Patient Targeting and Outreach: A Statistical Approach

David Kartchner, Andy Merrill, Jonathan Wrathall
Population Health Analytics, Intermountain Healthcare

DESCRIPTION

According to the US Goverment Accountability Office, the top 5% of medical patients account for nearly 50% of the $3.3 trillion in annual healthcare spending. Identifying future high-cost patients allows healthcare organizations to take preventative measures to both reduce future patient costs and lessen the burden of illness. By using advanced machine learning techniques on readily available data, we develop a predictive model that substantially increases predictive accuracy while allowing providers to identify risky patients within three months of their first encounter.

CARE MANAGEMENT

Intermountain Healthcare is an integrated health system located in Utah and Southern Idaho and employs care managers to help patients with complex health problems.

Care managers seek to prevent avoidable utilization and slow the progression of chronic diseases through:

- Financial, nutritional, pharmaceutical, and mental health consultations
- Personalized care plans
- Disease-management education

Care management services are free to uninsured and SelectHealth patients, since these are patients for which Intermountain has assumed financial responsibility.

EVALUATION METRICS

Ranking helps care managers focus their limited resources on patients with the highest need. We evaluate our model's usefulness using a **retrospective cohort study** on 375,876 Intermountain patients over five years, where patient health in the last three years is predicted using patient health observed in the first two.

We evaluated logistic regression, random forest, and XGBoost models based on the **positive predictive value** (PPV) each produced when predicting between 2,000 and 3,000 patients.Using PPV as a metric gives our model high specificity while making it robust to patients with unpredictable high costs, e.g. because of an accident. Patients were deemed recurrently high cost if they were in the top 15% of costs in at least 2 of 3 follow-up years.

WHICH COSTS ARE PREVENTABLE?

Not all healthcare costs are preventable. Accordingly, we focus on preventable costs by categorically excluding patients with on chemotherapy, IV therapy, or dialysis, as well as patients who have had a knee or hip replacement, spinal fusion, or organ transplant. We additionally excluded maternity costs on an item-by-item basis.

Costs were calculated based on insurance claims for insured patients and on Intermountain medical costs for uninsured patients. These populations differed enough that we trained separate models for each.

VARIABLE SELECTION

We considered three criteria in selecting variables to use for cost prediction (see ***table 1*** for variables***)***:

- Variables should be easily calculable from an individual's Electronic Health Record (EHR) or insurance claims history
- Variables should focus on proven indicators of patient health to make the model more intuitive and improve physician uptake
- Variables should contribute meaningfully to model performance (evaluated using mean decrease in accuracy importance for each variable)

Table 1: Summary of Predictor Variables for SelectHealth Patients

Variable	Mean	St. Dev.
Prev. Year Monthly Cost	336.5	973.2
Prev. Year Cost Percentile	0.496	0.303
Charlson Comorbidities	0.621	1.077
Psychological Conditions	0.760	1.337
Other Comorbidities	0.541	0.904
Obesity Flag	0.126	0.332
Age	41.2	14.64
Historical Mean Monthly Cost*	281.8	690.4
Cost Standard Deviation*	528.1	1,393
Cost Coefficient of Variation*	1.983	1.243
% Months with Nonzero Cost*	0.479	0.340
% Months with Above Average Cost*	0.217	0.147
Cost Trend**	0.017	0.271

*Calculated over entire cost history rather than previous year
**Calculated as 1, -1, or insignificant

RESULTS

We compared our models to two baselines: a physician ranking model and a patient demographic model that omitted patient cost history (marked with a * or ** in ***table 1***). This comparison, given in ***table 2***, reveals that our model substantially outperforms past models, nearly doubling PPV while simultaneously increasing the number of identified patients. We report results for our XGBoost model, which achieved the best results in all tests.

Table 2: Comparison of Ranking, Demographic, and XGBoost Patient Targeting Models

Statistic	Ranking	Demographic	XGBoost
Top 15% in 1+ Years	63.0%	79.0%	96.1%
Top 15% in 2+ Years	31.0%	48.1%	87.7%
Identified Patients	1,400	1,800	2,800
C-Statistic	0.54	0.71	0.88

We also evaluated our model with and without insurance data for the 88,228 patients for whom it was available. This comparison, given in ***table 3***, shows that an individual's insurance cost history is much more consistent indicator of future utilization. We thus propose merging EHR and insurance data when possible to obtain better predictive results.

Table 3: Comparison of XGBoost Model with and without Insurance Cost Data

Statistic	Without Insurance Costs	With Insurance Costs
Top 15% in 1+ Years	82.5%	96.1%
Top 15% in 2+ Years	52.8%	87.7%
Identified Patients	2,000	2,800

REFERENCES

[1] M. A. Cucciare and W. O'Donohue, "Predicting future healthcare costs: how well does riskadjustment work?" Journal of Health Organization and Management, vol. 20, no. 2, pp. 150–162, 2006, pMID: 16869351. [Online].
[2] "Medicaid: A small share of enrollees consistently accounted for a large share of expenditures," no. 15-460. Government Accountability Office, May 2015.
[3] S. T. Moturu, W. G. Johnson, and H. Liu, "Predicting future highcost patients: A real-world risk modeling application," in 2007 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007), Nov 2007, pp. 202–208.
[4] M. E. Charlson, R. E. Charlson, J. C. Peterson, S. S. Marinopoulos, W. M. Briggs, and J. P. Hollenberg, "The charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients," Journal of Clinical Epidemiology, vol. 61, no. 12, pp. 1234 – 1240, 2008. [Online].
[5] L. E. Egede, D. Zheng, and K. Simpson, "Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes," Diabetes Care, vol. 25, no. 3, pp. 464–470, 2002. [Online].
[6] M. Charlson, P. Pompei, K. Ales, and C. MacKenzie, "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation," Journal of Chronic Diseases, vol. 40, pp. 373–383, 1987.
[7] D. Dunn and D. Celebi, "System and method for assessing healthcare risks," patent US 9 058 629, Jun 16, 2015.
[8] D. Eddy, J. Adler, and M. Morris, "The 'global outcomes score': a quality measure, based on health outcomes, that compares current care to a target level of care," Health Affairs, vol. 31, pp. 2441–2450, Nov 2012.
[9] J. A. Wrathall and T. Belnap, "Reducing healthcare costs through patient targeting: Risk adjustment modeling to predict patients remaining high-cost," eGEMs (Generating Evidence and Methods to improve patient outcomes), vol. 5, 2017. [Online].
[10] G. K. Singh and M. Siahpush, "Widening socioeconomic inequalities in us life expectancy, 1980–2000," International journal of epidemiology, vol. 35, no. 4, pp. 969–979, 2006.
[11] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," ArXiv e-prints, Mar. 2016.
[12] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online].
[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[14] J. A. Fleishman and J. W. Cohen, "Using information on clinical conditions to predict high-cost patients," Health Services Research, vol. 45, no. 2, pp. 532–552, 2010. [Online].
[15] D. W. Frost, S. Vembu, J. Wang, K. Tu, Q. Morris, and H. B. Abrams, "Using the electronic medical record to identify patients at high risk for frequent emergency department visits and high system costs," The American Journal of Medicine, vol. 130, no. 5, pp. 601.e17 – 601.e22, 2017. [Online].
[16] G. Simon, J. Ormel, M. VonKorff, and W. Barlow, "Health care costs associated with depressive and anxiety disorders in primary care," The American Journal of Psychiatry, vol. 152, no. 3, pp. 352–7, 03 1995. [Online].
[17] A. Hochlehnert, D. Niehoff, B. Wild, J. Junger, W. Herzog, and B. Lowe, "Psychiatric comorbidity in cardiovascular inpatients: Costs, net gain, and length of hospitalization," Journal of psychosomatic research, vol. 70, no. 2, pp. 135–139, 02 2011. [Online]. Available: https://search.proquest.com/docview/855899207?accountid=4488