4-20-2017

Reducing Healthcare Costs Through Patient Targeting: Risk Adjustment Modeling to Predict Patients Remaining High-Cost

Jonathan A. Wrathall
Intermountain Healthcare, jonathan.wrathall@imail.org

Tom Belnap
Intermountain Healthcare, tom.belnap@imail.org

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.

Reducing Healthcare Costs Through Patient Targeting: Risk Adjustment Modeling to Predict Patients Remaining High-Cost

Abstract

Context: The transition to population health management has changed the healthcare landscape to identify high risk, high cost patients. Various measures of patient risk have attempted to identify likely candidates for care management programs. Pre-screening patients for outreach has often required several years of data. Intermountain Healthcare relied on cost-ranking algorithms which had limited predictive ability. A new risk-adjusted algorithm shows improvements in predicting patients' future cost status to facilitate identifying patient eligibility for care management.

Case Description: A retrospective cohort study design was used to evaluate high-cost patient status for two of the next three years. Modeling was developed using logistic regression and tested against other decision tree methods. Key variables included those readily available in electronic health records supplemented by additional clinical data and estimates of socio-economic status.

Findings: The risk-adjusted modeling correctly identified 79.0% of patients ranking among the top 15% of costs in one of the next three years. In addition, it correctly estimated 48.1% of the patients in the top 15% cost group in two of the next three years. This method identified patients with higher medical costs and more comorbid conditions than previous cost-ranking methods.

Major Themes: This approach improves the predictive accuracy of identifying high cost patients in the future and increases the sensitivity of identifying at-risk patients. It also shortened data requirements to identify eligibility criteria for case management interventions.

Conclusion: Risk-adjustment modeling may improve management programs' interface with patients thus decreasing costs. This method may be generalized to other healthcare settings.

Acknowledgements

Acknowledgements: The authors would like to thank Andy Merrill, MS for his contributions.

Keywords

Value/Cost, Care Coordination, Population Health

Disciplines

Other Medicine and Health Sciences | Social Statistics

eGEMs
Generating Evidence & Methods to improve patient outcomes

Reducing Health Care Costs Through Patient Targeting: Risk Adjustment Modeling to Predict Patients Remaining High Cost

Jonathan Wrathall, PhD,[i] Tom Belnap, MS

ABSTRACT

Context: The transition to population health management has changed the healthcare landscape to identify high risk, high cost patients. Various measures of patient risk have attempted to identify likely candidates for care management programs. Pre-screening patients for outreach has often required several years of data. Intermountain Healthcare relied on cost-ranking algorithms which had limited predictive ability. A new risk-adjusted algorithm shows improvements in predicting patients' future cost status to facilitate identifying patient eligibility for care management.

Case Description: A retrospective cohort study design was used to evaluate high-cost patient status for two of the next three years. Modeling was developed using logistic regression and tested against other decision tree methods. Key variables included those readily available in electronic health records supplemented by additional clinical data and estimates of socio-economic status.

Findings: The risk-adjusted modeling correctly identified 79.0% of patients ranking among the top 15% of costs in one of the next three years. In addition, it correctly estimated 48.1% of the patients in the top 15% cost group in two of the next three years. This method identified patients with higher medical costs and more comorbid conditions than previous cost-ranking methods.

Major Themes: This approach improves the predictive accuracy of identifying high cost patients in the future and increases the sensitivity of identifying at-risk patients. It also shortened data requirements to identify eligibility criteria for case management interventions.

Conclusion: Risk-adjustment modeling may improve management programs' interface with patients thus decreasing costs. This method may be generalized to other healthcare settings.

[i]Intermountain Healthcare

Introduction

In the face of rising health care costs, many voices within the health care industry have called for changes toward a more sustainable approach to health care with emphasis on population health management.[1] In this paper, we describe modeling techniques used to improve identification of high-cost patients likely to benefit from care management interventions. The modeling techniques described below do not require a resource intensive approach and may provide a means for other health systems to improve their own patient-intervention targeting.

One objective of population health management at Intermountain Healthcare is to facilitate the transition from a traditional "fee-for-service" system that compensates providers for services rendered, with a "fee-for-value" approach in which providers promote health among a defined patient cohort. This approach emphasizes improving outcomes and quality of service, and lowering overall health care costs.[2] This new health care climate requires changes to existing delivery systems in order to meet the needs of the community in ways that focus on the triple aim of improving the experience of care, the health of the population, and the cost of health care.[3]

Case Description

Intermountain Healthcare is an integrated delivery system based in Salt Lake City, Utah consisting of 22 hospitals and over 185 clinics. Intermountain has been actively engaged in developing programs designed to improve outcomes for defined patient populations that may require additional resources beyond the standard of care provided through a patient-centered medical home. One of these programs, known as Community Care Management (CCM), is designed to provide high intensity care management to high-cost, complex patients. This program is designed to help patients navigate the health care system with the goal of preventing avoidable utilization and slowing the progression of chronic conditions. The CCM teams specialize in in-home assessments, interdisciplinary care, intensive care coordination, and community integration. This program was designed to decrease catastrophic health episodes through patient education, disease management, and connection to community resources. To accomplish this, CCM teams are expected to improve the timeliness of care, improve medical coordination to reduce complications, and foster community relationships. These initiatives are intended to decrease overall health care costs primarily through avoiding unnecessary care or overutilizations.

In order for CCM programs to be successful, it is critical to identify and target the right patient population. To accomplish this, the stakeholders originally created a list of eligible patients via a ranking methodology, or Rank Algorithm, centered on reasonably simple inclusion criteria. In order to be eligible for the program, patients must be at least 18 years old, live within 30 miles of the program location, not already be enrolled in a care management program, be insured by Intermountain's affiliated health plan or be uninsured, and have health care costs in the top 10 percent of patients for the last year and in the top 15 percent of patients in one of the preceding two years. Patients meeting the inclusion criteria were then ranked equally based on the four following inclusive factors; prior year health care spending, the Charlson Comorbidity Index Score,[4] and two proprietary risk scores available within the organization—the IndiGO Expected Benefit Score[5] and the Optum Prospective Risk Score[6]. Patients were ranked independently by each factor, then rankings were averaged across the factors to get an overall rank. The patient with the lowest overall score was prioritized first, and the CCM staff was expected to invite patients into the program based on the order of the prioritized list.

Volume 5 (2017) Issue Number 2

The goal of this approach was to provide an objective enrollment process that was likely to enroll patients who would both benefit from the program and have enough cost savings potential to make the program viable. While the original approach was largely based on past health care spending, it did provide an objective approach to enrolling patient in the CCM program. These elements were used to rank patients based on historical data in order to guide patient outreach in the upcoming year. As a result, there were limitations to the Rank Algorithm that became apparent in the program over time.

The implementation team worked closely with the CCM clinical staff to implement the use of the Rank Algorithm. Over time there was ongoing feedback and refinement to the tool in order to ensure it was meeting the program's needs. The Rank Algorithm resulted in clinical staff taking significant time to review patient charts and appraise potential candidates. Many patients were considered ineligible, they declined to participate or their high cost episodes had resolved. As a result, there was a need to revisit the approach and methods used to identify patients and put in place something that better identified patients for the CCM program.

The team undertook an evaluation of the original patient selection process and tried to identify how the process had been used and how it could be improved moving forward. This evaluation identified several drawbacks to the ranking method, which held two important consequences. First, retrospective patient identification meant the system was less able to introduce appropriate health care interventions until after a health crisis, thus patients were able to be candidates for care management only when they had already experienced an acute episode. Second, a retrospective targeting method required significant time before patients accumulated enough utilization and cost to be identified as eligible for additional services. Additionally, this ranking method relied somewhat on opaque, third-party proprietary algorithms to establish clinical risk. These algorithms could not be calculated on all patients and were difficult for the clinical staff to interpret. Going forward, a predictive algorithm was needed to identify rising risk patients before they became medically complex and high cost. To accomplish this, a new algorithm has been developed to incorporate an approach that better predicts future patient costs and refines patient targeting. With these changes, there is an increasing ability to identify at-risk patients and to better engage them in their care.

Recent discussions of high-cost patient prediction have included debate as to the importance of administrative or clinical data sources.[7] As part of the recommendations made by Cucciare et al., the revised prediction methodology was modified to take advantage of gains introduced by both administrative and clinical data. In recent years, high-cost patient prediction has increasingly included an element of prior years' cost data as a means of predicting future patient costs. Doing so leads to better predictions than those obtained by patient demographics alone.[8,9] Alternative studies have shown that a combination of clinical and demographic data has also proved useful as a means to predict future patient costs.[10,11,12,13]

A retrospective cohort study design was used with logistic regression to evaluate high-cost patient status for two of the next three years, and was termed the "Logistic Model." The study sample consisted of patients in the top 15 percent of health care costs from January 1 to December 31, 2011 comprising 26,173 unique patients. Training data consisted of a random selection of 75 percent of the total sample while the remainder were reserved for the test data set. Because of the emphasis on patient enrollment in a Care Management program, similar inclusion criteria were adopted from the Rank Algorithm that included living adults over age 18,

patients not already enrolled in a care management program, uninsured patients or those covered by Intermountain Healthcare's insurance arm, SelectHealth, and patients living within 30 miles of a care management clinic. SelectHealth customers and the uninsured were included as a group of patients for which Intermountain Healthcare has assumed financial risk.

Health care costs for the study excluded chemotherapy, dialysis, intravenous (IV) therapy, spinal fusion, and knee and hip replacement. However, patients with these procedures could still be included if they had significant health care costs in other areas. It was determined that these conditions could not be impacted by the interventions provided by care management teams. Key predictors used in logistic regression modeling included age with gender and marital status derived from patient records. Socioeconomic factors included Average Household Income in the patient ZIP code based on the 2010 U.S. Census and the Area Deprivation Index (ADI) score in the patient Census block.[14] Dummy variables were used for ADI values greater than 115. Supplementary indicators were used for behavioral health conditions, additional comorbidities including obstructive sleep apnea, morbid obesity, coronary artery disease, hyperlipidemia, hypertension, and the count of Charlson Comorbidities.[15,16] Charlson Comorbidities and behavioral health conditions included in the analysis are shown in Table 1. Summary statistics on the training sample are included in Table 2. All analyses were performed using R software for statistical modeling and computing.[17]

Table 1. Charlson Comorbidities and Behavioral Health Conditions Included In Logistic Regression Modeling

CHARLSON COMORBIDITIES	BEHAVIORAL HEALTH CONDITIONS
Myocardial Infarction	Schizophrenic Disorders
Cancer	Depression Disorders
Connective Tissue Disease-Rheumatic Disease	Bipolar Disorders
Chronic Pulmonary Disease	Affective Disorders
Cerebrovascular Disease	Organic Psychotic Conditions
Metastatic Carcinoma	Nonorganic Psychoses
Dementia	Neurotic Disorders
Moderate or Severe Liver Disease	Personality Disorders
Diabetes with complications	Alcohol/Drug Dependence
Diabetes without complications	Eating Disorders
Mild Liver Disease	Childhood/Adolescence Disorders
Peripheral Vascular Disease	Intellectual Disability
AIDS/HIV	
Peptic Ulcer Disease	
Congestive Heart Failure	
Renal Disease	
Paraplegia and Hemiplegia	

Volume 5 (2017) Issue Number 2

Table 2. Summary Statistics of Training Sample

VARIABLE	PERCENT	VARIABLE	MEAN (SD)
Percent Female	66.34%	Comorbidity Count	1.29 (1.26)
Percent White	92.48%	Age	46.8 (14.9)
Percent Married	71.21%	Mean ZIP Income	$63,311 ($16,372)
Percent with Hypertension	40.35%	Prior Year Health Care Costs	$13,213 ($16,711)
Percent with Obesity	30.22%		
Percent with Behavioral Health Condition	55.95%		
Percent with Area Deprivation Index > 115	10.57%		

Findings

The results presented here were aimed at predicting the likelihood of a patient being in the highest 15th percentile for costs in two of the next three years for patients already in the top 15th percentile in the last year as described in the Logistic Model. Many of these metrics for the number of identified patients are somewhat arbitrary. If we identify the likelihood of being a high-cost patient in the future, there could be a range of optimal likelihoods that could be used. For example, determining the likelihood of a 50 percent chance of being in the top 15 percent of costs in the next two years would result in a larger patient cohort compared to those with a 95 percent chance. Due to constraints of the CCM teams, the number of manageable patients was estimated to be about 2,000. These results reflect an optimal match between the predicted likelihood of patient targeting and the number of patients with whom CCM teams might intervene. We report the results of targeting patients with likelihood score greater than 0.85 based on the fitted population being in the top 15th percentile of high-cost patients in two of the next three years. The original Rank Algorithm utilized by CCM predicted 63 percent of patients from the prior year would remain in the top 15th cost percentile for one of the next three years. Using logistic regression and additional sociodemographic covariates, the Logistic Model increased the predicted likelihood from 63 to 79 percent. Additionally, the Logistic Model demonstrated increases in the predicted likelihood of prescreening patients remaining in the top 15th percentile of cost for two of the next three years from 31 to 48 percent. The C-statistic, representing the "goodness of fit" of each model, also increased from .54 under the ranking model to .71 using logistic regression. Estimates indicate the patient cohort overlap to be less than 10 percent between the two models. Additional results of patient targeting methods are presented in Table 3. The Logistic Model shows gains in identifying medically complex patients, namely among those with additional chronic comorbidities, behavioral health conditions, obesity, and hypertension.

Alternative validation analyses were also conducted using decision tree methods including Classification and Regression Tree (CART) and Random Forest methodologies. CART is built on logical if-then

Table 3. Results of Patient Targeting Methods

MEASURE	RANK ALGORITHM	LOGISTIC MODEL
Average, SD Prior Year Cost	$38,700 ($27,256)	$44,000 ($61,125)
Average, SD Number of Charlson Comorbidities	3.6 (1.9)	5.0 (2.4)
Average, SD Number of Behavioral Health Conditions	1.7 (1.3)	2.2 (1.84)
Average, SD Number of Other Comorbidities	1.4 (.98)	2.3 (1.2)
Percent of Patients with Area Deprivation Index > 115 (Top Quintile)	16.9	18.0
Percent of Patients Diagnosed With Behavioral Health Condition	63.2	82.8
Percent of Patients Diagnosed With Obesity	27.8	54.9
Percent of Patients Diagnosed With Hypertension	59.3	80.3

conditions that partition data based on different predictors. Predictions in CART are based on stratifying the predictor space into regions and making predictions based on the mean of the total observations in each region. Random Forest methodology utilizes bootstrapping to stabilize the pathways of possible alternative outcomes. For this test, the number of bootstrap iterations was 500. Both methods are considered alternatives to regression methods in tuning variable importance and selection used in predictive modeling. However, neither test was found to increase the C-statistic of outcome prediction.

Major Themes

In the search for sustainable health care, many health care systems are turning to data for help in understanding the health of their population. The approaches used here demonstrate gains in identifying the patients most likely to benefit from patient intervention programs. The Logistic Model described above relies primarily on patient demographics, including the socioeconomic context of the patient and patient health care cost in the last year, to predict the future likelihood of being a high-cost patient in two of the next three years. We claim that the use of widely available patient demographic information in combination with rudimentary clinical data may be more predictive of high-cost patients beyond alternative ranking methods such as the Rank Algorithm, which rely on lengthy accumulated cost history and third-party clinical risk-adjustment indices.

Because of the cyclical nature of care episodes, many high-cost patients will have decreasing health care spending over time. As episodes resolve, there is significant "regression to the mean" that occurs within this patient population. Consequently it becomes increasingly important to identify the subset of the population that is likely to remain high cost in the future. The Rank Algorithm relied too heavily on past cost and was not designed to effectively predict future health care spending beyond relying on past trends. Since the Logistic Model has been implemented, CCM clinic staff have become more efficient in selecting the right patients,

Volume 5 (2017) Issue Number 2

which has resulted in a reduced overall burden of vetting patients.

Additionally, the gains from a regression-based patient targeting model provide the advantage that engagement with future high-risk patients could occur in multiple ways. For example, patient outreach could happen at the point of care, in proactive outreach settings such as the CCM case setting described above, or by delivery systems or payers with access to the necessary data used in the statistical modeling itself. These data are relatively common to most electronic medical record systems and reduce the data requirements from three years to one year of retrospective patient history. Using one year of data to make predictions is beneficial because it allows systems to more accurately target the segment of the at-risk population most likely to benefit from additional services and support. More precise allocation of services can reduce waste and improve access to care, which is particularly valuable throughout the population health transition many health systems are currently facing. Conversely, in the era of "big data" there may be common acceptance of the assumption that more data is better to use in predicting overall health outcomes. In this instance, health systems struggling to make use of emergent data systems need not feel overwhelmed by a lack of large or highly fine-tuned data systems. Our Logistic Model was developed on relatively few predictors on open-sourced software. Furthermore, we found, at least for the time being, that regression tree methods that rely on large data sets were less effective in obtaining greater modeling accuracy than traditional regression methods.

This study has several limitations. First, we claim to have increased the ability to target high-cost patients by using predictive methods over a rudimentary ranking system in the pursuit of reducing health care costs and improving patient outcomes. We do not claim to show that predictive methods can account for all these changes. Because the study relied upon retrospective data for the use of future cost prediction, we merely speak to methodological updates in patient identification and leave additional research to quantify how much downstream interventions may be able to reduce costs. Second, this modeling may not account for all the health conditions that may cause patients to be high risk in the future. The approach shown here represents a parsimonious prediction strategy, having compared multiple predictor variables and methods. Due to its parsimony, the Logistic Model may prove to be a useful starting point for alternative health care systems to engage in their own high-cost patient targeting intervention strategies. However, data training and testing was performed on a sample of patients with relatively homogenous demographics living in the intermountain western United States. This sample may represent a patient population with inherently different risk factors and health care needs than patients in another geographic location. While the Logistic method was not explicitly tested against the IndiGO or Optum indices directly, the lower performance of both indices combined as included in the Rank Algorithm did not warrant additional independent testing. The unanticipated finding surrounding the limited utility of third-party algorithms underscores the need for health systems pursuing population health initiatives to be sensitive to the unique characteristics of their population. In the present study, we found that third-party predictive algorithms trained on other populations were less helpful than training data on our own population.

Conclusion

Many strategies have been implemented in the search for health care delivery strategies that help patients manage illness and reduce waste. High-cost patient targeting can aid care management teams to effectively focus their efforts on those

in the most need of intervention. Compared to alternative modeling techniques, our Logistic Model, based on administrative and basic socioeconomic context data as well as information on chronic health conditions, increases the predictive ability to target at-risk patients. Using this model can shorten the time requirements to identify patients who are most likely to benefit from case management interventions, thus decreasing cost burdens to hospitals and patients alike. It is possible that this approach may prove helpful to other health care settings seeking to establish patient intervention programs of their own.

Acknowledgements

The authors would like to thank Andy Merrill, MS for his contributions.

References

1. Washington A, Coye MJ, Boulware L. Academic health systems' third curve: Population health improvement. JAMA. 2016; 315(5): p. 459-460.
2. Intermountain Healthcare. Shared Accountability and Population Health. Salt Lake City: Intermountain Healthcare, Decision Support; 2016.
3. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. Health Affairs. 2008 May; 27(3): p. 759-769.
4. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Diseases. 1987; 40(5): p. 373-383.
5. Eddy DM, Adler J, Morris M. The 'Global Outcomes Score': a quality measure, based on health outcomes, that compares current care to a target level of care. Health Affairs. 2012 Nov; 31(11): p. 2441-2450.
6. Dunn DL, Celebi D. System and method for assessing healthcare risks. 2015 Jun 16. US Patent 9,058,629.
7. Cucciare MA, O'Donohue W. Predicting future healthcare costs: how well does risk-adjustment work? Journal of health organization and management. 2006: p. 150-162.
8. Chechulin Y, Nazerian A, Rais S, Malikov K. Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). Health Policy. 2014 Feb; 9(3): p. 68-79.
9. Lieberthal RD. Analyzing the Health Care Cost Curve: A Case Study. Population Health Management. 2013 May; 16(5): p. 341-348.
10. Nyce S, Grossmeier J, Anderson DR, Terry PE, Kelley B. Association between changes in health risk status and changes in future health care costs: a multiemployer study. Journal of Occupational and Environmental Medicine. 2012 Nov: p. 1364-73.
11. Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keeffe Rosetti MC. Risk Adjustment Using Automated Ambulatory Pharmacy Data: The RxRisk Model. Medical Care. 2003; 41(1).
12. Otani K, Baden WW. Healthcare Cost and Predictive Factors: High- and Low-Utilization Model Development. Health Marketing Quarterly. 2009 Oct; 26(3): p. 198-208.
13. Moturu ST, Johnson WG, Liu H. Predictive risk modelling for forecasting high-cost patients: a real-world application using Medicaid data. Int. J. Biomedical Engineering and Technology. 2010: p. 114-132.
14. Singh GK. Area Deprivation and Widening Inequalities in US Mortality, 1969-1998. Am J Public Health. 2003 Jul; 93(7): p. 1137-1143.
15. Sundararajan V, Quan H, Halfon P, Fushimi K, Luthi JC, Burnand B, et al. Cross-National Comparative Performance of Three Versions of the ICD-10 Charlson Index. Medical Care. 2007; 45(12): p. 1210-1215.
16. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. Journal of Clinical Epidemiology. 2004 Dec; 57(12): p. 1288-1294.
17. R Core Team. R: A Language and Environment for Statistical Computing Vienna, Austria: R Foundation for Statistical Computing; 2015.