Report: Framework for Designing Evidence-Based Patient-Physician Matching Protocols

Siva Chaduvula, Adam Dachowicz, Mohammad Rahman, Mikhail Atallah, Jitesh Panchal

Purdue University
February 7, 2019

(originally dated December 4, 2018)

Abstract

This document presents and evaluates a framework for designing an intelligent system for matching physicians with patients seeking elective procedures. The evaluation consists of using, as a training set, data on surgical orthopedic procedures from one time period and evaluating the system’s utility by comparing its matching recommendations with what actually unfolded in the next time period. The relevant testing and training datasets were obtained from AHRQ’s HCUP program [1] and New York state’s SPARCS inpatient database [2] for the years 2013 and 2014. The contributions of this paper are threefold: (i) we explore metrics that are useful to consider for physicians operating within health systems, especially those health systems switching to new payment models such as value-based payment, (ii) we show the prescriptive power of these metrics over two years for orthopedic surgeons in New Jersey state, and (iii) we show preliminary results supporting the benefit of an intelligent matching system that integrates machine learning techniques, statistical techniques, and stochastic process modeling to propose personalized physician-patient matches. These benefits are expressed in terms of cost and length of stay, which is considered in this report to be a sign of care quality. Technical details of this analysis and the intelligent matching system will be presented in future publications.

1 Motivation

In this report, we comment on the variation observed in medical outcomes (as observed through length of stay) and cost for orthopedic procedures at facilities in New Jersey and New York state. Big data analysis, machine learning, and stochastic process modeling are leveraged to design an intelligent system for matching physicians with patients seeking elective procedures; the resulting framework is used to study this variation and estimate the effects of evidence-driven patient-provider matching on reducing this variation, improving outcomes, and lowering costs. In this report, we comment on the results of this implementation and its implications for healthcare moving forward; the technical details of the framework will be discussed in forthcoming publications.

Previous work has shown the presence and impact of cost variation in many common medical procedures [3]. The methods leveraged in this report are motivated by the related observation that physicians treating patients of varying medical history and demographics see variation in outcomes between these patient groups, and this variation contributes to uncertainty in the quality and cost of care [4]. An intelligent system for matching physicians to patients for whom they are best suited could reduce this variation in quality and cost; this is valuable for patients, physicians, and healthcare facilities as organizations.

To further motivate this study, consider variation in length of stay for total knee replacement procedures for hospitals across New York state as represented in the New York SPARCS inpatient database [2]. The relative variance across several dozen New York hospitals is given in Figure 1;
the wide range of variances suggests many hospitals do not achieve the consistency in length of stay that is possible. The framework used to generate results in this report may be used to help hospitals reduce this variation in quality of care. In the next section, we comment briefly on the methods important to this framework. This is followed by a discussion of our preliminary results for datasets relating to New Jersey and New York state hospitals [1, 2].

Figure 1: Length of stay variance for total knee replacement procedures at New York State hospitals. Generated with SPARCS data [2].

2 Brief Overview on the Methodology

The system used to generate the results presented in the next section is built upon a foundation of big data analysis, machine learning, and stochastic process modeling. By leveraging patient-related and physician-related data, machine learning tools may be deployed to match patients with the most-appropriate care providers in an evidence-driven, automated fashion. Stochastic process modeling may be introduced in the matching to account for randomness associated with patients and patient behavior.

An algorithm built upon this foundation was tested and validated using data available from New York state, namely the SPARCS de-identified hospital inpatient databases [2], and from ARHQ HCUP State Inpatient Databases for New Jersey [1]. These datasets provide hospital episode-level data that is sufficient as input to the algorithm from both the patient and provider sides of the matching problem. This testing allows for the study of variation among physicians at health facilities, across patient groups segmented according to the user. In this report, we only consider patients seeking arthroplasty services, as this is a common in-patient elective procedure that provides a reasonably large set of training data for each year in the datasets.

3 Preliminary Results

The motivations for presenting results of this analysis are twofold: we wish to demonstrate that the variation between physicians treating patient of different subgroups exist, and we wish to demonstrate that information regarding this variation can be leveraged to lower cost and increase quality for each procedure, relative to the outcomes expected if this framework is not used.
3.1 Experimental Setup

We use osteoarthritis elective surgery as a test procedure, as this is a commonly occurring elective surgery across the US. We chose facilities from New York and New Jersey to demonstrate the benefits of patient assignment under an instantiation of this framework. Specifically, we chose Facility IDs: 150, 221 and 10 (all pseudonyms) from the New Jersey AHRQ SID data set, and a New York facility from the New York SPARCS data set to illustrate our observations, as these facilities have more than 1400 osteoarthritis-related surgeries per year. The datasets have data on patient cases from 2012 to 2014.

In this study, we assume that there is a flexibility in physician’s availability. For simulation purposes, we assume that this is 5% across all physicians. In other words, a physician treating 100 patients in 2013 may treat up to 105 patients in 2014. This assumption is necessary as availability is a constraint on the potential matches, and as such the probability that a highly appropriate physician will be unavailable for a future patient must be accounted for during the matching process.

As a motivating case, consider two physicians at a New Jersey facility. Their variation in terms of length of stay and total cost across two patient groups, differing only in one comorbidity provided in the New Jersey data set, is as high as 20%, as seen in Figure 2. This variation leads to wasteful spending and decreased care quality at a patient population level, and is targeted and driven down through deployment of this framework. In the analysis presented here, we consider four patient subgroups segmented on gender and age (above and below 70 years old); the names for each subgroup are pseudonymized in the figures that follow.

Figure 2: Variation for two physicians in cost and length of stay across similar patient subgroups for knee replacement procedures. In each plot, the demographics and comorbidity profiles of patients are held constant, save for one comorbidity. In this case, we vary the presence of hypertension.

Now, we discuss our key preliminary findings. First, we illustrate the results from New Jersey Facility 221 for the test time frame as year 2014 using reference time frame as year 2013. Further, we validate the benefits of patient assignment under this framework across different years for different facilities.

Note first that, in our example algorithm instantiation, we choose to classify physicians as best-, better-, or worse-case matches for each patient subgroup based on their physician performance score (as calculated by a physician-scoring function within the algorithm), relative to the mean score for all physicians at the facility. Figure 3 shows three different patient assignments. The pie chart in the middle shows the actual assignments that occurred in the test time frame i.e., 2014. As per our performance-based classifier, we recognized that about 31.1% of the matches at Facility 221 fell in the “worse” class. Note that this does not necessarily reflect that 31.1% of physicians at the facility are poor, but only that matches between patients and physicians for which the matching function output a “worse” result occurred 31.1% percent of the time. The algorithm instantiation
learns about the classification strategy from the reference time frame (i.e., patient cases from year 2013) and proposes matches for the test time frame in such a way that it reduces the number of these “worse” matches. This is evident in the algorithm-proposed patient assignment captured in the right-most pie chart in Figure 3.

Figure 3: The proposed patient assignment framework attempts to reduce the number of “worse” matches through intelligent patient-physician assignments. The facility had > 1,000 knee replacement procedures each year from 2013 to 2014; these were considered in the analysis.

This reduction in the number of worse matches can help health facilities improve quality and achieve operational efficiency.

3.2 Backtesting

Backtesting was performed to validate performance for several facilities. The algorithm was trained using data from 2013 for each facility (Figure 3 left), and then applied to classify physicians as “best,” “better,” or “worse” case matches for patients seeking procedures at that hospital in 2014. (Figure 3 center). For each physician-patient pair in 2014, the physician class, cost, and length of stay was recorded, with the results for each class, for each metric reported in Figure 5. As expected, “best” case matches achieve lower cost and length of stay, with greater differentiation for cost than for length of stay for the facility considered.

3.3 Facility Impact

At a facility level, the proposed patient assignment framework instantiation attempts to utilize physicians in a way that improves operational efficiency while reducing the cost of care. Figure 4 illustrates the actual and proposed patient assignment at Facility 221 for 2014 for three physicians. Note that there are minor changes to the patient volume for the considered procedure (< 2.5% of the total cases considered at the facility) to each individual physician. Figure 5 shows outcomes in terms of cost and length of stay (quality) for cases that, as assigned historically, lined up with the corresponding RightFit category.

3.4 Physician Impact

At a Physician level, the proposed patient assignment alters the case mix in a way while maintaining a similar patient volume for each individual physician. This alteration to the patient case mix is based on the performance of a physician in the reference time frame, their expected case load, and the expected arrival of patient subgroup types to the facility. Figure 6 illustrates the case mix of two physicians (50196 and 71237). The algorithm instantiation learns from the reference year that “Physician 50196” is better in patient subgroup 0 than the other physicians such as “Physician 71237” and assigns more of these cases. However, “Physician 71237” is comparatively better in patient subgroup 1. So, a majority of the patients falling into subgroup 1 are assigned to “Physician 71237.”
Figure 4: Improved outcomes while maintaining physician load balance for three physicians at a New Jersey facility. For each physician, left bars indicate the proportion of cases assigned in 2014 that were classified as “best,” “better,” and “worse” case matches, while right bars indicate the algorithm-recommended proportion in 2014, learning from 2013 data. The heights of the right bars indicate the relative total case volume change for the physician.

Figure 5: Comparing “best”, “better,” and “worst” case matches in cost and quality (length of stay) for a New Jersey facility in 2014. Results are shown for cases where the as-observed matches lined up with the corresponding RightFit classification.

3.5 Validation

We back-tested our patient assignment approach using past data for NJ Facility 150, analyzed over a time period of two years. In both years, the proposed patient assignment identified that the worse matches are high as 21% of all cases. With the proposed matches, this facility could have saved 6% while reducing length of stay by half-a-day per re-assigned patient. The New York facility had about 6000+ cases per year. At this facility as well, the proposed assignment identified 6-8% of the patient cases as worse-case matches.
Figure 6: Improved outcomes through evidence-driven case mix adjustments. Note the “total patient volume” reported only represents the patients for this analysis that fell within the patient subgroups considered, and not necessarily the total number of all patients seen by the physician that year.

References


