Measuring Care Variation In Congestive Heart Failure Patients

---

A Thesis

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Master of Science

by

Callie R. Kwiatkowski

August

2014

APPROVAL SHEET

The thesis

is submitted in partial fulfillment of the requirements

for the degree of

Master of Science

Callie Kuratkowski
AUTHOR

The thesis has been read and approved by the examining committee:

Dr. Laura Barnes

Advisor

Dr. Matthew Gerber

Dr. James H. Harrison

Accepted for the School of Engineering and Applied Science:

James H. Ay

Dean, School of Engineering and Applied Science

August

2014

Abstract

The current state of healthcare is prime for the development of new methodologies utilizing the wealth of electronic data regarding patient history to measure and track quality of care automatically. According to the Institute of Medicine, 30% of US health care services are inappropriate or wasteful, only 55% of needed health care services are delivered, and care can vary significantly by geographic location [1]. In light of these facts, the importance of developing tools to assess whether patients are receiving consistent and quality care is paramount. This study addresses an approach to measuring treatment variation between congestive heart failure patients based on their clinical orders and diagnoses. The data used for this research consists of procedure and medication order records collected using the Electronic Medical Record System (EMR) in place at the University of Virginia Health Center. Through the use of a vector space model, we provide a visual and quantitative method for measuring treatment similarity between patients based on treatments received. For this study, we investigated the variation of care provided to a selection of congestive heart failure patient groups including inpatient visits, outpatient visits, and combined inpatient and outpatient visits. Additionally we investigated patients who were diagnosed with one of the following comorbidities: chronic kidney disease, diabetes mellitus, or chronic obstructive pulmonary disease. Calculating similarity distributions for the general inpatient, outpatient, and combined IP/OP groups revealed that patients receive less varied care as an inpatient than as an outpatient. Similarity distributions calculated for comorbidity groups suggest that treatment variation decreases for these comorbidity groups as compared to the general CHF population. Additionally, our results show that CHF patients with COPD or chronic kidney disease are divided in such a way that a distinct portion of patients receive care to that similar of CHF patients with diabetes and a distinct portion of patients receive care which is dissimilar from the treatment provided to CHF patients with diabetes. Finally, as a first step toward validating our methods, two subject matter experts were asked to provide their own assessment of the treatment similarity between a selection of patients and

chosen reference patients. Taking into account a Kappa statistic of 0.284 obtained for inter-rater agreement, our results suggest that our approach to measuring care variation does provide some insight on patients which are being treated with lesser or greater similarity. Thus, the techniques discussed in this study may support a framework for monitoring treatment variation automatically in electronic health record systems.

Table of Contents

List of Figures

List of Tables

1. Background and Statement of Problem

Health care services and health care cost are a growing concern in the United States today. According to the Agency for Healthcare Research and Quality, the total number of hospital discharges has increased from an estimated 34,678,703 in 1997 to 38,590,733 in 2011. Moreover, the average cost of hospital visits has increased from an estimated $11,281 in 1997 to $35,403 in 2011 [2]. These trends can be viewed in Figures 1 and 2 below. Clearly the need for and cost of health care services is a great burden on United States Citizens which will continue for years to come. Consequently, measures should be taken to ensure the highest quality of care possible is provided and that waste of valuable resources, time, and money are minimized. Toward this aim, many approaches have been used in an attempt to monitor and assess variation of care.
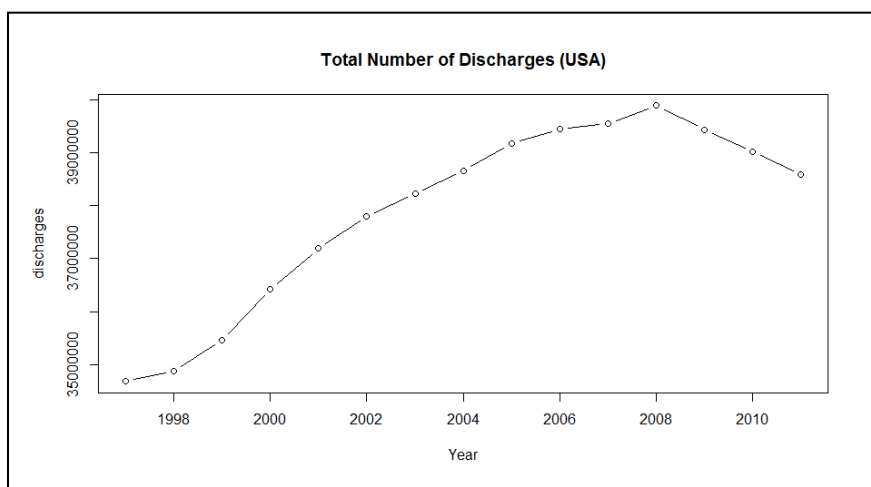
Figure 1: Hospital Discharges
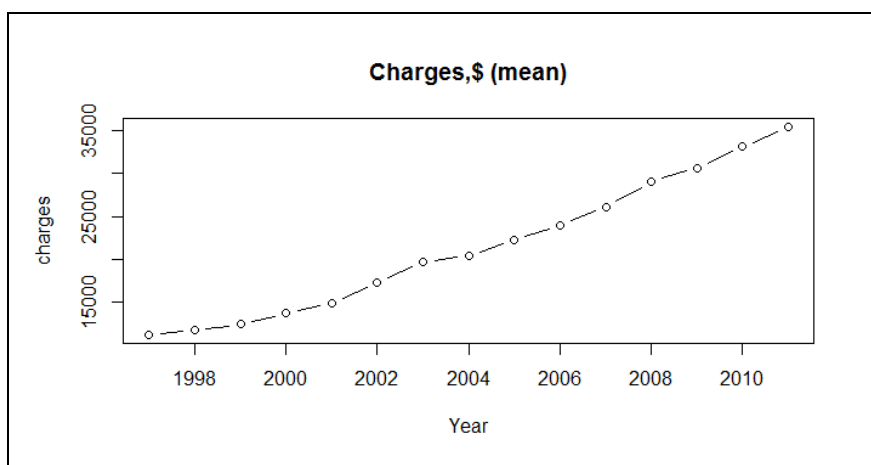
Figure 2: Hospital Visit Charges

The development of best practice guidelines has helped to set a higher standard of quality of care to hospitals across the United States. Throughout most of the 20th century, many people believed that the instances of poor quality of care were merely outliers, and that the majority of healthcare services provided were of high quality. After investigation of this topic, systematic failures were revealed to cause poor quality of care throughout the health care system. Quality of care studies leading into the 21st century can be broken down into 3 main categories: structures, processes and outcomes of care. Structures include personnel, facilities, training, certification and other factors which characterize the health care environment. Processes include the prescription of medication and the performance of procedures, which characterize the care provided to patients. Health outcomes measured include death rate and hospital readmission [3].

Many studies have been done in an attempt to capture some insight on the processes category of care variation. In most cases, the study involves a particular type of medical condition, a select few treatment types that are determined based on best practice guidelines to be beneficial for the condition, and something the researchers would like to compare, such as the rates at which the chosen treatments are used at hospitals in different locations, or the rates at which these chosen treatments

are used for different race, gender, or age groups. The general presumption is that the groups of patients which are provided the suggested treatments at a lower rate are receiving a lower quality of care [4-43]. However, the methods used in these studies do not take into account that multiple treatments might be used in combination, thus influencing the treatment choices for each patient.

Another consideration in quality of care studies is the method of data collection used. For years, the standard method for storing patient information has been to document on paper and store the medical records in a filing cabinet. Therefore, in order to access this information for analysis, one would have to physically go into the files and read through them. This approach can be costly and time consuming. In addition, if a researcher wanted to compare data from hospitals in different locations they would have to travel there or hire someone to do so which can greatly limit their access to data [4-10].

In an attempt to gain the same information more quickly and without the need to travel, some researchers have switched to creating "case scenario" surveys in which physicians are asked to answer questions and report what they would have done if they were helping the hypothetical patients. This method still limits the quantity of data that can be collected and has the disadvantage that you are relying on the physician's honesty and accuracy. We cannot guarantee that the physician's reaction to the situation in the case scenario will be exactly the same as in reality. Furthermore, the usefulness of the physicians' responses is dependent on the thoroughness of the surveys. There may be influential details that occur in reality that are not captured in the case scenarios [11, 12, 17, 36, 37, 44, 45].

The advancement of Electronic Medical Record Systems (EMR) being implemented in many hospitals for medical record storage offers multiple advantages for care variation studies [3]. These systems allow for quick access to large quantities of data, which can be easily viewed in an organized format. The cost to obtain such data is lower and the information collected more accurately reflects what actually happened during a patient's visit, rather than hypothetical responses to case scenarios.

EMR systems also allow data to be compared from distant locations. Through the use of EMR systems and the development of automated programs for measuring care variation, the performance of all hospitals could be monitored regularly, rather than only catching a glimpse of hospital performance through sporadic studies. The data for this study is extracted from the University of Virginia Medical Center EMR system. However, the examples below show that many care variation studies do not use this type of system for data collection.

For example, in [11], a study was conducted in which the goal was to determine whether primary care physicians or cardiologists more consistently treat congestive heart failure patients according to best practice guidelines. To accomplish this goal they sent case scenario surveys to 2,250 physicians and cardiologists, received 994 responses, and analyzed the rates at which certain treatments, including evaluation of left ventricular function, the use of angiotensin-converting enzyme inhibitors, and the use and dosage of certain prescriptions, were provided using multivariate models [11]. An assumption for this study was that the use of surveys sent to physicians provided accurate insight into primary care physician and cardiologist practices. In addition, the researchers assume that a more frequent use of the suggested treatments actually leads to better patient outcomes. Therefore, the rates at which these treatments are provided is the principal factor used to compare the performance of primary care physicians and cardiologists.

In [28], another study was conducted where researchers wanted to determine whether there was a racial difference in the rates at which hospice care was used for heart failure patients. For this study they used a national sample of 98,258 Medicare beneficiaries with a heart failure diagnosis that had not enrolled in hospice care during the year 2000. Subsequently, they observed the rates at which patients of different ethnic backgrounds entered hospice care in 2001 after adjusting for socio-demographic, clinical, and geographic factors. They presented their results using adjusted odds ratios and did not indicate the use of more robust statistical models [28]. Although they may have provided

insight into the rates at which hospice care is used for different ethnic backgrounds, this particular study didn't take into account the variety of factors which may lead to choosing hospice care (or not) and does not provide a suffice explanation as to why a higher rate of hospice care use is preferable.

In [23], a third study was conducted where the researchers wanted to determine whether treatment choices for various acute coronary syndromes differed across regions and hospitals in 14 countries. Data for 11,543 patients at 95 hospitals collected by the Global Registry of Acute Coronary Events (GRACE) was analyzed using statistical regression models [23]. The purpose of this study was primarily aimed at assessing how the practices differ between locations and types of hospitals and less concerned with whether similar patients were receiving similar treatment within a hospital setting. In addition, little insight was provided as to whether the practice differences led to better or worse patient outcomes.

These are only a few examples, among numerous studies, which aim to determine if the rates of certain treatments used for a particular condition vary across a factor of interest, such as location, ethnic background of patients, or physician type. Each of these studies use the assumption that a more frequent use of treatments mentioned in best practice guidelines leads to better patient outcomes. However, one frequently occurring issue in quality of care studies is that they focus completely on the underuse of a guideline recommended practice [3].

Instead of focusing on the use of a particular treatment, this study will attempt to measure the overall treatment similarity between patients within a population of interest. The methods applied in this study will determine whether the combination of treatments used for one patient is similar to the combination of treatments used for another. This approach is more insightful, since treatments are often chosen in combination in response to a patient having multiple comorbidities.

2. Research Design and Methods

2.1. Data Summary

The dataset used for this study consists of records of procedure and medication orders for patients at the University of Virginia Medical Center. The data, in its entirety, includes all patient orders that were made since the ambulatory installment of the EMR system in September 2010 and inpatient installment in March 2011 through February 2014. Although many subgroups of patients of interest may be studied using this data, this research will focus on patients who were given one of the 16 congestive heart failure; non-hypertensive (CHF) related diagnoses as defined by the AHRQ Clinical Classifications system. CHF related diagnoses include the 16 ICD-9 codes defined in Table 1 below. Additionally, this study will focus on the procedure order data. Since treatment may vary depending if a patient is given "inpatient," "outpatient," or "observation" status, the collection of CHF patients has been divided into 3 subsets, respectively.

Table 1: Congestive Heart Failure ICD-9 Diagnoses

ICD-9 Diagnosis Code	Description
398.91	Rheumatic Heart Failure (Congestive)
428.0	Congestive Heart Failure (CHF) Unspecified
428.1	Left Heart Failure
428.2	Unspecified Systolic Heart Failure
428.21	Acute Systolic Heart Failure
428.22	Chronic Systolic Heart Failure
428.23	Acute on Chronic Systolic Heart Failure
428.3	Unspecified Diastolic Heart Failure
428.31	Acute Diastolic Heart Failure
428.32	Chronic Diastolic Heart Failure
428.33	Acute on Chronic Diastolic Heart Failure
428.4	Unspecified Combined Systolic and Diastolic Heart Failure
428.41	Acute Combined Systolic and Diastolic Heart Failure
428.42	Chronic Combined Systolic and Diastolic Heart Failure
428.43	Acute on Chronic Combined Systolic and Diastolic Heart Failure
428.9	Heart Failure Unspecified

The inpatient subset contains 2,853,526 orders, making this the largest subset. In contrast, the outpatient subset contains 169,300 orders and the observation subset contains 54,914 orders. Within the inpatient dataset, 2,827 different types of procedures are ordered for 4,375 patients across 6,750 hospital visits. For the outpatient dataset, 1,565 types of procedures are ordered for 5,662 patients over 19,039 visits. Finally for the observation dataset, 851 types of procedures are ordered for 612 patients over 699 visits. As expected, the majority of patients only require a few visits. The average number of visits is 1.543 for patients with inpatient status, 3.363 for patients with outpatient status, and 1.142 for patients with observation status. In addition, the average number of procedure orders per visit is 422.745 for the inpatient dataset, 8.892 for the outpatient dataset, and 78.561 for the observation dataset. These data summary statistics can be found in Table 2 below. In addition, a sample observation from the data can be found in Table 3.

Table 2: Data Summary Statistics

Status	Inpatient	Outpatient	Observation
# Orders	2,853,526	169,300	54,914
# Procedure Types	2,827	1,565	851
# Patients	4,375	5,662	612
# Visits	6,750	19,039	699
Mean # Visits per Patient	1.543	3.363	1.142
Mean # Orders per Visit	422.745	8.892	78.561

Table 3: Sample Observation

Feature	Sample Observation
proc_order_num	23811801
PROC_ID	34584
PROC_NAME	MAMMO DIAGNOSTIC - DIGITAL – LT
PROC_CAT	IMG MAMMOGRAPHY ORDERABLES
PROC_CAT_ID	17
ORDER_DISPLAY_NAME	MAMMO DIAGNOSTIC - DIGITAL – LT
ptid	397
age	37
sex	F
caseid	22455444
VisitDXs	610.1 - DIFFUS CYSTIC MASTOPATHY,V16.3 - FAMILY HX-BREAST MALIG
timeline	2
anon_adm_date	0
anon_dis_date	569
cdrIPorOP	OP
physid_deid	32340

To better understand the data, the rates at which different diagnoses types occur were determined for the inpatient, outpatient, and observation datasets. Figure 3 compares the percentages of patients that, at some point in time, received one of the diagnoses from Table 1. There is much overlap here as patients frequently receive more than 1 of these diagnoses. The rates that each diagnosis type occurs vary if a patient was categorized as inpatient, outpatient, or observation; however, overall some of these diagnoses occur more frequently than others regardless of patient status. This is to be expected as the more frequently occurring diagnoses are also more general.
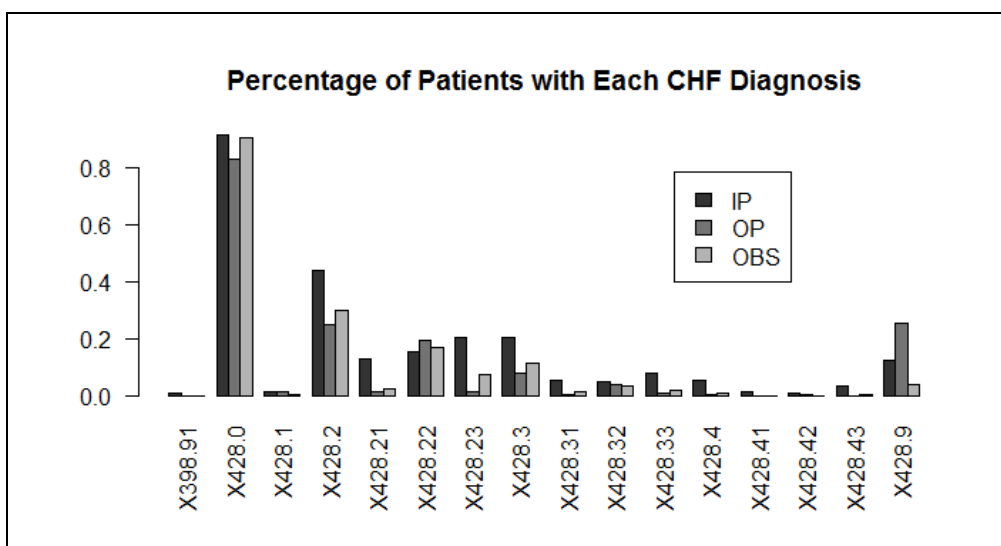
Figure 3: Percentage of Patients with Each CHF Diagnosis

To present a different perspective, the number of CHF diagnoses given during each visit was determined for each patient category. Figures 4, 5, and 6 show the percent of visits with various numbers of CHF diagnoses. Patients with inpatient, outpatient, or observational status do not often receive the same number of CHF diagnoses. Patients categorized as outpatient are most frequently given only one CHF diagnosis. Only receiving one CHF diagnosis is understandable since the state of a patient's health tends to be less complicated during an outpatient visit and the treatment provided tends to be more specific. In contrast, an inpatient visit tends to cover a longer period of time as the treatment and diagnoses during an inpatient visit are often more complicated and severe. Thus, patients categorized as inpatient are almost as likely to be given three CHF diagnoses during a visit as to be given a single CHF diagnosis. Finally, for those patients categorized as observational, although they are most likely to be given only 1 CHF diagnosis during a visit, there is also a significant percentage of visits in which they are given two or three CHF diagnoses. Again, this increase in CHF diagnoses for patients with observational status is as expected because visits with observational status tend have treatment and diagnoses which are more complicated than during outpatient visits but not quite as complicated as during inpatient visits. Regardless, any variation in diagnosis may lead to different treatment choices for

the patients. The number of CHF diagnoses provided should be taken into account when investigating care variation because the fact that outpatient visits are frequently provided only one CHF diagnosis increases the chance that two patients may have slightly different diagnoses which may contribute to increased variation of care compared to inpatient visits.
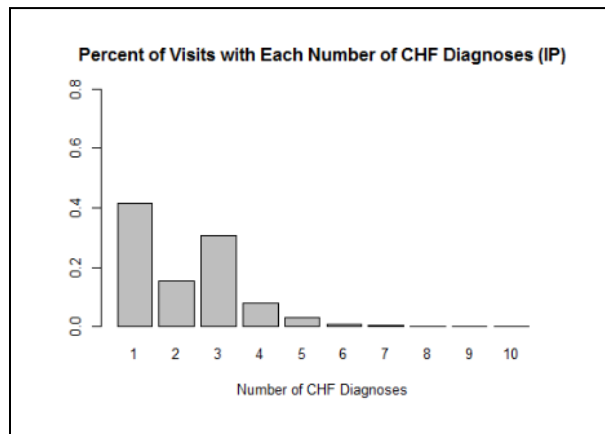
Figure 4: Frequency that Different Numbers of CHF Diagnoses Occur (IP)
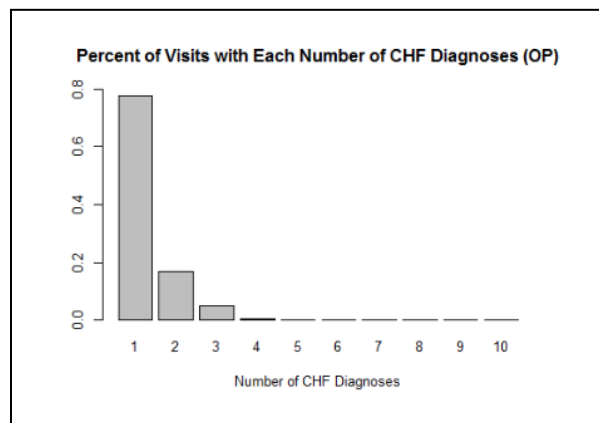
Figure 5: Frequency that Different Numbers of CHF Diagnoses Occur (OP)
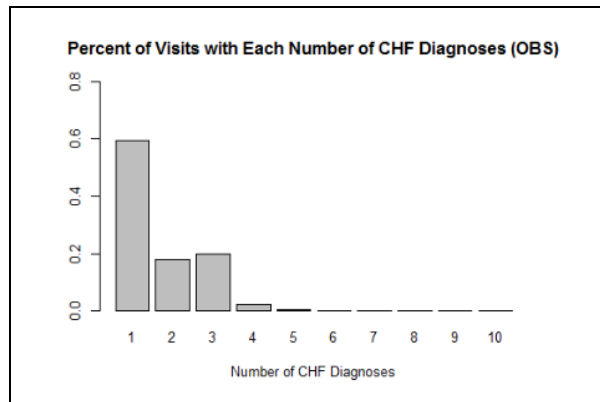
Figure 6: Frequency that Different Numbers of CHF Diagnoses Occur (OBS)

Finally, the rates at which different procedure types were ordered were determined for each patient category. Some of the procedure types are ordered frequently while a large number of the procedure types are ordered only a few times for a small number of patients. For the inpatient dataset, out of 2,827 possible procedure order types, the 13 most frequently ordered procedures cover 50% of all orders, the top 54 procedures cover 75% of all orders, and the top 159 procedures cover 90% of all orders. For the outpatient dataset, out of 1,565 procedure order types, the 18 most frequently ordered procedures cover 50% of all orders, the top 55 procedures cover 75% of all orders, and the top 143 procedures cover 90% of all orders. Finally, for the observation dataset, out of 851 procedure order types, the 16 most frequently ordered procedures cover 50% of all orders, the top 47 procedures cover 75% of all orders, and the top 124 procedures cover 90% of all orders. When comparing the inpatient dataset to the observation dataset, the observation dataset requires a larger fraction of its possible procedures in order to cover the same percentage of orders simply because the observation dataset has much fewer patients. Table 4 summarizes these findings. In general, the procedures which occur most frequently will influence our measure of treatment similarity between patients most significantly. Therefore future refinements of our methods should consider creating weightings for each procedure type which do not allow less important procedures which are ordered frequently to influence our treatment similarity measure.

Table 4: Number of Procedures Which Cover a High Percentage of Orders

Status	Inpatient	Outpatient	Observation
# Possible Procedure Types	2,827	1,565	851
# Procedures that Cover 50% of Orders	13	18	16
# Procedures that Cover 75% of Orders	54	55	47
# Procedures that Cover 90% of Orders	159	143	124

2.2. Samples

In order to measure variation in the population, random samples were extracted to represent various patient groups of interest. Patients with only observational visits were excluded as these patients are likely to have very unique circumstances allowing them to acquire this status. Patients whose visits have observational status have a severe enough condition such that an outpatient visit would not be appropriate; however, their diagnosis may not yet be clear enough to justify an inpatient visit. These circumstances make assessment of treatment variation using observational visits difficult to interpret. In general, the differences between the details of an inpatient or outpatient visit are more apparent, therefore comparing the treatment similarity within these 2 patient groups is more easily interpretable. Hence, a sample of 250 patients with only their outpatient visits was extracted, a sample of 450 patients with only their inpatient visits was extracted, and a sample of 550 patients with both their inpatient and outpatient visits was extracted. These sample sizes were chosen to be proportional to the number of patients which fall under each category within the overall population.

In addition to samples from the general population, samples were also extracted to highlight certain comorbidity groups. We hypothesize that patients within these comorbidity groups will show increased treatment similarity as a result of having an additional illness in common. The comorbidity groups chosen were based on the clinical guidelines and include Chronic Kidney Disease, Diabetes Mellitus, and Chronic Obstructive Pulmonary Disease (COPD). These 3 conditions are among the top ten most frequently co-occurring comorbidities with heart failure [46]. Additionally, [47-51], indicate that

careful consideration should be taken when treating heart failure patients who have these comorbidities.

Inpatient, outpatient, and a combined inpatient and outpatient samples were extracted for each of these comorbidity groups with comparable numbers of patients as before. Patients within these comorbidity groups were identified as having appropriate ICD-9 codes according to the AHRQ Clinical Classifications system. The full list of ICD-9 codes used can be found in Appendix A. The exact number of patients, which are included in each sample, can be found in Table 5.

Certain restrictions were applied to the samples in order to create more refined similarity distributions. Firstly, orders for certain procedures were excluded that were determined to be common across hospital patients and not as important for the treatment of CHF, diabetes mellitus, chronic kidney disease, or COPD. These categories of procedures can be found in Table 6. Second, outpatients who had fewer than 3 visits were excluded, since the orders made for an outpatient in only 1 or 2 visits may be quite varied and not provide a good view of a patient's medical history. The exact threshold of how many visits are necessary for capturing a patient's history is uncertain; however 3 visits were suggested by a subject matter expert to be sufficient. Third, patients with fewer than 20 orders across their history were excluded. This exclusion is again due to the fact that the treatment provided to a patient with only a few orders is likely to be quite varied and the few orders that did occur are not likely to characterize a patient's entire history for the treatment of congestive heart failure. However, the restriction for patients to have 20 or more orders was not enforced for comorbidity samples. This is because patients included in the comorbidity samples had the additional restriction that they could not have one of the other comorbidities of interest. For example, patients within the diabetes mellitus sample could not have COPD, or chronic kidney disease. This additional restriction greatly reduced the number of patients available for sampling, so therefore the restriction of having 20 or more orders was not applied. We should note, however, that not requiring 20 or more orders may cause the similarity distributions for

comorbidity groups to not be comparable to the similarity distributions for the general population. This is especially problematic for outpatient samples since outpatient visits do not have a lot of orders in general so there are likely to be a number of patients in the outpatient comorbidity groups which have fewer than 20 orders, resulting in increased treatment variation as compared to the general population. Therefore the difference in treatment similarity for patients which do have 20 or more orders in the comorbidity groups as compared to the general population will not be apparent.

Two additional categories of samples were also extracted. The first was a sample of patients with inpatient, outpatient, and combined inpatient and outpatient visits which included both procedure and medication orders. The motivation here was to see how including the medication orders would influence the similarity distributions. Certain medication categories were also excluded for the same reasons as above. Samples of this type were collected for both the general population and for comorbidity groups. The last category of samples collected was a sample of patients with inpatient, outpatient, and combined inpatient and outpatient visits which did not have any of the comorbidities of interest.

Table 5: Number of Patients in Each Sample

Sample Category	# Patients
Inpatient General	450
Outpatient General	250
Combined Inpatient/Outpatient General	550
Inpatient Chronic Kidney Disease	450
Outpatient Chronic Kidney Disease	250
Combined Inpatient/Outpatient CKD	550
Inpatient Diabetes Mellitus	450
Outpatient Diabetes Mellitus	250
Combined IP/OP Diabetes Mellitus	550
Inpatient COPD	450
Outpatient COPD	178
Combined Inpatient/Outpatient COPD	550
Procedure + Medication IP General	200
Procedure + Medication OP General	200
Procedure + Medication IP/OP General	200
Procedure + Medication IP CKD	200
Procedure + Medication OP CKD	101
Procedure + Medication IP/OP CKD	200
Procedure + Medication IP Diabetes Mellitus	200
Procedure + Medication OP Diabetes Mellitus	200
Procedure + Medication IP/OP Diabetes Mellitus	200
Procedure + Medication IP COPD	200
Procedure + Medication OP COPD	73
Procedure + Medication IP/OP COPD	200
"Leftover" Inpatients Without Comorbidities	450
"Leftover" Outpatients Without Comorbidities	250
"Leftover" IP/OP Without Comorbidities	550

Table 6: Categories of Orders Excluded

Categories of Procedures Excluded	Categories of Medications Excluded
Admission Orderables	Chemicals
Discharge Orderables	Unknown
Nursing Information/Communication Orderables	Med_Pharm_Cls
Code Status Orderables	Pharmaceutical Adjuvants
Nursing Assessment Orderables-Once Or At Intervals	Antiseptics & Disinfectants
Nursing Assessment ORderables-Until Discontinued	Dietary Products
Nursing Activity Orderables	Mouth & Throat (local)
Nursing Treatment Orderables	Assorted Classes
Nursing Education Orderables	Medical Devices
Transfer Orderables	Medi-Span Reserved or Unknown (95)
Diet Orderables	Diagnostic Products
Nursing Diet Orderables	Minerals & Electrolytes
Inpatient Ancillary Consults	Multivitamins
Outpatient Referral Orderables	Vitamins
Outpatient Ref Orderables UVA	Nutrients
General Equiptment Orderables	
IV Therapy Orderables	

2.3. Contingency Table and Vector Space Model

Before the treatment similarity between patients can be measured, a framework for capturing a patient's treatment history was designed. A straightforward approach is to create a contingency table. Each row in the contingency table corresponds to a patient and the columns in the table correspond to each of the procedures that may possibly be ordered for a patient. The elements of the table indicate the number of times that a particular procedure was ordered for a patient over their entire medical history. The contingency table allows us to survey a patient's complete treatment history in a way that can be easily compared to other patients.

Treatment similarity is then calculated using a Vector Space Model. From the contingency table, each row forms a patient vector. To calculate the treatment similarity between patients, we first define a function $sim(P_1, P_2)$ which takes patient vectors P_1 and P_2 as inputs and produces a scalar value output indicating similarity. This may be calculated as the cosine of the angle between 2 patient vectors, defined as:

$$sim(P_1, P_2) = \cos(\theta)$$

$$= \frac{\sum_{j=1}^{n} P_{1j} * P_{2j}}{\|P_1\| * \|P_2\|} \qquad (1)$$

where P_{ij} indicates the jth entry in patient vector P_i and $\|P_i\|$ is the Euclidean length of patient vector P_i. This measure gives the convenient range of [0,1] where patient vector pairs with a smaller angle, and thus higher treatment similarity, will result in a value closer to 1, and patient vector pairs with a larger angle, and thus lower treatment similarity, will result in a value closer to 0. Figures 7 and 8 give a graphical depiction of this model using 3 hypothetical patients and 3 procedure types.
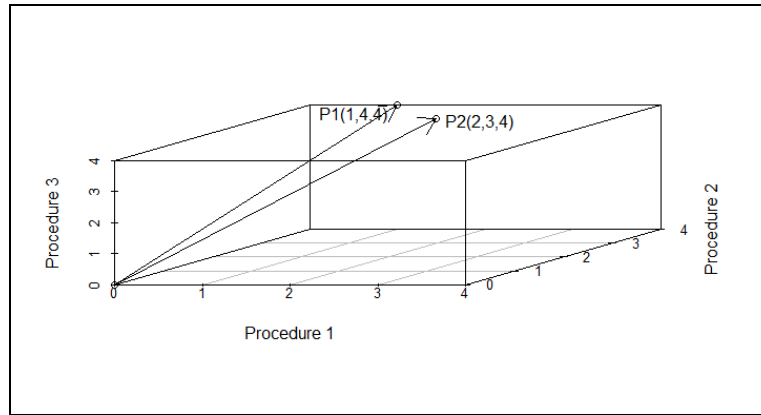
Figure 7: Graphical Depiction of VSM $(sim(P_1, P_2) = 0.97)$
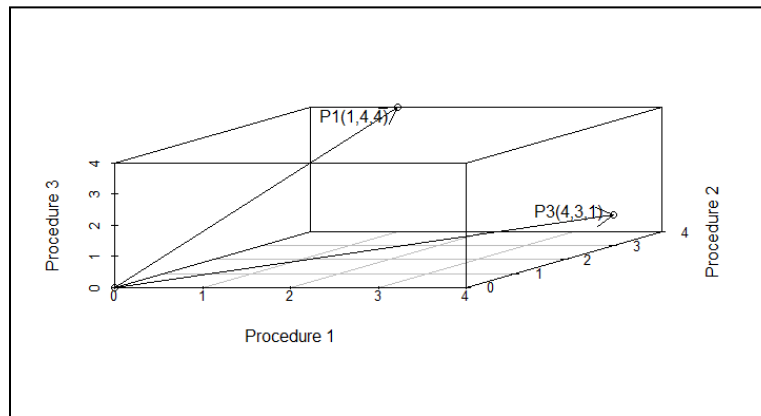
Figure 8: Graphical Depiction of VSM $(sim(P_1, P_3) = 0.68)$

For all vector space model calculations a package in Python called Gensim was used. Gensim is a vector space modeling and topic modeling toolkit originally developed for natural language processing studies. This package is designed to take a collection of documents and create a vector for each document which includes the number of times each possible word occurred for that document. Documents may be compared to patients in our study, words may be compared to procedures, and the collection of document vectors forms a corpus which is analogous to our contingency table in this study. In its simplest form, genism then calculates the similarity between all documents and a reference document using the cosine of the angle between each document vector. Gensim also provides easy implementation of a techniques such as term frequency–inverse document frequency (TF-IDF), latent

semantic analysis (LSA), and latent Dirichlet allocation (LDA) which transform documents from one vector representation into another in an attempt to reveal hidden structure in the corpus. These techniques and our investigation of their potential use for our problem are discussed further in Appendix B. Generally we found that these techniques take advantage of characteristics which are common in documents but do not provide valuable insight for measuring treatment variation between patients.

2.4. Index Patients

For this study, we calculated the similarity between all of the patients within a particular patient category and a single comparison patient referred to as an index patient. Once this similarity measure is calculated for each patient pair, the distribution of treatment similarity between patients can be visualized graphically. Examples of these visualizations are presented in Section 5.3. These visualizations can provide insight as to the degree of variation present in the treatment provided to the patient population in question.

There are many ways an index patient can be chosen. For this study, index patients are identified by computing the treatment similarity between all possible pairs of patients within the population under investigation; the patient who is most similar to all other patients on average is the index patient:

$$IndexPatient = max(mean(sim(P_i, P_j)) \qquad (2).$$

We then plot the distribution of all the similarity scores for each patient as compared to the chosen index patient.

2.5. Similarity Distributions

Calculating the similarity for all patients within a patient group as compared to an index patient chosen from that group allows us to see which patient groups are consistently being treated at a certain degree of similarity, and which patient groups are being treated with greater variation. For example, if a particular similarity distribution forms a narrow curve, then we may say that this patient group is being treated consistently with a certain degree of similarity. However, if a different similarity distribution forms a very broad curve then we may say that, even compared to a patient determined to be most similar to all other patients, the treatments provided to these patients varies tremendously.

Choosing a unique index patient for each patient group does not provide a complete understanding of how the treatment similarity compares between groups, however. Similarity distributions formed in this way are dependent on the number of patients and the number of different orders in the population. Therefore, the shape of the curves provide insight on the range of similarity obtained for each patient group, but the modes of the distributions cannot be directly compared across different patient groups. For example, if we compared the treatment similarity distributions for 2 different populations using their own unique index patient and the shapes of the curves were exactly the same except that one obtained a mode similarity value of 0.7 and the other mode was at a similarity value of 0.9, this does not necessarily tell us that the population whose similarity distribution has a mode at 0.9 shows greater treatment similarity. Since each population is being compared to a different index patient we cannot compare the absolute level of similarity obtained but only how consistently the treatment similarity stays at some level for each population.

We can however investigate the treatment similarity between different patient groups. The way index patients were chosen suggests that the set of orders for a particular index patient loosely reflects typical treatment choices for patients within the corresponding group. We should note, however, that if there is a lot of variation within a population then the index patient may not necessarily be a good

representation of a typical patient. Regardless, if we calculate the treatment similarity of one group of patients to an index patient from a different population we can see whether the treatment choices for the patients are similar to those for the alternate index patient. In this case, not only does the width of the curve indicate how consistently patients receive similar treatments to the other patient group of interest, but the location of the peak also provides insight into the degree of similarity to the other group. A similarity curve formed in this way with a peak at a larger similarity value would indicate a greater similarity to the compared population, or at least to the alternate index patient. If two similarity curves appear to overlap substantially, quantitative analysis, such as a Anderson-Darling test, may be applied to determine whether the difference between the 2 curves is significant. Basic statistical measures such as mean, standard deviation, and median may also provide some insight depending on the shape of the curve. In this study, we hypothesized that populations compared to their own index patient would show greater similarity than when compared to the index patients from other populations.

Considering that index patients may not be a sufficient representation of a typical patient within a patient group, we present an alternate way to compare patient groups. First, a hybrid inpatient sample is created which joins multiple patient groups together. For our study we chose to join the comorbidity samples as well as samples of patients which had no comorbidity of interest. All patients within this hybrid group are then compared to the same index patient. The index patient chosen for this study was an index patient chosen for one of the comorbidity groups. By calculating the similarity within this hybrid population as compared to one index patient we are able to maintain the same dimensions in our vector space models for each patient group. Finally, if we segregate the resulting similarity values for each patient group and then plot the similarity distributions for each group together on the same graph then we can directly compare not only the spread of similarity values obtained, evident by the width of

the distributions, but also the absolute similarity values obtained as compared to the chosen index patient, represented by the location of the modes.

2.6. Method Validation

In order to validate the similarity measure used in this study, we elicited the assistance of 2 medical doctors to serve as subject matter experts. They were provided a list of procedure orders for index patients corresponding to 4 different patient groups and procedure orders for 9 comparison patients from each of the 4 groups. They were then asked to compare each of the comparison patients to their corresponding index patient and indicate whether they were treated with high, medium, or low similarity to the index patient. The patient groups used were the general outpatient population, the general inpatient population, the diabetes mellitus outpatient population and the diabetes mellitus inpatient population. The general outpatient and general inpatient populations were chosen as our results showed that one of these 2 groups fairly consistently receives a particular level of similarity while the other group receives more varied degrees of treatment similarity. The 2 diabetes mellitus populations were chosen because the similarity distributions obtained using our methods were the most different from the general populations in contrast with the other comorbidity groups.

Three of the 9 comparison patients for a particular group were those which obtained a low similarity value using our methods, 3 were those which obtained a medium similarity value using our methods, and 3 were those which obtained a high similarity value using our methods. The intention was to see whether subject matter experts would rate each patient as obtaining the same similarity level which utilized methods in Section 2.3. Since the similarity distributions we calculated were different for each population, different conditions were used to define "high," "medium," and "low" similarity for each population. A summary of these conditions can be found in Table 7 below.

Table 7: Similarity Level Conditions

General Inpatient Population		General Outpatient Population	
High:	Similarity Value >= 0.9	High:	Similarity Value >= 0.75
Medium:	0.7 <= Similarity Value <= 0.75	Medium:	0.5 <= Similarity Value <= 0.6
Low:	Similarity Value <= 0.4	Low:	Similarity Value <= 0.2
Diabetes Mellitus Inpatient Population		**Diabetes Mellitus Outpatient Population**	
High:	Similarity Value >= 0.95	High:	Similarity Value >= 0.7
Medium:	0.7 <= Similarity Value <= 0.75	Medium:	0.4 <= Similarity Value <= 0.5
Low:	Similarity Value <= 0.5	Low:	Similarity Value <= 0.2

3. Results

3.1. Patient Groups Compared to Own Index Patients

In the first stage of this study treatment similarity was calculated for patient groups compared to their own index patients. First, similarity distributions were computed for the general inpatient, outpatient and combined IP/OP populations. As depicted in Figures 9-11, the similarity distribution curve for the general inpatients is much tighter than for the general outpatients. This suggests that the inpatients are more consistently treated at some level of similarity while the general outpatients receive more varied treatment. We suspect this is because in order to obtain inpatient status a patient must have a certain severity of illness which requires fairly specific treatment. Therefore, the variety of treatments, which may occur during an inpatient visit, is much less so patients are more likely to be treated similarly. Outpatient visits, however, could include a larger variety of illness severity and reasons the patient needs medical services; therefore, there is a larger variety of treatments that could potentially be provided. The similarity distribution for the combined IP/OP population shows 2 or 3 peaks and has a standard deviation of 0.210 indicating a wider curve with more variation. This is caused by the fact that the corresponding index patient is one that needs to be most similar to all patients on average. Some patients have only inpatient visits, some have only outpatient visits, and some have a mixture of both. There are a higher percentage of patients with inpatient visits so the types of treatments occurring at an inpatient visit are more strongly influencing the choice of index patient. Thus, the patients with only inpatient visits or a mixture of both are likely to fall at the higher end of the

similarity distribution and the patients with only outpatient visits are likely to fall at the lower end of the distribution.

As previously mentioned, all analysis presented in this study uses a collection of samples rather than using data for the entire population. To verify whether there was any sampling bias in our results, a Anderson-Darling (A-D) test was applied to see whether similarity distributions formed using inpatient, outpatient, and combined IP/OP samples compared to their own index patients and similarity distributions formed using the entire inpatient, outpatient and combined IP/OP populations can be viewed as samples taken from the same overlying distribution. With a p-value of 0.2385 for the outpatient group, the A-D test shows that at an α = 0.05 significance level we can presume that the sample used and the overall population came from the same overlying distribution. Similarly, a p-value of 0.0364 for the combined IP/OP group shows that at an α = 0.01 significance level the sample used and the overall population came from the same overlying distribution. Performing the A-D test for the inpatient group obtained a p-value of 0.019843, which by an α = 0.01 significance level suggests that the sample and overall population came from the same overlying distribution. We should note that the similarity distributions for the samples were calculated using the index patient chosen from the sample and the dimensions in the vector space model only included procedures provided to the patients in the sample. Since the similarity distributions calculated for the overall population and samples do not use the same number of dimensions for the vector space models and different index patients were used, the verification that the samples were comparable using the A-D test may not be sufficient. For better verification, the similarity distributions for the samples should be calculated using the same dimensions as the overall population and the same index patient. For this study, the sample-based results may be used for demonstrative purposes, though similarity distributions for the entire population will be calculated in the future.
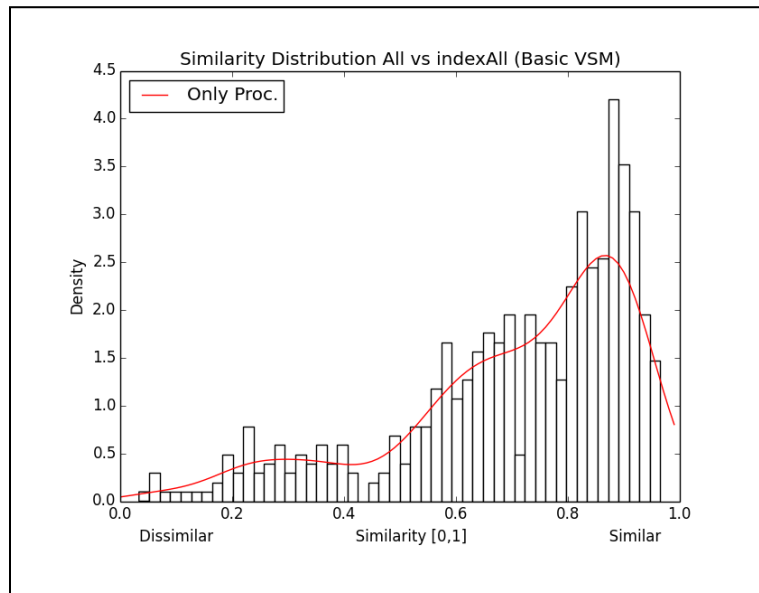
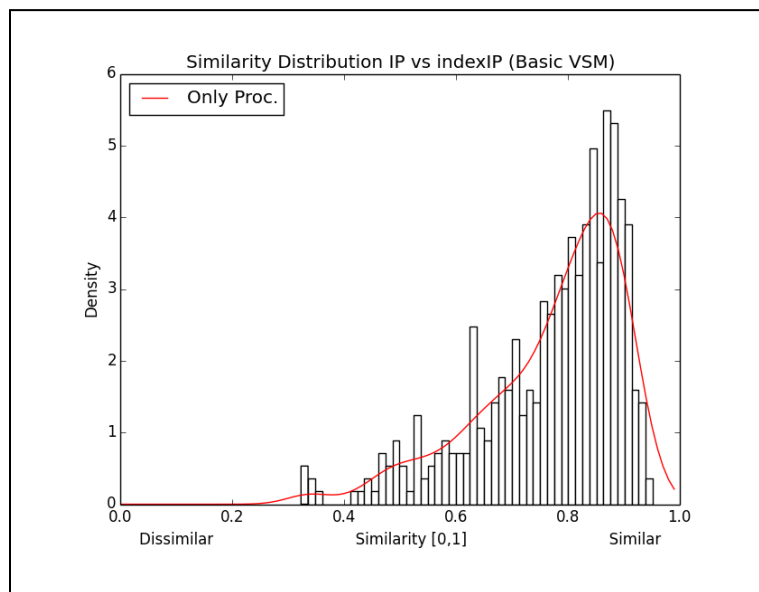**Figure 9: General Combined IP/OP vs. Own Index Patient**

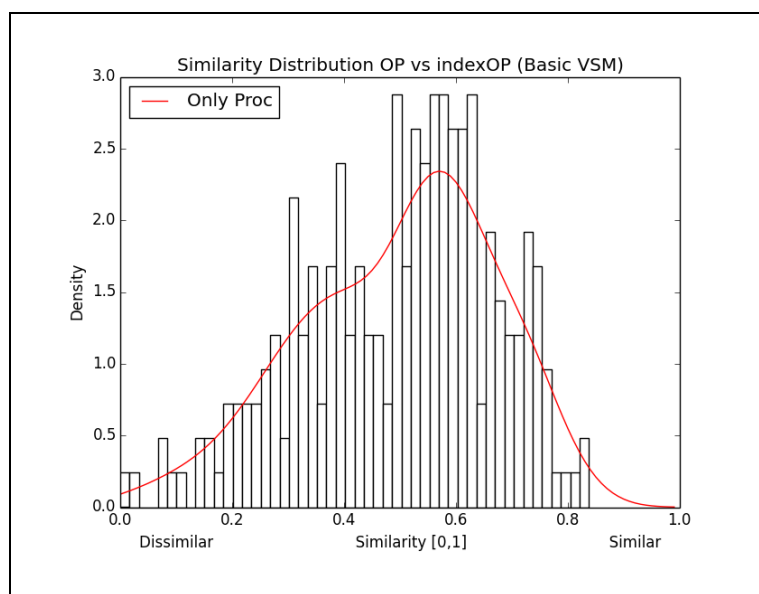**Figure 10: General Inpatients vs. Own Index Patient**

Figure 11: General Outpatients vs. Own Index Patient

Similarity distributions were also calculated for patients with inpatient, outpatient, and combined IP/OP visits for the 3 comorbidity groups of interest (Chronic Kidney Disease, Diabetes Mellitus, and COPD). Comparing the inpatient, outpatient, and combined IP/OP curves for each comorbidity, we found similar trends to the general population. Additionally, Figures 12-17 show that the inpatient and combined IP/OP distributions have narrower curves suggesting that patients who also have a particular comorbidity in common in addition to having CHF more consistently receive treatments at a particular level of similarity than the general populations. This method does not allow us to compare the levels of similarity but observe how consistently patients are treated at some level of similarity to the index patient. Figures 18-20, on the other hand, show that the outpatient distributions for comorbidity groups are broader, reflecting more varied treatment. One reason this may be occurring is because we chose not to remove patients who had fewer than 20 orders within the comorbidity groups. Since the treatment provided at outpatient visits tends to be more varied in general, the patients with very few orders are likely causing much variation in the similarity distributions. For this reason, we cannot directly compare the similarity distributions obtained for the general outpatient

population and comorbidity groups unless we maintain the same restrictions for all samples extracted. Based on the widths of the inpatient and combined IP/OP distributions, it seems that patients with diabetes mellitus are most consistently receiving treatment at some level of similarity to the index patients. Since the peaks of the similarity curves for the diabetes mellitus group are shifted to the right we can suspect that these patients are obtaining a high level of treatment similarity to the index patient. However, since we don't know the quality of the treatment being provided to the index patient in this group we cannot say whether these patients are receiving an optimal quality of treatment.
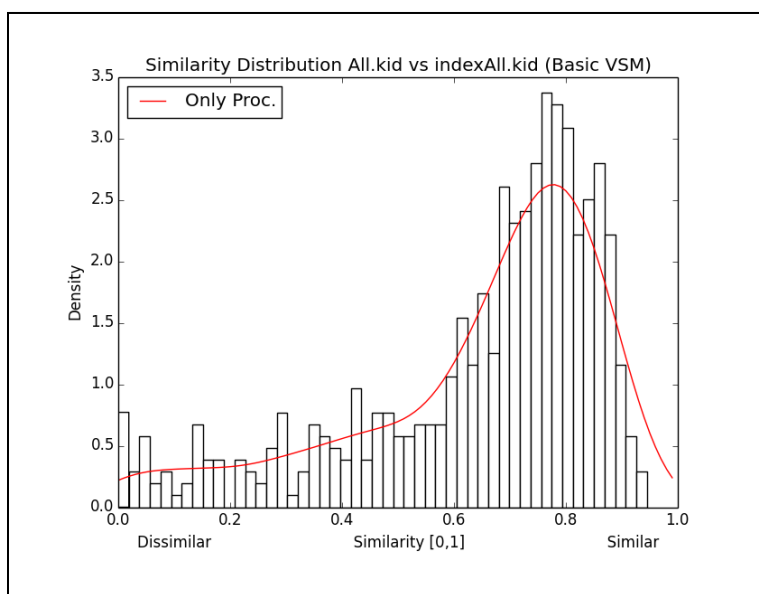
Figure 12: CKD Combined IP/OP vs. Own Index Patient
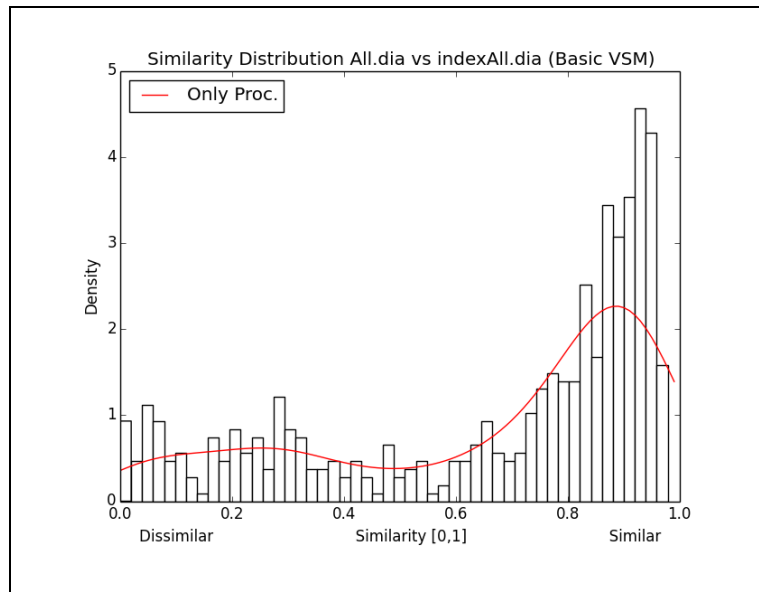
Figure 13: Diabetes Combined IP/OP vs. Own Index Patient
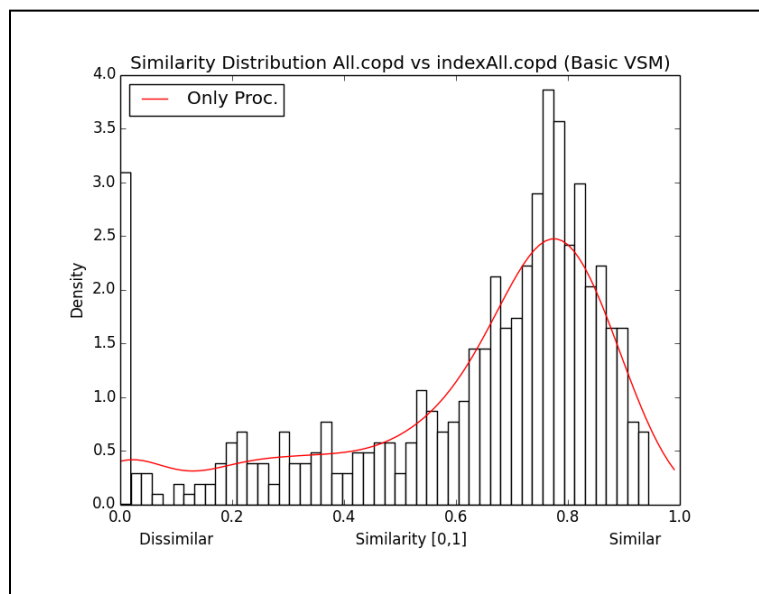
Figure 14: COPD Combined IP/OP vs. Own Index Patient

Figure 15: CKD Inpatients vs. Own Index Patient
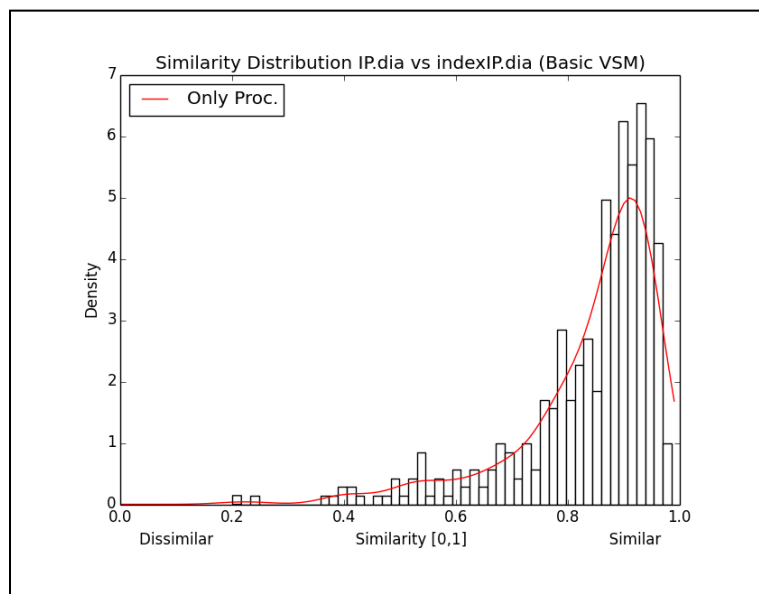
Figure 16: Diabetes Inpatients vs. Own Index Patient
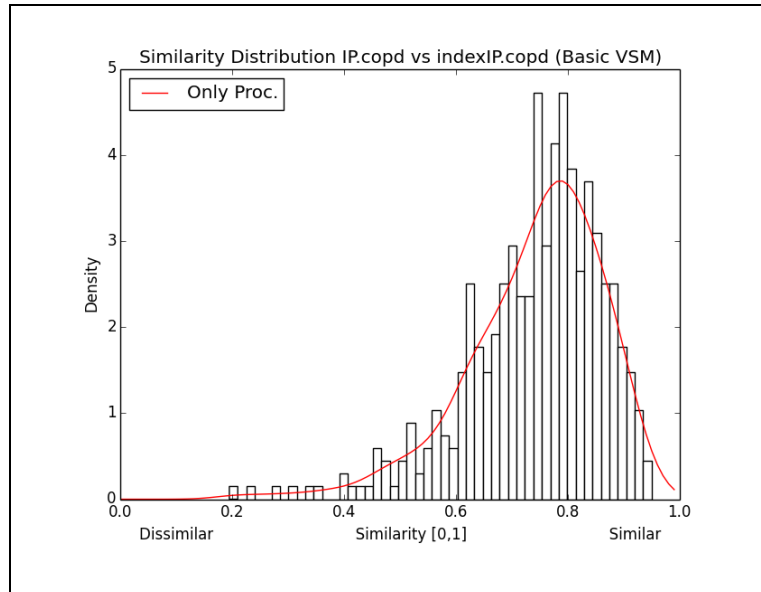
Figure 17: COPD Inpatients vs. Own Index Patient

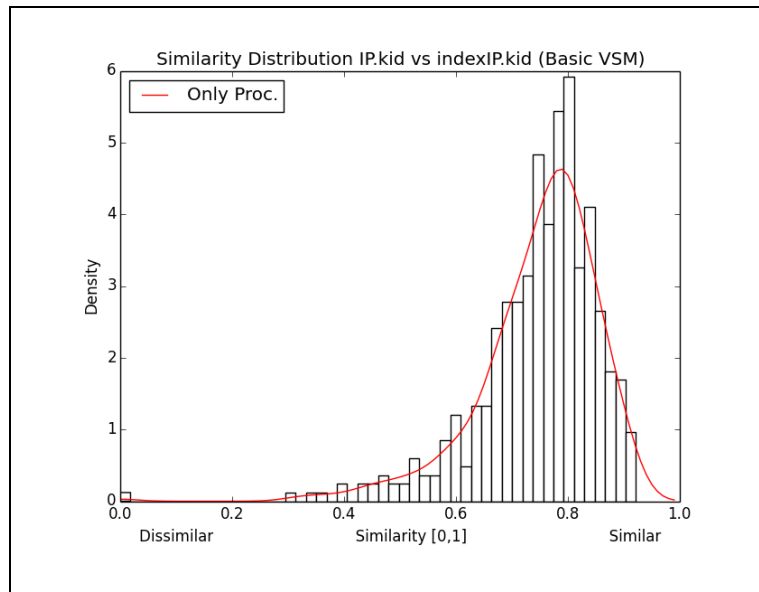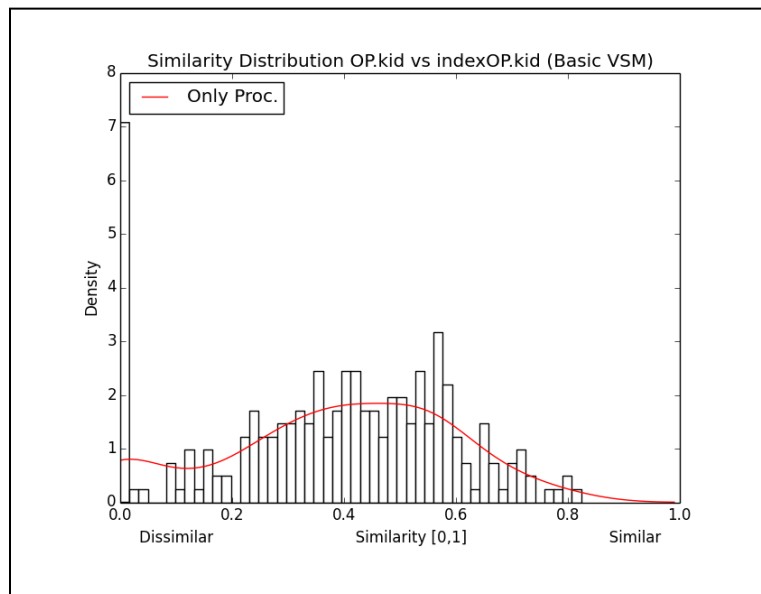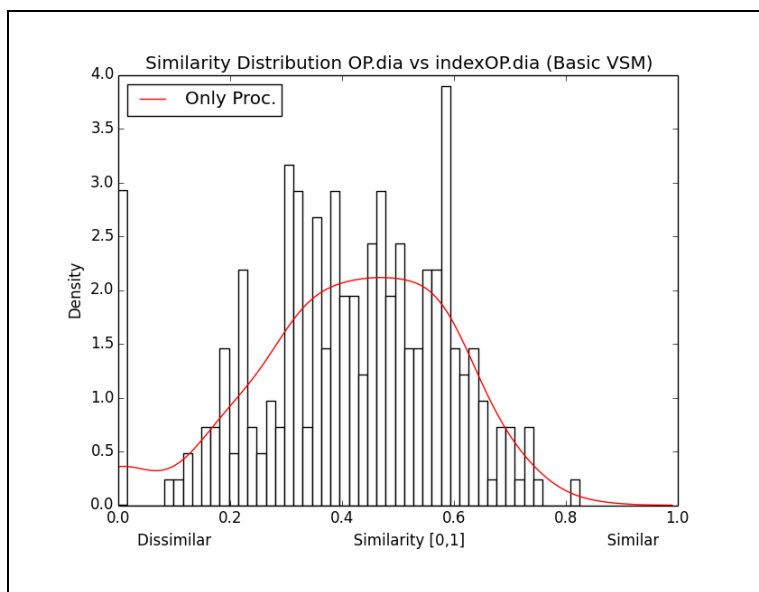**Figure 18: CKD Outpatients vs. Own Index Patient**

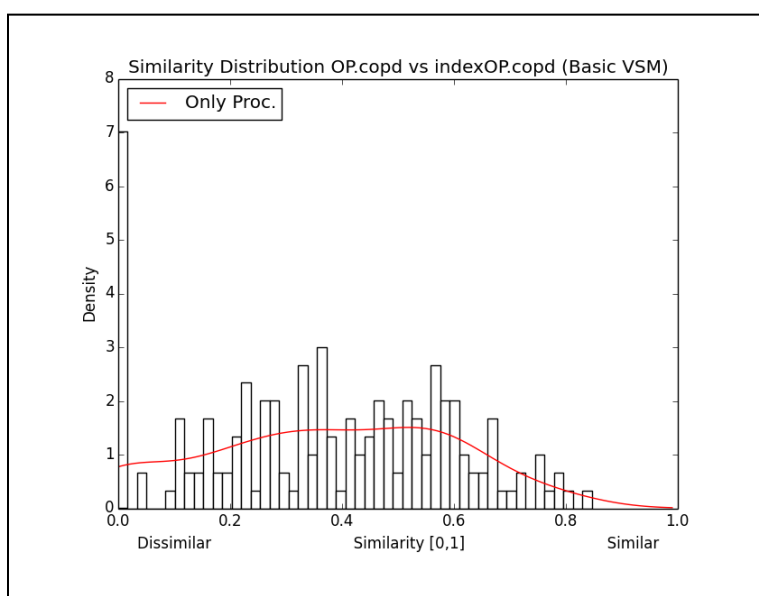**Figure 19: Diabetes Outpatient vs. Own Index Patient**

Figure 20: COPD Outpatients vs. Own Index Patient

3.2. Patient Groups Compared to Alternate Index Patients

Comparing patients within a population to an index patient chosen from that population provides a base level understanding of the treatment similarity within that patient group. We may also

want to determine how similar the treatments for patients in one population are to treatments provided to another population. To accomplish this, one approach is to calculate the treatment similarity between all patients in one population to an index patient chosen from another population. If the index patient is chosen as the patient who is provided the most similar treatments on average to the population he belongs to, then we may loosely view this index patient as representative of a typical patient within that group. Therefore, when we compare this index patient to all other patients in a different population we can investigate how similar the treatment choices are for patients in 2 different groups. We should again note that if there is much variation within a population then the index patient may not be a good representation of a typical patient. Nonetheless, we hypothesize that if the similarity is calculated for a population of patients as compared to an alternate index patient, then we should observe a left shift of the similarity distribution indicating decreased similarity.

For this study we applied the suggested approach to all possible comparisons of diabetes, CKD, and COPD comorbidity groups as well as general population groups. The first set of comparisons highlights the fact that our approach may reveal interesting characteristics of care provided to different groups of patients. Figure 21 shows the population of patients who have COPD as a comorbidity and both inpatient and outpatient visits are considered compared to the index patients for other comorbidity groups and the general population. From this graph, we can see that the similarity distributions formed when this population of patients is compared its own index patient or to the index patient for patients with chronic kidney disease are almost identical. The resemblance between these two distributions occurs because the treatment provided to CHF patients with COPD is comparable to the treatment provided to CHF patients with chronic kidney disease. Investigating the orders made for the COPD and CKD index patients supports this as well. Second, we notice that there is a very pronounced double peak in the similarity curve when this population is compared to the index patient from the diabetes group. The index patient for the diabetes group likely has some distribution of CHF

(and other non-uniquely diabetes) related orders and one portion of the COPD population receives these same orders in similar quantities (producing the peak on the right) while another portion of the COPD population receives these orders in more varying quantities (producing the peak towards the left). Factors influencing this division are currently uncertain and require further investigation, though it is apparent that the treatment for congestive heart failure may be divided into at least 2 categories, where many of the CHF patients with diabetes happen to receive treatment under one of these categories.
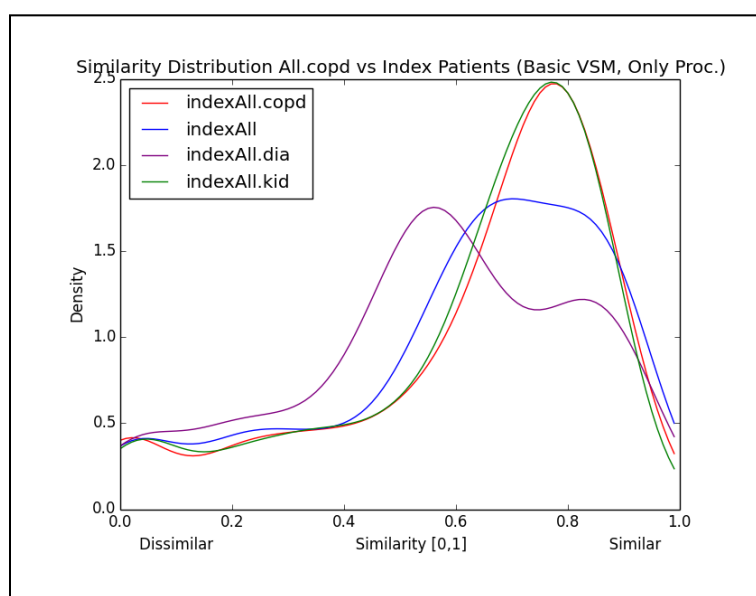
Figure 21: Combined IP/OP with COPD vs. Alternate Index Patients

In Figure 22, we compared all patients with combined IP/OP visits who had diabetes as comorbidity to the index patients from all other populations that considered combined IP/OP visits. While considering orders for both inpatient and outpatient visits, we wish to determine whether treatment choices for patients who have diabetes as a comorbidity are similar to treatment choices for patients who have COPD or CKD as a comorbidity. These comparisons show that when patients are compared to index patients from other patient groups the similarity curves are shifted to the left and are somewhat wider than when compared to the index patient from their own group. A curve shifted to the left indicates a decrease in treatment similarity to the alternate population and a wider curve means

that the patients are less consistently receiving similar treatments as the alternate group. We also expect that the smallest decrease in similarity will occur when a population is compared to the index patient from the general population rather than compared to index patients from a different comorbidity group because the general population includes patients that have the original comorbidity. The results for this patient group in Figure 21 support our hypothesis.
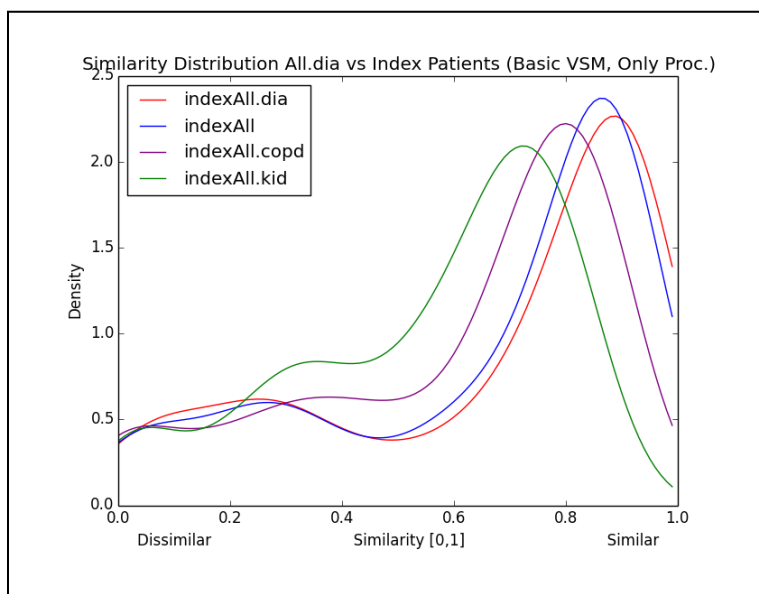
Figure 22: Combined IP/OP with Diabetes vs. Alternate Index Patients

Figure 23 shows the similarity distributions of patients with Chronic Kidney Disease as a comorbidity and only outpatient visits considered compared to the index patients from other patient groups. This plot also highlights the fact that treatment for CHF patients with certain comorbidities may be more distinguishable than for other comorbidities. In general this plot shows the expected outcome that when compared to index patients from other populations there is a shift to the left indicating less similarity. An interesting outcome in this plot is that when compared to the index patient for the diabetes group not only is there a shift to the left but the curve is very narrow indicating that the outpatient orders for chronic kidney disease patients are consistently dissimilar to the outpatient orders

for a diabetes patient. This dissimilarity is again due to the fact that many CHF patients with diabetes seems to fall under one category of treatment while CHF patients with other comorbidities, such as chronic kidney disease, sometimes receive treatment under the same category as the patients with diabetes and sometimes receive treatment under a different category.
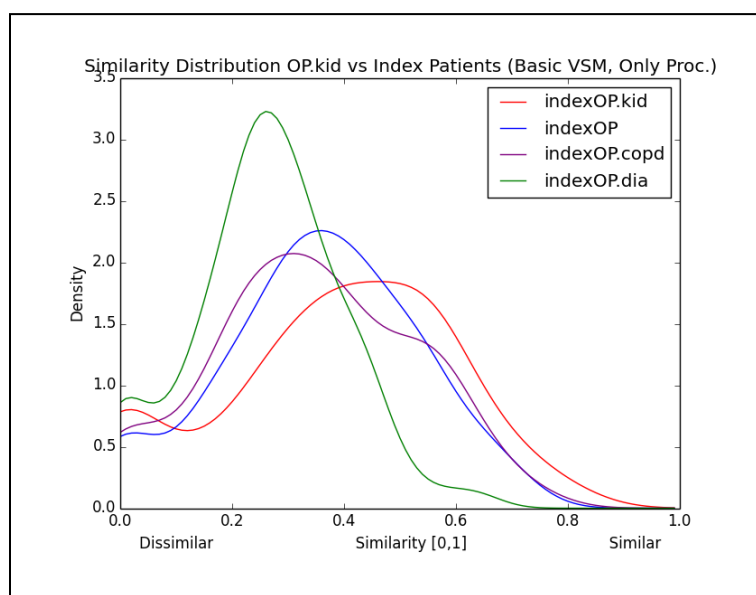
Figure 23: Outpatients with CKD vs. Alternate Index Patients

Figure 24 emphasizes further that patients with various comorbidities sometimes receive treatment, which is comparable to the treatment provided to diabetes patients, and sometimes receive treatment which falls under a different category. In this plot, a sample of patients with orders considered for both inpatient and outpatient visits, which do not have one of the 3 comorbidities of interest, is being compared to the index patients for comorbidity groups as well as for the general population. Here it seems that these "leftover" patients are most consistently similar to both the index patient for the chronic kidney disease and COPD groups. Since these patients do not have chronic kidney disease or COPD this suggests that patients with various comorbidities are treated for CHF in the same way that patients with CKD or COPD are treated. When compared to the index patient for the diabetes group these "leftover" patients seem to be divided into 2 levels of similarity. The same phenomena must

be occurring here as was occurring in Figure 22 when patients with COPD were compared to the index patient for the diabetes group. Finally, when the "leftover" patients are compared to the index patient from the general population the similarity distribution shows a moderate peak level of similarity with much variation. This is likely due to the fact that the general population has patients with diabetes in it, so the population of "leftover" patients will show less similarity to this general index patient in the same way that they were less similar to the diabetes index patient.
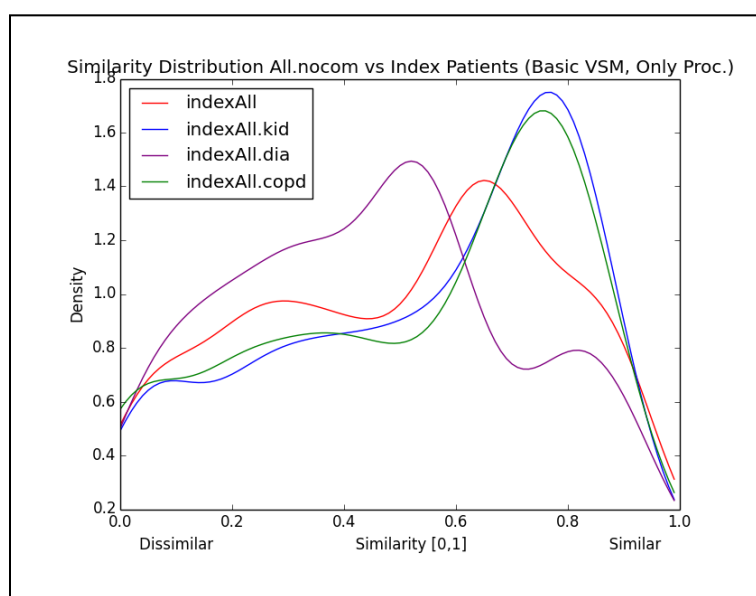
Figure 24: "Leftover" IP/OP vs Alternate Index Patients

3.3. Hybrid Inpatient Population

As an alternate way to compare patient groups, a hybrid inpatient sample was created which joined the inpatient comorbidity samples as well as the sample of inpatients which had no comorbidity of interest together. All patients within this hybrid group were then compared to a particular index patient from one of the comorbidity groups. Finally, the resulting similarity values were segregated according to which comorbidity group the patients originally belonged to so that we may plot the similarity distributions for each comorbidity group as compared to one index patient together on the same graph. For example, Figure 25 shows each comorbidity group compared to the chronic kidney

disease index patient. All patient groups show a comparable degree of similarity to this index patient with the exception of the chronic kidney disease group which shows a slightly narrower curve, indicating more consistent similarity, and the "leftover" patient group which shows more variation when compared to this index patient. The consistent similarity of the chronic kidney disease patients is to be expected since this is a CKD index patient. We are uncertain why there is greater variation among the patients with no comorbidity of interest when compared to the CKD index patient. There is no obvious reason why not having CKD, diabetes, or COPD would lead to more variation.

For verification of whether the similarity distributions obtained for each patient group are comparable to the similarity distribution for the chronic kidney disease patient group, a Anderson-Darling test was used. The COPD group obtained a p-value of 0.0509, which at the $\alpha = 0.05$ significance level suggests that the COPD similarity distribution and the CKD similarity distribution came from the same overlying distribution. The diabetes group obtained a p-value of 0.52309 which at the $\alpha = 0.05$ significance level suggests that the diabetes similarity distribution and the CKD similarity distribution came from the same overlying distribution. Finally, the "leftover" patient group obtained a p-value of 4.9072e-08 which at the $\alpha = 0.05$ significance level suggests that the COPD similarity distribution and the CKD similarity distribution did not come from the same overlying distribution. This A-D test result is consistent with what we interpret from the graph.

In general when compared to different index patients, all of the patient groups generally produce very comparable similarity distributions. The only exception is when each patient group is compared to the diabetes mellitus index patient. As expected, the diabetes group shows a high degree of similarity with the diabetes index patient. All other patient groups, however, have some proportion of patients with high similarity and some proportion of patients with low similarity when compared to the diabetes index patient as is evident by the double peaks on the curves found in Figure 26. These graphs only re-emphasize the fact that patients with diabetes as a comorbidity receive treatment under one

category while patients with other comorbidity sometimes receive treatment under the same category as diabetes and sometimes receive treatment under a different category.
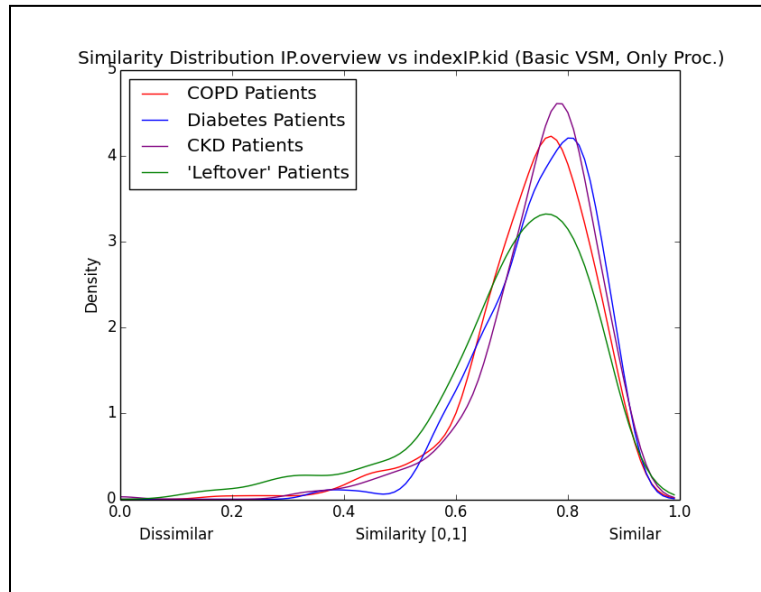
Figure 25: Inpatient Comorbidity Groups vs. CKD Index Patient
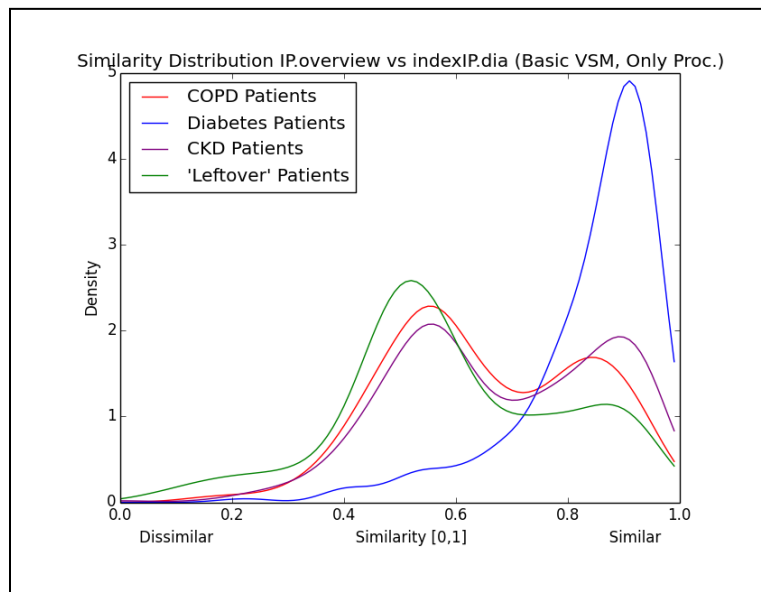
Figure 26: Inpatient Comorbidity Groups vs. Diabetes Index Patient

3.4. Similarity Distributions with Medication Orders Included

Finally, we investigated whether including medication orders in parallel with procedure orders would produce any new trends or noticeable difference in similarity distributions than when only procedure orders were considered. For all similarity distributions calculated, an index patient unique to the particular patient group was chosen. The key observations that can be made are reoccurring across different comorbidity groups and the general population. Therefore, to avoid repetition we will only discuss the similarity distributions for the inpatient, outpatient, and combined IP/OP groups of patients who have Chronic Kidney disease as a comorbidity. As shown in Figures 27-29, the resulting similarity distributions are very comparable whether medication orders are included or not. This is a reasonable outcome because of two reasons: 1.) Procedure orders were considered in both cases so there should be a base level of similarity between the curves, and 2.) If patients within a particular group are receiving procedures at a certain level of consistency then it is likely that they are receiving medications at a similar level of consistency. We cannot say that this will always be the case but we can say this result is reasonable. Additionally, when comparing these distributions we may say that adding the medication order information causes the curves to widen slightly. This suggests that the population is less consistently receiving similar orders. This difference is most likely due to the fact that there are more features considered in the similarity computation (procedures plus medications) which leaves room for variation. Ultimately our results suggest that the variability of medication and procedure orders provided to congestive heart failure patients is comparable.
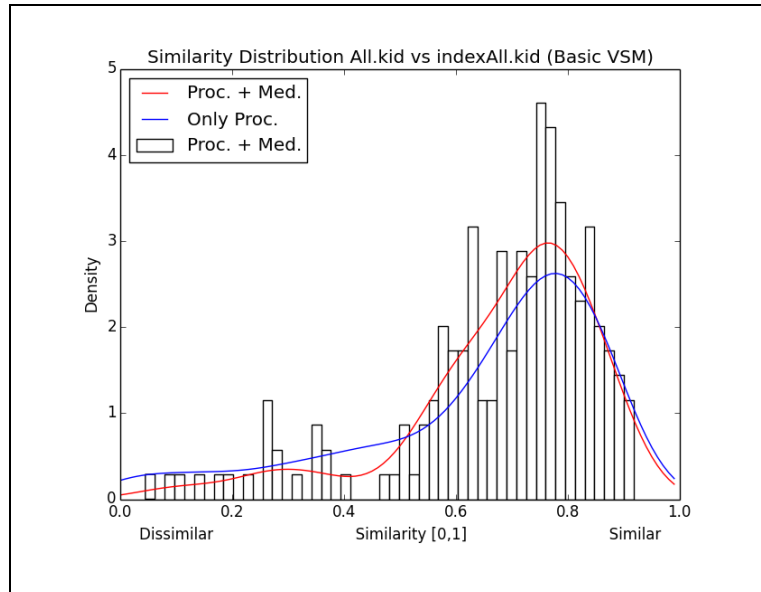
Figure 27: CKD Combined IP/OP with Medication Orders
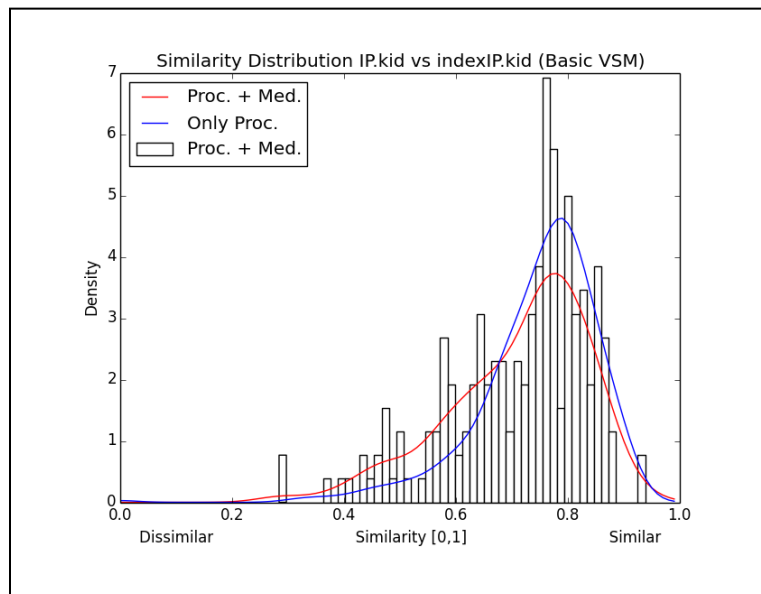
Figure 28: CKD Inpatients with Medication Orders

Similarity Distribution OP.kid vs indexOP.kid (Basic VSM)
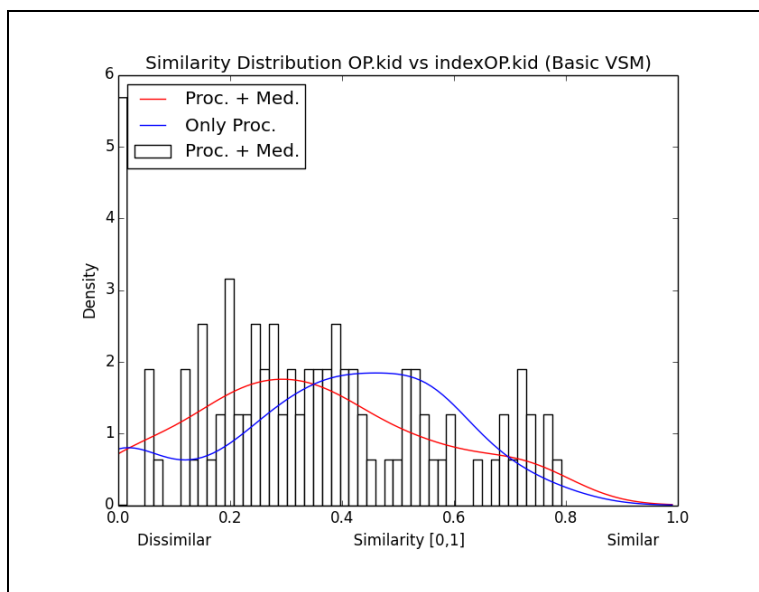
Figure 29: CKD Outpatients with Medication Orders

3.5. Subject Matter Expert Similarity Assessment

Table 8 provides an overview of the results of the subject matter expert study. Though we obtained a noticeable amount of disagreement between the subject matter expert's responses and the original rankings, the SME's themselves showed a decent amount of disagreement between each other, as apparent by the Kappa statistic value of 0.284 for the overall study. Additionally, if we consider a 2-step response difference to be when one party provides a rank of high while another party provides a rank of low, both SME's infrequently obtain a 2-step difference compared to the original rankings, or compared to each other. Thus the response disagreements are not often completely diverging. Taking these study outcomes into consideration, we concluded that the results are sufficient for initial validation of our methods.

Table 8: Subject Matter Expert Results Overview

	SME's Disagree	SME Agreement Kappa Statistic	SME 1 Disagree w/ Original	SME 2 Disagree w/ Original	SME's 2-Step Difference	SME 1 2-Step Diff. vs Original	SME 2 2-Step Diff. vs Original
General IP	5/9	0.210	5/9	2/9	0/11	0/13	0/10
General OP	5/9	0.062	7/9	7/9	1/8	2/9	0/11
Diabetes IP	1/9	0.608	8/9	6/9	0/15	3/14	3/13
Diabetes OP	4/9	0.25	4/9	3/9	0/10	1/11	0/11
Total	15/36	0.284	24/36	18/36	1/44	6/47	3/45

Table 9 below summarizes the subject matter expert responses. Surveying the results for the general inpatient group, SME 1 seems to be quite reserved when ranking a patient as having a greater amount of similarity to the index patient. SME 1 ranks many patients as having low similarity, 2 as having medium similarity and 1 as having high similarity. However, the general trend of ranking certain patients as having greater similarity and certain patients as having less similarity is comparable to our original rankings. For example, those patients whom we ranked as having low similarity SME 1 also ranked as having low similarity. Those patients whom we ranked as having medium similarity SME 1 ranked as having medium or low similarity. Finally those patients whom we ranked as having high similarity SME 1 ranked as having medium or high similarity. SME 2 also maintained the general trend of ranking the same patients as having higher similarity as our method did as well as ranking the same patients as having less similarity as our method did. SME 2, however, was more willing to rank a patient as having medium similarity than SME 1 was.

The subject matter expert responses for the general outpatient group are less consistent with our original rankings. SME 1 is again reluctant to rank any patient as having high similarity. This reluctance is understandable for the outpatient group because our results indicate high variation within the outpatient population. Those patients which we ranked as having high similarity do not actually have quite as high of a similarity score as those patients which we ranked as having high similarity within the inpatient group. In contrast to the results for the general inpatient group, SME 1 is more willing to rank patients as having medium similarity within the general outpatient group. Interestingly, however, there are 2 patients which were ranked as having high similarity using our method that SME 1 ranked as having low similarity. Looking at notes SME 1 provided on factors which contributed to his/her decisions, it seems that these 2 patients had orders for the treatment or monitoring of certain comorbidities which stood out to SME 1, distinguishing these patients from the index patient. This is important to consider because the distinguishing orders which stood out to SME 1 likely did not occur frequently which is why our method did not rank these patients as having less similarity to the index patient. The responses from SME 2 were closer to the rankings provided by our method in the sense that SME 2 did not rank any patients which we ranked as having high similarity as having low similarity or vice versa. In general both SME 1 and SME 2 show a preference toward ranking outpatients as having medium similarity. This is likely because the outpatient group shows much variation, so distinguishing between patients which have slightly greater or slightly less similarity to the index patient may be difficult.

Table 9: Subject Matter Expert Assessment Results

General Inpatient Group

	Original Rank	SME 1	SME 2
Patient 1	Medium	Low	Medium
Patient 2	High	Medium	Medium
Patient 3	High	High	High
Patient 4	Low	Low	Low
Patient 5	Low	Low	Low
Patient 6	High	Medium	High
Patient 7	Medium	Low	Medium
Patient 8	Medium	Low	Medium
Patient 9	Low	Low	Medium

General Outpatient Group

	Original Rank	SME 1	SME 2
Patient 1	High	Medium	Medium
Patient 2	Medium	Medium	Low
Patient 3	Low	Medium	Low
Patient 4	High	Low	High
Patient 5	Low	Medium	Medium
Patient 6	Medium	Medium	High
Patient 7	Low	Medium	Medium
Patient 8	High	Low	Medium
Patient 9	Medium	Low	Low

Diabetes Inpatient Group

	Original Rank	SME 1	SME 2
Patient 1	High	Low	Low
Patient 2	High	Low	Low
Patient 3	Medium	Low	Low
Patient 4	Low	Low	Low
Patient 5	Medium	Low	Medium
Patient 6	Low	Medium	Medium
Patient 7	High	Low	Low
Patient 8	Low	Low	Low
Patient 9	Medium	Low	Low

Diabetes Outpatient Group

	Original Rank	SME 1	SME 2
Patient 1	Low	Low	Low
Patient 2	Low	Low	Low
Patient 3	Medium	High	Medium
Patient 4	Medium	Medium	Low
Patient 5	High	Medium	Medium
Patient 6	Low	Low	Low
Patient 7	Medium	Medium	Medium
Patient 8	High	Low	Medium
Patient 9	High	Medium	High

As an alternate way to assess the subject matter expert results, Table 10 shows the distribution

of high, medium, and low rankings provided conditioned upon the original ranking obtained using our

method. For example, out of the 3 patients from the general inpatient group which we ranked as having

high similarity using our approach, 1 of these patients SME 1 ranked as having high similarity and 2 of these patients SME 1 ranked as having medium similarity to the index patient. More importantly, Table 9 shows average the distribution of high, medium, and low rankings provided by both subject matter experts conditioned upon the original ranking obtained using our method. Reiterating the observations made earlier, patients from the general inpatient group which we ranked as having high similarity are often ranked by the SME's as having medium or high similarity. Patients which we ranked as having medium similarity are often ranked by the SME's as having medium or low similarity. Finally, patients which we ranked as having low similarity are most frequently ranked by the SME's as having low similarity and occasionally as having medium similarity.

For the general outpatient group, patients which we ranked as having high similarity are most often ranked by the SME's as having medium similarity but occasionally are ranked as having high or low similarity. Patients which we ranked as having medium similarity are most often ranked by the SME's as having low similarity but occasionally are ranked as having medium or high similarity. Finally patients which we ranked as having low similarity are frequently ranked by the SME's as having medium similarity and occasionally as having low similarity. Based on these results, subject matter experts provide rankings which are somewhat comparable to the original rankings, however, the physicians are less willing to rank patients as having high similarity to the index patients. Additionally, their responses regarding the general inpatient group were more consistent with both our original rankings and with each other, than their responses regarding the general outpatient group. Both physicians indicated that they considered the types of procedures ordered and various groups of procedures ordered which characterize aspects of the patient's health when making their judgments. The number of times each procedure was ordered was a secondary consideration.

Table 10: Subject Matter Expert Rank Distributions

General Inpatient Group			
Original Rank	SME 1	SME 2	Average
High	33.33% High, 66.66% Medium	66.66% High, 33.33% Medium	50% High, 50% Medium
Medium	100% Low	100% Medium	50% Low, 50% Medium
Low	100% Low	66.66% Low, 33.33% Medium	83.33% Low, 16.66% Medium
General Outpatient Group			
Original Rank	SME 1	SME 2	Average
High	33.33% Medium, 66.66% Low	66.66% Medium, 33.33% High	33.33% Low, 50% Medium, 16.66% High
Medium	66.66% Medium, 33.33% Low	66.66% Low, 33.33% High	50% Low, 33.33% Medium, 16.66% High
Low	100% Medium	66.66% Medium, 33.33% Low	83.33% Medium, 16.66% Low
Diabetes Inpatient Group			
Original Rank	SME 1	SME 2	Average
High	100% Low	100% Low	100% Low
Medium	100% Low	66.66% Low, 33.33% Medium	83.33% Low, 16.66% Medium
Low	66.66% Low, 33.33% Medium	66.66% Low, 33.33% Medium	66.66% Low, 33.33% Medium
Diabetes Outpatient Group			
Original Rank	SME 1	SME 2	Average
High	66.66% Medium, 33.33% Low	66.66% Medium, 33.33% High	66.66% Medium, 16.66% Low, 16.66% High
Medium	66.66% Medium, 33.33% High	66.66% Medium, 33.33% Low	66.66% Medium, 16.66% Low, 16.66% High
Low	100% Low	100% Low	100% Low

4. Limitations and Future Work

The computational methods developed in this study provide a framework to begin understanding the treatment variation within a population of patients using an EMR system. However, this study has a number of limitations which must be addressed. As an initial strategy for understanding care variation, we calculated the treatment similarity of a population of patients as compared to an

index patient chosen from within that population. The width of the similarity curve indicates how consistently patients within that population are receiving treatment with some level of similarity. However, since we chose this index patient from the population we are investigating it is hard to say whether that "level of similarity" is optimal for that population. To account for this in the future, one approach is to create an artificial patient to use an as index patient which captures the treatment choices suggested in clinical guidelines. Another approach, which may be easier to attempt, is to compare the order history of the current index patients to clinical guidelines. This may provide some insight as to whether the population as a whole is being provided treatment similar to guideline recommendations.

Second, we investigated the treatment similarity of patient groups to alternate index patients chosen from other populations of interest. Our results highlight the fact that treatment for certain patient groups may be more distinguishable than for other patient groups. The results of our study revealed that the treatment provided to congestive heart failure patients with diabetes as a comorbidity is more distinguishable than the treatment provided to congestive heart failure patients with COPD or chronic kidney disease as a comorbidity. In the future we may want to compare other patient groups such as CHF patients at the UVA Hospital and CHF patients at competing hospital, or patients which are being provided care by two different physicians.

Another limitation of this study is that it ignores factors such as order sequence and patient demographics. Though we might hypothesize that the most significant types of treatments always occur in a particular order and so incorporating sequence information may not change the resulting similarity distributions, we cannot assume that this is the case. There are techniques which may be considered, such as n-gram models used in natural language processing, which replace the dimensions in the vector space model with sequences of procedures, rather than individual procedure types. Additionally, since we do not account for patient demographics or any other details specific to a patient's state of health,

our results may give the impression that the treatment choices provided to a group of patients are more varied than is justifiable. More investigation should be done on how order sequence and more detailed patient information should be incorporated into our approach.

Also, we have not conclusively investigated whether a simple contingency table or one with alternative weighting should be used. With the simple contingency table, all procedures are weighted equally, so if orders pertaining to the treatment of a certain comorbidity appear with a high frequency then those orders will have a large influence on the similarity calculation. It may be beneficial to weigh orders pertaining to a particular illness of interest higher, or possibly to remove all orders completely which do not pertain to that illness. Weightings learned from approaches such as Latent Semantic Analysis, Latent Dirichlet Allocation, and the Term Frequency–Inverse Document Frequency (TF-IDF) approach were briefly investigated (see Appendix B), but it did not seem apparent that the weightings produced were appropriate for the study of treatment similarity.

Finally, further validation of method utilizing subject matter experts is necessary. Further insight could be obtained from having more than 2 SME's participate and presenting a larger number of patients for each distribution. In addition, the subject matter experts could provide more accurate rankings if they had enough time to only work on the rankings for a set of patients as compared to one index patient per day. Keeping track of which index patient is being compared when the SME has looked at several can be a challenge. Finally, a study could be designed which presents patient information to the subject matter experts in a way which is more natural for them to interpret than a list of orders.

5. Discussion

The results of this study provide much insight on care variation within the congestive heart failure population at the UVA hospital and the methods presented here provide much opportunity for expansion and future applications. The first set of similarity distributions calculated by comparing

patients within each group to an index patient uniquely chosen for each group provides a base level understanding of the treatment similarity within each group. The width of each similarity curve reflects the spread of similarity values obtained, with a wider distribution indicating more treatment variation. Though the peaks of each distribution cannot be compared between populations, each index patient may be compared to clinical guidelines. If a particular index patient is receiving treatment which is consistent with guideline recommendations, then obtaining a corresponding similarity distribution with a peak at a high similarity value would suggest that the population of patients are receiving care which is consistent with the guidelines. On the other hand, if an index patient is not receiving treatment which is consistent with guidelines, then obtaining a similarity distribution with a high mode value suggests that patients within this population, though receiving similar care, are not receiving care which is consistent to clinical guideline recommendations.

A second approach to comparing different patient groups presented in this study is to calculate the treatment similarity for one group of patients as compared to the index patient chosen from another patient group of interest. If a particular patient group does not have much treatment variation, then the index patient chosen from that group may be viewed as representative of a typical patient within that group. Therefore, when treatment similarity is calculated for another population as compared to this index patient, we may interpret from our results the degree of treatment similarity between the two populations. Even if the index patient is not representative of a typical patient we can still expect that the treatment similarity distributions of patient groups as compared to an alternate index patient should show a decrease in similarity if the two patient groups in question are reasonably different in terms of diagnosis. When we calculated the similarity of comorbidity groups as compared to the index patient from other comorbidity groups we observed this decrease in similarity.

A final approach to comparing the treatment similarity of different patient groups is to join all patient groups in question into one hybrid population and calculate the treatment similarity of all

patients compared to the same index patient. Using this approach, similarity calculations are made for each patient group using the same number of dimensions for the vector space models. Plotting the resulting similarity values for each patient group on the same graph allows us to compare not only the spread of treatment similarity within each patient group but also the absolute similarity values obtained, represented by the modes of the distributions. This approach of comparing multiple patient groups to the same index patient is quite insightful. Not only can we use this approach to compare the treatment variation between different comorbidity groups, as shown in this study, but additionally, if an index patient were chosen or created to somehow represent a patient who receives care according to guidelines, then different patient groups could simultaneously be compared to guideline recommendations. In theory we could provide a way to compare patients at different hospitals or who are being cared for by different physicians to see which groups of patients are being treated more consistently with clinical guidelines. In this way, our methods may contribute to healthcare quality assessment.

6. Conclusions

The approach presented in this study clearly offers much potential for understanding care variation within a population of patients. The use of a simple contingency table allows us to capture a patient's entire treatment history in a way that is relatable to other patients while not putting excessive emphasis on only a few specific treatments. Then using this contingency table and an index patient chosen from within the population we can calculate a vector space model to obtain a base level understanding of the treatment similarity within that population. Additionally, the treatment similarity between different populations can be investigated by comparing patients within one population to an index patient chosen from another population of interest. Finally, if we join different patient populations of interest into one hybrid population and calculate the treatment similarity of all patients as compared

to the same index patient, then this potentially allows us to directly compare how consistent treatment for each patient group is to clinical guidelines. In general, our results show that the methods presented in this study can successfully distinguish between populations we expect to receive less similar treatment. Using the approach developed in this study as a foundation and the convenience of data collected from EMR systems, the treatment choices of future generations may be monitored regularly to ensure consistent and quality care is provided.

References

[1]   *Institute of Medicine. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary*. Washington, DC: The National Academies Press, 2010.

[2]   "HCUPnet, Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality." [Online]. Available: http://hcupnet.ahrq.gov/.

[3]   A. a Fitzgerald, L. a Allen, and F. a Masoudi, "The evolving landscape of quality measurement for heart failure.," *Annals of the New York Academy of Sciences*, vol. 1254, pp. 131–9, Apr. 2012.

[4]   J. Z. Ayanian, J. S. Weissman, S. Chasan-Taber, and a M. Epstein, "Quality of care by race and gender for congestive heart failure and pneumonia.," *Medical care*, vol. 37, no. 12, pp. 1260–9, Dec. 1999.

[5]   L. D. DiMartino, A. M. Shea, A. F. Hernandez, and L. H. Curtis, "Use of guideline-recommended therapies for heart failure in the Medicare population.," *Clinical cardiology*, vol. 33, no. 7, pp. 400–5, Jul. 2010.

[6]   J. C. Luthi, W. M. McClellan, D. Fitzgerald, J. Herrin, R. J. Delaney, H. M. Krumholz, D. W. Bratzler, K. Elward, C. B. Cangialose, and D. J. Ballard, "Variations among hospitals in the quality of care for heart failure.," *Effective clinical practice : ECP*, vol. 3, no. 2, pp. 69–77, 2000.

[7]   J. Mitchell and D. Ballard, "Racial variation in treatment for transient ischemic attacks: impact of participation by neurologists.," *Health Services Research Journal*, 2000.

[8]   A. C. Plint, D. W. Johnson, N. Wiebe, B. Bulloch, M. Pusic, G. Joubert, P. Pianosi, T. Turner, G. Thompson, and T. P. Klassen, "Practice Variation among Pediatric Emergency Departments in the Treatment of Bronchiolitis," *Academic Emergency Medicine*, vol. 11, no. 4, pp. 353–360, Apr. 2004.

[9]   S. Rathore, J. Foody, and Y. Wang, "Race, quality of care, and outcomes of elderly patients hospitalized with heart failure," *JAMA: the journal of the American Medical Association*, vol. 289, no. 19, pp. 2517–2524, 2003.

[10]   G. C. Fonarow, C. W. Yancy, N. M. Albert, A. B. Curtis, W. G. Stough, M. Gheorghiade, J. T. Heywood, M. L. McBride, M. R. Mehra, C. M. O'Connor, D. Reynolds, and M. N. Walsh, "Heart failure care in the outpatient cardiology practice setting: findings from IMPROVE HF.," *Circulation Heart failure*, vol. 1, no. 2, pp. 98–106, Jul. 2008.

[11]   M. E. Edep, N. B. Shah, I. M. Tateo, and B. M. Massie, "Differences between primary care physicians and cardiologists in management of congestive heart failure: relation to practice guidelines.," *Journal of the American College of Cardiology*, vol. 30, no. 2, pp. 518–26, Aug. 1997.

[12]   D. W. Baker, R. P. Hayes, B. M. Massie, and C. a Craig, "Variations in family physicians' and cardiologists' care for patients with heart failure.," *American heart journal*, vol. 138, no. 5 Pt 1, pp. 826–34, Nov. 1999.

[13]   A. Majeed, K. Moser, and R. Maxwell, "Age, sex and practice variations in the use of statins in general practice in England and Wales.," *Journal of public health medicine*, vol. 22, no. 3, pp. 275–9, Sep. 2000.

[14]   K. Thomas, A. Hernandez, and D. Dai, "Association of race/ethnicity with clinical risk factors, quality of care, and acute outcomes in patients hospitalized with heart failure," *American heart journal*, vol. 161, no. 4, pp. 746–754, 2011.

[15]   D. E. Lanfear, T. N. Hrobowski, E. L. Peterson, K. E. Wells, T. V Swadia, J. a Spertus, and L. K. Williams, "Association of β-blocker exposure with outcomes in heart failure differs between African American and white patients.," *Circulation Heart failure*, vol. 5, no. 2, pp. 202–8, Mar. 2012.

[16]   P. Jong, Y. Gong, P. P. Liu, P. C. Austin, D. S. Lee, and J. V Tu, "Care and outcomes of patients newly hospitalized for heart failure in the community treated by cardiologists compared with other specialists.," *Circulation*, vol. 108, no. 2, pp. 184–91, Jul. 2003.

[17]   J. Veloski, S. Tai, A. S. Evans, and D. B. Nash, "Clinical vignette-based surveys: a tool for assessing physician practice variation.," *American journal of medical quality : the official journal of the American College of Medical Quality*, vol. 20, no. 3, pp. 151–7, 2005.

[18]   J. Tu, D. Ko, H. Guo, and J. Richards, "Determinants of variations in coronary revascularization practices," *Canadian Medical Association Journal*, vol. 184, no. 2, pp. 179–186, 2012.

[19]   B. Xhyheri and R. Bugiardini, "Diagnosis and Treatment of Heart Disease: Are Women Different From Men?," *Progress in Cardiovascular Diseases*, vol. 53, no. 3, pp. 227–236, Nov. 2010.

[20]   R. Sainsbury, "Does it matter where you live ? Treatment variation for breast cancer in Yorkshire," pp. 1275–1278, 1995.

[21]   G. T. O'Connor, H. B. Quinton, N. D. Traven, L. D. Ramunno, T. a Dodds, T. a Marciniak, and J. E. Wennberg, "Geographic variation in the treatment of acute myocardial infarction: the Cooperative Cardiovascular Project.," *JAMA : the journal of the American Medical Association*, vol. 281, no. 7, pp. 627–33, Feb. 1999.

[22]   A. Nattinger and M. Gottlieb, "Geographic variation in the use of breast-conserving treatment for breast cancer," *New England Journal of Medicine*, 1992.

[23]   K. Fox, "Management of acute coronary syndromes. Variations in practice and outcome. Findings from the Global Registry of Acute Coronary Events (GRACE)," *European Heart Journal*, vol. 23, no. 15, pp. 1177–1189, Aug. 2002.

[24]   F. Peters-Klimm, G. Laux, S. Campbell, T. Müller-Tasch, N. Lossnitzer, J.-H. Schultz, A. Remppis, J. Jünger, and C. Nikendei, "Physician and patient predictors of evidence-based prescribing in heart failure: a multilevel study.," *PloS one*, vol. 7, no. 2, p. e31082, Jan. 2012.

[25]   a Zink, J. Listing, S. Ziemer, and H. Zeidler, "Practice variation in the treatment of rheumatoid arthritis among German rheumatologists.," *The Journal of rheumatology*, vol. 28, no. 10, pp. 2201–8, Oct. 2001.

[26]   M. E. Patterson, A. F. Hernandez, B. G. Hammill, G. C. Fonarow, E. D. Peterson, K. a Schulman, and L. H. Curtis, "Process of care performance measures and long-term outcomes in patients hospitalized with heart failure.," *Medical care*, vol. 48, no. 3, pp. 210–6, Mar. 2010.

[27]   S. S. Rathore, a. K. Berger, K. P. Weinfurt, M. Feinleib, W. J. Oetgen, B. J. Gersh, and K. a. Schulman, "Race, Sex, Poverty, and the Medical Treatment of Acute Myocardial Infarction in the Elderly," *Circulation*, vol. 102, no. 6, pp. 642–648, Aug. 2000.

[28]   J. L. Givens, J. Tjia, C. Zhou, E. Emanuel, and A. S. Ash, "Racial and ethnic differences in hospice use among patients with heart failure.," *Archives of internal medicine*, vol. 170, no. 5, pp. 427–32, Mar. 2010.

[29]   J. Whittle and J. Conigliaro, "Racial differences in the use of invasive cardiovascular procedures in the Department of Veterans Affairs medical system," *The New England Journal of Medicine*, 1993.

[30]   A. Sonel, C. Good, J. Mulgund, and M. Roe, "Racial Variations in Treatment and Outcomes of Black and White Patients With High-Risk Non–ST-Elevation Acute Coronary Syndromes Insights From CRUSADE," *Circulation*, 2005.

[31]   S. Stewart and C. Demers, "Substantial between-hospital variation in outcome following first emergency admission for heart failure," *European Heart Journal*, pp. 650–657, 2002.

[32]   M. Komajda, "The EuroHeart Failure Survey programme—a survey on the quality of care among patients with heart failure in Europe Part 2: treatment," *European Heart Journal*, vol. 24, no. 5, pp. 464–474, Mar. 2003.

[33]   S. Sporer, "The geographic incidence and treatment variation of common fractures of elderly patients," *Journal of the American Academy of Orthopaedic Surgeons*, 2006.

[34]   M. R. Cooperberg, J. M. Broering, and P. R. Carroll, "Time trends and local variation in primary treatment of localized prostate cancer.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 28, no. 7, pp. 1117–23, Mar. 2010.

[35]   S. H. Kataoka, L. Zhang, and K. B. Wells, "Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status.," *The American journal of psychiatry*, vol. 159, no. 9, pp. 1548–55, Sep. 2002.

[36]   J. W. Peabody, F. Tozija, J. a Muñoz, R. J. Nordyke, and J. Luck, "Using vignettes to compare the quality of clinical care variation in economically divergent countries.," *Health services research*, vol. 39, no. 6 Pt 2, pp. 1951–70, Dec. 2004.

[37]   F. L. Lucas, B. E. Sirovich, P. M. Gallagher, A. E. Siewers, and D. E. Wennberg, "Variation in cardiologists' propensity to test and treat: is it associated with regional variation in utilization?," *Circulation Cardiovascular quality and outcomes*, vol. 3, no. 3, pp. 253–60, May 2010.

[38]   S. L. Krein, T. P. Hofer, E. a Kerr, and R. a Hayward, "Whom should we profile? Examining diabetes care practice variation among primary care providers, provider groups, and health care facilities.," *Health services research*, vol. 37, no. 5, pp. 1159–80, Oct. 2002.

[39]   D. Farrow, W. Hunt, and J. Samet, "Geographic variation in the treatment of localized breast cancer," *New England Journal of Medicine*, 1992.

[40]   R. Giugliano, "Elderly patients receive less aggressive medical and invasive management of unstable angina: potential impact of practice guidelines," *Archives of Internal Medicine*, pp. 1113–1120, 1998.

[41]   H. M. Krumholz, J. Chen, S. S. Rathore, Y. Wang, and M. J. Radford, "Regional variation in the treatment and outcomes of myocardial infarction: investigating New England's advantage.," *American heart journal*, vol. 146, no. 2, pp. 242–9, Aug. 2003.

[42]   G. T. O. Connor, N. D. Traven, L. D. Ramunno, T. A. Dodds, T. A. Marciniak, and J. E. Wennberg, "Geographic Variation in the Treatment of Acute Myocardial Infarction The Cooperative Cardiovascular Project," *JAMA: the journal of the American Medical Association*, vol. 281, no. 7, pp. 627–633, 2013.

[43]   L. Pilote, R. Califf, and S. Sapp, "Regional variation across the United States in the management of acute myocardial infarction," *New England journal of Medicine*, vol. 333, no. 9, pp. 565–572, 1995.

[44]   P. Davis, B. Gribben, a Scott, and R. Lay-Yee, "The 'supply hypothesis' and medical practice variation in primary care: testing economic and clinical models of inter-practitioner variation.," *Social science & medicine (1982)*, vol. 50, no. 3, pp. 407–18, Feb. 2000.

[45]   T. R. Dresselhaus, J. W. Peabody, J. Luck, and D. Bertenthal, "An evaluation of vignettes for predicting variation in the quality of preventive care.," *Journal of general internal medicine*, vol. 19, no. 10, pp. 1013–8, Oct. 2004.

[46]   C. W. Yancy, M. Jessup, B. Bozkurt, J. Butler, D. E. Casey, M. H. Drazner, G. C. Fonarow, S. a Geraci, T. Horwich, J. L. Januzzi, M. R. Johnson, E. K. Kasper, W. C. Levy, F. a Masoudi, P. E. McBride, J. J. V McMurray, J. E. Mitchell, P. N. Peterson, B. Riegel, F. Sam, L. W. Stevenson, W. H. W. Tang, E. J. Tsai, and B. L. Wilkoff, "2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines.," *Journal of the American College of Cardiology*, vol. 62, no. 16, pp. e147–239, Oct. 2013.

[47]   C. Y. Wong, S. I. Chaudhry, M. M. Desai, and H. M. Krumholz, "Trends in comorbidity, disability, and polypharmacy in heart failure.," *The American journal of medicine*, vol. 124, no. 2, pp. 136–43, Feb. 2011.

[48]   U. Dahlström, "Frequent non-cardiac comorbidities in patients with chronic heart failure.," *European journal of heart failure*, vol. 7, no. 3, pp. 309–16, Mar. 2005.

[49]   A. Mosterd and A. W. Hoes, "Clinical epidemiology of heart failure.," *Heart (British Cardiac Society)*, vol. 93, no. 9, pp. 1137–46, Sep. 2007.

[50]   A. Marengoni, D. Rizzuto, H.-X. Wang, B. Winblad, and L. Fratiglioni, "Patterns of chronic multimorbidity in the elderly population.," *Journal of the American Geriatrics Society*, vol. 57, no. 2, pp. 225–30, Feb. 2009.

[51]   C. C. Lang and D. M. Mancini, "Non-cardiac comorbidities in chronic heart failure.," *Heart (British Cardiac Society)*, vol. 93, no. 6, pp. 665–71, Jun. 2007.

[52]   M. a. Vedomske, M. S. Gerber, D. E. Brown, and J. H. Harrison, "Scalable and Locally Applicable Measures of Treatment Variation That Use Hospital Billing Data," *2013 12th International Conference on Machine Learning and Applications*, pp. 422–427, Dec. 2013.

Appendix A: Comorbidity ICD-9 Codes

Comorbidity	ICD-9 Codes
Chronic Kidney Disease	585, 585.1, 585.2, 585.3, 585.4, 585.5, 585.6, 585.9, 792.5, V42.0, V45.1, V45.11, V45.12, V56.0, V56.1, V56.2, V56.31, V56.32, V56.8
Diabetes Mellitus	
Without complication	249.00, 250.00, 250.01, 790.2, 790.21, 790.22, 790.29, 791.5, 791.6, V45.85, V53.91, V65.46
With complications	249.01, 249.10, 249.11, 249.20, 249.21, 249.30, 249.31, 249.40, 249.41, 249.50, 249.51, 249.60, 249.61, 249.70, 249.71, 249.80, 249.81, 249.90, 249.91, 250.02, 250.03, 250.10, 250.11, 250.12, 250.13, 250.20, 250.21, 250.22, 250.23, 250.30, 250.31, 250.32, 250.33, 250.40, 250.41, 250.42, 250.43, 250.50, 250.51, 250.52, 250.53, 250.60, 250.61, 250.62, 250.63, 250.70, 250.71, 250.72, 250.73, 250.80, 250.81, 250.82, 250.83, 250.90, 250.91, 250.92, 250.93
Chronic Obstructive Pulmonary Disease (and Bronchiectasis)	490, 491.0, 491.1, 491.2, 491.20, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 494, 494.0, 494.1, 496

Appendix B: Alternate Contingency Table Weightings

A number of alternatives to the simple contingency table were investigated. The first was to replace the contingency table with a standardized procedure count (SPC) matrix. Similar to the contingency table, the rows of this matrix correspond to patients and the columns correspond to procedures, however, instead of referring to the raw counts of each procedure over a patient's history, each element of this standardized procedure count matrix, $s_{q,p}$, captures the "standardized count," or average number of times a particular procedure was ordered per day for each visit that it was ordered for a given patient [52]. In mathematical form

$$s_{q,p} = \frac{\sum_{v=1}^{V} |p^v|/l^v}{n}, \qquad (3)$$

where "$v = 1, \ldots, V$ is the set of all visits that contain procedure p, n is the number of patient q's visits containing procedure p, $|p^v|$ is the number of times procedure p was used in visit v, and l^v is the length

of stay (number of days) for visit v" [52]. The motivation behind using this matrix was that looking at average rates of use per day for each procedure might be more informative than just using raw counts. Figure 30, which shows the similarity of a sample of general inpatients compared to an index patient chosen from within this group, reveals that the resulting similarity distribution is very comparable to the distribution created using the raw counts. For this reason it did not seem beneficial to add an additional complication to our approach.
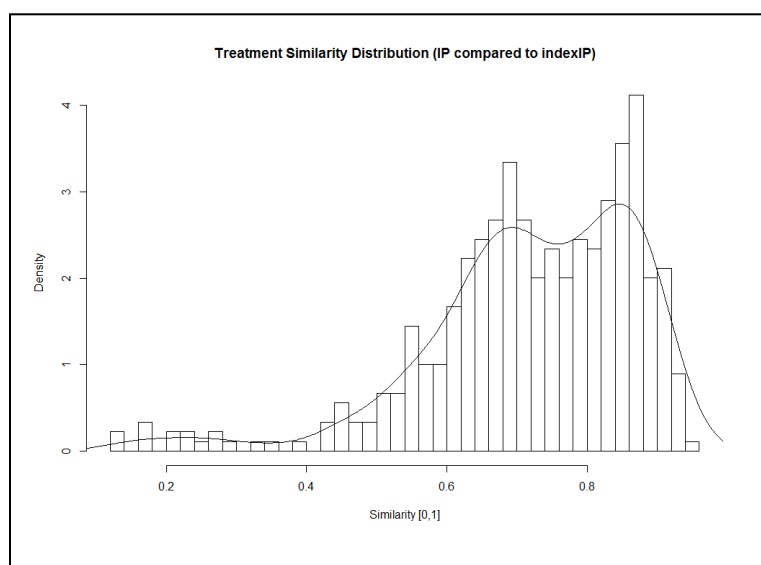

Figure 30: Inpatients vs. Own Index Patient Using SPC Matrix

A second approach investigated was to re-weight the columns of the simple contingency table according to a method called Term Frequency–Inverse Document Frequency (TF-IDF). This method was originally developed for the purposes of natural language processing and is designed to weight attributes which occur in many observations lower with the intention of highlighting attributes which are more unique to specific observations. For our purposes this means that procedures which occur for many patients receive a lower weighting. The result, as shown in Figure 31, is that the overall similarity when inpatients are compared to their own index patient is reduced. This is to be expected since the

procedures which many patients have in common (and would thus lead to increased similarity) have a lower weighting. This approach did not seem appropriate for our study, however, because procedures related to the treatment of congestive heart failure are likely to occur for many patients (since all patients we studied have this diagnosis) and so we would not want the weightings for these orders to be lower.
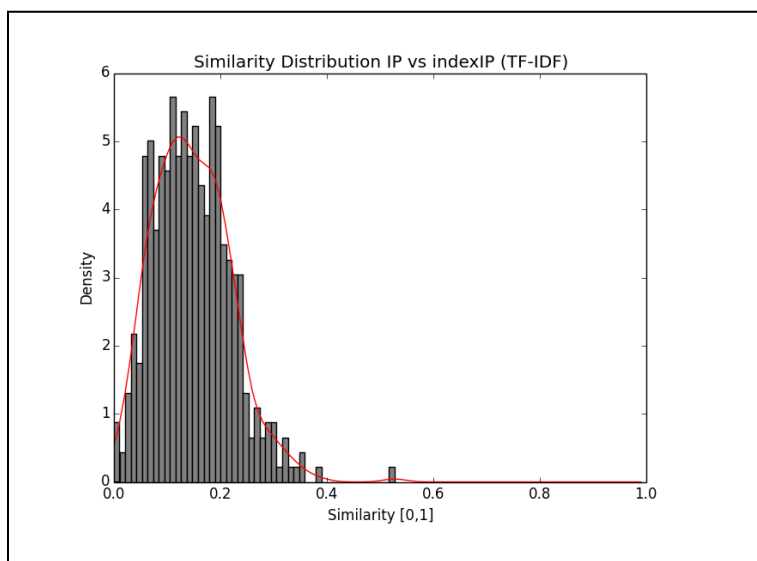

Figure 31: Inpatients vs. Own Index Patient Using TF-IDF Approach

A different approach investigated, which is also used for natural language processing, is called Latent Semantic Analysis (LSA). This approach uses techniques from linear algebra to identify an alternate set of attributes, referred to as topics, which are a linear combination of the original attributes. A particular attribute may appear as part of more than one topic, though the weight applied to that attribute may be different. The chosen topics are meant to reduce the dimensionality of your problem while still capturing the most relevant information characterizing your data. We presumed that this might be helpful for our problem if procedures which served similar purposes were grouped into topics together. That way the treatment for 2 patients would not be measured as dissimilar if they were each provided a different procedure which actually provided the same benefit. After investigating this

method it was not clear whether the resulting topics were meaningful for our problem. The obtained topics were linear combinations of the most frequently ordered procedures. We expected that if we searched for fewer topics that the resulting topics would group more procedures together (or provide weightings which are close in value for more attributes), forming more generalized topics. However, what actually occurs using this method is that the same set of topics are created, but a smaller subset of topics are used for the similarity calculation. The resulting similarity distribution, shown in Figure 32, is comparable to the similarity distribution obtained using the simple contingency table, although there is a slight increase in similarity. It is still a possibility that grouping procedures into topics may be useful for this study; however, we are not currently certain whether the topics created using LSA are appropriate.
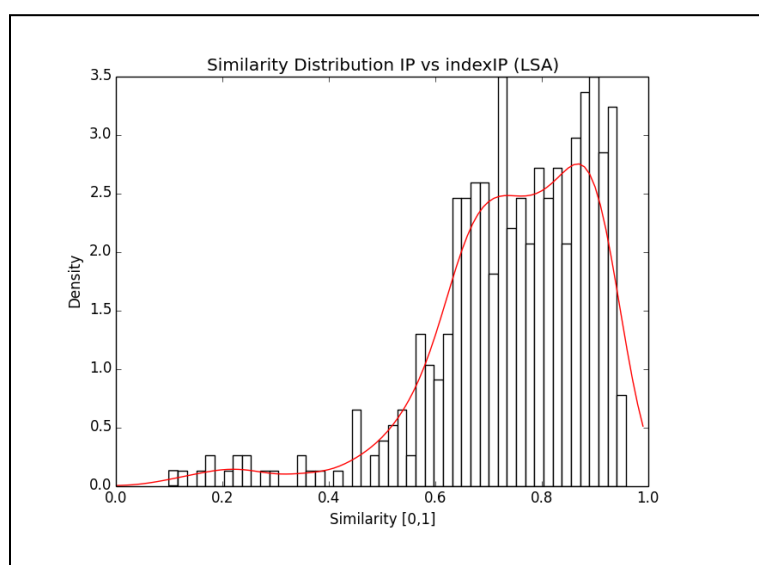

Figure 32: Inpatients vs. Own Index Patient Using LSA Approach

The last alternate approach investigated was Latent Dirichlet Allocation (LDA). Also used for natural language processing, this approach uses a different strategy for generalizing the complete list of attributes into broader topics. The idea behind LDA is that we assume that there is a distribution of topics which characterize your data and that each topic includes a distribution of attributes. These 2 distributions are assumed to be a part of the Dirichlet distribution family. The LDA approach aims to

estimate the parameters of these 2 distributions based on your data. Once these parameters are estimated, the resulting distributions are used to transform your original observations into equivalent observations which use the created topics as attributes instead of the original set of attributes. We then calculated the similarity between these transformed observations. When applied to our problem, a patient's set of procedure orders are translated into an equivalent set of topics using the estimated Dirichlet distributions. This again reduces the dimensionality of our problem but it is unclear whether the resulting topics are appropriate. Figure 33 shows that the overall similarity reduces dramatically. More investigation is needed to see whether the similarity distributions formed using this approach are more relevant than those formed using the simple contingency table.
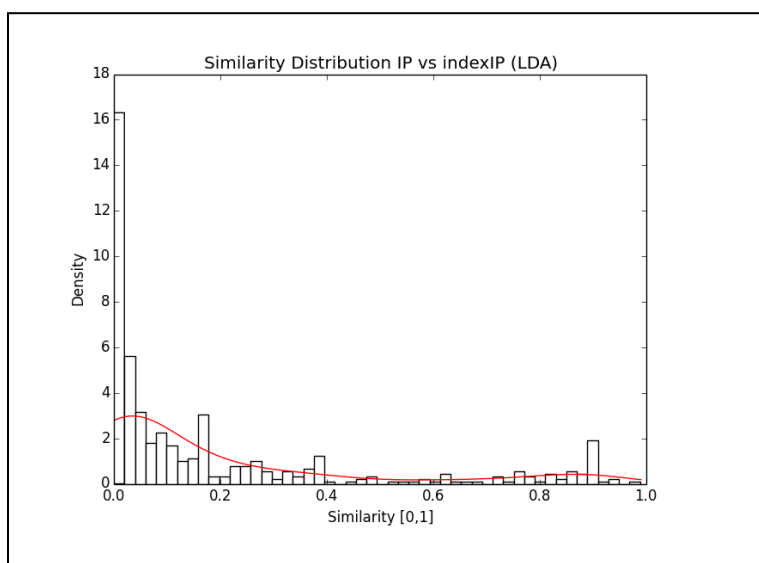

Figure 33: Inpatients vs. Own Index Patient Using LDA Approach