MEDINFO 2019: Health and Wellbeing e-Networks for All L. Ohno-Machado and B. Séroussi (Eds.) © 2019 International Medical Informatics Association (IMIA) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHTI190595

# Can Solo Practitioners Survive in Value-Based Healthcare? Validating a Predicative Model for ED Utilization

# Pamella Howell<sup>a</sup>, Peter L. Elkin<sup>a</sup>

<sup>a</sup> Department of Biomedical Informatics, State University of New York at Buffalo, Buffalo, New York, United States,

#### Abstract

The health industry will see increased implementations of value-based models. This study validates a predictive model for determining emergency room utilization. Data from 2991 records are used for the analysis. To validate the model we used Poisson and random forest models. The results indicate that patients with one of six chronic conditions, who missed scheduled appointments or had higher body mass indexes were more likely to utilize the emergency department.

### Keywords:

Value-based Healthcare; Machine Learning; Resource Allocation

### Introduction

The value-based healthcare model is poised to change the practice of medicine in the United States. Value-based care addresses two concerns: are the available resources distributed to those who need it and how clinicians use the resources optimally [1]. Participating insurers and providers have aligned valued based programs with broader strategic objectives on quality. Value-based care serves three primary functions to reduce cost, improve outcomes and improve quality. As participation expands, patients, payers, and providers will experience a wide range of changes due to its effect. The most significant impact for patients is the improvement in the quality of care and outcomes. Both public and private sector insurers will benefit from any reductions in the cost of care. Insurers are yet to propose methods for decreasing premiums, which is a salient topic; however, it is beyond the scope of this study.

Clinicians serve dual roles in the value-based model. Firstly, they are primarily responsible for improving the quality of care and outcomes of patients. Secondly, they must assist payors by lowering the cost to care for patients. This study aims to provide a predictive model for identifying patients associated with high cost. Using the Health Cost Guidelines (HCG) as a reference, we have identified five categories of cost that a physician may need to reduce, including inpatient, outpatient, prescription drugs, physician, and other. Historical estimates show that inpatient, outpatient and pharmacy cost are highest representing 62 % of the national expenditure for 2016 [2]. The emphasis here will be outpatient cost specifically those incurred in the emergency department. Approximately 3.3% of all emergency department (ED) visits are avoidable [3]; however, statistics also show increased utilization for the treatment of chronic diseases [4].

How does the clinician identify patients who are likely to incur high ED cost? Can the factors that cause high ED usage be eliminated? In answering these question, we attempt to address two of many challenges. One, individuals might argue that the insurers will provide the information needed to monitor cost. The information provided by insurers is generally delayed due to claims processing; this significant time lag reduces a provider's ability to act. Delayed responses when attempting to improve patient adherence can have a negative impact. In this research, we utilize secondary data from the electronic health record (EHR) to reduce data lags and the resource intensity associated with value-based analytics.

The second challenge is that providers need technology and analytics infrastructure to explore cost. Though organizations like hospitals and large health systems have met some of these requirements, the outlay of capital has been significant. For solo and medium-size practices, the ability to participate effectively in the value-based ecosystem will be limited by finite human and financial resources. Clinicians need resources to implement value-based programs such as advanced analytic platforms. Given the current interoperability issues between EHRs and the limited efforts by some vendors to provide analytics support providers must find a way to supplement the need for these services. By implementing our proposed model physicians can reallocate current health information technology resources including team members without significant expenditure.

# Methods

Using predictive analytics and machine learning models, we validate a model for predicting patient utilization of the outpatient services. To evaluate the research objective, we utilize a two-stage combination model. The first stage uses the outcomes of a model developed with regression-based machine learning techniques and standardized statistical techniques [5]. The second stage also used standardized regression techniques in the form a Poisson regression and a classification machine learning technique in the form of a random forest model. The feedback loop uses data from the EHR to create a predictive model, then in this study we validate the model using a new patient cohort covering a different period and with more observations. The dataset contains a total of 2991 records with six de-identified variables collected between 2017 and 2018.

### Variable Selection and Description

The model variables include age, a polynomial age variable, body mass index (BMI), depression screen indicator, number of no-show appointments, gender and a chronic disease indicator – diabetes, heart failure, COPD, Hypertension, Hyperlipidemia and asthma. The variables can be subdivided into process variables including no-show visits and depression screens indicator and clinical outcomes such as chronic conditions and BMI. All the variables were established as significant in a prior study. The dependent variable is emergency department visits.

#### Statistical Methodology

We use two approaches to analyzing the dependent variable. For one method the dependent variable is a count of the discrete non-negative integer ED visits. The Poisson or negative binomial regression models are typically the first recommended models. For the second method we dichotomize the count variable and use the random forest for the following benefits: it gives estimates of what variables are important, offers a technique for detecting variable interactions and has methods for balancing error in unbalanced datasets. Classification and Regression Trees (CART) like random forest are powerful predictive tools, they allow for the combining of predictors [6] [7]. The random forest model generates many CART models in which it chooses that model classification with the most votes.

# Results

Using SAS 9.4 and R, we validate the model using Poisson and random forest models. We analyzed the response variable first as a count, then as a dichotomous indicator for the random forest model. We examine the association of ED count with age, gender, no show visits, depression screening, six chronic diseases, and body mass index. The model results suggest that four variables have a significant impact on the frequency of emergency room visits at p < 0.05.

The relationship between the response variable and age is convex when compared to the previous study; this likely due to the increased distribution of age. For a given level of all x variables, the expected number of emergency visits will decrease by a factor of exp(-0.055)=0.946 to its lowest, at this point the expected number of emergency department (ED) visits would increase by a factor of exp(0.004)=1.004. The number of ED visits for patients one of 6 chronic diseases is exp(1.362) = 3.90 times more than patients without one of the chronic diseases. No-show visits positively impact the expected number of ER visits for patients by a factor of exp(0.051)=1.420. Body mass index positively impacts the expected number of ER visits for patients by a factor of exp(0.099)=1.104. In this cohort, depression screen indicator and gender did not impact the utilization of the ED.

The preliminary model validation results using random forest, show that the area under the curve (AUC) 0.825 indication good discrimination. The accuracy of the test model is 0.983, sensitivity - 0.884 and specificity - 0.996. The variable importance results showed that the chronic disease indicator held the highest rank followed by no-show visits, BMI, gender, depression, and age.

## Discussion

The movement of medical practice away from fee for service toward a value-based system will take time. Insurers have not changed their current products instead — for example, the Center for Medicare and Medicaid Services (CMS) as piloted some value-based programs for diseases such as end-stage renal disease. Private insurers have taken a different approach by contracting with health care providers to implement valuebased services for population subsets. Solo and medium-sized practices will face many challenges when making decisions on how best to utilize limited human and financial resources to meet value-based objectives. Physicians who intend to benefit from participating in value-based quality initiatives must leverage their limited resources by using predictive modeling to define parameters that enable the reduction of patient cost like emergency rooms utilization.

The regression model indicates that chronic diseases indicator, no-show visits, BMI, and age can be used to predict the count of ED visits. By combining these finding with the CART prioritization estimates, we can provide suggestions for physicians on how to allocate resources. We suggest creating electronic patient registries that capture parameters outlined in

the model. Providers should limit registries to patients with the six diseases indicated in the model due to the higher likelihood of predicting ED use. With a focus on a subset of the population, the practice can assign specific team members to work with the registry. The random forest importance also indicated that physicians should monitor patients with high noshow rates and body mass index this can also be achieved with a registry in the electronic health record. An automated alert and calling feature is available on most EHR; this feature permits the increases monitoring of patients who no-show with little human intervention. Clinical decision support systems parameters can be developed to encompass all the predicated variables for assisting physicians on which patients to counsel for increased risk of ED use. Otherwise, patients who do not meet these criteria should be monitored lightly freeing up human and technological resources.

# Conclusions

Utilizing minimal data from the electronic medical record physicians can use existing resources to meet cost related valuebased metrics. This method ensures that providers can implement the model by merely using registries in the EHR. Partnering with the insurer and utilizing our predictive model will delay a significant outlay of capital for specialized analytic technology and team members. To better understand the research question, future research will use additional machine learning techniques to improve the prediction capability of the model.

### Acknowledgements

None.

# References

[1] M. Gray, Value based healthcare, in, British Medical Journal Publishing Group, 2017.

[2] National Health Expenditure Projections 2013-2023, in: C.f.M.M. Services, ed., Centers for Medicare & Medicaid Services

[3] R.Y. Hsia and M. Niedzwiecki, Avoidable emergency department visits: a starting point, *International Journal for Quality in Health Care* 29 (2017), 642-645.
[4] J.G. Trogdon, L.B. Murphy, O.A. Khavjou, R. Li, C.M. Maylahn, F.K. Tangka, T.A. Nurmagambetov, D.U. Ekwueme, I. Nwaise, D.P. Chapman, and D. Orenstein, Costs of Chronic Diseases at the State Level: The Chronic Disease Cost Calculator, *Preventing chronic disease U6 - Journal Article* 12 (2015), E140.

[5] P.C. Howell, Clinical Communication and Collaboration: Three Essays Examining the Impact of IT Interventions on At-Risk Populations Using Healthcare Analytics, State University of New York at Buffalo, 2018.
[6] L. Breiman and P. Spector, Submodel selection and evaluation in regression. The X-random case, International statistical review/revue internationale de Statistique (1992), 291-319.

[7] L. Breiman, *Classification and regression trees*, Routledge, 2017.

### Address for correspondence

Pamella Howel; pamellah@buffalo.edu