Research Plan

A) Background & significance

Volume to Value Based Imaging

Electronic medical record alerts and clinical decision support are increasingly used to better align routine clinical practice with evidence and guidelines. However, these systems typically use a blunt approach that fails to recognize differences between clinicians and clinician-patient encounters. This lack of customization leads to frustration among clinicians and reduces their effectiveness as clinicians view them more as a hindrance rather than a helpful tool. At a time when we are pushing to make medicine more personalized and when consumer-facing computer interfaces in customer service and other fields increasingly mimic human-to-human interaction, we are missing an opportunity to improve the acceptability and effectiveness of alerts and decision support tools. Our project will pioneer a novel approach to mining patient records that could improve the ability of electronic alert systems to accurately identify the overuse of imaging tests in the ICU. Advanced imaging is costly and often ordered in the absence of evidence that it improves patient outcomes, leading many medical societies to identify overuse of imaging as a target for improvement in Choosing Wisely recommendations.

In 2014, the U.S. NHE increased 5.3% to $3.0 trillion dollars, representing 17.5% of U.S. Gross Domestic Product (GDP). ICU stays account for $80 billion per year in healthcare spending, making up 3% of all healthcare costs and 1% of GDP. Additionally, up to 27% of all hospitalizations are ICU stays, which are usually 3 times as costly as general hospital stays. Associated diagnostic imaging is being increasingly scrutinized due to its high costs and poorly defined contribution to patient outcomes. Now, more than ever, there is a greater need for innovative and objective methods of assessing ICU diagnostic imaging utilization, costs, and associated patient outcomes. Sentiment analysis is a potential method for understanding healthcare provider ordering patterns and their association with patient outcomes and costs.

MIMIC-III Database

The MIMIC-III database contains structured and unstructured anonymized data from ~60,000 admissions to the medical and surgical ICU’s at the Beth Israel Deaconess Medical Center in Boston Massachusetts. The open-source dataset contains medical data, including radiology reports and associated CPT and International Classification of Disease (ICD) codes for all services (https://mimic.physionet.org/gettingstarted/overview/).

Sentiment Analysis

Sentiment analysis refers to the use of natural language processing (NLP) to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. Over the past few years, sentiment analysis has found wide applicability to automated monitoring of online reviews and social media for a variety of applications, ranging from marketing to customer service. We plan to use a method of sentiment analysis based on the work of Ghassemi et al., which has been previously validated on a large clinical database. In the first step of analysis, words in a training data set (such as a unstructured progress reports or discharge summaries) are mapped into a set of numeric vector representations (or a vector-space), using an approach known as Word2Vec, first described by Mikolov et al. in 2013 at Google (Mountain View, CA). The Word2Vec algorithm is a simple neural network that learns a set of numeric vector representations for a collection of words such that words that are determined, statistically, to be more related are positioned closer in the vector-space than words that are unrelated (See Figure 1).

Figure 1. An Illustration of word vector space for patients who did not survive their hospital stay in MIMIC. Note that related terms are clustered more closely in the word space than unrelated terms.
This property of the constructed vector-space allows one to determine words and concepts that are related according to the provided training data. Next, using a list of over 6,500 positive and negative keywords in the English language, gauging the average distance between positive and negative keywords assesses the general sentiment of each clinical note and all words in the note, to arrive at a sentiment score as described in Ghassemi et al (see Figure 2).

![Diagram](image)

**Figure 2.** An Illustration of the proposed sentiment analysis Method (21). Notes were extracted from the MIMIC database, pre-processed, and categorized by time, and demographic features. Each group of notes was then vectorized using the Word2vec tool, analyzed to identify clusters and visualized.

### Specific Aims

1. Evaluate potential relationships between ICU provider sentiment of patient medical condition and diagnostic imaging orders using information derived from the MIMIC-III database.
   a. **Hypothesis:** After adjusting for underlying comorbidities, we predict that negative (rather than positive) provider sentiment will be associated with higher diagnostic imaging orders.
   b. **Rationale:** Patients with a poor prognosis, overall poor health status, and negative changes in their clinical condition (associated with negative provider sentiment) may prompt providers to use more diagnostic imaging as a means to look for any potential imaging finding that could lead to a management and treatment changes that could save a patient. This would reflect heroic measures to keep patients with worse prognoses alive.

2. Compare costs of diagnostic imaging orders for positive vs. negative sentiment patient medical conditions using CPT code data from the MIMIC-III database.
   a. **Hypothesis:** Negative provider sentiments will be associated with higher imaging costs compared to positive provider sentiments.
   b. **Rationale:** Due to the high cost of performing and interpreting diagnostic imaging studies, great imaging use will translate to higher costs, especially for cross-sectional studies (e.g. CT scan, MRI scan).

### B) Significance and Innovation

This project will bring together the burgeoning field of sentiment analysis with the evaluation of actual patient related healthcare data, particularly as they relate to imaging utilization and costs. It is likely to lead to new insights on how healthcare provider sentiment relates to diagnostic imaging ordering patterns and utilization. Discovery of a link between provider sentiment and diagnostic imaging utilization may lead to development of a new generation of clinical decision making software tools that leverage machine learning to “guide” providers in appropriate and judicious use of high cost healthcare resources. For example, we envision a developing a software program that integrates with existing electronic medical record (EMR) programs and continuously monitors sentiment scores as part of a decision support system that alerts providers and team leaders when ordering of resources seems to be based on sentiment, more than objective data. Such a system could trigger a “red flag” warning that would prompt providers to pause and re-assess their next medical decision, rather than reflexively ordering additional imaging in an effort to continue to “do something” for a patient with declining health status.

### C) Study design

**Overview**

In this retrospective observational cohort study, we will extract 1,237,977 medical notes from 60,000 unique ICU stays from the publicly available MIMIC-III database. **We will only include nursing notes and physician notes.** Radiology
reports will be excluded, as they are rendered after resources are utilized (i.e., an order submitted for a patient to undergo a given imaging exam). Only patients with ICU hospitalizations of 5 days or more will be included, similar to the initial paper by Ghassemi et al. Two cohorts will be constructed: 1) patients alive at day 5, and 2) patients deceased at day 5. **Sentiment scores will be computed from physician and nursing medical notes only for both cohorts based on positive and negative words in the notes examined (described below).** Based on CPT codes also available in the MIMIC-III database, the frequency and modality type (e.g., x-ray, ultrasound, computed tomography [CT], MRI) of diagnostic imaging exams will be computed for both cohorts. We will only use non-invasive exams, as these are the most common type of diagnostic exams performed on ICU patients (as opposed to invasive procedures performed by interventional radiologists, not dissimilar to those performed by surgeons or other proceduralists).

**Demographics & Measures of Comorbidity**

We will tabulate age, gender, race, day-of-admission, admitting diagnosis, marital status, and outcome (alive or deceased). Race will be grouped as white (Caucasian), black, Hispanic, non-Hispanic, Asian, and other. We will account for baseline patient comorbid conditions using the Elixhauser comorbidity index and individual conditions included in the index. Also, the Sepsis-related Organ Failure Assessment (SOFA) score will be tabulated for each patient daily. The SOFA score is a clinically accepted and highly objective way to assess the ICU comorbidity of patients on an individual and group basis. It incorporates blood clotting, heart, kidney, lung, and liver function and is independent of CPT and ICD codes. Although ICD-10 was just enacted as the national standard for diagnosis coding in October 2015, the MIMIC-III data set is based on ICD-9 coding which for many years was the standard for diagnosis coding and widely used for claims-based research.

**Outcomes**

**Aim 1:** Our primary outcome is the correlation between the number and type of imaging exams ordered and the computed sentiment score for both groups over 5 ICU days. Secondary outcomes include: 1) Per day (days 1-5) correlation of number of diagnostic imaging exams ordered with sentiment score in both groups, and 2) Influence of sociodemographic factors (e.g., age, gender etc.) & SOFA score on sentiment and diagnostic imaging orders.

**Aim 2:** The primary outcome is national average Medicare allowable payment (which will be used as a surrogate for costs to society) for diagnostic imaging exams. The secondary outcome is per day correlation of cost of imaging exams ordered with sentiment score in both groups.

**Note Text Processing**

We will pre-process the notes from the MIMIC-III database by removing all numbers, stop-words, punctuation and white-space characters (new line, tabs, etc.) from the extracted text. We will also cast all words into lower-case and removed any single-character words from the text (‘a’ and ‘I’, for instance) or words that appear less than five times. Lastly, we will replace all positive sentiment terms (such as ‘good’, ‘happy’, ‘better’, etc.) in the text with the single term ‘POSITIVE’, and all negative sentiment terms with the single term ‘NEGATIVE’. We will use the 6500 negative and positive terms based on a 2005 paper by Liu. Following pre-processing of the notes, we will separate the text into groups according to patient age, ethnicity, gender, marital status, outcome and the hospital stay day. In Table 1 we list the extracted note categories and the corresponding total word-count for each derived from the work by Ghassemi et al. Note that even with this stratification the word count for each category is in the millions, yielding adequate statistical power for the analyses.
After grouping notes we will apply the Word2vec tool to each of the categories in Table 1. The first five days of clinical notes will be further separated by patient outcome and analyzed. Word2vec is a tool that analyzes a corpus of text and generates vector representations of the words in the text using the skip-gram and continuous bag-of-words approaches. The tool has several parameters that affect the nature of the embedding. For our analysis, we will use a continuous bag-of-words approach.

CPT Codes

Each patient's diagnostic imaging exams will be identified, counted, and characterized based on billed CPT codes for specific exams. Under the Health Insurance Portability and Accountability Act (HIPAA), CPT is the preferred code set used by Medicare and private payers for processing claims for physician and other provider services. The code set is updated annually to ensure compatibility with contemporaneous medical services. Tens of thousands of codes distinctly and uniquely identify specific medical services, making CPT the preferred basis for retrospective service identification. Most radiology services are identified by their number (range from 70000 – 79999) and thus will be readily extractable from the database for our analyses.

Statistical Analysis

For both aims we will start with bubble plot visualization and a correlation analysis to determine whether there is a linear or quadratic relationship (positive or negative) between the sentiment analysis scores and the dependent variable of number of diagnostic exams ordered (Aim 1) and cost based on CPT code (Aim 2) for the 2 cohorts of patients (alive vs. deceased). Subsequently, A Poisson Generalized Estimating Equation Regression Analysis will then be performed to assess the impact of the following variables (covariates) on the number of tests ordered: sentiment analysis score?, Age, Ethnicity (Race), Gender, OASIS score, Elixhauser index, Diabetes status, HIV status, Metastatic Cancer status, Obesity status, ICU type, type of provider (physician vs. nurse), time of medical note (daytime vs. nighttime), and time of week (weekday vs. weekend). With 1,237,977 notes and associated demographic and CPT code data available on 60,000 ICU admissions for analysis, we will have at a minimum 35,000 diagnostic imaging exam orders as nearly every ICU patient has at least 1 exam during their stay. With this amount of data, power will be over 0.90 with even a moderate correlation of 0.50. Separately, we will used a generalized linear model to estimate the impact of sentiment analysis score and other covariates on imaging costs using Medicare reimbursements as a proxy for costs. Generalized linear models account for the (typically) skewed distribution of health care costs.

Preliminary Results

The results from Ghassemi et al. (21) serve as the proof-of-concept work that inspired this proposal. This work found differences in the sentiment of clinical notes over time, outcome, and demographic features. The analysis revealed a decrease in the homogeneity and complexity of language clusters over time for patients with poor outcomes (i.e. death) at day 5 of ICU hospitalization (Figure 3). Additionally, there was greater positive sentiment for females, unmarried patients, and patients of African American race (Table 2). However, this proof of concept analysis did not adjust for confounding factors and comorbidity indices, which is one of the objectives of the current study. Using the refined methods and preliminary data from this Health Innovation Seed-Grant, we will apply for an NIH R-level grant to validate the methods on Emory ICU Clinical data.
D) Research team

Falgun H. Chokshi, M.D., M.S. – Assistant Professor of Radiology & Imaging Sciences; Neuroradiology and Health Services/Biomedical Informatics Research. Dr. Chokshi has clinical and research expertise in the radiology health services research space and he will direct and be responsible for all aspects of this pilot grant project and subsequent NIH grant application and manuscript preparation.

Shamim Nemati, Ph.D. – Assistant Professor of Biomedical Informatics; Machine Learning: Dr. Nemati will lead the extraction of the MIMIC III data, and using the post-processing of the data using machine learning software, post-process the data, and subsequent tabulation to make the data ready for statistical analysis can be performed. The below-mentioned Graduate Student Research Assistant will work under Dr. Nemati’s guidance.

Graduate Student Research Assistant (to be named), Biomedical Informatics: The graduate student research assistant will work in Dr. Nemati’s lab.

David Howard, PhD, Associate Professor, Health Policy and Management: Dr. Howard is an experienced health economist and facile with cost analyses. He will conduct the analysis in our economics aim 2 and assist with statistical analysis.

E) Future funding plans

We hope this work will ultimately contribute to objective integration of sentiment analysis in clinical decision support systems. After completion of this project we will use the data to submit an R01 to the NIH that will examine this important issue in more depth.

F) Timeline

Months 1-4: Extraction and Processing of Data
Months 5-7: Post-Processing of Data and Statistical Analysis
Months 7-12: Manuscript and NIH R01 Grant Application Preparation


