Improving Physician Decision-Making & Patient Outcomes Using Analytics: A Case Study with The World's Leading Knee Replacement Surgeon

Anish Pahwa
Krannert School of Management
Purdue University
West Lafayette, United States
pahwaa@purdue.edu

Shikhar Jamuar
Krannert School of Management
Purdue University
West Lafayette, United States
sjamuar@purdue.edu

Varun Kumar Singh
Krannert School of Management
Purdue University
West Lafayette, United States
singh675@purdue.edu

Matthew Lanham
Krannert School of Management
Purdue University
West Lafayette, United States
lanhamm@purdue.edu

Abstract**-- Every year in the United States, more than 300,000 knee replacements are performed. According to Time magazine, this number is expected to increase by 525 percent by the year 2030. Although knee surgeries are highly effective, patients are still prone to post-surgery complications. We address this problem in collaboration with one of the world's leading knee replacement technology company. We show how analysis of insurance codes, patient demographics, and health data can help better support physicians in the diagnosis phase, by assessing patients' risk of developing complications. We identified the factors that led to successful knee surgeries by utilizing classification algorithms. We developed a composite KPI to track surgery failure rates by combing 3 important factors. Namely, a number of post-op visits, direct complications from ICD codes, and whether a revision surgery has been carried out. Our study found factors such as BMI, smoking, blood pressure, and age were statistically significant parameters for determining a surgery outcome. The surgeon performing the surgery was also a significant factor in ascertaining the outcome. This could be due to the different techniques used by different surgeons. Our model can save millions of dollars per year by detecting two-thirds of actual complications that would occur. We believe healthcare providers and consulting firms who are developing analytics-driven solutions in the healthcare space will find our study novel and inspiring.**

Keywords—Classification, Data cleaning and Preparation, Sampling and rebalancing, Model Interpretation

I. INTRODUCTION

Every year, there are thousands of cases of chronic knee pain and disability due to different types of arthritis in people from different age groups. Athletes, accident victims, and aged people are among the most vulnerable. On top of this, the severity of an arthritis case tends to increase over time causing the patient more pain. Knee replacement is the only effective and long-lasting course of treatment. Knee replacement has become so common that America itself get new knees at a rate of more than 600,000 per year. In brief, the aim of the surgery is to minimize pain and restore mobility. However, there are many risks associated with the surgery, and a few complications like infection, blood clots, implant loosening, continued pain can make matters even worse. These complications generally require revision surgery. Revision surgeries often have risk associated and might even add to complications, on top of the fact that there are extra costs associated. Forbes in its articles titled as "Medicare's Bundled Fees Hit Knee, Hip Replacement"[1], talks about how Medicare's bundling of fees hit knee replacement surgeries, affecting not only the patients but also physicians and insurance companies. Therefore, it becomes crucial to understand the drivers of these complications and control them to mitigate risk and minimize the complications. Finding statistically significant but controllable factors that could lead to complications is an industry-wide challenge.

This study tries to address this problem, in association with a world's leading knee replacement surgeon who has performed more than ten thousand knee replacement surgeries, using cutting edge minimally invasive technologies. The study tries to build a system that predicts the most important factors leading to post-operative complications by utilizing using patient demographics, patient comorbidity data, surgery details, doctor details, and procedure specifics. The Wall Street Journal in an article talks about a study that found that exercise and increased muscle strength lead to better surgical outcomes. Building upon that we see how factors like age, the height of a patient, distance from the clinic, marital status affects the outcome of the surgical procedure. Surgery is classified as a failure depending on various parameters some of which are the number of post-operation visits, direct complication codes. The output of this model can give us insight regarding potential complications with a surgery and what complications can be expected. This can further control and minimize issues by warning doctors and patients in advance and even giving them recommendations to avoid these complications. The study also tries to map patient clusters to various doctors for improving the success rate and minimizing complications. Since the data we had was historical (ranging from over 10 years back), we had limited surgical specifics data. We believe that these surgical and medical data could be significant in finding various controllable factors that affect the outcome. To that end, we believe our analysis could be finetuned with more data.

The remainder of the paper talks about the following: The next section has a literature review on the factors considered and the measure of a successful total knee replacement surgery. We looked into the literature to find the significant factors in the previous studies, and how the categorized surgeries as failures or successes. Section three presents our methodology and criteria formulation. We summarize our steps chronologically and provide reasonings for our assumptions in our analysis. In section four, we talk about the different models that we built and tested. Since failure rate is less than fifty percent, we are dealing with a class imbalance problem. We will talk about how we deal with this problem in the healthcare context. Section five summarizes the performance of all the models and the last section concludes the paper with a discussion of the implications of this study, our recommendations from the clinical perspective, the future scope, and concluding remarks.

II. LITERATURE REVIEW

Many recent researches, have focused on the problem of post-operative total knee replacement surgery complications. However, most of the researchers have only taken pain scores as indicators of surgery outcomes. While this method has its merits, collecting and using this data requires additional efforts and time such as surveys and the responses are subjective, or perception based.

When predictors such as various pain or knee function scores are not present in the historical data, we must find a new approach to define surgery failures or complications. With an exponential increase in the number of knee replacement surgeries, the total number of surgery failures or complications are also increasing with the same magnitude. On average, at least five out of a hundred total knee replacement surgeries develop complications. This not only leads to added costs to the health insurance providers and surgeons but also decreases the quality of life of the patient. Therefore, it becomes important to address this orthopedic industry-wide challenge. A successful predictive model based on historical patient data could help us predict post-op complications before they occur and thus save costs and improve the success rate of these types of surgeries. The researches we based our study focused on some factors such as different success metrics, complications, use of post-op data, and building predictive models.

A 5-year prospective study by A.K. Nilsdotte of patient outcomes after Total Knee Arthroplasty takes into consideration the Knee injury and osteoarthritis outcome score (KOOS) preoperatively once and then 6 months, 12 months, and 5 years postoperatively[2]. The result showed significant improvement in all KOOS and scores 6 months post-op. The best postoperative result was reported at the 1-year follow-up. The 5-yearfollow-up again showed a decline in KOOS scores. Age, comorbid conditions, the sex was some of the factors that affected post-operative KOOS pain scores.

In the study titled "Predicting the outcome of knee arthroplasty"[3] relief of pain and the restoration of functional activities are used as the outcome parameters for primary total knee arthroplasty. Preoperative predictors of pain and functional outcomes atone and two years following the surgery are used. The study recruited patients from three different countries. The study employed hierarchical regression models and found that the most significant predictors of failures were low preoperative pain scores, a higher number of comorbid conditions, and a low mental health score. The country was also a significant factor for the functional status of patients.

A similar study titled "Development and validation of a clinical prediction model for patient-reported pain and function after primary total knee replacement surgery"[4] aims to build a prediction model of patient-reported pain and function after undergoing total knee replacement (TKR). Pre-operative predictors such as patient characteristics and clinical factors were considered. The study employed bootstrap backward linear regression analysis. Low preoperative Knee score, living in poor areas, high BMI, and anxiety or depression were associated with worse outcomes. This is the first clinical prediction model for predicting self-reported pain and function 12 months after total knee replacement surgery.

In the study "Predicting total knee replacement pain: a prospective, observational study"[5] excessive postoperative pain, clinical and radiographic variables were used as predictors for total knee arthroplasty outcomes. Measures were VASP pain index, patient health, psychological state, and surgical component reliability. Greater preoperative pain, depression and anxiety were associated with greater postoperative pain. These factors also corresponded to more home therapy and postoperative manipulations.

A study by J.J. Tolk created and validated models that predict residual symptoms on 10 specific outcome

parameters at 12-month follow-up for patients undergoing primary TKA for knee osteoarthritis.[6]
The predictive algorithms showed acceptable discriminative values (AUC 0.68–0.74) for predicting complications/ residual symptoms. A study by Christoffer C Jorgensen is about construction of a preoperative risk score for patients in high risk of potentially preventable complications[7]. The study concluded that preoperative identification of patients at risk of preventable 'medical' complications was statistically possible. The study 'Predicting individual knee range of motion, knee pain, and walking limitation outcomes following total knee arthroplasty' by Yong-Hao PUA[8], concludes that statistically significant predictors were for TKA outcomes were age, sex, race, education level, diabetes mellitus, preoperative use of gait aids, contralateral knee pain, and psychological distress.

Our study aims to build a predictive model and a clinical recommender system using patient characteristics, patient health data and doctor data, to predict a failure of a total knee replacement surgery. We strive to identify the drivers to success of a surgery in order to minimize complications and revision surgeries thereby reducing total costs associated. Our study is novel because we have developed a tailored outcome variable that is a combination of number of post-op visits, direct complications from ICD codes, and whether a patient underwent a knee surgery. If any of the required conditions are met, then the surgery is flagged as a failure. Most of the other studies used regular or hierarchical linear regression models in their studies, since their predictor variable (pain score or knee function score) was a continuous variable. This study however had a binary predictor variable. Therefore, logistic regression and other classification models were utilized.

Author	Target variable used	Model type	Data Used
Lingard, Elizabeth A. *et al*	WOMAC Pain score	Hierarchical linear regression	Mental health, pain score, comorbid conditions
Brander, Victoria, A. *et al*	Pain score	Multiple linear regression	pain scale, health data, psychological state, device reliability
Sanchez-Santos, M.T. *et al*	Pain score	General linear model	pain score, previous history, weight data
Our Study	Derived metric using 3 conditions	Logistic Regression	patient demographics, comorbid conditions, doctor data

Table 1: Literature analysis and comparison

III. DATA

Variable	Type	Description
Sex	Categorical	The sex of the patient (either M or F or U)
age	Numeric	Age of patient at the time of surgery
BMI	Numeric	Body mass index of patient
Financial status	Categorical	The financial class for this employer
smoke	Binary	1 if patient smokes, 0 if patient does not smoke
height	Numeric	Height of patient
weight	Numeric	Weight of Patient
bp	Binary	If patient has high blood pressure
Surgery outcome	Binary	Derived variable. 1 if surgery is a failure, 0 if it is a success

Table 2: Data Used in the study

The data that we used was extracted from a database of 400 tables. We extracted patient related demographics, health and billing data from twelve different tables. We also used surgeon data in our analyses. We had patient age, sex, weight and height data. We also had patient financial class data. Further, we extracted patient smoking history and their blood pressure data.

We derived variables such as BMI and surgery failure or success (the predictor variable) from the existing data. BMI was calculated using patient weight and height data. The outcome variable was created based on three conditions: the number of post-operative visits, direct complications using ICD codes (medical codes corresponding to specific complications) and whether a patient had a revision surgery after the initial surgery. Any number of post-operative visits greater than six were considered as a failure in the surgery outcome. If any of these three conditions were met, the surgery was considered a failure.

IV. METHODOLOGY

In this whole research we tried to follow Cross Industry Process for Data Mining (CRISP-DM), although our business case was unique so had to de-tour from the standard process at some places. We had over 400 data sets from a unique data base, collected by our client over past 2 decades. The most time taking process of this whole study was to parse all these 400 tables to gather relevant data. Finally, we narrowed down to 400 paraments with more than 6000 observation. While going through all this data we found scope for some derived variable like BMI and Age, which could make our analysis more comprehensive. As our whole can be summed up as predicting if a surgery will be successful or not, it becomes crucial to define what is success in this scenario. For our model to robust we used three different paraments to define success:

1. Number of post operation visits that a patient has pertaining to a surgery
2. If a patient has a revision surgery corresponding to a surgery
3. If a patient has direct complication recorded (according to ICD-9 and ICD-10 codes)

Ones we were final on which variables to use in the model, we did some data cleaning. We treated outliers depending on case and variable for example in age we used capping and flooring and BMI we used a range [Q1 - 1.5(IQR), Q3 + 1.5(IQR)]. For imputing we used decision tree models and checked for any abrupt changes in the distribution of variables before and after imputation, to make sure our imputations are as realistic and there are no synthetic spikes in the data. We split our data into 75% train and 25% test, since this was the optimal split in our case. We did not have a very huge dataset, and since there was some noise in the data. Ones we were confident that the scope of analysis and the scope of data were in line, we proceeded with some initial investigation.

As the data corresponds to surgery success and failure, it was obvious that the number of failures would be very less compared to surgery success. To encounter the class imbalance, we generated some synthetic observations using upscale techniques (SMOTE), increasing the minority class from a mere ~11% to 25%, which is an industry wide practice.

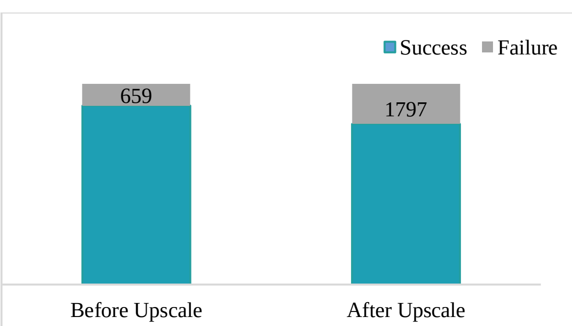

Figure 1: Data Used in the study

The business problem when converted to a statistical problem turned out to a binary classification challenge. As we are trying to predict failure or no failure. In this business case, model interpretability was prioritized above accuracy. So, we employed simpler classification models like Logistic Regression and Decision Tree. Due to a huge class imbalance in the dataset, using accuracy did not make a lot of sense as it does not highlight falsely predicted minority class. To encounter this problem, we used metrics like F-1 score, as it accounts for class imbalance.

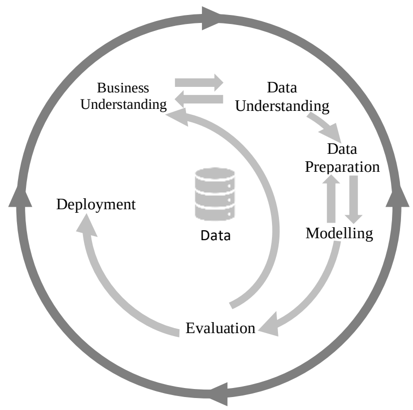

Figure 2: concise approach to our problem

V. MODEL

For our case model interpretability was prioritized over model predictability and model accuracy. We have developed an early warning system that would air doctors in decision making. For a doctor to take some insights form the model it was crucial that the model is simple and explainable. In the light of all this we used models simpler and easy to interpret models.

	Decision Tree	**Random Forest**	**Logistic Regression**
Graphically Interpretable	*****	***	**
Nominal Predictor Handling	*****	*****	***
Low Variance	**	***	*****
Highly Interpretable	**	*	*****
No Hyper Parameter Tuning	**	**	*****
Easy to Implement	***	***	****
Test Error Estimation	***	****	***

Table 3: Pros and cons of each model

Model	Hyper Parameter
Decision Tree	• Maximum depth: 6 • Interval Target Criterion: Reduction Gini index Leaf size: 5
Random Forest	• Number of trees: 200 • Number of variables at each split: 4 • Maximum depth: 10 Proportion of sample in each sample:75%

Table 4: Models' Hyper Parameters

Logistic regression models by nature are highly interpretability in our case proved to be very accurate as well. Therefore, we chose to use it.

$$1)\ \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x \quad => \quad P = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$2)\ \ln\left(\frac{Failure}{Success}\right) = -0.92 + 0.99 * Doctorid16 + 0.8 * Doctorid496 + 0.27 * Doctorid1478 + 0.34 * Doctorid1933 - 3.22 * Age + 0.29 * Smoke + 0.85 * BP + 0.9 * BMI$$

Formula 1 descries the logit expression of logistics regression for estimating probability of an event happening or not.
Formula 2 has all the beta values plugged into the previous equation.

VI. RESULT

Figure 3: Model performances

We already had enough reasons to select Logistic model. Now, in addition, we also found that Logistic model was the least overfitting model compared to others. To further investigate, we created a ROC chart as shown in the figure below

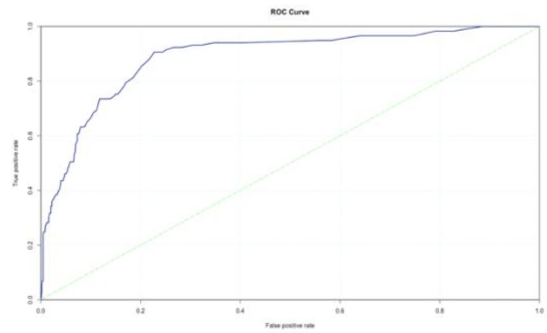

Figure 4: ROC curve for the final model

The Receiver operating Characteristics (ROC) is a graph between False Positive rate and True Positive Rate, which indicates models precative power across both the classes. The Graph shows that the model in in good shape this is also backed by a good AUC value of 87.8. Below is the confusion matrix that shows the True negative and false positive cases in our test set:

		Actual	
		Success	**Failure**
Prediction	**Success**	1297	47
	Failure	15	152

Table 5: Confusion matrix for final model

- Accuracy: 95%
- Precision: 82%
- Recall: 97 %

A good precision and recall show that the model can correctly classify failures as failure and success as success. A brief discussion of the parameters significantly effecting the complication or the failure rate for Knee Surgery are discussed below:

BMI of a person is a significant factor contributing to the post-operative failure of surgeries. Higher BMI is associated with increased risks of developing complications after a surgery. Age of the patient significantly contributes towards failure of a surgery. People with a higher age are less likely to develop complications post-surgery. Having a smoking habit and/or having a higher blood pressure significantly contributes towards failure of a surgery. Smoking and high blood pressure increases the risk of infections in the artificial joints and increases the risk of developing blood clots or causing deep vein thrombosis (DVT) which could lead to a life-threatening situation. Additionally, surgeons and insurance carriers were also found to be significant factors impacting the outcome of a surgery.

VII. CONCLUSIONS

There has been a constant rise in the number of total knee replacement surgeries all over the world, and consequently the number of post-operative complications added to the upward spiralling healthcare costs. It is therefore important to discover the statistically significant drivers of surgery outcomes to address this challenge. This study tried to do this by analyzing six thousand plus knee replacement surgeries data and building a predictive model based on the drivers. This could help decrease the associated avoidable industry wide costs and bring down the failure rate of surgeries.

The average cost of a total knee replacement surgery is fifty-seven thousand US dollars typically. There will be approximately 550,000 total knee arthroplasty procedures in the year 2021. Our model predicts up to two-thirds of complications accurately. When we factor in two-thirds of the complications requiring a revision, even if we ignore the extra costs associated with complications such as hospitalization, medications, imaging and radiology costs, and just accounting for the raw cost of a revision surgery more than 2.1 billion USD could be saved in the US alone in one year, if predicted complications/failures could be prevented. In addition to the economic costs, patient quality of life can improve drastically, and surgeon and patient time could be saved. By extrapolating we can conclude that our model can use patient demographics data, EHR and clinic data to predict complications before a surgery help clinics and insurance companies save up to eleven billion USD every year by the year 2030.

There are however some underlying assumptions based on the business context. The number of post-operative visits above which a surgery could be classified as a failure is a subjective number. We are assuming that this sample dataset has the same characteristics as would the entire population in the US.

We also assume that the components used in the replacement surgeries are the same or have the same performance, since we did not have data on this.

Further, more investigation could be done on surgical factors, such as the technique used in the surgery (whether the ligaments were cut), the angle of cut, the gait of the knee etc. to see if any of these factors significantly affect the outcome of the procedure. Collecting and using these data could improve our model further.

VIII. REFERENCES

1. Japsen, B. (2016, April 03). Medicare's Bundled Fees Hit Knee, Hip Replacements. Retrieved June 11, 2020, from https://www.forbes.com/sites/brucejapsen/2016/04/01/starting-today-medicares-bundled-fees-hit-hospitals-hard/
2. Nilsdotter, A.-K., Toksvig-Larsen, S., & Roos, E. (2009). A 5-year prospective study of patient-relevant outcomes after total knee replacement. Osteoarthritis and Cartilage, 17(5), 601–606. doi: 10.1016/j.joca.2008.11.007
3. Lingard, E. A., Katz, J. N., Wright, E. A., Sledge, C. B., & Kinemax Outcomes Group. (2004, October). Predicting the outcome of total knee arthroplasty. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15466726
4. Sanchez-Santos, M. T., Garriga, C., Judge, A., Batra, R. N., Price, A. J., Liddle, A. D., … Arden, N. K. (2018, February 21). Development and validation of a clinical prediction model for patient-reported pain and function after primary total knee replacement surgery. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29467465
5. Brander, V. A., Stulberg, S. D., Adams, A. D., Harden, R. N., Bruehl, S., Stanos, S. P., & Houle, T. (2003, November). Predicting total knee replacement pain: a prospective, observational study. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14646737
6. Tolk, J., Waarsing, J., Janssen, R., Steenbergen, L. V., Bierma-Zeinstra, S., & Reijman, M. (2019). Thu0462 Outcome Prediction For Treatment Of Knee Osteoarthritis With A Total Knee Arthroplasty. Development And Validation Of A Prediction Model For Pain And Functional Outcome Using The Dutch Arthroplasty Register (Lroi) Data. Poster Presentations. doi: 10.1136/annrheumdis-2019-eular.1654
7. Jørgensen, C. C., Petersen, M. A., & Kehlet, H. (2016). Preoperative prediction of potentially preventable morbidity after fast-track hip and knee arthroplasty: a detailed descriptive cohort study. BMJ Open, 6(1). doi: 10.1136/bmjopen-2015-009813
8. Pua, Y.-H., Poon, C. L.-L., Seah, F. J.-T., Thumboo, J., Clark, R. A., Tan, M.-H., … Yeo, S.-J. (2019). Predicting individual knee range of motion, knee pain, and walking limitation outcomes following total knee arthroplasty. Acta Orthopaedica, 90(2), 179–186. doi: 10.1080/17453674.2018.1560647