 402.01, 402.91, 404.03, 404.13, 404.93, 425.[145789]*

CHF, Elixhauser	Congestive Heart Failure, Diagnosis	398.91, 402.[019]1, 404.[019][13], 425.[456789], 428.*
CHF	CHF, Condition	398.91, 402.11, 404.01, 404.11, 404.91, 428*, 402.01, 402.91, 404.13, 404.93, 425.[145789]*
Cirrhosis	Cirrhosis,Condition	571.2, 571.5
Cirrhosis_Risk_1	Combination of Cirrhosis Risk Factors	291.[0123589], 303.*, 305.0*, 571.[013], 070.[23][0-3], V02.6[12], 070.[45][14], 070.7[01], 275.0[1-3], 571.42, 571.6
CKD	Chronic Kidney Disease,Condition	585*, 403*, 404*
Colitis	Colitis, Diagnosis	555.[0129], 556., 556.[0-6]
Cons_Condition	Constipation, Condition	564.0*, 560.3[029], 560.89, 560.9*, 564.7*, 787.99
Cons_Proc	Constipation, Procedure	45915, E035[02], A4458, E0740, 96.3[789], 96.09
COPD	COPD, Condition	49[126]*, 493.2*
COPDAsthma	COPD/Asthma,Condition	491.*, 492.*, 493.*, 496.*, V17.5*, V81.3*
CPD, Elixhauser	Chronic Pulmonary Disease	416.[89], 49[0123456789].*, 50[012345].*, 506.4, 508.[18]
CS	Cardiogenic Shock,Condition	785.51
CVA	Cerebrovascular Disease,Condition	43[0-8]*, 362.34
DA, Elixhauser	Deficiency Anemia	280.[123456789], 281.*
DecALD	Decompensated Cirrhosis, Condition	456.[012], 571.*, 572.[1-8], 789.5
Dementia	Dementia w/o Delirium, Condition	290.*, 294.[1]*, 331.[012]*
Dermatomyositis	Dermatomyositis, Diagnosis	710.3
Deyo's CHF	Congestive Heart Failure, Diagnosis	428*
Deyo's CPD	Chronic Pulmonary Disease, Diagnosis	49[0-24-9]*, 493.[2-8]*, 50[0-5]*, 506.4*
Deyo's CVD	Cerebrovascular Disease, Diagnosis	43[0-8]*

Deyo's Dementia	Dementia, Diagnosis	290*
Deyo's DM w/ chronic comp	Diabetes with chronic complication, Diagnosis	250.[4-6]*
Deyo's DM w/o chronic comp	Diabetes without chronic complication, Diagnosis	250.[0-37]*
Deyo's Hemiplegia or Paraplegia	Hemiplegia or Paraplegia, Diagnosis	344.1*, 342*
Deyo's HIV	AIDS/HIV, Diagnosis	04[2-4]*
Deyo's Malignancy	Any Malignancy (except of skin), Lymphoma, or Leukemia	1[4-6][0-9]*, 17[0-24-9]*, 18[0-9]*, 19[0-5]*, 20[0-8]*
Deyo's MI	Myocardial Infarction, Diagnosis	410*, 412*
Deyo's Mild Liver Disease	Mild Liver Disease, Diagnosis	571.[24-6]*
Deyo's Moderate to Severe Liver Disease	Moderate or Severe Liver Disease, Diagnosis	456.[01]*, 456.2[01], 572.[2-8]*
Deyo's PUD	Peptic Ulcer Disease, Diagnosis	53[1-4]*
Deyo's PVD	Peripheral Vascular Disease	443.9*, 441*, 785.4*, V43.4*, 38.48
Deyo's Renal	Renal Disease, Diagnosis	58[2568]*, 583.[0-7]*
Deyo's Rheumatic	Rheumatic Disease, Diagnosis	710.[014]*, 714.[0-2]*, 714.81, 725*
Deyo's Tumor	Metastatic Solid Tumor, Diagnosis	19[6-9]*
DiabetesC, Elixhauser	Diabetes, Complicated	250.[456789]*
DiabetesU, Elixhauser	Diabetes, uncomplicated	250.[0123]*
DIAL	Dialysis,Procedure	90921, 90925, 90935, 90937, 90945, 9096[0126], G8956, 90947, 90989, 9099[39], 585.6, V39.27, V39.42, V39.43, V45.1, V56.0, V56.2, V56.31, V56.32, V56.8, 39.9[35], 54.98
DIAR	Diarrhea,Condition	009.2, 009.3, 564.5, 787.91
DKA	Diabetic Ketoacidosis, Condition	249.1*, 250.1*
DM	Diabetes,Condition	249*, 250*, 357.2*, 362.0*, 366.41, V45.85, V53.91
DMNeuropathy	Diabetic Neuropathy, Diagnosis	357.2
DMOsm	Diabetic w/ Hyperosmolarity, Condition	249.2*, 250.2*

DP, Elixhauser	Depression	296.[235], 300.4, 309.*, 311
Drug Abuse, Elixhauser	Drug Abuse	292.*, 304.*, 305.[23456789], V65.42
DYS	Dyslipidemia,Condition	272.*
ECMO	ECMO, Procedure	3396[01], 37.62
ETOH	Alcohol Use,Condition	291.*, 303.*, 305.0*, 535.3*, 292.21, 357.5, 425.5, 571.0, 571.1, 571.2, 571.3, 760.71, 790.3, 977.3, 980.[012456789]*, E947.3, E860.[012356789]*, V11.3
Etoh_Abuse	Alcohol Abuse	291.[0123589], 303.*, 305.0*
Fatigue	Chronic Fatigue, Diagnosis	780.71
FED, Elixhauser	Fluid and Electrolyte Disorder	253.6, 276.*
Fibromyalgia	Fibromyalgia, Diagnosis	729.1
Gastroparesis	Gastroparesis, Diagnosis	536.3
GI	GI Bleeding,Condition	530.82, 53[1-4].[0246]0, 535.[045]1, 578.*
GIPerf	GI Perforation, Diagnosis	569.83
GLOM	Acute GLOMERULONEPHRITIS, Condition	580.*
GLOMNephEx	Glomerular Nephritis (Exclusion), Condition	580.[049], 580.8[19], 581.[0123], 582.[01249], 582.8[19], 583.[0124], 581.89
HBC	Cirrhosis Risk Cohort without NAFLD	070.[23][0-3], V02.6[12], 070.[45][14], 070.7[01]
HCC	Hepatocellular Carcinoma, Condition	155
HE	Hepatic Encephalopathy, Condition	572.2*, 070.00, 070.2*, 070.40, 070.41, 070.44, 070.49, 070.60
Headache	Migraine & Headache, Diagnosis	784.0, 339.0[0123459], 339.[12][0-2], 339.3, 339.4[1-4], 339.8[123459], 346.[0-5][0-3], 346.[0-2], 346.[7-9][0-3], 346.[89]
HEMOCH	Cirrhosis Risk Cohort without NAFLD	275.0[1-3]

Hemorrhoid	Hemorrhoids, Diagnosis	455.[0-9]
HEP	Hepatitis,Condition	070.*, 072.71, 091.62, 130.5, 571.1, 571.4*, 573.1, 573.2, 573.3, V02.6*, V05.3
Hep_B_C	Hepatitis B and C, acute and chronic	070.[23][0-3], V02.6[12], 070.[45][14], 070.7[01]
HF, Elixhauser	Hypertension, uncomplicated, Diagnosis	401.*
HFC, Elixhauser	Hypertension, complicated, Diagnosis	40[2345].*
HIV	HIV,Condition	04[234]*, 079.53, 795.71, V08*
HOSP	Hospice,Condition	99377, 99378
HOTN	Hypotension,Condition	458.*
HRS	Hepatorenal Syndrome	572.4
HrtTx	Heart Transplant,Procedure	V42.1, 37.5[1-5]
HSVNeuralgia	Post Herpetic Neuralgia, Diagnosis	53.19
HTD, Elixhauser	Hypothyroidism	240.9, 24[34].*, 246.[18]
HTN	Hypertension,Condition	401*, 402*, 403*, 404*, 405*, 437.2*
HTNEmer	Hypertension Emergency, Condition	40[1-5].0, 40[2-5].01, 404.0[23], 405.0[19]
Hydronephrosis	Hydronephrosis	591*
HYPC	Hypercalcemia,Condition	275.42
Hyperkalemia	Hyperkalemia, Diagnosis	276.7
Hyperparathyroidism	Hyperparathyroidism, Diagnosis	252.0*
IABP	Intra-Aortic Balloon Pump, Procedure	3396[78], 3397[0134], 37.61
IBS	Irritable Bowel Syndrome, Diagnosis	564.1
iDIAL	Dialysis Inpatient	v45.1, v56.0, v56.1, 39.95
Impaction	Impaction, Diagnosis	560.32
IVD	Intravascular Volume Disease,Condition	276.5, 276.5[01]
JAUD	Jaundice,Condition	282.00, 774.[0123567]*, 782.4, , ,
LD, Elixhauser	Liver Disease	070.[23][23], 070.[45]4, 070.[69], 456.[012], 57[01].*, 572.[2345678], 573.[3489], V42.7

LIV	Liver,Condition	070.22, 070.23, 070.33, 070.44, 070.54, 456.0, 456.1, 456.20, 456.21, 571.0, 571.2, 571.3, 571.40, 571.41, 571.42, 571.49, 571.5, 571.6, 571.8, 571.9, 572.3, 572.8, 573.5, V42.7
LKM	Leukemia,Condition	202.4*, 203.1*, 20[4-8].*, V10.6*
LM, Elixhauser	Lymphoma	20[012].*, 203.0, 238.6
LngTx	Lung Transplant,Procedure	V42.6, 33.5*
LUP	Systemic Lupus Erythematosus,Condition	286.5, 323.81, 517.8, 58[023].81, 695.4, 710.0
LvrTx	Liver Transplant,Procedure	4713[56], V42.7, 50.5[19]
MC, Elixhauser	Metastatic Cancer	19[6789].*
MECHVENT	Mechanical Ventilation, Procedure	93.92, 96.0[45], 96.7[012]
Megacolon	Megacolon, Diagnosis	564.7
MEN	Multiple Endocrine Neoplasia, Diagnosis	258.0*
MetabolicSyndrome	Metabolic Syndrome using sub-conditions	usp_Build_Metabolic_Syndrome
MI	Myocardial Infarction,Condition	410*
MM	Multiple Myeloma,Condition	203.0*
MultScler	Multiple Sclerosis, Diagnosis	340
MVR	Mitral Regurgitation,Procedure	396.3, 424.0, 746.6
Myopathies	Myopathies, Diagnosis	359.8, 359.89, 425.4
NAFLD	Non-alcoholic fatty liver disease	571.8, 571.9
NAS	Nausea,Condition	787.0, 787.01, 787.02
NephGLOM	Nephritis Glomerular Not Specified,Condition	580.81, 58[03].9, 583.8[19]
Nephrtmy	Nephrectomy, Procedure	5022[05], 5023[046], 50240, 50300, 50320, 50340, 50370, 5054[35-8], 55.4, 55.5[1-4]
NFXss	Infection (sepsis sup),Condition	00[1-589]*, 01[0-8]*, 02[0-7]*, 03*, 04[01]*, 09[0-8]*, 10[0-4]*, 11[0-24-8]*, 32[0245]*, 42[01]*, 451*, 46[1-5]*, 48[1256]*, 491.21, 494*, 51[03]*, 54[012]*, 562.[01][13], 56[67]*,

		569.5*, 569.83, 572.[01]*, 575.0*, 59[07]*, 599.0*, 601*, 61[456]*, 68[1236]*, 711.0*, 730*, 790.7*, 996.6*, 998.5*, 999.3*
NSTEMI	NSTEMI, Condition	410.7*
OA	Osteoarthritis, Diagnosis	715.0[049], 715.1[0-8], 715.2[0-8], 715.3[0-8], 715.8[09], 715.9[0-8], V13.4
Obesity	Obesity	278.0, 278.0[01], 649.1, 278.03
Obesity, Elixhauser	Obesity	278
OFss	Organ Failure (sepsis sup),Condition	785.5*, 458*, 348.3*, 293*, 348.1*, 286.[69]*, 287.[45]*, 570*, 573.4*, 584*, 96.7*
OLG	Oliguria,Condition	788.5
OND, Elixhauser	Other Neurological Disorders	331.9, 332.[01], 333.[45], 333.92, 33[45].*, 336.2, 34[015].*, 348.[13], 78[04].3
OrganTrans	Organ Transplant,Procedure	50320, 50360, 50365, 50370, 50380, 33935, 33940, 33945, 32851, 32852, 32853, 32854, 47135, 47136, 38240, 38241, 48554, 48556, V42.0*, V42.1*, V42.6*, V42.7*, V42.81*, V42.83*
PALL	Pallative Care,Condition	V66.7*
PANTx	Pancreas Transplant,Procedure	48554, 99686, V42.83, 52.8[0-3]
Paracentesis	Paracentesis, procedure	4908[0-3], 54.91
Paralysis, Elixhauser	Paralysis, Diagnosis	334.1, 34[23].*, 344.[01234569]
Parkinsons	Parkinson's Disease, Diagnosis	332
PCD, Elixhauser	Pulmonary Circulation Disorders, Diagnosis	415.[01], 416.*, 417.[089]
PCI	PCI, Procedure	9297[34], 9298[0124], 9299[56], G029[01], V45.82, 00.66, 36.0[125679], 92.27

PCR	Pancreatitis,Condition	577.[01]*
PFT	Spirometry,Procedure	94010, 94060, 94375, 94150, 94200, 93720, 93721, 93722, 94726, 94727, 94728, 94729, 94240, 94260, 94350, 94360, 94370, 94720, 94725
Plegia_CC	Hemiplegia or Paraplegia (Charlson Comorbidity Definition)	334.1*, 342.*, 343.*, 344.[01234569]*
Porphyria	Porphyria, Diagnosis	277.1
PREG	Pregnancy,Condition	V22.*, 6[3-7]*
ProctFugax	Proctalgia Fugax, Diagnosis	564.6
PUD_CC	Peptic Ulcer Disease (Charlson Comorbidity Definition)	531.*, 532.*, 533.*, 534.*
PVD	Peripheral Vascular Disease,Condition	440*, 441*, 442*, 444.2*, V43.4
PVD, Elixhauser	Peripheral Vascular Disorders, Diagnosis	093.0, 437.3, 44[01].*, 443.[123456789], 447.1, 557.[19], V43.4
PY, Elixhauser	Psychoses	293.8, 295.*, 296.[0145]4, 29[78].*
RA	Rheumatoid Arthritis, Diagnosis	714
RA, Elixhauser	Rheumatoid Arthritis, Collagen vascular diseases	446.*, 701.0, 710.[0123489], 711.2, 714.*, 719.3, 72[05].*, 728.5, 728.89, 729.30
RectalProlapse	Rectal prolapse, Diagnosis	569.1
Rectocele	Rectocele, Diagnosis	618.04
RenalTrans	Renal Transplant,Procedure	50365, 50360, 996.81, V42.0, 55.69*, 00.9[123]
RF, Elixhauser	Renal Failure	403.[019]1, 404.0[23], 404.[19][23], 58[56].*, 588.0, V42.0, V45.1, V56.*
RHBD	Rhabdomyolysis,Condition	728.88
Rheum_CC	Rheumatic Disease (Charlson Comorbidity Definition)	446.5*, 710.[01234]*, 714.[0128]*, 725.*
PUD, Elixhauser	Peptic Ulcer Disease, excluding bleeding	53[1234].[79]*
SBP	Spontaneous Bacterial Peritonitis - extra general, Condition	567.23, 567.[0289]0, 567.2[19], 567.89, 567.[0289]
Scleroderma	Scleroderma, Diagnosis	701

Sickle	Sickle Cell Disease, Condition	282.4[1-4]*, 282.6*
SIRS	SIRS,Condition	995.9*
SpinalCord	Spinal Cord Injury / Tumors, Diagnosis	349.39, 806.[0-3][0-9], 806.[4589], 806.[67][0129], 907.2, 952.[01][0-9], 952.[23489]
SPS	Sepsis,Condition	785.52, 995.92
SSTS	Schistosomiasis,Condition	120*
ST, Elixhauser	Solid Tumor without Metastasis	1[456][0123456789].*, 17[012].*, 17[456789].*, 18[0123456789].*, 19[012345].*
StableAngina	Stable Angina, Diagnosis	413.[019]
STEMI	STEMI, Condition	410.[012345689]*
STROKE	Stroke, Condition	43[01]*, 434.[019], 434.[019]1, 436*, 997.02
TB	Tuburculosis,Condition	01[0-8].*, 137.*, V12.01
TIA	TIA, Condition	435.[89]
TIPS	Transjugular Intrahepatic Portosystemic Shunt	3718[23], 39.1
TOBC	Tobacco Use,Condition	305.1*, V15.82
TRAU	Injury-Trauma,Condition	349.39, 716.1*, 717.*, 718.0*, 718.3*, 806.*, 80[0134].[12346789]*, 83_.*, 85[01234].*, 86_.*, 90[012347].*, 905.6, 906.4, 908.[01234]*, 92[56789].*, 952.*, V15.52
UNAN	Unstable Angina, Condition	411.1*, 413*
UrinaryObst	Urinary Obstruction, Condition	592.1, 593.4, 594.[29], 596.0, 598.[1289], 599.6, 599.69, 599.82, 600.[0129]1, 753.[26], 753.2[129], 788.2, 788.29, V44.6, V55.6
VAD	Ventricular Assist Device, Procedure	3397[5-9], 33980, 37.41, 37.5[2-5], 37.6[0356]
Valvular, Elixhauser	Valvular Disease, Diagnosis	093.2, 39[4567].*, 424.*, 746.[3456], V42.2, V43.3
Varices	Varices, Condition or Procedure	4324[34], 4320[45], 456.[012][01], 456.[012], 42.33

VascProc	All Vascular Procedures	3332[012], 3333[025], 3386[013], 3387[057], 3388[01], 34001, 34051, 34101, 34111, 34151, 3420[13], 34401, 34421, 34451, 34471, 34490, 3480[02-5], 3482[56], 3483[0-2], 34900, 3500[12], 3501[13], 3502[12], 35045, 3508[12], 3509[12], 3510[23], 3511[12], 3512[12], 3513[12], 3514[12], 3515[12], 3516[12], 3521[16], 3522[16], 3523[16], 3524[16], 3525[16], 3526[16], 3527[16], 3528[16], 3530[1-6], 35311, 35321, 35331, 35341, 3535[15], 3536[13], 3537[12], 35381, 35390, 3545[024689], 3547[0-5], 3548[0-5], 3549[0-5], 3550[16-9], 3551[1568], 3552[16], 3553[135-9], 3554[01689], 3555[168], 3556[0356], 3557[01], 3558[2357], 3560[16], 3561[26], 3562[136], 3563[12678], 3564[12567], 3565[0146], 3566[1356], 35671, 3569[45], 35700, 35820, 35840, 35860, 3587[0569], 35881, 3590[1357], 37799, 38.0[0-9], 38.1[0-68], 38.2[1-69], 38.3[0-9], 38.4[0-9], 38.5[0-3579], 38.6[0-9], 38.7, 38.8[0-9], 38.9[1-5789], 39.0, 39.1, 39.2[1-9], 39.3[012], 39.4[1239], 39.5[1-9],

		39.7[1-9], 39.8[1-9], 39.9[0-46-9]
VascSurg	Major Vascular Surgery, Procedure	3332[012], 3333[025], 3386[013], 3387[057], 3388[01], 34151, 34201, 3480[02-5], 3482[56], 3483[0-2], 3508[12], 3509[12], 3510[23], 3511[12], 3512[12], 3521[16], 35221, 3524[16], 35251, 3527[16], 35281, 35331, 35341, 3536[13], 3545[02], 3547[12], 3548[01], 3549[01], 3553[15-9], 3554[01689], 35551, 35560, 35582, 3563[12678], 3564[167], 35651, 35820, 35840, 35870, 3590[57], 38.0[456], 38.1[4568], 38.2[123], 38.[3468][456], 38.9[1-5789], 39.0, 39.2[13-6], 39.5[45], 39.7[138]
VHD	Valvular Heart Disease, Condition	424.[0-3]
VLP	Valvulopathy,Condition	39[4567].*, 424.*, 745.*, 746.[0-7]*, 746.8[134], 747.3, 785.[23]*, V42.2, V43.3
VMT	Vomiting,Condition	078.82, 307.54, 536.2, 564.3, 569.87, 578.0, 643*, 787.0, 787.00, 787.01
WL, Elixhauser	Weight Loss	26[0123].*, 783.2, 799.4

Appendix Table A.3: Variables used in the multiple imputation of laboratory values. Multiple imputation was carried out by the mi package for the R statistical programming software. Imputation was carried out for 30 iterations using 4 separate chains. Imputations were carried out to convergence. Values from the four separate chains were averaged together for the final imputed values used in the dataset.

Domain	TimePeriod	Variable
Condition	Home	Atrial Fibrillation
Condition	Home	Anemia
Condition	Home	Ascites
Condition	Home	AZ_Cancer
Condition	Home	Biliary Cirrhosis
Condition	Home	Coronary Artery Disease
Condition	Home	Congestive Heart Failure
Condition	Home	Chronic Kidney Disease
Condition	Home	Dialysis
Condition	Home	Diabetes Mellitus
Condition	Home	Etoh Abuse
Condition	Home	GI Bleed
Condition	Home	Hepatocellular Carcinoma
Condition	Home	Hepatic Encephalopathy
Condition	Home	Viral Hepatitis
Condition	Home	HIV
Condition	Home	HTN
Condition	Home	Nonalcoholic Fatty Liver Disease
Condition	Home	Spontaneous Bacterial Peritonitis
Condition	Home	Transjugular Intrahepatic Portosystemic Shunt
Condition	Home	Varices
Medication	Home	Rifaxmin
Medication	Home	Lactulose
MedClass	Home	Quinolones
MedClass	Home	Anticoagulants
MedClass	Home	Platelet Aggregation Inhibitors
MedClass	Home	Opioids
MedClass	Home	Sedative Hypnotics
MedClass	Home	Anticonvulsants
MedClass	Home	Antidepressants
MedClass	Home	Digitalis glycosides
MedClass	Home	Beta Blockers
MedClass	Home	Alpha Blockers
MedClass	Home	Calcium Channel Blockers
MedClass	Home	Antiarrhythmics

MedClass	Home	Antilipemic Agents
MedClass	Home	Thiazides
MedClass	Home	Loop Diuretics
MedClass	Home	Potassium Sparing Combination Diuretics
MedClass	Home	ACE Inhibitors
MedClass	Home	Angiotensin II Inhibitors
Procedure	Home	Paracentesis
Medication	Inpt	Trimethoprim
Medication	Inpt	Vancomycin
Medication	Inpt	Lactulose
Medication	Inpt	Octreotide
Medication	Inpt	Midodrine
Medication	Inpt	Human Albumin
Medication	Inpt	Norepinephrine
Medication	Inpt	Vasopressin
Medication	Inpt	Rifaximin
Medication	Inpt	Septra
Medication	Inpt	Dobutamine
Medication	Inpt	Phenylephrine
Medication	Inpt	Octreotide and Midodrine
Medication	Inpt	Octreotide and Norepinephrine
Medication	Inpt	Octreotide and Vasopressin
Medication	Inpt	Any IV Vasopressor
MedClass	Inpt	NSAIDs
MedClass	Inpt	Aminoglycosides
MedClass	Inpt	Beta Blockers
MedClass	Inpt	ACE Inhibitors
MedClass	Inpt	Angiotensin Receptor Blockers
MedClass	Inpt	Glucocorticoids
MedClass	Inpt	Potassium Sparing Diuretics
MedClass	Inpt	Statins
MedClass	Inpt	Insulin
MedClass	Inpt	Extended Spectrum Penicillins
MedClass	Inpt	3rd Generation Cephalosporins
MedClass	Inpt	Anticoagulants
MedClass	Inpt	Platelet Aggregation Inhibitors
Condition	Inpt	Dialysis
Condition	Inpt	Chronic Kidney Disease
Condition	Inpt	Diabetes Mellitus
Condition	Inpt	Coronary Artery Disease
Condition	Inpt	Hypertension

Condition	Inpt	Hospice
Condition	Inpt	Congestive Heart Failure
Condition	Inpt	Cancer
Condition	Inpt	Sepsis
Condition	Inpt	Diarrhea
Condition	Inpt	SIRS
Condition	Inpt	Hepatitis
Condition	Inpt	Vomiting
Condition	Inpt	Alcohol Use
Condition	Inpt	Anemia
Condition	Inpt	GI Bleed
Condition	Inpt	Cardiovascular Disease
Condition	Inpt	COPDAsthma
Condition	Inpt	Glomerulonephritis
Condition	Inpt	Acute Tubular Necrosis
Condition	Inpt	Decompensated Liver Disease
Condition	Inpt	Urinary Obstruction
Condition	Inpt	Glomerularnephritis NOS
Condition	Inpt	Atrial Fibrillation
Condition	Inpt	Acute Renal Failure
Condition	Inpt	Nephrectomy
Condition	Inpt	Palliative Care
Condition	Inpt	Hydronephrosis
Condition	Inpt	Etoh Abuse
Condition	Inpt	Ascites
Condition	Inpt	Hepatorenal Syndrome
Condition	Inpt	Transjugular Intrahepatic Portosystemic Shunt
Condition	Inpt	Varices
Condition	Inpt	Hepatic Encephalopathy
Condition	Inpt	Spontaneous Bacterial Peritonitis
Condition	Inpt	Hepatocellular Carcinoma
Condition	Inpt	Nonalcoholic Fatty Liver Disease
Condition	Inpt	Elixhauser: CHF
Condition	Inpt	Elixhauser: Arrhythmias
Condition	Inpt	Elixhauser: Valvular Heart Disease
Condition	Inpt	Elixhauser: Pulmonary Circulation d/o
Condition	Inpt	Elixhauser: Peripheral Vascular Dz
Condition	Inpt	Elixhauser: HTN uncomplicated
Condition	Inpt	Elixhauser: HTN complicated
Condition	Inpt	Elixhauser: Paralysis
Condition	Inpt	Elixhauser: Other Neurological d/o

Condition	Inpt	Elixhauser: Pulmonary Disease
Condition	Inpt	Elixhauser: Diabetes uncomplicated
Condition	Inpt	Elixhauser: Diabetes Complicated
Condition	Inpt	Elixhauser: Hypothyroidism
Condition	Inpt	Elixhauser: Renal Failure
Condition	Inpt	Elixhauser: Liver Disease
Condition	Inpt	Elixhauser: Peptic Ulcer Disease
Condition	Inpt	Elixhauser: AIDS/HIV
Condition	Inpt	Elixhauser: Lymphoma
Condition	Inpt	Elixhauser: Metastatic Cancer
Condition	Inpt	Elixhauser: Solid Tumor without mets
Condition	Inpt	Elixhauser: Collagen vascular diseases
Condition	Inpt	Elixhauser: Coagulopathy
Condition	Inpt	Elixhauser: Weight Loss
Condition	Inpt	Elixhauser: Fluid and Electrolyte Disorder
Condition	Inpt	Elixhauser: Blood Loss Anemia
Condition	Inpt	Elixhauser: Deficiency Anemia
Condition	Inpt	Elixhauser: Alcohol Abuse
Condition	Inpt	Elixhauser: Drug Abuse
Condition	Inpt	Elixhauser: Psychoses
Condition	Inpt	Elixhauser: Depression

Appendix Table A.4: List of semantic types used to filter variables for semantic

Semantic Type Identifier	Description
T020	acquired abnormality
T190	anatomical abnormality
T053	behavior
T031	body substance
T201	clinical attribute
T060	diagnostic procedure
T047	disease or syndrome
T033	Finding
T131	Hazardous or Poisonous Substance
T058	Health Care Activity
T129	Immunologic Factor
T037	Injury or Poisoning
T048	mental or behavioral dysfunction
T191	neoplastic process
T046	pathologic function
T184	sign or symptom
T061	therapeutic or preventive procedure

Appendix Table A.5: List of a-priori CUIs

CUI	STR	Assigned Variable Category
C0035222	Acquired respiratory distress syndrome	Acquired respiratory distress syndrome
C0003460	ANURIA	Anuria
C0000731	abdomen distended	Ascites
C0003962	abdominal dropsy	Ascites
C0426682	fluid wave	Ascites
C0741244	Tense ascites	Ascites
C0333501	Acute necrosis	Acute Tubular Necrosis
C0344391	granular cast	Casts
C0011175	body water dehydration	Dehydration
C0013604	dropsies	Edema
C0017658	Glomerulonephritis NOS	Glomerulonephritis
C0017662	GLOMERULONEPHRITIS, MEMBRANOPROLIFERATIVE	Glomerulonephritis
C0156221	acute glomerulonephritis	Glomerulonephritis
C0017665	glomerulonephritis membranous	Glomerulopnephritis
C0018926	Hematemeses	Hematemesis
C0019151	coma hepaticum	Hepatic Encephalopathy
C0019212	Hepatorenal syndrome	Hepatorenal Syndrome
C1708271	HRS	Hepatorenal Syndrome
C0020295	Hydronephrosis	Hydronephrosis
C0020649	hypotension	Hypotension
C0027707	Interstitial Nephritis	Nephritis
C0041349	Tubulo-interstitial nephritis	Nephritis
C0149937	acute interstitial nephritis	Nephritis
C0347129	AIN	Nephritis
C1514118	nephrotoxic	Nephrotoxic
C0003211	Agents, Nonsteroidal Antiinflammatory	NSAIDs
C0011991	bowel loose movements	Nausea/Vomiting/Diarrhea (NVD)
C0027497	nausea	Nausea/Vomiting/Diarrhea (NVD)
C0027498	N&V - Nausea and vomiting	Nausea/Vomiting/Diarrhea (NVD)
C0042963	Emesis	Nausea/Vomiting/Diarrhea (NVD)
C0034115	paracentesis	Paracentesis
C0031154	Peritonitis	Peritonitis
C0014772	blood cell count red	RBCs

C0014772	Red Cell Count	RBCs
C0036690	SEPSIS	Sepsis
C0243026	SEPSIS	Sepsis
C0036974	Shock	Shock
C0036982	Haemorrhagic shock	Shock
C0036983	SHOCK SEPTIC	Shock
C0242966	SIRS	Systemic Inflammatory Response Syndrome (SIRS)
C0552639	tubular cell	Tubular Cells
C0553257	Epithelial cell of renal tubule	Tubular Cells
C1261248	urinary sediment	Urine Sediment
XXXXXXX	XXXXXXX	Prerenal

Appendix Table A.6: Evaluation of model performance after exclude "Maybe HRS" cases from model building and evaluation using the *a priori* CUIs.

	AUC (95% CI)	Slope (95% CI)	Intercept (95% CI)	Brier Score (95% CI)
Logistic Regression	0.94 (0.93, 0.94)	0.36 (0.31, 0.40)	-0.24 (-0.32, -0.16)	0.08 (0.08, 0.09)
Gradient Boosting	0.94 (0.93, 0.94)	1.70 (1.64, 1.75)	0.16 (0.08, 0.24)	0.10 (0.09, 0.10)
Naïve Bayes	0.77 (0.74, 0.79)	0.20 (0.19, 0.22)	-3.84 (-4.28, -3.40)	0.44 (0.40, 0.47)
Random Forest	0.94 (0.94, 0.94)	2.54 (2.44, 2.65)	0.45 (0.36, 0.54)	0.10 (0.10, 0.11)
Support Vector Machine	0.94 (0.93, 0.94)	0.86 (0.82, 0.89)	-0.06 (-0.14, 0.02)	0.09 (0.09, 0.10)

Appendix Table A.7. Discrimination and calibration performance of the five models to phenotype Hepatorenal Syndrome using the SAFE concept unique identifiers.

(Note: Slope and Intercept refer to the parameters of the best-fit line through the observed-to-predicted probability plot; AUC: Area Under the Curve)

	AUC (95% CI)	**Slope (95% CI)**	**Intercept (95% CI)**	**Brier Score (95% CI)**
Logistic Regression	0.93 (0.92, 0.93)	0.47 (0.43, 0.50)	-0.16 (-0.22, -0.11)	0.11 (0.11, 0.11)
Gradient Boosting	0.91 (0.91, 0.92)	1.60 (1.54, 1.65)	0.09 (0.04, 0.14)	0.12 (0.12, 0.13)
Naïve Bayes	0.70 (0.68, 0.73)	0.14 (0.12, 0.15)	-2.60 (-2.91, -2.29)	0.42 (0.40, 0.45)
Random Forest	0.91 (0.91, 0.92)	1.97 (1.92, 2.02)	0.32 (0.27, 0.37)	0.13 (0.13, 0.13)
Support Vector Machine	0.91 (0.90, 0.91)	0.85 (0.82, 0.87)	-0.01 (-0.06, 0.04)	0.12 (0.12, 0.13)

Appendix A.8: CUIs selected based on the Automated Feature Extraction for Phenotyping (AFEP) and the Surrogate-Assisted Feature Extraction (SAFE) method.

Candidate CUIs for consideration in the AFEP algorithm [93] were extracted from Medscape and Wikipedia articles on HRS. The SAFE algorithm [92] also considered CUIs extracted from the HRS entries in Merck Manuals and Medline Plus Medical Encyclopedia. Mayo Clinic Disease and Conditions did not include relevant content on HRS to allow inclusion in the SAFE approach. We limit candidate CUIs lists for both methods based on semantic type and grouped drug concepts by generic names and drug classes. Brand name concept to generic concept mapping was performed using the "has_tradename" and "tradename_of" relationships in the UMLS hierarchy, similar to the AFEP paper. Drug class identification was performed using the RxNorm hierarchy and the RxNorm WebAPI.[303] The remaining candidates were filtered for rarity and commonality as recommended. A list of filtered candidate CUIs based on the public knowledge sources is available in the Online Appendix. We adjusted the SAFE silver standard thresholds to ensure reasonable sample sizes in the extreme subsets (L_{ICD}=0, L_{NLP}=0, U_{ICD}=1, and U_{NLP}=3). We implemented 50 iterations of the elastic net models for each of the three silver standards in the SAFE approach, selecting CUIs included in 50% of the models overall. The final list of CUIs selected by AFEP and SAFE are listed in Appendix Table A.8.

Appendix Table A.8: List of Concept Unique Identifiers (CUI) used in the automated feature extraction for phenotyping (AFEP) and surrogate-assisted feature extraction (SAFE) methods.

The sources are: Wikipedia, Medscape, Merck Manual, MedlinePlus, and MayoClinic.

	Included in public knowledge source					AFEP		SAFE	
CUI	**Wiki.**	**Med.**	**Merck**	**Medline**	**Mayo**	**Candidate**	**Selected**	**Candidate**	**Selected**
C0000731				X					
C0000970		X				X			
C0001047	X	X				X			
C0001128		X				X	X		
C0001306	X	X		X		X		X	
C0001443		X				X			
C0001648		X				X			
C0001924	X	X	X	X		X	X	X	
C0002006	X					X			
C0002170		X				X			
C0002210		X				X			
C0002556		X				X			
C0002772		X				X			
C0002792		X				X			
C0003009		X				X			
C0003018	X	X				X			
C0003211				X					
C0003232	X	X		X		X		X	
C0003402		X				X			
C0003448	X					X			
C0003779		X				X			
C0003962	X	X	X	X		X	X	X	
C0004610		X				X			
C0004623	X	X				X			
C0005771		X				X			
C0005779		X				X	X		
C0006318		X				X			
C0007430		X				X			
C0007554		X				X			
C0007584		X				X			
C0007955		X				X			
C0008370		X				X			

	Included in public knowledge source					AFEP		SAFE	
C0008679		X		X		X			
C0008809		X				X			
C0009319		X				X			
C0009421				X					
C0009429			X						
C0009450	X	X		X		X		X	
C0009555		X				X			
C0009566	X	X		X		X		X	
C0009676				X					
C0009905				X					
C0009924	X	X				X			
C0010403		X				X			
C0010404		X				X			
C0010592	X					X			
C0010957	X		X	X		X		X	
C0011276		X				X			
C0011710		X				X			
C0011744		X				X			
C0011923		X				X			
C0011946	X	X		X		X	X	X	
C0011947		X				X			
C0012169		X				X			
C0012237		X				X			
C0012299		X				X			
C0012359	X					X			
C0012582		X				X			
C0012772		X				X			
C0012798	X	X		X		X	X	X	
C0012854		X				X			
C0013030		X				X			
C0013103		X				X			
C0013221	X					X			
C0013378		X				X			
C0013516		X				X			
C0013604		X				X	X		
C0013819				X					
C0013862				X					
C0013983		X				X			

	Included in public knowledge source					AFEP		SAFE	
C0014245	X					X			
C0014264	X					X			
C0014442		X				X			
C0014745		X				X			
C0014772		X				X	X		
C0014867	X	X				X			
C0015672		X				X			
C0015950		X				X			
C0016059	X					X			
C0016107	X	X	X			X		X	
C0017181	X	X		X		X		X	
C0017654	X	X	X			X		X	
C0017658		X				X			
C0017662		X				X			
C0017665		X				X			
C0017675		X				X			
C0017817		X				X			
C0018418		X		X		X			
C0018801				X					
C0018935		X				X			
C0018941		X				X			
C0018965	X	X				X			
C0019004	X	X				X			
C0019014		X				X			
C0019080	X	X		X		X		X	
C0019151	X			X		X	X		
C0019158	X	X	X	X		X	X	X	
C0019163	X	X				X			
C0019187	X	X		X		X		X	
C0019214		X				X			
C0019270		X				X			
C0019311		X				X			
C0019868		X				X			
C0019932	X	X				X			
C0020295		X				X			
C0020488		X				X			
C0020538	X	X				X			
C0020541	X	X				X			

	Included in public knowledge source					AFEP		SAFE	
C0020625		X				X	X		
C0020649	X	X		X		X	X	X	
C0020651				X					
C0020683		X				X			
C0020740				X					
C0021081	X	X				X			
C0021368	X					X			
C0021440	X					X			
C0021936		X				X			
C0021968		X				X			
C0022116	X	X				X			
C0022346	X	X		X		X	X	X	
C0022658	X	X		X		X		X	
C0022660	X	X				X	X		
C0022671	X					X			
C0022672		X				X			
C0023175	X					X			
C0023518		X				X			
C0023545		X				X			
C0023890	X	X	X	X		X		X	
C0023891	X					X	X		
C0023895	X	X		X		X	X	X	
C0023899		X				X			
C0023901		X				X			
C0023911	X	X	X			X		X	
C0024337		X				X			
C0025424		X				X			
C0026018	X					X			
C0026078	X	X	X			X	X	X	X
C0026160	X	X				X			
C0026846		X				X			
C0027310		X				X			
C0027479	X					X			
C0027481	X					X			
C0027497				X					
C0027769		X				X			
C0028128	X	X				X			
C0028158				X					

	Included in public knowledge source					AFEP		SAFE	
C0028259	X					X			
C0028351		X				X			
C0028365		X				X			
C0028778		X				X			
C0028833	X	X	X			X	X	X	
C0028961	X	X	X			X		X	
C0029276	X	X				X			
C0029944		X				X			
C0030125	X					X			
C0030899	X	X				X			
C0030946		X				X			
C0031001		X				X			
C0031154	X	X				X			
C0031448		X				X			
C0032017	X	X				X			
C0032042		X				X			
C0032181		X				X			
C0033085		X				X			
C0033095	X	X				X			
C0033554	X					X			
C0033567	X					X			
C0033684	X					X	X		
C0033687	X	X				X			
C0033707		X		X		X			
C0034783		X				X			
C0034933				X					
C0035078	X	X	X	X		X	X	X	
C0035139	X	X				X			
C0036140		X				X			
C0036193		X				X			
C0036974	X	X				X			
C0037473	X	X				X			
C0037494		X				X			
C0037659	X	X				X			
C0038257	X					X			
C0038689		X				X			
C0038999				X					
C0039052		X				X			

	Included in public knowledge source					AFEP		SAFE	
C0039082	X	X	X	X		X	X	X	X
C0039796		X				X			
C0040057	X					X			
C0040061	X					X			
C0040125		X				X			
C0040549	X					X			
C0040732	X	X	X			X	X	X	
C0040808	X	X				X			
C0040958	X					X			
C0041041		X				X			
C0041044		X				X			
C0041834		X				X			
C0041942		X				X			
C0042029		X				X			
C0042345	X	X				X	X		
C0042373		X				X			
C0042397	X	X	X			X		X	
C0042402		X				X			
C0042769	X					X			
C0042963				X					
C0043047		X		X		X			
C0043094				X					
C0066480		X				X			
C0066563	X					X			
C0072471		X				X			
C0078077		X				X			
C0079284		X				X			
C0079595	X					X			
C0080059		X				X			
C0080274		X				X			
C0082420		X		X		X			
C0085128		X				X			
C0085149	X					X			
C0085174	X	X				X			
C0085584	X	X		X		X	X	X	
C0085590	X	X				X			
C0085605	X	X		X		X	X	X	
C0085649		X				X			

	Included in public knowledge source					AFEP		SAFE	
C0086761		X				X			
C0087111	X	X	X	X		X		X	
C0105421	X	X		X		X		X	
C0125644		X				X			
C0127400	X					X			
C0145185	X	X	X			X		X	
C0149651		X				X			
C0150041		X				X			
C0150077		X				X			
C0151578		X				X			
C0152451		X				X			
C0155210		X				X			
C0155789	X					X			
C0156221		X				X			
C0156246		X				X			
C0156247		X				X			
C0161959		X				X			
C0162529		X				X			
C0162557	X					X			
C0175661		X				X			
C0179802		X				X			
C0181074		X				X			
C0181805	X					X			
C0184486	X					X			
C0194133		X				X			
C0198497				X					
C0199176	X	X				X			
C0200396		X		X		X			
C0200679	X					X			
C0200949		X				X			
C0201803	X	X				X			
C0201838	X	X				X	X		
C0201849	X					X			
C0201879				X					
C0201888		X				X			
C0201913		X				X	X		
C0201975	X	X				X			
C0201976	X	X				X			

	Included in public knowledge source					AFEP		SAFE	
C0201989		X				X			
C0202145		X				X			
C0202195		X				X			
C0217843		X				X			
C0221135		X				X			
C0221198				X					
C0221226		X				X			
C0221239		X				X			
C0221423	X					X			
C0221752	X					X			
C0231176		X		X		X			
C0231187		X				X			
C0231218	X	X				X			
C0232342	X					X			
C0232766		X				X			
C0232831				X					
C0233494	X					X			
C0235395	X	X				X			
C0235618		X				X			
C0239571		X				X			
C0240182		X				X			
C0240962		X				X	X		
C0242528		X	X	X		X		X	
C0242656	X					X			
C0242889		X				X			
C0242903				X					
C0242937		X				X			
C0243026		X				X			
C0243071	X	X				X			
C0262926	X	X				X			
C0266258		X				X			
C0275551	X	X				X			
C0277787		X				X			
C0278252		X				X	X		
C0279033	X					X			
C0282090		X				X			
C0282151		X				X			
C0282638		X				X			

	Included in public knowledge source					AFEP		SAFE	
C0302353		X				X			
C0302809		X	X			X			
C0303753	X	X				X			
C0304550		X				X			
C0304551		X				X			
C0304925		X				X			
C0309872		X				X			
C0311392	X	X		X		X		X	
C0333121	X					X			
C0333501	X	X				X			
C0337443	X	X	X	X		X	X	X	
C0338237		X				X			
C0339897	X	X	X			X		X	
C0341503	X	X				X			
C0344441	X					X			
C0348042		X		X		X			
C0350056	X					X			
C0353714		X				X			
C0355614		X				X			
C0368721		X				X			
C0373535	X	X				X			
C0373595	X	X				X			
C0373719	X	X				X			
C0392148	X	X				X			
C0403416		X				X			
C0418967	X					X			
C0422768		X				X			
C0426396				X					
C0427944		X				X			
C0428279	X	X				X			
C0428283	X					X			
C0428437				X					
C0428601		X				X			
C0428642		X				X			
C0428886		X				X			
C0429119		X				X			
C0430397	X					X			
C0439775	X					X			

	Included in public knowledge source					AFEP		SAFE	
C0440102	X	X				X			
C0441513	X					X			
C0441610		X				X			
C0442811	X			X		X			
C0442856		X				X			
C0442886				X					
C0445115	X					X			
C0449970		X				X			
C0450442	X	X				X			
C0450458		X				X			
C0456378	X	X				X			
C0457422		X				X			
C0460139	X	X		X		X		X	
C0472683		X				X			
C0475371	X			X		X			
C0475806		X				X			
C0487602	X					X			
C0520819	X					X			
C0520890		X				X			
C0520891	X					X			
C0521302		X				X			
C0523444	X	X				X			
C0523891	X	X				X			
C0542331	X	X				X			
C0545131		X				X			
C0546817				X					
C0546866	X					X			
C0546884		X				X			
C0554309		X	X			X			
C0554756		X				X			
C0558148				X					
C0559546		X				X			
C0572025		X				X			
C0574032	X			X		X			
C0577060				X					
C0577118		X				X			
C0580859				X					
C0581142		X				X			

	Included in public knowledge source					AFEP		SAFE	
C0585109		X				X			
C0587081		X				X			
C0587355		X				X			
C0587362		X				X			
C0591050		X				X			
C0596170		X				X			
C0597198		X				X			
C0597357	X	X				X			
C0600061	X	X				X			
C0600688	X					X			
C0677039	X					X			
C0677582		X				X			
C0679861	X					X			
C0681827		X				X			
C0684336	X	X				X			
C0700308		X				X			
C0700445		X				X			
C0724649		X				X			
C0728940	X					X			
C0732165		X				X			
C0740085	X	X				X			
C0740469				X					
C0741244	X	X				X			
C0742724		X				X			
C0859036		X				X			
C0868945	X	X				X			
C0949378		X				X			
C1137947	X	X				X			
C1140999	X					X			
C1145640		X				X			
C1171398		X				X			
C1256585	X	X		X		X	X	X	X
C1261287	X					X			
C1261720		X				X			
C1263666			X						
C1266240		X				X			
C1268852		X				X			
C1271104	X			X		X			

	Included in public knowledge source					AFEP		SAFE	
C1272641		X				X			
C1272919		X				X			
C1278293	X	X				X			
C1287298	X	X		X		X	X	X	
C1292856		X				X			
C1293861	X	X				X			
C1299583	X					X			
C1302112	X					X			
C1318478	X					X			
C1366678		X				X			
C1396851	X					X			
C1442858	X	X	X	X		X		X	
C1443036	X					X			
C1444662		X				X			
C1511237		X				X			
C1522240		X				X			
C1533693		X				X			
C1533734	X	X				X			
C1536696		X				X			
C1559265	X	X		X		X		X	
C1705480	X	X				X			
C1710425	X	X	X	X		X		X	
C1718097		X				X			
C1832073		X				X			
C1874188		X				X			
C1874190		X				X			
C1874191		X				X			
C1874271	X	X				X			
C1874288		X				X			
C1874289		X				X			
C1874292		X				X			
C1874295		X				X			
C1874882		X				X			
C1874911		X				X			
C1874953		X				X			
C1874955		X				X			
C1874970		X				X			
C1875099		X				X			

	Included in public knowledge source					AFEP		SAFE	
C1875100		X				X			
C1875146		X		X		X			
C1875186		X				X			
C1875409		X				X			
C1875410		X				X			
C1875417		X				X			
C1875522		X				X			
C1875542				X					
C1875577		X				X			
C1875579		X				X			
C1875643		X		X		X			
C1875728		X		X		X			
C1875761	X					X			
C1875865		X	X			X			
C1881049		X				X			
C1882365	X					X			
C1882443	X	X				X			
C1962945	X					X			
C1970989		X				X			
C2239176		X				X			
C2242979		X				X			
C2266920	X					X			
C2266943	X					X			
C2266959		X				X	X		
C2266960		X				X			
C2266971		X				X			
C2266972		X				X			
C2347023	X					X			
C2347080	X					X			
C2348813	X					X			
C2746010		X		X		X			
C2825032		X				X			
C2825050	X					X			
C2825091	X	X				X			
C2826616		X				X			
C2917331		X				X			
C2917342	X	X				X			
C2917344		X				X			

	Included in public knowledge source					AFEP		SAFE	
C2917403	X	X				X			
C2917419		X				X	X		
C2919641		X				X			
C2986592	X					X			
C2986642	X	X				X			
C2987634	X	X				X			
C3263722		X				X			
C3275118		X				X			
C3514012	X					X			
C3532188		X				X			
C3536742		X				X			
C3536752		X				X			
C3536808	X	X				X			
C3536825	X	X	X			X	X	X	
C3536828	X	X	X			X	X	X	
C3536840				X					
C3536843	X	X				X			
C3536888		X				X			
C3537198		X				X			
C3537226		X				X			
C3537240	X	X				X			

APPENDIX B: DETAILS ON RISK PREDICTION ALGORITHMS

Appendix Table 1: List of initial candidate predictor variables and ultimate variables chosen for the LASSO model after elimination of low prevalence and collinear variables.

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Gender	34785 (98.23%)	0 (0%)	X
Race		950 (2.68%)	
0 UNKNOWN	2037 (5.75%)		X
1 WHITE	25168 (71.07%)		X
2 BLACK	7191 (20.31%)		X
3 ASIAN-HAWAIIAN-PACIFIC ISLANDER	538 (1.52%)		X
4 AMERICAN INDIAN-ALASKAN NATIVE	478 (1.35%)		X
Age	61 (57.%)	0 (0%)	X
KDIGO Renal Failure Stage			
1	23920 (67.55%)	0 (0%)	X
2	5913 (16.7%)	0 (0%)	X
3	5579 (15.75%)	0 (0%)	X
Admit MELD	19.28 (15.13,23.67)	5313 (15.%)	X
Baseline Creatinine	1.13 (0.9,1.59)	0 (0%)	X
Admit Avg Creatinine	1.86 (1.39,2.8)	29 (0.08%)	X
Admit Avg Sodium	135. (131.,138.)	162 (0.46%)	X
Admit Avg Chloride	102. (97.5,106.5)	60 (0.17%)	X
Admit Avg Bicarbonate	23. (20.,26.)	59 (0.17%)	X
Admit Avg Calcium	8.45 (8.,8.88)	1467 (4.14%)	X
Admit Avg Blood Urea Nitrogen	34. (22.,52.)	2193 (6.19%)	X
Admit Avg Glucose	121.6 (101.7,157.)	138 (0.39%)	X
Admit Avg Hemoglobin	10.6 (9.2,12.15)	1153 (3.26%)	X
Admit Avg Hematocrit	31.35 (27.2,35.8)	189 (0.53%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit Avg White Blood Cell	7.7 (5.5,10.9)	202 (0.57%)	X
Admit Avg Platelet	116.3 (73.5,176.)	312 (0.88%)	X
Admit Avg Mean Corpuscular Volume	94.1 (88.75,99.7)	217 (0.61%)	X
Admit Avg Mean Corpuscular Hemoglobin Conc.	33.93 (33.1,34.65)	215 (0.61%)	X
Admit Avg Mean Corpuscular Hemoglobin	31.9 (29.7,34.15)	348 (0.98%)	X
Admit Avg Albumin	2.5 (2.,3.05)	3309 (9.34%)	X
Admit Avg Aspartate Aminotransferase	54.5 (34.,92.)	3490 (9.86%)	X
Admit Avg Alanine Aminotransferase	30. (18.,50.27)	2991 (8.45%)	X
Admit Avg Direct Bilirubin	0.42 (0.11,1.3)	21438 (60.54%)	X
Admit Avg Total Bilirubin	1.6 (0.8,3.4)	2751 (7.77%)	X
Admit Avg Alkaline Phosphatase	118. (82.,173.)	2716 (7.67%)	X
Admit Avg Prothrombin Time	16.55 (14.5,19.5)	5907 (16.68%)	X
Admit Avg Partial Thromboplastin Time	34.7 (31.3,39.1)	10577 (29.87%)	X
Admit Avg International Normalized Ratio	1.4 (1.22,1.7)	4706 (13.29%)	X
Admit Systolic Blood Pressure	116. (104.,129.)	613 (1.73%)	X
Admit Diastolic Blood Pressure	66. (59.,73.)	617 (1.74%)	X
Admit Mean Arterial Pressure	99. (90.,110.)	617 (1.74%)	X
Admit Temperature	97.72 (97.22,98.16)	692 (1.95%)	X
Admit Pulse	81.4 (71.33,92.)	598 (1.69%)	X
Admit Respirations	19. (18.,20.)	662 (1.87%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit Weight	193.6 (167.,209.9)	7707 (21.76%)	X
Admit Min Systolic Blood Pressure	100. (90.,113.)	613 (1.73%)	X
Admit Max Temperature	98.6 (98.,99.1)	692 (1.95%)	X
Admit Max Pulse	93. (81.,106.)	598 (1.69%)	X
Admit Max Resp	20. (20.,24.)	662 (1.87%)	X
Admit Weight Change	-2.44 (-9.28,4.52)	7885 (22.27%)	X
# Paracentesis in last 90 days	1. (1.,1.)	0 (0%)	X
Atrial Fibrillation	6233 (17.6%)	0 (0%)	X
Amyloidosis	84 (0.24%)	0 (0%)	X
Angina	3235 (9.14%)	0 (0%)	X
Anemia	24253 (68.49%)	0 (0%)	X
Arrhythmia	12836 (36.25%)	0 (0%)	X
Ascites	27836 (78.6%)	0 (0%)	
Asthma	2022 (5.71%)	0 (0%)	X
Autonomic Neuropathy	62 (0.18%)	0 (0%)	
Cancer	10729 (30.3%)	0 (0%)	X
Biliary Cirrhosis	181 (0.51%)	0 (0%)	X
Bone Marrow Transplant	9 (0.03%)	0 (0%)	
CABG	3083 (8.71%)	0 (0%)	X
Coronary Artery Disease	12410 (35.04%)	0 (0%)	X
Carotid Disease	133 (0.38%)	0 (0%)	X
Congestive Heart Failure	10172 (28.72%)	0 (0%)	X
Chronic Kidney Disease	13585 (38.36%)	0 (0%)	X
Colitis	300 (0.85%)	0 (0%)	X
Chronic Obstructive Pulmonary Disease	12683 (35.82%)	0 (0%)	X
Cerebrovascular Accident	4159 (11.74%)	0 (0%)	X
Dementia	1025 (2.89%)	0 (0%)	X
Dermatomyositis	9 (0.03%)	0 (0%)	
Dialysis	4191 (11.83%)	0 (0%)	X
Diabetes Mellitus	17980 (50.77%)	0 (0%)	X
Diabetic Neuropathy	4225 (11.93%)	0 (0%)	X
Dyslipidemia	14992 (42.34%)	0 (0%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Etoh Abuse	19001 (53.66%)	0 (0%)	X
Gastrointestinal Bleed	12073 (34.09%)	0 (0%)	X
Acute Glomerulonephritis	330 (0.93%)	0 (0%)	X
Glomerular Nephritis	911 (2.57%)	0 (0%)	X
Hepatocellular Carcinoma	4134 (11.67%)	0 (0%)	X
Hepatic Encephalopathy	11646 (32.89%)	0 (0%)	X
Viral Hepatitis	18548 (52.38%)	0 (0%)	X
HIV	722 (2.04%)	0 (0%)	X
Hospice	27 (0.08%)	0 (0%)	
Heart Transplant	63 (0.18%)	0 (0%)	
Hypertension	28271 (79.83%)	0 (0%)	X
Hyperparathyroidism	433 (1.22%)	0 (0%)	X
Leukemia	307 (0.87%)	0 (0%)	X
Lung Transplant	28 (0.08%)	0 (0%)	
Lupus	2122 (5.99%)	0 (0%)	X
Multiple Endocrine Neoplasia	4 (0.01%)	0 (0%)	
Myocardial Infarction	2674 (7.55%)	0 (0%)	X
Multiple Myeloma	202 (0.57%)	0 (0%)	X
Multiple Sclerosis	47 (0.13%)	0 (0%)	
Myopathies	3248 (9.17%)	0 (0%)	X
NAFLD	5390 (15.22%)	0 (0%)	X
Glomerular Nephritis, NOS	2800 (7.91%)	0 (0%)	X
Nephrectomy	165 (0.47%)	0 (0%)	X
Osteoarthritis	8919 (25.19%)	0 (0%)	X
Obesity	7607 (21.48%)	0 (0%)	X
Palliative Care	4159 (11.74%)	0 (0%)	X
Parkinsons	273 (0.77%)	0 (0%)	X
Pancreatitis	3964 (11.19%)	0 (0%)	X
Hemi- or Paraplegia	882 (2.49%)	0 (0%)	X
Porphyria	104 (0.29%)	0 (0%)	X
Peptic Ulcer Disease	3879 (10.95%)	0 (0%)	X
Peripheral Vascular Disease	3621 (10.23%)	0 (0%)	X
Rheumatoid Arthritis	461 (1.3%)	0 (0%)	X
Renal Transplant	169 (0.48%)	0 (0%)	X
Rheumatic Disease (Charlson Comorbidity Index Definition)	715 (2.02%)	0 (0%)	X
Spontaneous Bacterial Peritonitis	4691 (13.25%)	0 (0%)	X
Scleroderma	28 (0.08%)	0 (0%)	

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Sickle Cell Anemia	48 (0.14%)	0 (0%)	
Spinal Cord Injury	247 (0.7%)	0 (0%)	X
STEMI	1783 (5.04%)	0 (0%)	X
Stroke	2657 (7.5%)	0 (0%)	X
Tuberculosis	416 (1.17%)	0 (0%)	X
Transient Ischemic Attack	902 (2.55%)	0 (0%)	X
TIPS	599 (1.69%)	0 (0%)	X
Tobacco Use	16956 (47.88%)	0 (0%)	X
Urinary Obstruction	4598 (12.98%)	0 (0%)	X
Varices	9129 (25.78%)	0 (0%)	X
Valvular Heart Disease	3736 (10.55%)	0 (0%)	X
Home Med Rifaximin	2166 (6.12%)	0 (0%)	X
Home Med Lactulose	8439 (23.83%)	0 (0%)	X
Home Med Antidotes Deterrents And Poison Control	995 (2.81%)	0 (0%)	X
Home Med Alcohol Deterrents	48 (0.14%)	0 (0%)	
Home Med Heavy Metal Antagonists	15 (0.04%)	0 (0%)	
Home Med Antidotes Deterrents And Poison Control Exchange Resins	276 (0.78%)	0 (0%)	X
Home Med Antidotes Deterrents Other	664 (1.88%)	0 (0%)	X
Home Med Antihistamines	4730 (13.36%)	0 (0%)	X
Home Med Antihistamines Phenothiazine	806 (2.28%)	0 (0%)	X
Home Med Antihistamines Ethanolamine	961 (2.71%)	0 (0%)	X
Home Med Antihistamines Alkylamine	49 (0.14%)	0 (0%)	
Home Med Antihistamines Piperazine	1876 (5.3%)	0 (0%)	X
Home Med Antihistamines Butyrophenone	49 (0.14%)	0 (0%)	
Home Med Antihistamines Piperidine	104 (0.29%)	0 (0%)	X
Home Med Antihistamines Other	1461 (4.13%)	0 (0%)	X
Home Med Antimicrobials	9133 (25.79%)	0 (0%)	X
Home Med Penicillin G Related Penicillins	40 (0.11%)	0 (0%)	
Home Med Penicillins Amino Derivatives	858 (2.42%)	0 (0%)	X
Home Med Penicillinase Resistant Penicillins	47 (0.13%)	0 (0%)	

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Home Med Extended Spectrum Penicillins	10 (0.03%)	0 (0%)	
Home Med Penicillins And Beta Lactam Antimicrobials	1553 (4.39%)	0 (0%)	X
Home Med Cephalosporin 1st Generation	449 (1.27%)	0 (0%)	X
Home Med Cephalosporin 2nd Generation	62 (0.18%)	0 (0%)	
Home Med Cephalosporin 3rd Generation	122 (0.34%)	0 (0%)	X
Home Med Cephalosporin 4th Generation	2 (0.01%)	0 (0%)	
Home Med Beta Lactams Antimicrobials Other	2 (0.01%)	0 (0%)	
Home Med Erythromycins Macrolides	382 (1.08%)	0 (0%)	X
Home Med Tetracyclines	370 (1.04%)	0 (0%)	X
Home Med Aminoglycosides	398 (1.12%)	0 (0%)	X
Home Med Lincomycins	249 (0.7%)	0 (0%)	X
Home Med Quinolones	4125 (11.65%)	0 (0%)	X
Home Med Antituberculars	93 (0.26%)	0 (0%)	X
Home Med Methenamine Salts Antimicrobials	9 (0.03%)	0 (0%)	
Home Med Nitrofurans Antimicrobials	85 (0.24%)	0 (0%)	X
Home Med Sulfonamide Related Antimicrobials	1123 (3.17%)	0 (0%)	X
Home Med Antifungals	490 (1.38%)	0 (0%)	X
Home Med Antivirals	774 (2.19%)	0 (0%)	X
Home Med Anti Infectives Other	3051 (8.62%)	0 (0%)	X
Home Med Antineoplastics	690 (1.95%)	0 (0%)	X
Home Med Antineoplastics Alkylating Agents	10 (0.03%)	0 (0%)	
Home Med Antineoplastic Antibiotics	0 (0.%)	0 (0%)	
Home Med Antineoplastics Antimetabolites	98 (0.28%)	0 (0%)	X
Home Med Antineoplastic Hormones	121 (0.34%)	0 (0%)	X
Home Med Protective Agents	0 (0.%)	0 (0%)	
Home Med Antineoplastic Other	502 (1.42%)	0 (0%)	X
Home Med Antiparasitics	284 (0.8%)	0 (0%)	X
Home Med Antiprotozoals	255 (0.72%)	0 (0%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Home Med Antimalarials	234 (0.66%)	0 (0%)	X
Home Med Antiprotozoals Other	21 (0.06%)	0 (0%)	
Home Med Anthelmintics	5 (0.01%)	0 (0%)	
Home Med Pediculicides	24 (0.07%)	0 (0%)	
Home Med Antiparasitics Other	0 (0.%)	0 (0%)	
Home Med Antiseptics Disinfectants	1 (0.%)	0 (0%)	
Home Med Autonomic Medications	995 (2.81%)	0 (0%)	X
Home Med Sympathomimetics .adrenergics.	19 (0.05%)	0 (0%)	
Home Med Sympatholytics	2 (0.01%)	0 (0%)	
Home Med Parasympathomimetics Cholinergics	762 (2.15%)	0 (0%)	X
Home Med Parasympatholytics	220 (0.62%)	0 (0%)	X
Home Med Autonomic Agents Other	9 (0.03%)	0 (0%)	
Home Med Blood Products Modifiers Volume Expanders	3921 (11.07%)	0 (0%)	X
Home Med Anticoagulants	1713 (4.84%)	0 (0%)	X
Home Med Thrombolytics	2 (0.01%)	0 (0%)	
Home Med Antihemorrhagics	26 (0.07%)	0 (0%)	
Home Med Platelet Aggregation Inhibitors	1222 (3.45%)	0 (0%)	X
Home Med Blood Formation Products	1190 (3.36%)	0 (0%)	X
Home Med Blood Derivatives	30 (0.08%)	0 (0%)	
Home Med Volume Expanders	1 (0.%)	0 (0%)	
Home Med Central Nervous System Medications	19537 (55.17%)	0 (0%)	X
Home Med Analgesics	14293 (40.36%)	0 (0%)	X
Home Med Opioid Analgesics	9892 (27.93%)	0 (0%)	X
Home Med Opioid Antagonist Analgesics	46 (0.13%)	0 (0%)	
Home Med Non Opioid Analgesics	7363 (20.79%)	0 (0%)	X
Home Med Antimigraine Agents	38 (0.11%)	0 (0%)	
Home Med Anesthetics	9 (0.03%)	0 (0%)	
Home Med General Anesthetics Other	1 (0.%)	0 (0%)	
Home Med Local Anesthetics Injection	8 (0.02%)	0 (0%)	
Home Med Anesthetic Adjuncts	0 (0.%)	0 (0%)	
Home Med Sedatives Hypontics	4019 (11.35%)	0 (0%)	X
Home Med Barbituric Acid Derivative Sedatives Hypnotics	29 (0.08%)	0 (0%)	

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Home Med Benzodiazepine Derivative Sedatives Hypnotics	2777 (7.84%)	0 (0%)	X
Home Med Sedatives Hypnotics Other	1457 (4.11%)	0 (0%)	X
Home Med Anticonvulsants	4746 (13.4%)	0 (0%)	X
Home Med Antiparkinson Agents	388 (1.1%)	0 (0%)	X
Home Med Antivertigo Agents	138 (0.39%)	0 (0%)	X
Home Med Antidepressants	9435 (26.64%)	0 (0%)	X
Home Med Tricyclic Antidepressants	1187 (3.35%)	0 (0%)	X
Home Med Monamine Oxidase Inhibitor Antidepressants	2 (0.01%)	0 (0%)	
Home Med Antidepressants Other	8764 (24.75%)	0 (0%)	X
Home Med Antipsychotics	2146 (6.06%)	0 (0%)	X
Home Med Phenothiazine Related Antipsychotics	110 (0.31%)	0 (0%)	X
Home Med Antipsychotics Other	2061 (5.82%)	0 (0%)	X
Home Med Lithium Salts	97 (0.27%)	0 (0%)	X
Home Med Cns Stimulants	70 (0.2%)	0 (0%)	
Home Med Amphetamines	6 (0.02%)	0 (0%)	
Home Med Amphetamine Like Stimulants	50 (0.14%)	0 (0%)	
Home Med Cns Stimulants Other	14 (0.04%)	0 (0%)	
Home Med Cns Medications Other	452 (1.28%)	0 (0%)	X
Home Med Cardiovascular Medications	27201 (76.81%)	0 (0%)	X
Home Med Digitalis Glycosides	1423 (4.02%)	0 (0%)	X
Home Med Beta Blockers Related	16825 (47.51%)	0 (0%)	X
Home Med Alpha Blockers Related	4450 (12.57%)	0 (0%)	X
Home Med Calcium Channel Blockers	5194 (14.67%)	0 (0%)	X
Home Med Antianginals	2718 (7.68%)	0 (0%)	X
Home Med Antiarrhythmics	546 (1.54%)	0 (0%)	X
Home Med Antilipemic Agents	6364 (17.97%)	0 (0%)	X
Home Med Antihypertensive Combinations	446 (1.26%)	0 (0%)	X
Home Med Antihypertensives Other	1944 (5.49%)	0 (0%)	X
Home Med Peripheral Vasodilators	2 (0.01%)	0 (0%)	
Home Med Diuretics	21146 (59.71%)	0 (0%)	X
Home Med Thiazides Related Diuretics	2698 (7.62%)	0 (0%)	X
Home Med Loop Diuretics	18419 (52.01%)	0 (0%)	X
Home Med Carbonic Anhydrase Inhibitor Diuretics	14 (0.04%)	0 (0%)	

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Home Med Potassium Sparing Combinations Diuretics	14719 (41.57%)	0 (0%)	X
Home Med Diuretics Other	0 (0.%)	0 (0%)	
Home Med Ace Inhibitors	7725 (21.81%)	0 (0%)	X
Home Med Angiotensin Ii Inhibitor	1504 (4.25%)	0 (0%)	X
Home Med Direct Renin Inhibitor	5 (0.01%)	0 (0%)	
Home Med Cardiovascular Agents Other	613 (1.73%)	0 (0%)	X
Home Med Dermatological Agents	7574 (21.39%)	0 (0%)	X
Home Med Anti Infective Topical	3054 (8.62%)	0 (0%)	X
Home Med Antibacterial Topical	1352 (3.82%)	0 (0%)	X
Home Med Antifungal Topical	1731 (4.89%)	0 (0%)	X
Home Med Antiviral Topical	18 (0.05%)	0 (0%)	
Home Med Anti Infective Topical Other	268 (0.76%)	0 (0%)	X
Home Med Anti Inflammatory Topical	2017 (5.7%)	0 (0%)	X
Home Med Anti Infective Anti Inflammatory Combinations Topical	82 (0.23%)	0 (0%)	X
Home Med Sun Protectants Screens Topical	55 (0.16%)	0 (0%)	
Home Med Emollients	2413 (6.81%)	0 (0%)	X
Home Med Soaps Shampoos Soap Free Cleansers	324 (0.91%)	0 (0%)	X
Home Med Deodorants Antiperspirants Topical	3 (0.01%)	0 (0%)	
Home Med Keratolytics Caustics Topical	30 (0.08%)	0 (0%)	
Home Med Antineoplastic Topical	27 (0.08%)	0 (0%)	
Home Med Analgesics Topical	787 (2.22%)	0 (0%)	X
Home Med Local Anesthetics Topical	562 (1.59%)	0 (0%)	X
Home Med Antiacne Agents	132 (0.37%)	0 (0%)	X
Home Med Antiacne Agents Systemic	0 (0.%)	0 (0%)	
Home Med Antiacne Agents Topical	132 (0.37%)	0 (0%)	
Home Med Antipsoriatic	185 (0.52%)	0 (0%)	X
Home Med Antipsoriatics Systemic	24 (0.07%)	0 (0%)	
Home Med Antipsoriatics Topical	163 (0.46%)	0 (0%)	X
Home Med Dermatologicals Systemic Other	1 (0.%)	0 (0%)	
Home Med Dermatologicals Topical Other	1543 (4.36%)	0 (0%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit IVF Total	350. (0.,1250.)	0 (0%)	X
Admit Octreotide	1803 (5.09%)	0 (0%)	X
Admit Midodrine	478 (1.35%)	0 (0%)	X
Admit Albumin Inf	6042 (17.06%)	0 (0%)	X
Admit Norepinephrine	712 (2.01%)	0 (0%)	X
Admit Vasopressin	153 (0.43%)	0 (0%)	X
Admit Rifaximin	2888 (8.16%)	0 (0%)	X
Admit Med Class Nsaids	806 (2.28%)	0 (0%)	X
Admit Med Class Aminoglycosides	796 (2.25%)	0 (0%)	X
Admit Med Class Betablockers	15805 (44.63%)	0 (0%)	
Admit Med Class Ace	3091 (8.73%)	0 (0%)	
Admit Med Class Arb	684 (1.93%)	0 (0%)	
Admit Med Class Glucocorticoids	2681 (7.57%)	0 (0%)	X
Admit Med Class Ksparingdiuretic	7890 (22.28%)	0 (0%)	
Admit Med Class Benzodiazepines	5798 (16.37%)	0 (0%)	
Admit Med Class Statins	5310 (14.99%)	0 (0%)	X
Admit Med Class Insulin	12004 (33.9%)	0 (0%)	X
Admit Med Class Fluoroquinolones	4147 (11.71%)	0 (0%)	X
Admit Med Class Opioids	19447 (54.92%)	0 (0%)	X
Admit Med Va Class Antidotes Deterrents And Poison Control	0 (0.%)	0 (0%)	
Admit Med Va Class Alcohol Deterrents	21 (0.06%)	0 (0%)	
Admit Med Va Class Heavy Metal Antagonists	6 (0.02%)	0 (0%)	
Admit Med Va Class Antidotes Deterrents And Poison Control Exchange Resins	3502 (9.89%)	0 (0%)	X
Admit Med Va Class Antidotes Deterrents Other	188 (0.53%)	0 (0%)	X
Admit Med Va Class Antihistamines	0 (0.%)	0 (0%)	
Admit Med Va Class Antihistamines Phenothiazine	974 (2.75%)	0 (0%)	X
Admit Med Va Class Antihistamines Ethanolamine	1895 (5.35%)	0 (0%)	X
Admit Med Va Class Antihistamines Alkylamine	12 (0.03%)	0 (0%)	
Admit Med Va Class Antihistamines Piperazine	991 (2.8%)	0 (0%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit Med Va Class Antihistamines Butyrophenone	23 (0.06%)	0 (0%)	
Admit Med Va Class Antihistamines Piperidine	61 (0.17%)	0 (0%)	
Admit Med Va Class Antihistamines Other	823 (2.32%)	0 (0%)	X
Admit Med Va Class Antimicrobials	0 (0.%)	0 (0%)	
Admit Med Va Class Penicillin G Related Penicillins	51 (0.14%)	0 (0%)	
Admit Med Va Class Penicillins Amino Derivatives	996 (2.81%)	0 (0%)	X
Admit Med Va Class Penicillinase Resistant Penicillins	125 (0.35%)	0 (0%)	X
Admit Med Va Class Extended Spectrum Penicillins	5186 (14.64%)	0 (0%)	X
Admit Med Va Class Penicillins And Beta Lactam Antimicrobials	0 (0.%)	0 (0%)	
Admit Med Va Class Cephalosporin 1st Generation	695 (1.96%)	0 (0%)	X
Admit Med Va Class Cephalosporin 2nd Generation	134 (0.38%)	0 (0%)	X
Admit Med Va Class Cephalosporin 3rd Generation	5955 (16.82%)	0 (0%)	X
Admit Med Va Class Erythromycins Macrolides	1337 (3.78%)	0 (0%)	X
Admit Med Va Class Tetracyclines	350 (0.99%)	0 (0%)	X
Admit Med Va Class Aminoglycosides	753 (2.13%)	0 (0%)	X
Admit Med Va Class Lincomycins	544 (1.54%)	0 (0%)	X
Admit Med Va Class Quinolones	5703 (16.1%)	0 (0%)	X
Admit Med Va Class Antituberculars	80 (0.23%)	0 (0%)	X
Admit Med Va Class Nitrofurans Antimicrobials	19 (0.05%)	0 (0%)	
Admit Med Va Class Sulfonamide Related Antimicrobials	597 (1.69%)	0 (0%)	X
Admit Med Va Class Antifungals	861 (2.43%)	0 (0%)	X
Admit Med Va Class Antivirals	774 (2.19%)	0 (0%)	X
Admit Med Va Class Anti Infectives Other	10306 (29.1%)	0 (0%)	X
Admit Med Va Class Antineoplastics	0 (0.%)	0 (0%)	

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit Med Va Class Antineoplastics Alkylating Agents	4 (0.01%)	0 (0%)	
Admit Med Va Class Antineoplastics Antimetabolites	33 (0.09%)	0 (0%)	
Admit Med Va Class Antineoplastic Hormones	6 (0.02%)	0 (0%)	
Admit Med Va Class Antineoplastic Other	227 (0.64%)	0 (0%)	X
Admit Med Va Class Antiparasitics	0 (0.%)	0 (0%)	
Admit Med Va Class Antiprotozoals	0 (0.%)	0 (0%)	
Admit Med Va Class Antimalarials	151 (0.43%)	0 (0%)	X
Admit Med Va Class Antiprotozoals Other	30 (0.08%)	0 (0%)	
Admit Med Va Class Pediculicides	30 (0.08%)	0 (0%)	
Admit Med Va Class Autonomic Medications	0 (0.%)	0 (0%)	
Admit Med Va Class Parasympathomimetics Cholinergics	1022 (2.89%)	0 (0%)	X
Admit Med Va Class Parasympatholytics	221 (0.62%)	0 (0%)	X
Admit Med Va Class Autonomic Agents Other	6 (0.02%)	0 (0%)	
Admit Med Va Class Blood Products Modifiers Volume Expanders	0 (0.%)	0 (0%)	
Admit Med Va Class Anticoagulants	11162 (31.52%)	0 (0%)	X
Admit Med Va Class Antihemorrhagics	25 (0.07%)	0 (0%)	
Admit Med Va Class Platelet Aggregation Inhibitors	1103 (3.11%)	0 (0%)	X
Admit Med Va Class Blood Formation Products	541 (1.53%)	0 (0%)	X
Admit Med Va Class Blood Derivatives	6062 (17.12%)	0 (0%)	X
Admit Med Va Class Volume Expanders	126 (0.36%)	0 (0%)	X
Admit Med Va Class Central Nervous System Medications	0 (0.%)	0 (0%)	
Admit Med Va Class Analgesics	0 (0.%)	0 (0%)	
Admit Med Va Class Opioid Analgesics	14915 (42.12%)	0 (0%)	X
Admit Med Va Class Opioid Antagonist Analgesics	222 (0.63%)	0 (0%)	X
Admit Med Va Class Non Opioid Analgesics	9986 (28.2%)	0 (0%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit Med Va Class Antimigraine Agents	4 (0.01%)	0 (0%)	
Admit Med Va Class Anesthetics	0 (0.%)	0 (0%)	
Admit Med Va Class Local Anesthetics Injection	362 (1.02%)	0 (0%)	X
Admit Med Va Class Sedatives Hypontics	1 (0.%)	0 (0%)	
Admit Med Va Class Barbituric Acid Derivative Sedatives Hypnotics	60 (0.17%)	0 (0%)	
Admit Med Va Class Benzodiazepine Derivative Sedatives Hypnotics	5798 (16.37%)	0 (0%)	X
Admit Med Va Class Sedatives Hypnotics Other	1232 (3.48%)	0 (0%)	X
Admit Med Va Class Anticonvulsants	4889 (13.81%)	0 (0%)	X
Admit Med Va Class Antiparkinson Agents	363 (1.03%)	0 (0%)	X
Admit Med Va Class Antivertigo Agents	48 (0.14%)	0 (0%)	
Admit Med Va Class Antidepressants	0 (0.%)	0 (0%)	
Admit Med Va Class Tricyclic Antidepressants	863 (2.44%)	0 (0%)	X
Admit Med Va Class Antidepressants Other	8512 (24.04%)	0 (0%)	X
Admit Med Va Class Antipsychotics	0 (0.%)	0 (0%)	
Admit Med Va Class Phenothiazine Related Antipsychotics	117 (0.33%)	0 (0%)	X
Admit Med Va Class Antipsychotics Other	2612 (7.38%)	0 (0%)	X
Admit Med Va Class Lithium Salts	82 (0.23%)	0 (0%)	X
Admit Med Va Class Cns Stimulants	0 (0.%)	0 (0%)	
Admit Med Va Class Amphetamine Like Stimulants	35 (0.1%)	0 (0%)	
Admit Med Va Class Cns Stimulants Other	5 (0.01%)	0 (0%)	
Admit Med Va Class Cns Medications Other	425 (1.2%)	0 (0%)	X
Admit Med Va Class Cardiovascular Medications	0 (0.%)	0 (0%)	
Admit Med Va Class Digitalis Glycosides	1118 (3.16%)	0 (0%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit Med Va Class Beta Blockers Related	15805 (44.63%)	0 (0%)	X
Admit Med Va Class Alpha Blockers Related	3815 (10.77%)	0 (0%)	X
Admit Med Va Class Calcium Channel Blockers	3892 (10.99%)	0 (0%)	X
Admit Med Va Class Antianginals	2018 (5.7%)	0 (0%)	X
Admit Med Va Class Antiarrhythmics	702 (1.98%)	0 (0%)	X
Admit Med Va Class Antilipemic Agents	5855 (16.53%)	0 (0%)	X
Admit Med Va Class Antihypertensive Combinations	93 (0.26%)	0 (0%)	X
Admit Med Va Class Antihypertensives Other	2365 (6.68%)	0 (0%)	X
Admit Med Va Class Diuretics	0 (0.%)	0 (0%)	
Admit Med Va Class Thiazides Related Diuretics	1074 (3.03%)	0 (0%)	X
Admit Med Va Class Loop Diuretics	12466 (35.2%)	0 (0%)	X
Admit Med Va Class Carbonic Anhydrase Inhibitor Diuretics	16 (0.05%)	0 (0%)	
Admit Med Va Class Potassium Sparing Combinations Diuretics	7890 (22.28%)	0 (0%)	X
Admit Med Va Class Ace Inhibitors	3091 (8.73%)	0 (0%)	X
Admit Med Va Class Angiotensin Ii Inhibitor	684 (1.93%)	0 (0%)	X
Admit Med Va Class Cardiovascular Agents Other	847 (2.39%)	0 (0%)	X
Admit Med Va Class Dermatological Agents	0 (0.%)	0 (0%)	
Admit Med Va Class Anti Infective Topical	0 (0.%)	0 (0%)	
Admit Med Va Class Antibacterial Topical	1116 (3.15%)	0 (0%)	X
Admit Med Va Class Antifungal Topical	1746 (4.93%)	0 (0%)	X
Admit Med Va Class Antiviral Topical	8 (0.02%)	0 (0%)	
Admit Med Va Class Anti Infective Topical Other	184 (0.52%)	0 (0%)	X
Admit Med Va Class Anti Inflammatory Topical	1260 (3.56%)	0 (0%)	X

Variable Name	n (%) or median (IQR)	Missing, n (%)	Candidate Variable in Final Model
Admit Med Va Class Anti Infective Anti Inflammatory Combinations Topical	43 (0.12%)	0 (0%)	
Admit Med Va Class Sun Protectants Screens Topical	3 (0.01%)	0 (0%)	
Admit Med Va Class Emollients	1540 (4.35%)	0 (0%)	X
Admit Med Va Class Soaps Shampoos Soap Free Cleansers	138 (0.39%)	0 (0%)	X
Admit Med Va Class Keratolytics Caustics Topical	12 (0.03%)	0 (0%)	
Admit Med Va Class Antineoplastic Topical	19 (0.05%)	0 (0%)	
Admit Med Va Class Analgesics Topical	364 (1.03%)	0 (0%)	X
Admit Med Va Class Local Anesthetics Topical	235 (0.66%)	0 (0%)	X
Admit Med Va Class Antiacne Agents	0 (0.%)	0 (0%)	
Admit Med Va Class Antiacne Agents Topical	57 (0.16%)	0 (0%)	
Admit Med Va Class Antipsoriatic	2 (0.01%)	0 (0%)	
Admit Med Va Class Antipsoriatics Systemic	11 (0.03%)	0 (0%)	
Admit Med Va Class Antipsoriatics Topical	114 (0.32%)	0 (0%)	X
Admit Med Va Class Dermatologicals Topical Other	1122 (3.17%)	0 (0%)	X
Admit Procedure Dialysis	2569 (7.25%)	0 (0%)	X
Admit Proc Liver Transplant	23 (0.06%)	0 (0%)	
Admit Procedure TIPS	84 (0.24%)	0 (0%)	X
Admit Procedure Variceal Control	672 (1.9%)	0 (0%)	X
Admit Procedure Paracentesis	7814 (22.07%)	0 (0%)	X
Admit Procedure Cardiac Cath	165 (0.47%)	0 (0%)	X
Admit Condition SBP	2654 (7.5%)	0 (0%)	X
3 Day Pre Admit Procedure Dialysis	0 (0.%)	0 (0%)	
3 Day Pre Admit Procedure Paracentesis	617 (1.74%)	0 (0%)	X

Appendix Table 1b: Code definitions for co-morbid conditions used in the model. (ICD-9: International Classification of Diseases-Version 9; CPT: Current Procedural Terminology; ICDProc: ICD Procedure Code)

Description	ICD-9	CPT	ICDProc
Acute Glomerulonephritis	580.*		
Etoh Abuse	291.[0123589],303.*,305.0*		
Amyloidosis	277.3,277.3[019]		
Anemia	280*,281*,282.01,282.2*,282.3*,282.4*, 282.71,282.8,282.9,283.[019]*,284*,285*, 648.2*,776.5*		
Angina	413*,411.1		
Arrhythmia	427*,785.0,785.1,779.81,426*,V45.0*,V53.3*,746.86		
Ascites	789.5*	4908[0-3]	54.91
Asthma	493.*		
Atrial Fibulation	427.3[12]		
Autonomic Neuropathy	337.9		
Cancer	1[4-9][0-9]*,20[0-8]*,209.[0-3]*,23[0-3]*		
Biliary Cirrhosis	571.6		
Bone Marrow Transplant	996.8[58],V42.8[12]	3824[012]	41.00*
CABG	V45.81,414.04	3351[012346789], 3352[123], 3353[3-6]	36.1*, 36.2*
Carotid Disease	433.1		38.12
Cerebrovascular Disease	43[0-8]*,362.34		
Congestive Heart Failure	398.91,402.11,404.01,404.11,404.91,428*, 402.01,402.91,404.13,404.93,425.[145789]*		
Chronic Kidney Disease	585*,403*,404*		
Colitis	555.[0129],556.,556.[0-6]		
Chronic Obstructive Pulmonary Disease	491.*,492.*,493.*,496.*,V17.5*,V81.3*		

Description	ICD-9	CPT	ICDProc
Coronary Artery Disease	410.*,411.*,412.*,413.*, 414.[02-9]*,V45.81,V45.82		
Dementia	290.*,294.[1]*,331.[012]*		
Dermatomyositis	710.3		
Diabetes Mellitus	249*,250*,357.2*,362.0*,366.41,V45.85, V53.91		
Diabetic Neuropathy	357.2		
Dialysis	585.6,V39.27,V39.42,V39.43,V45.1,V56.0 , V56.2,V56.31,V56.32,V56.8	90921,90925, 90935,90937, 90945,9096[012 6],G8956,90947, 90989,9099[39]	39.9[35], 54.98
Dyslipidemia	272.*		
Gastrointestinal Bleeding	530.82,53[1-4].[0246]0,535.[045]1,578.*		
Glomerular Nephritis (Exclusion)	580.[049],580.8[19],581.[0123],582.[01 249], 582.8[19],583.[0124],581.89		
Heart Transplant	V42.1		37.5[1-5]
Hemiplegia or Paraplegia	334.1*,342.*,343.*,344.[01234569]*		
Hepatic Encephalopathy	572.2*,070.00,070.2*,070.40,070.41,070 .44, 070.49,070.60		
Viral Hepatitis	070.[23][0-3],V02.6[12],070.[45][14],070.7[01]		
Hepatocellular Carcinoma	155.0		
HIV	04[234]*,079.53,795.71,V08*		
Hospice		9,937,799,378	
Hyperparathyroidism	252.0*		
Hypertension	401*,402*,403*,404*,405*,437.2*		
Leukemia	202.4*,203.1*,20[4-8].*,V10.6*		
Lung Transplant	V42.6		33.50*
Multiple Endocrine Neoplasia	258.0*		
Multiple Myeloma	203.0*		
Multiple Sclerosis	340.		
Myocardial Infarction	410*		
Myopathies	359.8,359.89,425.4		

Description	ICD-9	CPT	ICDProc
Nephrectomy		5022[05], 5023[046],50240, 50300,50320, 50340,50370, 5054[35-8]	55.4, 55.5[1-4]
Glomerular Nephritis, NOS	580.81,58[03].9,583.8[19]		
NAFLD	571.8,571.9		
Obesity	278.0,278.0[01],649.1,278.03		
Osteoarthritis	715.0[049],715.1[0-8],715.2[0-8], 715.3[0-8],715.8[09],715.9[0-8],V13.4		
Pallative Care	V66.7*		
Pancreatitis	577.[01]*		
Parkinson's Disease	332.0		
Peptic Ulcer Disease	531.*,532.*,533.*,534.*		
Peripheral Vascular Disease	440*,441*,442*,444.2*,V43.4		
Porphyria	277.1		
Renal Transplant	996.81,V42.0	5,036,550,360	55.69*, 00.9[123]
Rheumatic Disease (Charlson Comorbidity Definition)	446.5*,710.[01234]*,714.[0128]*,725.*		
Rheumatoid Arthritis	714.0		
Scleroderma	701.0		
Sickle Cell Disease	282.4[1-4]*,282.6*		
Spinal Cord Injury	349.39,806.[0-3][0-9], 806.[4589], 806.[67][0129],907.2, 952.[01][0-9], 952.[23489]		
Spontaneous Bacterial Peritonitis	567.23,567.[0289]0,567.2[19],567.89, 567.[0289]		
STEMI	410.[012345689]*		
Stroke	43[01]*,434.[019],434.[019]1,436*,997.02		
Lupus	286.5,323.81,517.8,58[023].81,695.4,710.0		
Transient Ischemic Attack	435.[89]		

Description	ICD-9	CPT	ICDProc
Tobacco Use	305.1*,V15.82		
TIPS		3718[23]	39.1
Tuberculosis	01[0-8].*,137.*,V12.01		
Urinary Obstruction	592.1,593.4,594.[29],596.0,598.[1289],599.6, 599.69,599.82,600.[0129]1,753.[26], 753.2[129],788.2,788.29,V44.6,V55.6		
Valvular Heart Disease	424.[0-3]		
Varices	456.[012][01],456.[012]	4324[34],4320[45]	42.33

Appendix Table 2: Variables used in the multiple imputation of laboratory values.

Multiple imputation was carried out by the mi package for the R statistical programming software. Imputation was carried out for 30 iterations using 4 separate chains. Imputations were carried out to convergence. Values from the four separate chains were averaged together for the final imputed values used in the dataset.

Variable Name
Gender
Race
Age
Admit MELD
Baseline Creatinine
Admit Avg Creatinine
Admit Avg Sodium
Admit Avg Chloride
Admit Avg Bicarbonate
Admit Avg Calcium
Admit Avg Blood Urea Nitrogen
Admit Avg Glucose
Admit Avg Hemoglobin
Admit Avg Hematocrit
Admit Avg White Blood Cell
Admit Avg Platelet
Admit Avg Mean Corpuscular Volume
Admit Avg Mean Corpuscular Hemoglobin Conc.
Admit Avg Mean Corpuscular Hemoglobin
Admit Avg Albumin
Admit Avg Aspartate Aminotransferase
Admit Avg Alanine Aminotransferase
Admit Avg Direct Bilirubin
Admit Avg Total Bilirubin
Admit Avg Alkaline Phosphatase
Admit Avg Prothrombin Time
Admit Avg Partial Thromboplastin Time
Admit Avg International Normalized Ratio
Atrial Fibrillation
Anemia
Ascites
Cancer
Biliary Cirrhosis
Coronary Artery Disease

Congestive Heart Failure
Chronic Kidney Disease
Dialysis
Diabetes Mellitus
Etoh Abuse
Gastrointestinal Bleed
Hepatocellular Carcinoma
Hepatic Encephalopathy
Viral Hepatitis
HIV
Hypertension
NAFLD
Spontaneous Bacterial Peritonitis
TIPS
Varices
Paracentesis in last 90 days
Home Med Rifaximin
Home Med Lactulose
Home Med Quinolones
Home Med Anticoagulants
Home Med Platelet Aggregation Inhibitors
Home Med Opioid Analgesics
Home Med Sedatives Hypontics
Home Med Anticonvulsants
Home Med Antidepressants
Home Med Digitalis Glycosides
Home Med Beta Blockers Related
Home Med Alpha Blockers Related
Home Med Calcium Channel Blockers
Home Med Antiarrhythmics
Home Med Antilipemic Agents
Home Med Thiazides Related Diuretics
Home Med Loop Diuretics
Home Med Potassium Sparing Combinations Diuretics
Home Med Ace Inhibitors
Home Med Angiotensin Ii Inhibitor
Admit Procedure Dialysis
Admit Procedure TIPS
Admit Procedure Variceal Control
Admit Procedure Paracentesis

Appendix Table 3: GEE Odds ratios and confidence intervals for the variables selected by LASSO yielding the minimum deviance.

Variable	Odds Ratio (95% CI)
(Intercept)	2.733
Race White	0.804 (0.67,0.965)
Race Black	0.639 (0.515,0.793)
Race Asian	0.481 (0.3,0.772)
Race Native American	0.644 (0.407,1.02)
Age	1.012 (1,1.02)
KDIGO Stage 2 Renal Failure	1.188 (1.04,1.35)
KDIGO Stage 3 Renal Failure	0.942 (0.774,1.15)
Admit IVF Total	0.942 (0.898,0.987)
Admit MELD	1.169 (1.15,1.19)
Baseline Creatinine	0.77 (0.711,0.834)
Admit Avg Creatinine	1.066 (1.02,1.11)
Admit Avg Sodium	0.971 (0.955,0.988)
Admit Avg Chloride	1.016 (0.999,1.03)
Admit Avg Bicarbonate	0.992 (0.974,1.01)
Admit Avg Calcium	0.978 (0.913,1.05)
Admit Avg Blood Urea Nitrogen	1.005 (1,1.01)
Admit Avg Glucose	0.998 (0.997,0.999)
Admit Avg Hematocrit	1 (0.99,1.01)
Admit Avg Platelet	1 (0.999,1)
Admit Avg Mean Corpuscular Volume	0.996 (0.99,1)
Admit Avg Mean Corpuscular Hemoglobin Concentration	1.068 (1.02,1.12)
Admit Avg Albumin	0.918 (0.836,1.01)
Admit Avg Alanine Aminotransferase	0.999 (0.999,1)
Admit Avg Total Bilirubin	0.988 (0.978,0.998)
Admit Avg Alkaline Phosphatase	1 (1,1)

Variable	Odds Ratio (95% CI)
Admit Avg Partial Thromboplastin Time	0.995 (0.989,1)
Admit Avg International Normalized Ratio	0.568 (0.491,0.657)
Admit Mean Systolic Blood Pressure	0.985 (0.979,0.992)
Admit Mean Diastolic Blood Pressure	1.014 (1.01,1.02)
Admit Mean Temperature	0.942 (0.889,1)
Admit Mean Pulse	1 (0.993,1.01)
Admit Mean Weight	1.002 (1,1)
Admit Min Systolic Blood Pressure	1.006 (1,1.01)
Admit Max Pulse	0.999 (0.993,1)
# Paracentesis in last 90 days	1.073 (1.05,1.1)
Anemia	1.121 (0.997,1.26)
Arrhythmia	0.951 (0.845,1.07)
Asthma	0.732 (0.572,0.938)
Cancer	0.902 (0.804,1.01)
CABG	0.767 (0.592,0.993)
Coronary Artery Disease	0.946 (0.824,1.09)
Congestive Heart Failure	0.94 (0.806,1.09)
Dialysis	1.04 (0.85,1.27)
Gastrointestinal Bleed	0.9 (0.802,1.01)
Hepatic Encephalopathy	1.446 (1.29,1.63)
Viral Hepatitis	0.951 (0.851,1.06)
Hypertension	0.922 (0.818,1.04)
NAFLD	1.083 (0.955,1.23)
Osteoarthritis	0.969 (0.858,1.09)
Obesity	1.301 (1.14,1.48)
Palliative Care	1.183 (1.03,1.36)
Peptic Ulcer Disease	0.856 (0.725,1.01)
Peripheral Vascular Disease	0.818 (0.661,1.01)

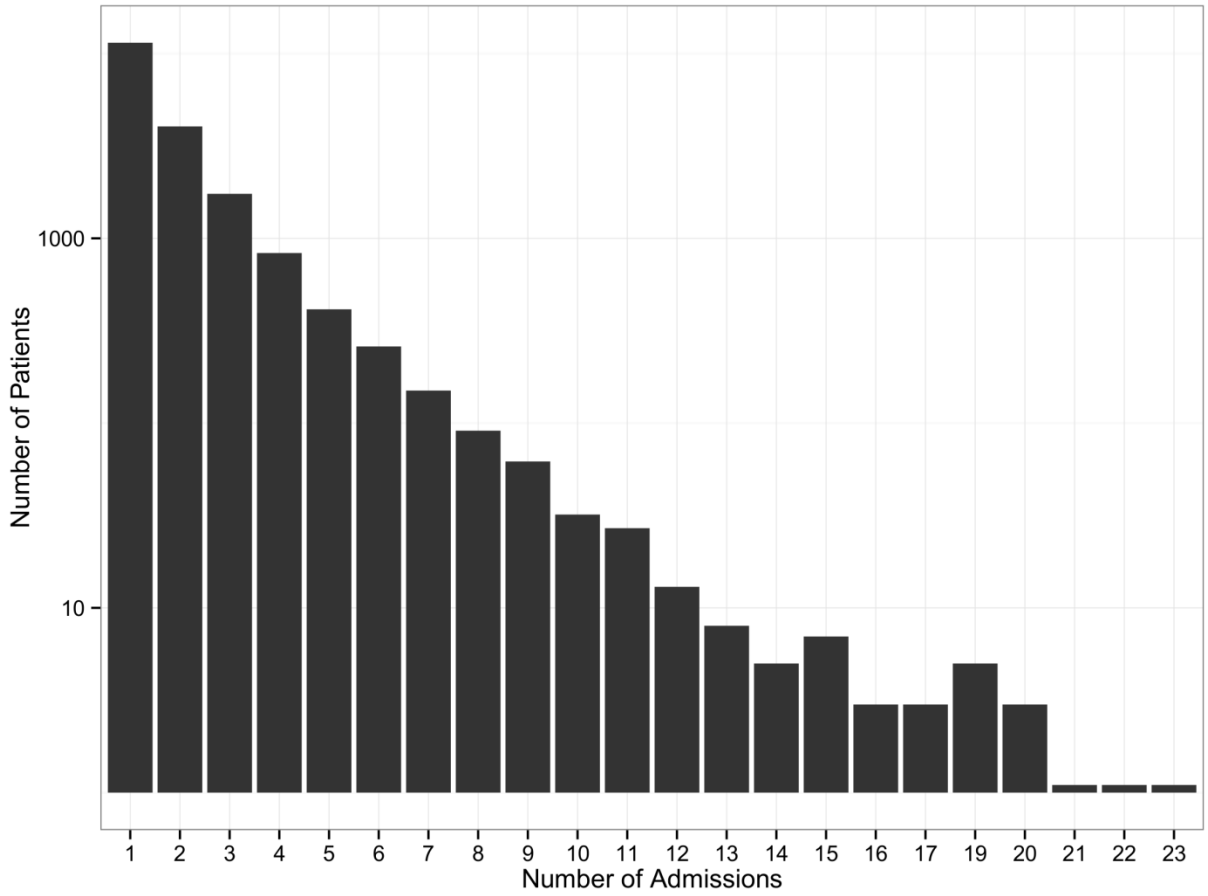
Variable	Odds Ratio (95% CI)
STEMI	1.118 (0.836,1.5)
Tobacco Use	1.126 (1.02,1.25)
Urinary Obstruction	0.864 (0.733,1.02)
Varices	1.03 (0.915,1.16)
Home Med Lactulose	0.942 (0.837,1.06)
Home Med Antihistamines Piperazine	0.847 (0.677,1.06)
Home Med Antihistamines Other	1.024 (0.787,1.33)
Home Med Cephalosporin 1st Generation	0.448 (0.243,0.826)
Home Med Quinolones	1.04 (0.91,1.19)
Home Med Antivirals	1.352 (1,1.82)
Home Med Anti Infectives Other	0.828 (0.693,0.989)
Home Med Antineoplastic Other	0.589 (0.365,0.952)
Home Med Autonomic Medications	0.805 (0.575,1.13)
Home Med Platelet Aggregation Inhibitors	2.195 (1.61,2.98)
Home Med Opioid Analgesics	0.915 (0.812,1.03)
Home Med Non Opioid Analgesics	0.876 (0.758,1.01)
Home Med Antidepressants Other	0.877 (0.763,1.01)
Home Med Cardiovascular Medications	1.114 (0.962,1.29)
Home Med Beta Blockers Related	0.831 (0.741,0.931)
Home Med Calcium Channel Blockers	1.065 (0.891,1.27)
Home Med Antianginals	0.837 (0.646,1.08)
Home Med Antilipemic Agents	1.028 (0.845,1.25)
Home Med Antihypertensives Other	0.592 (0.406,0.862)
Home Med Potassium Sparing Combinations Diuretics	1.052 (0.93,1.19)
Home Med Ace Inhibitors	0.859 (0.733,1.01)
Home Med Dermatologicals Topical Other	1.184 (0.937,1.5)
Admit Midodrine	1.459 (0.937,2.27)
Admit Albumin Infusion	1.37 (1.22,1.54)

Variable	Odds Ratio (95% CI)
Admit Norepinephrine	0.913 (0.652,1.28)
Admit Rifaximin	1.42 (1.16,1.73)
Admit Med Class Glucocorticoids	0.76 (0.607,0.951)
Admit Med Class Statins	0.913 (0.569,1.47)
Admit Med Class Insulin	0.88 (0.769,1.01)
Admit Med Class Opioids	1.103 (0.994,1.22)
Admit Med Va Class Antidotes Deterrents And Poison Control Exchange Resins	1.306 (1.13,1.51)
Admit Med Va Class Antihistamines Phenothiazine	1.126 (0.851,1.49)
Admit Med Va Class Extended Spectrum Penicillins	0.751 (0.644,0.876)
Admit Med Va Class Anti Infectives Other	0.924 (0.805,1.06)
Admit Med Va Class Anticoagulants	1.013 (0.9,1.14)
Admit Med Va Class Non Opioid Analgesics	0.759 (0.664,0.868)
Admit Med Va Class Benzodiazepine Derivative Sedatives Hypnotics	1.007 (0.87,1.17)
Admit Med Va Class Anticonvulsants	1.134 (0.957,1.34)
Admit Med Va Class Antidepressants Other	1.153 (1,1.33)
Admit Med Va Class Antipsychotics Other	0.647 (0.508,0.824)
Admit Med Va Class Beta Blockers Related	1.179 (1.05,1.32)
Admit Med Va Class Antilipemic Agents	0.829 (0.531,1.3)
Admit Med Va Class Potassium Sparing Combinations Diuretics	1.007 (0.887,1.14)
Admit Med Va Class Ace Inhibitors	0.731 (0.524,1.02)
Admit Med Va Class Cardiovascular Agents Other	1.446 (1,2.08)
Admit Med Va Class Emollients	0.68 (0.514,0.901)
Admit Procedure Dialysis	0.765 (0.555,1.05)
Admit Procedure Paracentesis	1.453 (1.3,1.63)
Admit Condition SBP	1.631 (1.41,1.88)
3 Day Pre Admit Procedure Paracentesis	1.573 (1.2,2.06)

Appendix Table 3b: Overall GEE model performance using the LASSO variables yielding the minimum deviance.

	Value
Intercept	-0.154 (-0.319, 0.012)
Slope	0.926 (0.860, 0.993)
Brier	0.052 (0.050, 0.055)
AUC	0.843 (0.833, 0.853)
O/E	0.998 (0.913, 1.083)

Appendix 4: Distribution of number of patients as a function of number of admissions.



Appendix Table 5: Discrimination and calibration statistics for each cluster along with 95% confidence intervals obtained by bootstrap sampling.

Clusters with "N/A" had ≤ 2 observations. When comparing the cluster number with the cluster map in Figure 3, cluster #1 starts in the bottom left corner and proceeds row-wise until cluster #49 in the top right corner.

Cl.	N	Intercept	Slope	Brier	AUC
1	423	-0.374 (-0.707,-0.04)	0.612 (0.387,0.837)	0.17 (0.15,0.191)	0.665 (0.606,0.725)
2	927	1.647 (-3.471,6.765)	1.555 (0.289,2.82)	0.008 (0.003,0.013)	0.803 (0.554,1.051)
3	1535	0.686 (-0.552,1.923)	1.217 (0.883,1.551)	0.016 (0.01,0.022)	0.859 (0.784,0.934)
4	2026	0.785 (-0.002,1.572)	1.214 (0.966,1.461)	0.027 (0.021,0.033)	0.821 (0.767,0.875)
5	973	0.458 (-0.637,1.553)	1.075 (0.731,1.42)	0.033 (0.022,0.044)	0.762 (0.678,0.846)
6	236	-0.166 (-0.644,0.312)	0.82 (0.507,1.132)	0.146 (0.116,0.176)	0.749 (0.672,0.826)
7	266	-1.001 (-1.747,-0.255)	0.381 (-0.017,0.779)	0.139 (0.11,0.168)	0.617 (0.52,0.714)
8	321	-0.615 (-1.144,-0.087)	0.682 (0.426,0.937)	0.124 (0.1,0.149)	0.757 (0.702,0.812)
9	860	0.366 (-0.196,0.928)	1.037 (0.807,1.266)	0.08 (0.066,0.094)	0.792 (0.745,0.839)
10	1332	2.172 (0.58,3.764)	1.665 (1.2,2.13)	0.02 (0.013,0.028)	0.865 (0.794,0.937)
11	2006	0.104 (-0.523,0.73)	1.004 (0.81,1.197)	0.045 (0.036,0.053)	0.775 (0.733,0.817)
12	1	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)
13	327	-0.262 (-1.273,0.749)	0.69 (0.281,1.1)	0.101 (0.071,0.13)	0.649 (0.555,0.743)
14	0	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)
15	66	4.122 (-47.616,55.859)	4.391 (-14.986,23.769)	0.029 (-0.001,0.059)	0.515 (-0.17,1.2)
16	425	-0.427 (-1.51,0.656)	0.865 (0.482,1.247)	0.05 (0.033,0.067)	0.819 (0.722,0.916)
17	1117	0.757 (0.207,1.306)	1.315 (1.092,1.539)	0.053 (0.043,0.063)	0.868 (0.832,0.904)
18	1223	0.053 (-0.327,0.433)	0.975 (0.807,1.144)	0.092 (0.08,0.104)	0.78 (0.743,0.817)
19	378	0.176 (-0.421,0.773)	0.906 (0.656,1.157)	0.109 (0.083,0.134)	0.782 (0.727,0.838)
20	457	-0.548 (-1.528,0.431)	0.963 (0.544,1.382)	0.035 (0.023,0.048)	0.801 (0.712,0.89)

Cl.	N	Intercept	Slope	Brier	AUC
21	811	-1.029 (-1.621,-0.437)	0.541 (0.313,0.769)	0.077 (0.061,0.092)	0.665 (0.603,0.728)
22	508	1.535 (-0.659,3.729)	1.55 (0.883,2.218)	0.02 (0.009,0.032)	0.835 (0.75,0.92)
23	892	-0.312 (-2.116,1.492)	1.022 (0.544,1.5)	0.013 (0.007,0.02)	0.768 (0.665,0.871)
24	788	-0.435 (-0.77,-0.1)	0.699 (0.496,0.901)	0.137 (0.122,0.152)	0.694 (0.649,0.739)
25	831	-0.518 (-1.105,0.07)	0.888 (0.647,1.13)	0.055 (0.041,0.068)	0.783 (0.728,0.837)
26	819	0.148 (-1.106,1.402)	1.126 (0.642,1.61)	0.037 (0.027,0.047)	0.785 (0.687,0.882)
27	121	7.581 (-13.038,28.201)	4.699 (-6.656,16.053)	0.034 (0.011,0.058)	0.969 (0.917,1.02)
28	1092	0.956 (0.24,1.672)	1.209 (0.955,1.463)	0.054 (0.043,0.066)	0.807 (0.761,0.854)
29	1256	1.191 (-1.188,3.571)	1.573 (0.826,2.321)	0.011 (0.006,0.016)	0.785 (0.661,0.908)
30	907	0.137 (-0.709,0.983)	0.947 (0.636,1.259)	0.061 (0.048,0.075)	0.719 (0.652,0.785)
31	261	-0.081 (-0.465,0.302)	1.02 (0.695,1.344)	0.16 (0.135,0.185)	0.761 (0.699,0.824)
32	506	-0.316 (-0.544,-0.089)	0.584 (0.377,0.791)	0.206 (0.185,0.226)	0.667 (0.612,0.722)
33	622	-0.043 (-0.981,0.894)	1.011 (0.669,1.353)	0.043 (0.03,0.056)	0.85 (0.778,0.923)
34	1142	1.51 (-0.911,3.931)	1.631 (0.868,2.394)	0.012 (0.008,0.017)	0.871 (0.76,0.981)
35	1493	-0.226 (-1.705,1.253)	1.128 (0.703,1.554)	0.012 (0.007,0.017)	0.83 (0.742,0.918)
36	815	-0.211 (-1.09,0.668)	0.87 (0.557,1.183)	0.057 (0.043,0.07)	0.714 (0.644,0.785)
37	449	-0.219 (-1.143,0.705)	1.029 (0.644,1.414)	0.039 (0.024,0.055)	0.829 (0.745,0.913)
38	813	-0.019 (-0.633,0.595)	0.889 (0.631,1.146)	0.083 (0.068,0.098)	0.749 (0.685,0.813)
39	253	-0.914 (-1.583,-0.246)	0.411 (0.162,0.66)	0.124 (0.091,0.156)	0.661 (0.575,0.748)
40	192	0.566 (-1.037,2.169)	1.535 (0.664,2.407)	0.058 (0.036,0.08)	0.936 (0.886,0.986)
41	562	0.275 (-3.947,4.496)	1.457 (0.262,2.653)	0.008 (0.001,0.014)	0.85 (0.636,1.064)
42	604	-0.296 (-0.707,0.116)	0.821 (0.607,1.036)	0.115 (0.098,0.132)	0.782 (0.736,0.827)

Cl.	N	Intercept	Slope	Brier	AUC
43	1456	0.756 (0.034,1.479)	1.23 (0.994,1.466)	0.036 (0.028,0.044)	0.819 (0.774,0.864)
44	308	1.862 (-0.798,4.523)	1.7 (0.767,2.633)	0.023 (0.01,0.036)	0.874 (0.786,0.962)
45	488	-0.619 (-1.257,0.018)	0.696 (0.378,1.014)	0.109 (0.086,0.133)	0.676 (0.602,0.749)
46	2	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)	N/A (N/A,N/A)
47	1537	0.181 (-0.325,0.687)	1.063 (0.841,1.285)	0.072 (0.062,0.082)	0.747 (0.706,0.788)
48	131	-1.849 (-9.556,5.859)	1.985 (-5.986,9.957)	0.031 (0.012,0.05)	0.91 (0.805,1.015)
49	854	-1.399 (-4.175,1.376)	1.168 (0.312,2.024)	0.006 (0.002,0.01)	0.908 (0.802,1.013)

Appendix 6: Sensitivity analysis taking patients with HRS who were admitted with acute decompensated heart failure or acute myocardial infarction and re-assigning them to the no-HRS group. Performed to assess the possibility of misdiagnosing cardiorenal syndrome.

As cardiorenal syndrome develops in the setting of acute (or acute on chronic) decompensated heart failure (ADHF) or acute myocardial infarction (AMI),[304–306] we performed a sensitivity analysis assigning patients with decompensated heart failure or acute myocardial infarction who had an HRS ICD9 code to the "No HRS" cohort.

METHODS: We identified patients with decompensated heart failure by an ICD9 code for acute or acute on chronic heart failure (428.21, 428.23, 428.31, 428.33, 428.41, 428.43) occurring any time during the admission or a primary discharge diagnosis of heart failure (regardless of chronicity). We identified patients with an acute myocardial infarction by an ICD9 code (410.*) any time during the admission.

There were 11 patients with ADHF and 19 patients with AMI who also had an ICD9 code for HRS (out of a total of 2258).

Appendix Table 6a: Odds Ratios for the GEE model predicting HRS for statistically significant variables.

Risk Factor	GEE Odds Ratio (95% CI)
Admit Intravenous Fluids / 1000 mL	0.899 (0.858,0.942)
Admit MELD	1.154 (1.139,1.169)
Baseline Creatinine	0.789 (0.749,0.831)
Admit Sodium	0.982 (0.974,0.99)
Admit Bicarbonate	0.98 (0.969,0.991)
Admit Blood Urea Nitrogen	1.006 (1.004,1.008)
Admit Glucose	0.997 (0.996,0.998)
Admit Mean Corpuscular Hemoglobin Conc.	1.076 (1.024,1.131)
Admit Mean Corpuscular Hemoglobin	1.003 (0.986,1.02)
Admit Alanine Aminotransferase	0.999 (0.999,1)
Admit Alkaline Phosphatase	1 (1,1)
Admit Partial Thromboplastin Time	0.994 (0.988,1)
Admit International Normalized Ratio	0.565 (0.493,0.647)
Admit Systolic Blood Pressure	0.995 (0.993,0.998)
Admit Temperature	0.954 (0.866,1.051)
Admit Weight	1.002 (1.001,1.003)
Admit Maximum Temperature	0.977 (0.907,1.053)
# Paracentesis in 90 days Pre-Admit	1.104 (1.084,1.125)
KDIGO Stage II (vs. KDIGO Stage I as baseline)	1.243 (1.099,1.406)
KDIGO Stage III (vs. KDIGO Stage I as baseline)	1.005 (0.863,1.171)
Hepatic Encephalopathy	1.58 (1.424,1.752)
Home Medication Analgesics	0.859 (0.776,0.95)
Home Medication Potassium Sparing Diuretics	1.072 (0.968,1.188)
Admit Medication Albumin Infusion	1.474 (1.315,1.651)
Admit Medication Non Opioid Analgesics	0.7 (0.616,0.796)
Admit Procedure Paracentesis	1.519 (1.359,1.698)
Admit Diagnosis SBP	1.56 (1.354,1.796)
Intercept	10.913

Appendix Table 6b: Overall GEE model performance.

	Value
Intercept	-0.039 (-0.206, 0.129)
Slope	0.983 (0.921, 1.046)
Brier	0.052 (0.049, 0.055)

AUC	0.841 (0.832, 0.850)
O/E	0.998 (0.913, 1.084)

Appendix 7: Hepatorenal Syndrome ICD-9 Code Accuracy Identification

Study Population

We analyzed a retrospective cohort of patients hospitalized from among 124 medical centers in the Department of Veterans Affairs (VA) between January 1, 2005 and December 31, 2013. We included all hospitalizations for patients who had a cirrhosis diagnosis (based on a history of two outpatient or one inpatient) ICD-9 code (571.2 or 571.5) and had AKI during their hospitalization with a maximum inpatient creatinine of at least 1.5 mg/dl. We excluded hospitalizations where the patient was on dialysis prior to admission, did not have at least one serum creatinine value within the year prior to admission or during the inpatient stay, who had a diagnosis of HRS prior to the hospitalization, who had a prior hospitalization with AKI, or who were discharged in less than forty eight hours.

We performed stratified sampling based on presence/absence of an ICD-9 code for Hepatorenal Syndrome, level of kidney injury, and level of liver disease. We sampled in blocks of twelve: six patients were selected if they had an ICD-9 code for Hepatorenal Syndrome anytime during their hospitalization; six patients (without an HRS ICD-9 code) were selected based on two levels of kidney injury (KDIGO Stage I versus KDIGO Stage II and III) and three levels of MELD (< 20, >= 20, and unable to calculate). We randomly selected a total of 42 blocks (504 inpatient admissions) to serve as the gold standard cohort.

Outcome

Two physician annotators reviewed the 504 hospitalizations reviewing all clinical notes, relevant laboratory values, medications, and radiology reports to assign each hospitalization into one of five categories: HRS Type I, HRS Type II, HRS Type Indeterminate, Maybe HRS, and Not HRS based on International Ascites Club criteria, with the exception that chronic kidney disease did not automatically preclude HRS. We employed a training phase where the two annotators worked in blocks of twelve patients until the inter-annotator agreement was >= 0.8. Disagreements on the 504 patient set were adjudicated by a board certified nephrologist.

Results

The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of a discharge ICD-9 code for HRS were 57.6%, 88.8%, 78.6%, and 74.6%.

Appendix 8: Point Score Model

A point-based scoring model was developed in line with the Framingham risk study.[260] Points ranged from -12 to 56 (median 13, IQR: 7 – 19). The points based model had an AUC of 0.835, intercept and slope of the observed-to-expected calibration line of -4.208 and 1.043 respectively, and a Brier score of 0.424.

Appendix Table 8a: Point allocation for statistically significant variables.

Variable	Categories	Points
Admit Intravenous Fluids	<500	1
Admit Intravenous Fluids	(500,1e+03]	0
Admit Intravenous Fluids	(1e+03,2e+03]	0
Admit Intravenous Fluids	> 2000	-2
Admit MELD	< 7	-2
Admit MELD	(7,10]	0
Admit MELD	(10,15]	4
Admit MELD	(15,20]	9
Admit MELD	(20,25]	13
Admit MELD	(25,30]	18
Admit MELD	> 30	26
Baseline Creatinine	<1	1
Baseline Creatinine	(1,1.5]	0
Baseline Creatinine	(1.5,2]	-1
Baseline Creatinine	(2,2.5]	-2
Baseline Creatinine	(2.5,3]	-3
Baseline Creatinine	(3,3.5]	-4
Baseline Creatinine	(3.5,4]	-4
Baseline Creatinine	> 4	-9
Admit Sodium	<125	2
Admit Sodium	(125,140]	0
Admit Sodium	> 140	-1
Admit Bicarbonate	< 10	1
Admit Bicarbonate	(10,15]	1
Admit Bicarbonate	(15,20]	0
Admit Bicarbonate	(20,25]	-1
Admit Bicarbonate	> 25	-2
Admit Blood Urea Nitrogen	< 10	0
Admit Blood Urea Nitrogen	(10,20]	0
Admit Blood Urea Nitrogen	(20,40]	1
Admit Blood Urea Nitrogen	(40,80]	2
Admit Blood Urea Nitrogen	> 80	4
Admit Glucose	< 100	0

Admit Glucose	(100,200]	-1
Admit Glucose	(200,300]	-3
Admit Glucose	(300,400]	-5
Admit Glucose	> 400	-7
Admit Mean Corpuscular Hemoglobin Conc.	< 25	-1
Admit Mean Corpuscular Hemoglobin Conc.	(30,32]	0
Admit Mean Corpuscular Hemoglobin Conc.	(32,34]	1
Admit Mean Corpuscular Hemoglobin Conc.	(34,36]	2
Admit Mean Corpuscular Hemoglobin Conc.	> 36	3
Admit Alanine Aminotransferase	< 20	0
Admit Alanine Aminotransferase	(20,50]	0
Admit Alanine Aminotransferase	(50,150]	0
Admit Alanine Aminotransferase	(150,300]	-1
Admit Alanine Aminotransferase	> 300	-4
Admit Partial Thromboplastin Time	< 30	0
Admit Partial Thromboplastin Time	(30,40]	0
Admit Partial Thromboplastin Time	(40,50]	0
Admit Partial Thromboplastin Time	(50,70]	-1
Admit Partial Thromboplastin Time	> 70	-2
Admit International Normalized Ratio	< 1.2	0
Admit International Normalized Ratio	(1.2,1.5]	-1
Admit International Normalized Ratio	(1.5,2]	-3
Admit International Normalized Ratio	(2,2.5]	-5
Admit International Normalized Ratio	> 2.5	-11
Admit Systolic Blood Pressure	< 90	1
Admit Systolic Blood Pressure	(90,120]	1
Admit Systolic Blood Pressure	(120,130]	0
Admit Systolic Blood Pressure	(130,140]	0
Admit Systolic Blood Pressure	> 140	-1
Admit Weight	< 45	-1
Admit Weight	(45,60]	0
Admit Weight	(60,100]	1
Admit Weight	> 100	2
# Paracentesis in 90 days Pre-Admit	< 2	0
# Paracentesis in 90 days Pre-Admit	(2,4]	2
# Paracentesis in 90 days Pre-Admit	(4,7]	3
# Paracentesis in 90 days Pre-Admit	> 7	7
Hepatic Encephalopathy	Present	3
Home Medication Analgesics	Present	-1
Admit Medication Albumin Infusion	Present	3
Admit Medication Non Opioid Analgesics	Present	-2

Admit Procedure Paracentesis	Present	3
Admit Diagnosis SBP	Present	3
KDIGO Stage 2 Renal Failure	Present	1
KDIGO Stage 3 Renal Failure	Present	0

Appendix Table 8b: Probability of HRS based on total points.

Points	Probability
-16	0.039398
-15	0.044875
-14	0.051073
-13	0.058074
-12	0.065969
-11	0.074851
-10	0.084821
-9	0.095981
-8	0.108435
-7	0.122287
-6	0.137635
-5	0.15457
-4	0.173171
-3	0.193498
-2	0.215588
-1	0.239451
0	0.265064
1	0.292363
2	0.321244
3	0.351561
4	0.383125
5	0.415705
6	0.449039
7	0.482837
8	0.516793
9	0.550595
10	0.583936
11	0.616526
12	0.648102
13	0.678433
14	0.707331
15	0.734648
16	0.760279

17	0.784162
18	0.806272
19	0.826617
20	0.845236
21	0.862189
22	0.877554
23	0.891422
24	0.903891
25	0.915064
26	0.925046
27	0.93394
28	0.941845
29	0.948856
30	0.955062
31	0.960546
32	0.965385
33	0.969649
34	0.973403
35	0.976703
36	0.979603
37	0.982148
38	0.984381
39	0.986338
40	0.988053
41	0.989555
42	0.99087
43	0.992021
44	0.993027
45	0.993908
46	0.994678
47	0.995351
48	0.995939
49	0.996453
50	0.996902
51	0.997295
52	0.997638
53	0.997937
54	0.998199
55	0.998427
56	0.998627

Appendix 9: Natural Language Processing pipeline to identify ascites from radiology reports.

To identify ascites, we constructed a natural language processing (NLP) pipeline. To develop the pipeline we assembled a gold standard of radiology reports for cirrhotic patients. We filtered all available radiology reports to include only computed tomography, magnetic resonance imaging, and ultrasound examinations of the abdomen, pelvis, or chest. Four hundred and fifty-six documents were randomly sampled and manually reviewed for assertion of ascites (either positive or negative). Of the reviewed documents, 124 were sampled for training and testing of the NLP pipeline (64 with at least one positive ascites assertion, 30 with at least one negative assertion and no positive assertions, and 30 with zero positive or negative assertions). The documents were split into a training (50%) and testing set (50%). We converted the documents into a string of concept unique identifiers (CUIs) mapped to the Unified Medical Language System (version 2013AB)[84] using the clinical Text Analysis Knowledge Extraction System (cTAKES) version 3.2.[85] A rule-based algorithm was devised based on the training set using CUIs related to ascites. Performance was assessed on the testing set; the sensitivity, specificity, positive predictive value, and negative predictive value were 96.5%, 94.0%, 93.3%, and 96.9%, respectively.