

4-20-2017

# Improved Risk Prediction Following Surgery Using Machine Learning Algorithms

Anne P. Ehlers

*University of Washington - Seattle Campus, apugel@uw.edu*

Senjuti Basu Roy

Sara Khor

Prathyusha Mandagani

*See next pages for additional authors*

Follow this and additional works at: <http://repository.edm-forum.org/egems>

---

## Recommended Citation

Ehlers, Anne P.; Basu Roy, Senjuti; Khor, Sara; Mandagani, Prathyusha; Maria, Moushumi; Alfonso-Cristancho, Rafael; and Flum, David R. (2017) "Improved Risk Prediction Following Surgery Using Machine Learning Algorithms," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 5: Iss. 2, Article 3.

DOI: <https://doi.org/10.13063/2327-9214.1278>

Available at: <http://repository.edm-forum.org/egems/vol5/iss2/3>

This Informatics Empirical Research is brought to you for free and open access by the the Publish at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

---

# Improved Risk Prediction Following Surgery Using Machine Learning Algorithms

## Abstract

**Background:** Machine learning is used to analyze big data, often for the purposes of prediction. Analyzing a patient's healthcare utilization pattern may provide more precise estimates of risk for adverse events (AE) or death. We sought to characterize healthcare utilization prior to surgery using machine learning for the purposes of risk prediction.

**Methods:** Patients from MarketScan Commercial Claims and Encounters Database undergoing elective surgery from 2007-2012 with  $\geq 1$  comorbidity were included. All available healthcare claims occurring within six months prior to surgery were assessed. More than 300 predictors were defined by considering all combinations of conditions, encounter types, and timing along with sociodemographic factors. We used a supervised Naïve Bayes algorithm to predict risk of AE or death within 90 days of surgery. We compared the model's performance to the Charlson's comorbidity index, a commonly used risk prediction tool.

**Results:** Among 410,521 patients (mean age 52,  $52 \pm 9.4$ , 56% female), 4.7% had an AE and 0.01% died. The Charlson's comorbidity index predicted 57% of AE's and 59% of deaths. The Naïve Bayes algorithm predicted 79% of AE's and 78% of deaths. Claims for cancer, kidney disease, and peripheral vascular disease were the primary drivers of AE or death following surgery.

**Conclusions:** The use of machine learning algorithms improves upon one commonly used risk estimator. Precisely quantifying the risk of an AE following surgery may better inform patient-centered decision-making and direct targeted quality improvement interventions while supporting activities of accountable care organizations that rely on accurate estimates of population risk.

## Acknowledgements

Dr. Ehlers was supported by a training grant from the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award Number T32DK070555. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This project was supported by grant number R21HS023080 from the Agency for Healthcare Research and Quality (AHRQ). The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ. Dr. Alfonso-Cristancho is employed by GlaxoSmithKline. His participation during the design and the conception of the study was done when he was still fully employed by the Surgical Outcomes Research Center, but has since changed employment (GSK) and continued his contributions in an advisory role.

## Keywords

Outcomes Assessment; Data Analysis Method; Methods;

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](http://creativecommons.org/licenses/by-nc-nd/3.0/).

---

**Authors**

Anne P Ehlers, *University of Washington - Seattle Campus*; Senjuti Basu Roy; Sara Khor; Prathyusha Mandagani; Moushumi Maria; Rafael Alfonso-Cristancho; David R Flum.



# Improved Risk Prediction Following Surgery Using Machine Learning Algorithms

Anne P. Ehlers, MD, MPH;<sup>i</sup> Senjuti Basu Roy, PhD;<sup>ii</sup> Sara Khor, MS;<sup>iii</sup> Prathyusha Mandagani; Moushumi Maria; Rafael Alfonso-Cristancho, MD, MSc, PhD;<sup>iv</sup> David R. Flum, MD, MPH<sup>v</sup>

## ABSTRACT

**Background:** Machine learning is used to analyze big data, often for the purposes of prediction. Analyzing a patient's healthcare utilization pattern may provide more precise estimates of risk for adverse events (AE) or death. We sought to characterize healthcare utilization prior to surgery using machine learning for the purposes of risk prediction.

**Methods:** Patients from MarketScan Commercial Claims and Encounters Database undergoing elective surgery from 2007-2012 with  $\geq 1$  comorbidity were included. All available healthcare claims occurring within six months prior to surgery were assessed. More than 300 predictors were defined by considering all combinations of conditions, encounter types, and timing along with sociodemographic factors. We used a supervised Naïve Bayes algorithm to predict risk of AE or death within 90 days of surgery. We compared the model's performance to the Charlson's comorbidity index, a commonly used risk prediction tool.

**Results:** Among 410,521 patients (mean age 52,  $52 \pm 9.4$ , 56% female), 4.7% had an AE and 0.01% died. The Charlson's comorbidity index predicted 57% of AE's and 59% of deaths. The Naïve Bayes algorithm predicted 79% of AE's and 78% of deaths. Claims for cancer, kidney disease, and peripheral vascular disease were the primary drivers of AE or death following surgery.

**Conclusions:** The use of machine learning algorithms improves upon one commonly used risk estimator. Precisely quantifying the risk of an AE following surgery may better inform patient-centered decision-making and direct targeted quality improvement interventions while supporting activities of accountable care organizations that rely on accurate estimates of population risk.

<sup>i</sup>University of Washington, Seattle Campus, <sup>ii</sup>Department of Computer Science, New Jersey Institute of Technology, <sup>iii</sup>University of Washington Surgical Outcomes Research Center, <sup>iv</sup>GlaxoSmithKline, <sup>v</sup>University of Washington School of Medicine

## Introduction

The use of machine learning and data science techniques is increasingly common.<sup>1-3</sup> Machine learning and data science refer to a family of methods that have been devised to handle complex data, often referred to as “big data.” The National Institutes of Health defines “biomedical big data” as being sources of information that “include the diverse, complex, disorganized, massive, and multimodal data being generated by researchers, hospitals, and mobile devices around the world...It includes imaging, phenotypic, molecular, exposure, health, behavioral, and many other types of data.”<sup>4</sup> The challenge is that biomedical big data can contain huge quantities of data with hundreds of variables that limit the use of traditional analytic techniques—and not just because of the size and complexity of the data sets themselves.<sup>3</sup> For example, in one analysis that sought to predict incidence of type 2 diabetes using claims data, investigators identified approximately 900 variables (out of an initial set of 42,000) that were predictive of disease.<sup>5</sup> This type of analysis would be nearly impossible using traditional regression-based methods given the size of the data set and the number of variables.

One particular area where machine learning may add value in health care is in the setting of perioperative risk stratification—i.e., prediction of adverse events (AE) and death following surgery. The proportion of patients ages 50 years or older with conditions that require medical treatments and surgical procedures are predicted to double in the next 50 years. These procedures are increasingly performed in patients with multiple chronic conditions, a group at greatest risk of serious complications and procedure-related deaths.<sup>6</sup> Although complication rates following surgery increase with age and multiple chronic conditions, the effect of specific combinations, sequences, and timing of comorbid conditions on outcomes has not been well described. Many current

methods for risk stratification are limited in that they rely on cross-sectional information available at the time of care that is then summed and weighted into a generic risk estimate,<sup>8-10</sup> which ignores the increasingly detailed and complex nature of health care data.

Machine learning techniques may provide the opportunity for more robust risk estimates by taking advantage of the rich supply of patient-level data pertaining not only to the types of comorbid conditions that patients have, but also the frequency for which they seek care, the sequence and timing of health care utilization, and the intensity of that care. We sought to analyze health care utilization claims data among a cohort of commercially insured adults in the United States for the purposes of risk prediction following elective surgery. We hypothesized that incorporating information about the sequence, timing, and intensity of health care utilization into a Naïve Bayes algorithm would improve risk estimates provided by one commonly used risk stratification tool.

## Methods

### Data Source and Patients

Eligible patients were 18 years and older with at least one comorbid condition who were enrolled in the Truven Health MarketScan Commercial Claims and Encounters Databases from 2007 to 2012 and were undergoing elective surgery in one of 10 categories (esophageal, bariatric, gastric, small bowel, colorectal, pelvic—e.g., hysterectomy, prostate, hip, knee, spine). The MarketScan database includes health care claim data for millions of privately insured U.S. employees and their dependents. Inpatient, outpatient, and emergency health claims are linked at the individual patient level to provide granular information about each patient’s unique health care utilization pattern. Each claim is identified by the associated diagnoses and procedures, codified using Current Procedural Terminology (CPT)



and International Statistical Classification of Diseases and Related Health Problems 9 (ICD-9) codes. We used previously published national data as a guide to determine the most frequently performed operation types<sup>11</sup> but included other surgical types, as well. We focused on elective surgery because it provides an opportunity in clinical practice to identify patients who would potentially benefit from quality improvement interventions aimed at risk reduction. The goal was to create a cohort of patients with a variety of surgical types that could be generalizable to the broader population, especially when thinking about using such an algorithm in a large health system. This analysis of existing, de-identified data is not considered human subjects research by the University of Washington Human Subjects Division and was exempt from review.

We identified surgical procedures using manually selected CPT codes. To define comorbidities, we used the coding algorithm developed by Deyo et al., adapted from the Charlson comorbidity index, a commonly used risk stratification tool.<sup>8,9</sup> This algorithm was validated among a cohort of patients undergoing lumbar spine surgery. Comorbid conditions defined include myocardial infarction, congestive heart failure, peripheral vascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, mild liver disease, moderate or severe liver disease, diabetes with or without chronic complication, hemiplegia or paraplegia, renal disease, any malignancy (including lymphoma and leukemia, except malignant neoplasm of skin), metastatic solid tumor, and HIV/AIDS.<sup>8,9</sup>

### Exposures

We recorded patient demographics (age, sex, insurance type, and region of residence in the United States) as defined by the MarketScan database. We identified all claims within the six-month period prior to the patient's elective surgery. While it is possible to identify claims that occurred more distantly, we

selected this time point because it represented a clinically meaningful period prior to elective surgery. We considered each claim based on the condition identified as the following: (1) the primary diagnosis (e.g., diabetes); (2) the health care utilization type (outpatient, emergency room, or inpatient), and (3) the timing (months before surgery). All claims were translated into a unique sequence that could be used as a predictor for AEs and death. For example, a claim for uncomplicated diabetes in the outpatient setting one month prior to surgery was denoted as "month1\_diab\_wo\_com\_outpatient," which could then be used as a predictor of AE or death in combination with other covariates. This claim categorization method led to the creation of more than 300 predictors, each containing these three separate components. For patients with inpatient claims, we also created an additional variable that described the average length of stay for each hospital stay in the six months prior to surgery.

### Outcomes

The outcomes of interest were a composite measure of AEs (classified as reoperation, respiratory, nervous system, urinary, cardiac, infection/sepsis, gastrointestinal, deep venous thrombosis/pulmonary embolus, adverse gastrointestinal outcome, or other—e.g., postoperative hemorrhage) and death. Death was defined by the MarketScan database and included deaths that occurred in the inpatient setting, either at the time of the initial operation or during a readmission within 90 days. Within each category of AE, we identified multiple ICD-9 codes to be as inclusive as possible. Events were captured up to 90 days following surgery. We selected this time point because we determined that it would maximize the number of eligible patients while identifying events not typically captured in the 30-day postoperative period. While the 30-day postoperative period is frequently used as the more conventional endpoint, prior work suggests that a

significant number of complications (such as venous thromboembolism) occur after 30 days.<sup>12</sup> All ICD-9 and CPT codes are available in the Appendix.

## Analysis

We formalized the task as a supervised learning problem with two possible classes or outcomes, where the objective was to predict whether each patient would experience an AE or death within 90 days after surgery. We designed machine-learning models that accounted for the more than 300 variables previously described as predictors in a supervised Naïve Bayes statistical classification algorithm.<sup>13</sup> Naive Bayes algorithms are based on the application of Bayes' theorem and assume that each individual feature or predictor (in this case each health care utilization claim) is independent of the value of other features or predictors.<sup>14</sup> Despite this somewhat simplified assumption and the agnostic design, this algorithm allows for analysis of complex, unstructured data such as those contained in the MarketScan database. An additional advantage is that it allows for estimation of classification parameters using only small amounts of training data. The algorithm is "supervised" in the sense that we selected the features that should be used for the prediction purposes. Under the Bayesian framework, we do not manually assign any specific weights to any specific variable (as occurs in logistic regression models); rather, the Naïve Bayes Algorithm derives the weight of each predictor (i.e., each encounter type) probabilistically by working through the data while computing the probability of an AE given the set of predictors.

To develop the predictive model, the data set was randomly partitioned into 10 equal-sized samples and the algorithms were run in 10 iterations. In each iteration, nine-tenths of the samples were used for training and one-tenth for testing, such that each patient was contained in both the training and testing data set, but no patient was used for both training and testing in the same iteration. Results

from each iteration of testing were then aggregated to give a final description of the model performance. This method avoids the problem of overfitting that is often seen in regression models. We evaluated the model's performance individually and compared it to the performance of Deyo's adaptation of the Charlson Comorbidity Index in this patient population.<sup>8</sup> We assessed performance by measuring the area under the curve (AUC), which describes the probability that the model will assign the correct outcome (i.e., AE/death or not) when compared to random chance. Typically tests with AUC greater than 0.70 are thought to have good performance characteristics, while AUC=0.50 is what would be expected given random chance or a coin flip.

To further identify characteristics of patients who were particularly high risk, after evaluating the model's performance in the training and testing sets we identified the 20 claim types that were most predictive of AE or death among all patients, as well as the clusters of claims that occurred most commonly in the one percent of patients at highest risk for AE or death.

The development and test environment uses RStudio (Studio, R., 2012), which is a free, open source, integrated development environment (IDE) for R, a programming language for statistical computing and visualizations.

## Results

Among 410,521 patients (mean age  $52 \pm 9.4$ , 56 percent female), the most common chronic conditions as defined by the Charlson comorbidity index were uncomplicated diabetes ( $n=136,767$ , 34.8 percent) and malignancy ( $n=120,950$ , 30.8 percent), while the least commonly reported were dementia ( $n=412$ , 0.1 percent) and severe liver disease ( $n=985$ , 0.2 percent). The most common surgical category type was pelvic surgery such as hysterectomy ( $n=94,201$ , 23 percent), while the least common



was esophageal surgery (n=980, 0.2 percent). The majority of patients had at least one outpatient claim in the six months prior to surgery (n=341,739, 83.2 percent), while approximately half of patients had

an inpatient claim in the six months prior to surgery (n=233,620, 56.9 percent). Very few patients had an emergency claim prior to surgery (n=16,003, 3.9 percent) (Table 1).

**Table 1. Baseline Characteristics of Study Population**

| Characteristic                         | N=410,521      |
|--|----------------|
| Mean Age                               | 52 ± 9.4       |
| Female Sex (%)                         | 231,115 (56.3) |
| <b>CHRONIC CONDITION (%)</b>           |                |
| Previous myocardial infarction         | 9,099 (2.3)    |
| Congestive heart failure               | 15,652 (3.9)   |
| Peripheral vascular disease            | 21,648 (5.5)   |
| Cerebrovascular disease                | 14,986 (3.8)   |
| Dementia                               | 412 (0.1)      |
| Chronic pulmonary disease              | 109,664 (2.8)  |
| Rheumatologic disease                  | 23,851 (6.0)   |
| Peptic ulcer disease                   | 8,188 (2.0)    |
| Mild liver disease                     | 40,195 (10.2)  |
| Severe liver disease                   | 985 (0.2)      |
| Hemiplegia or paraplegia               | 6,345 (1.6)    |
| Renal disease                          | 13,245 (3.3)   |
| Any malignancy                         | 120,950 (30.8) |
| Metastatic solid tumor                 | 18,970 (4.8)   |
| HIV/AIDS                               | 1,364 (0.3)    |
| Diabetes without chronic complication  | 136,767 (34.8) |
| Diabetes with chronic complication     | 15,528 (3.9)   |
| Adverse event prior to surgery         | 49,387 (12.0)  |
| <b>SURGERY TYPE (%)</b>                |                |
| Esophageal surgery                     | 980 (<1)       |
| Bariatric surgery                      | 38,761 (9.4)   |
| Gastrectomy                            | 2,505 (<1)     |
| Small bowel surgery                    | 12,331 (3)     |
| Colorectal surgery                     | 41,835 (10)    |
| Pelvic surgery                         | 94,201 (23)    |
| Prostate surgery                       | 36,310 (8.8)   |
| Hip surgery                            | 27,414 (6.7)   |
| Knee surgery                           | 67,292 (16.3)  |
| Spine surgery                          | 88,892 (21.6)  |
| <b>CLAIM TYPE PRIOR TO SURGERY (%)</b> |                |
| Inpatient                              | 233,620 (56.9) |
| Outpatient                             | 341,739 (83.2) |
| Emergency                              | 16,003 (3.9)   |



Within the 90 day period following surgery, 19,266 patients (4.7 percent) experienced an AE and 46 patients (0.01 percent) died. Rates of AE were highest following esophagectomy (20.6 percent) and small bowel surgery (13.5 percent), and were lowest for knee surgery (3.1 percent), hip surgery (3.4 percent), and pelvic surgery (3.4 percent). With the exception of esophageal surgery, which had a death rate of 0.2 percent, all other surgical types had a death rate of <0.1 percent (Table 2).

The machine learning model predicted AE and death more accurately than the Charlson comorbidity index did. For AE, the machine learning model accurately predicted 79 percent of AE (AUC=0.79,) while the Charlson comorbidity index accurately predicted 57 percent (AUC=0.57). Results were similar for prediction of death: the machine learning model accurately predicted 78 percent (AUC=0.78) of deaths, while the Charlson score accurately predicted only 59 percent (AUC=0.59) (Figure

1a-1d). The predictive capability varied by operation type. For prediction of AE, the model was most accurate for patients undergoing colorectal surgery (AUC 0.80), while it was least accurate for patients undergoing esophageal surgery (AUC 0.52). For death, the model was most successful for patients undergoing spine surgery (AUC 0.99) and least successful for patients undergoing hip surgery (AUC 0.48). For the purposes of prediction of adverse events, the machine learning model had a sensitivity of 70 percent and a specificity of 75 percent. For prediction of death, the model had a sensitivity of only 39 percent, but a very high specificity of 98 percent.

Among all patients, the feature that was most strongly predictive of AE or death following surgery was a claim for cancer in the outpatient setting within one month prior to surgery; the next most predictive was a claim for kidney disease in the outpatient setting two months prior to surgery;

**Table 2. Adverse Event and Death Rates Within 90 Days Following Surgery Stratified by Surgical Type**

| SURGERY TYPE    | ADVERSE EVENT | DEATH     |
|-----------------|---------------|-----------|
| Esophageal (%)  | 202 (20.6)    | 2 (0.2)   |
| Bariatric (%)   | 1,662 (4.3)   | 2 (<0.1)  |
| Gastrectomy (%) | 278 (11.1)    | 1 (<0.1)  |
| Small bowel (%) | 1,664 (13.5)  | 10 (<0.1) |
| Colorectal (%)  | 4,174 (10.0)  | 16 (<0.1) |
| Pelvic (%)      | 3,218 (3.4)   | 1 (<0.1)  |
| Prostate (%)    | 1,648 (4.5)   | 0 (0)     |
| Hip surgery (%) | 931 (3.4)     | 1 (<0.1)  |
| Knee (%)        | 2,093 (3.1)   | 3 (<0.1)  |
| Spine (%)       | 3,396 (3.8)   | 10 (<0.1) |
| Total           | 19,266 (4.7)  | 46 (<0.1) |



Figure 1a. The Charlson Score Accurately Predicts 57 Percent of Adverse Events Following Surgery

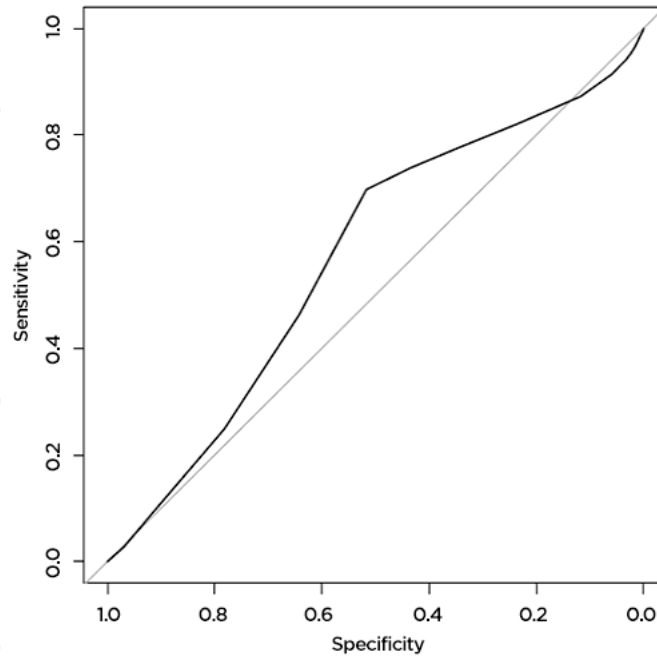


Figure 1b. The Machine Learning Model Accurately Predicts 79 Percent of Adverse Events Following Surgery

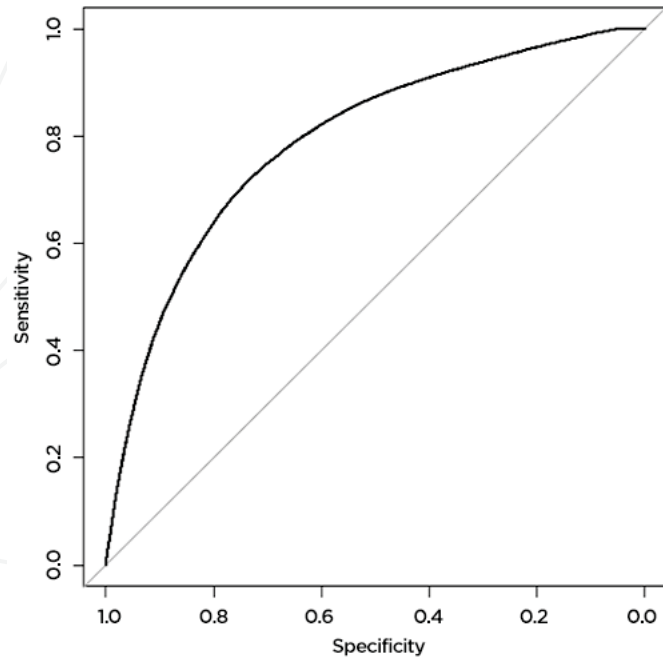


Figure 1c. The Charlson Score Accurately Predicts 59 Percent of Deaths Following Surgery

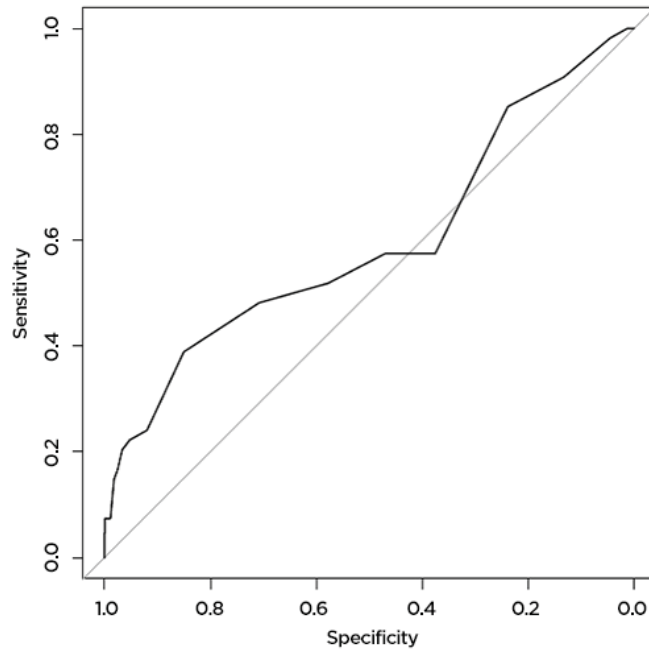


Figure 1d. The Machine Learning Model Accurately Predicts 78 Percent of Deaths Following Surgery

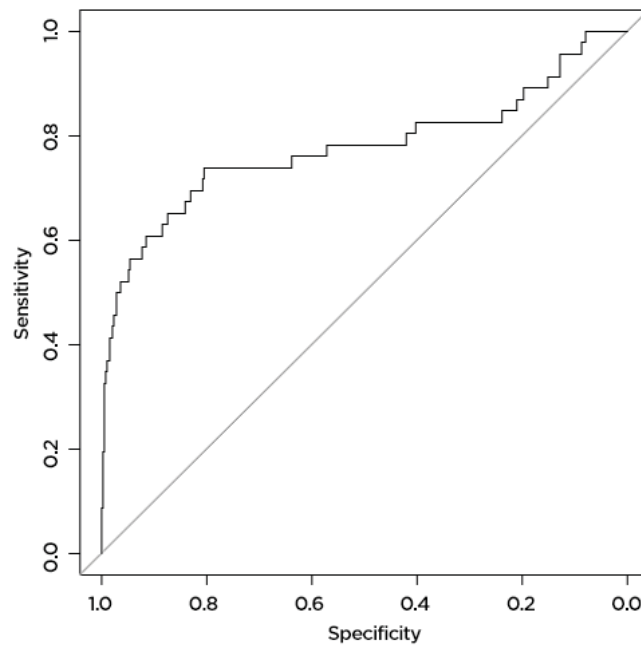




Table 3. Claim Types Most Predictive of Adverse Event or Death Within 90 Days Following Surgery

| MONTHS PRIOR TO SURGERY | CHRONIC CONDITION                      | CLAIM TYPE |
|-------------------------|--|------------|
| 1                       | Malignancy                             | Outpatient |
| 2                       | Renal Disease                          | Outpatient |
| 1                       | Peripheral Vascular Disease            | Outpatient |
| 4                       | Renal Disease                          | Outpatient |
| 2                       | Malignancy                             | Emergency  |
| 1                       | Renal Disease                          | Outpatient |
| 1                       | Malignancy                             | Emergency  |
| 2                       | Metastatic Solid Tumor                 | Emergency  |
| 3                       | Renal Disease                          | Outpatient |
| 4                       | Malignancy                             | Emergency  |
| 3                       | Malignancy                             | Emergency  |
| 5                       | Renal Disease                          | Outpatient |
| 1                       | Metastatic Solid Tumor                 | Emergency  |
| 1                       | Renal Disease                          | Emergency  |
| 5                       | Malignancy                             | Emergency  |
| 3                       | Metastatic Solid Tumor                 | Emergency  |
| 1                       | Hemiplegia/Paraplegia                  | Outpatient |
| 2                       | Diabetes without Chronic Complications | Emergency  |
| 1                       | Myocardial Infarction                  | Outpatient |
| 5                       | Renal Disease                          | Emergency  |

the third most predictive was a claim for peripheral vascular disease in the outpatient setting one month before surgery. Other claims that were highly predictive included claims for hemiplegia or myocardial infarction (Table 3). Among the one percent of patients with the highest predicted risk of AE or death following surgery, the most common

cluster of claims prior to surgery was an outpatient claim for cancer and an inpatient claim for cancer in the one month prior to surgery. Of the top 20 most commonly appearing clusters of claims in this highest risk group, 16 clusters were for cancer in the inpatient and outpatient setting at various time points before surgery, while the other 4 were for

diabetes without complications in the inpatient and outpatient setting, again at various time points.

## Discussion

In this analysis, we found that inclusion of information describing the type, timing, sequence, and intensity of health care utilization prior to surgery provided superior predictive capability when compared to common risk predictions tools such as the Charlson comorbidity index. Our model accurately predicted 79 percent of AEs and 78 percent of deaths, while the Charlson comorbidity index only predicted 57 percent and 59 percent, respectively. Modeled analyses suggest that approximately 3 million preventable AEs, which may cost more than \$16 billion annually, occur every year in the United States.<sup>15</sup> At these rates, the Charlson comorbidity index would potentially identify 1.71 million preventable AEs, while our model would potentially identify an additional 670,000 events for a total of 2.37 million preventable AEs. This improved identification may translate into significant cost savings.

Current risk stratification tools are limited in that they provide a somewhat cross-sectional view of patient disease. Machine learning techniques such as the Naïve Bayes algorithm provide a novel approach to data analysis in surgical research. Traditionally, surgical researchers approached risk prediction using logistic regression models that identified specific factors that were highly predictive of the outcome. While many current risk stratification models assign certain diagnoses a higher weight in the resulting score (e.g., liver disease is weighted more heavily than diabetes),<sup>8</sup> they are limited in that they fail to account for the fact that two patients with the same constellation of diagnoses might have very different levels of risk. For example, a patient with a 10-year history of diabetes as well as mild hemiplegia from a stroke in the distant past

would be given the same risk score as a patient with recently diagnosed diabetes in the setting of hospitalization for acute stroke shortly before surgery. The Naïve Bayes approach is different in that the algorithm itself derives the weight of the predictors probabilistically by looking into the data while computing the likelihood of the outcome given the predictors. Again, this avoids the problem of overfitting that is concerning when using logistic regression approaches with a high number of predictors. While clinicians in the real world can look back through patients' medical records to evaluate their risk of specific events, previous work has shown that physicians do poorly when asked to predict the likelihood of clinical events such as survival following cardiopulmonary resuscitation<sup>16</sup> or readmission to the hospital.<sup>17</sup>

The model we describe here approached the performance of other more resource-intensive models such as the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) Calculator,<sup>18</sup> which predicts AEs and death with 80–90 percent accuracy among a cohort of patients undergoing surgery in the fields of general surgery, gynecology, neurosurgery, orthopedics, otolaryngology, plastic surgery, cardiothoracic surgery, urology, and vascular surgery.<sup>10</sup> It is important to note that we achieved these results using existing claims data that are generated on an ongoing basis as patients interact with the health care system without the need for manual chart abstraction. In the case of NSQIP, hospitals must pay between \$10,000 and \$29,000 annually to participate in the program and employ a Surgical Clinical Reviewer and abstractors to prospectively abstract data from the medical record.<sup>19</sup> While initial development of a machine learning algorithm does require some resources such as an analyst and a data set, once in place it can be automatically updated using incoming information



from the electronic health record or claims database without the need for manual data abstraction.

This study has several limitations. First, we compared our model performance to the Charlson score, which was originally developed to predict death.<sup>8</sup> Although Deyo adapted the Charlson score for use with ICD-9 codes for the purposes of AE prediction,<sup>9</sup> this may explain its poor performance. However, we found similar results when comparing the two models for the purpose of predicting death. Second, our model was trained and tested on a population of commercially insured adults in the United States, which may not be representative of the general population, and it has not been validated on an external population that may limit its applicability. Nonetheless, these methods could be applied to any population for the purpose of accurate risk estimation. It may also be that the variety of operations represented in our cohort are not representative of the population as a whole, however we did test the model within each surgical type and found it to be superior to the Charlson score in most cases. Additionally, the time window we selected may not have captured all conditions. However, future iterations could address this by changing the time frame of condition and event identification both before and after surgery. Finally, we were able to capture only deaths that occurred while a patient was in the hospital, as the MarketScan database does not capture deaths occurring in the outpatient setting.

One general concern in any study using claims data is the reliance on diagnostic and procedural codes to identify cases and events. One challenge is that we may have failed to identify some chronic conditions, operations, or events based on our selection of ICD-9 or CPT codes, which may reduce the accuracy of prediction by being restricted to a subset of patients not truly reflective of the population at large. Another is that ICD-9 codes are often not granular enough to distinguish patients with varying levels

of illness. For example, in this cohort we found that kidney disease was associated with a higher risk of AE or death, but clinicians in practice know that there are varying levels of kidney disease among the general population. The advantage to an algorithm such as ours is that it could identify patients at high risk based on their overall pattern of health care utilization and then allow for a directed team of experts to more carefully review those individuals' health to determine if they would benefit from additional intervention. A final limitation of using claims data and billing codes is that they cannot measure items such as provider expertise that may affect outcomes such as the occurrence of AE or death.

From a health systems perspective, accurately predicting the risk of AEs and death for each patient is critically important. While not all AEs or deaths are necessarily preventable,<sup>15</sup> payers and policymakers rely on the availability of accurate risk estimation to guide decision-making and to determine reimbursement. For those events that are preventable, it is important to identify the patients who would benefit most from potentially burdensome interventions. In such a scenario, a similar algorithm could be implemented for a population of patients (such as within an accountable care organization) to “flag” patients at highest risk based on their health care claims. Patients identified as the highest risk patients could then be given special attention prior to undergoing an invasive procedure to try to minimize their risk of AE or death. This would also allow for careful allocation of resources to the patients most in need. When caring for large populations of patients, providers need to identify the highest risk patients. Often there is not one single feature that makes the patients high risk; rather, it is their pattern of health care utilization as a whole, their health care “fingerprint” that gives a more accurate assessment

of their risk. Perhaps what needs to be addressed is not one discrete item (e.g., controlling a patient's blood glucose) but patients' health as a whole. As the United States moves toward alternative payment models,<sup>20</sup> health care systems are assuming the responsibility for the health of whole patient populations, rather than just individual patients, which requires the availability of accurate and reliable information about underlying risk.<sup>21</sup> Many reimbursement schemes will also take into account the risk-adjusted rates of AE and death,<sup>21</sup> especially in pay for performance models.<sup>22</sup> Failing to account for the fact that one population is higher risk than others may have significant financial consequences for health care systems.

In conclusion, we found that incorporation of a patient's health care utilization pattern prior to undergoing elective surgery provided superior risk estimation when compared to a commonly available model. In the future, granular information that characterizes an individual's health care utilization pattern may serve as a health care fingerprint, of sorts. Just as individual genomic data have been leveraged into the field of genomics to make highly personalized predictions regarding risk of disease (e.g., risk of breast cancer among patients with BRCA mutations), perhaps this health care utilization fingerprint can be similarly leveraged into a new field of "utilomics" that would take advantage of the unique way that each patient moves through the health system, and the ways in which this is increasingly captured in big data.

## Acknowledgements

Dr. Ehlers was supported by a training grant from the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award Number T32DK070555. The content is solely the responsibility of the authors and does not necessarily represent the official views

of the National Institutes of Health. This project was supported by grant number R21HS023080 from the Agency for Healthcare Research and Quality (AHRQ). The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ. Dr. Alfonso-Cristancho is employed by GlaxoSmithKline. His participation during the design and the conception of the study was done when he was still fully employed by the Surgical Outcomes Research Center, but has since changed employment (GSK) and continued his contributions in an advisory role.

## References

1. Parikh RB, Kakad M, Bates DW. Integrating Predictive Analytics Into High-Value Care: The Dawn of Precision Delivery. *JAMA*. 2016 Feb 16;315(7):651-2.
2. Darcy AM, Louie AK, Roberts LW. Machine Learning and the Profession of Medicine. *JAMA*. 2016 Feb 9;315(6):551-2.
3. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013 Apr 3;309(13):1351-2.
4. Big Data to Knowledge (BD2K) [Internet].; 2015 []. Available from: <http://datascience.nih.gov/bd2k>.
5. Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors. *Big Data*. 2015;3(4):277-87.
6. Baldwin LM, Klabunde CN, Green P, Barlow W, Wright G. In search of the perfect comorbidity measure for use with administrative claims data: Does it exist? *Med Care* 2006;44(8):745.
7. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*. 2014 Jul;33(7):1123-31.
8. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-83.
9. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992 Jun;45(6):613-9.
10. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmieciak TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013 Nov;217(5):833,42.e1-3.
11. Weiss AJ, Elixhauser A. Trends in Operating Room Procedures in U.S. Hospitals, 2001-2011: Statistical Brief #171. In: *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville (MD): ; 2006.



12. Nelson DW, Simianu VV, Bastawrous AL, Billingham RP, Fichera A, Florence MG, Johnson EK, Johnson MG, Thirlby RC, Flum DR, Steele SR. Thromboembolic Complications and Prophylaxis Patterns in Colorectal Surgery. *JAMA Surg.* 2015 Aug; 150(8):712-20.
13. Murphy KP. Naive Bayes Classifiers. University of British Columbia; 2006.
14. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med.* 1975 Jul 31;293(5):257.
15. Jha AK, Chan DC, Ridgway AB, Franz C, Bates DW. Improving safety and eliminating redundant tests: cutting costs in U.S. hospitals. *Health Aff (Millwood).* 2009 Sep-Oct;28(5):1475-84.
16. Ebell MH, Bergus GR, Warbasse L, Bloomer R. The inability of physicians to predict the outcome of in-hospital resuscitation. *J Gen Intern Med.* 1996 Jan;11(1):16-22.
17. Allaudeen N, Schnipper JL, Orav EJ, Wachter RM, Vidyarthi AR. Inability of providers to predict unplanned readmissions. *J Gen Intern Med.* 2011 Jul;26(7):771-6.
18. ACS NSQIP Surgical Risk Calculator [Internet]. []. Available from: <http://riskcalculator.facs.org/>.
19. Frequently Asked Questions about ACS NSQIP [Internet]. []. Available from: <https://www.facs.org/quality-programs/acs-nsqip/joinnow/joinfaq>.
20. H.R.2. - Medicare Access and CHIP Reauthorization Act of 2015, 2015).
21. Devore S, Champion RW. Driving population health through accountable care organizations. *Health Aff (Millwood).* 2011 Jan;30(1):41-50.
22. Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res.* 2010 Aug 23;10:247,6963-10-247.