Risk prediction for future 6-month healthcare resource utilization in Maine

Shiying Hao, Karl G. Sylvester, Xuefeng B. Ling*
Departments of Surgery
Stanford University
Stanford, CA, USA

Zhongkai Hu, Bo Jin, Chunqing Zhu, Dorothy Dai, Frank Stearns, Eric Widen
HBISolutions Inc.
Palo Alto, CA, USA

Andrew Young Shin
Departments of Pediatrics
Stanford University
Stanford, CA, USA

Devore S. Culver, Shaun T. Alfreds, Todd Rogow
HealthInfoNet
Portland, ME, USA

* Correspondence: Xuefeng B. Ling, Department of Surgery, Stanford University; S370 Grant Building, 300 Pasteur Drive, Stanford, CA 94305, USA; (office) 650-427-9198, (fax) 650-723-1154, (Email) bxling@stanford.edu.

***Abstract*— Understanding the future costs of the healthcare service utilization in patients can benefit the resource allocation management. The aim of this study is to develop a risk stratification model for healthcare resource utilization in future 6 months of patients in the state of Maine. A retrospective cohort of 1,273,114 patients was constructed to derive a decision-tree-based model to estimate the risk of resource utilization between January 1, 2013 and June 30, 2013, using the preceding 12-month clinical historical data. The model was validated with a prospective cohort of 1,358,153 patients by testing total costs between July 1, 2013 and December 31, 2013. Prospective results showed that the sensitivities of the model varied between 0.057 and 0.800, with confidence levels varying between 0.858 and 0.937 at all risk levels. Potential economic impacts of the model on healthcare resource utilization were explored. Future applications of our model will enable a more efficient resource allocation and targeted care intervention.**

Keywords—healthcare cost; risk model; prediction; prospective validation; economic impact

I. Introduction

The healthcare spending in the United States has been undergoing a dramatically rapid growth since 1980s [1]. The trend of the increasingly high spending in national health demands a comprehensive understanding of the cost-of-care strategies, which can be approached by analyzing the healthcare cost drivers and predicting the future costs. An effective prediction of future costs can benefit healthcare providers in both business and clinical plan makings.

A few studies have been developed to analyze the factors associated to future healthcare expenditures, for clinical or financial purposes [2-6]. However, many of those studies focused on the cost prediction of special patient groups [2,4,6], or lack prospective validation [7]. All these limitations prevent many of the existing cost predictive models from a broad utilization.

In this study, we set to develop a population-based predictive model to stratify the patients in the state of Maine into 3 subgroups, representing the different risk levels of resource utilization in future 6 months. All the patient information were gathered from health information exchanges (HIE) in the US, which contained clinical histories stored in forms of electric medical records (EMR) for more than 1 million patients, with attempts to cover the information of all payers, all diseases, all test results and all ages. The predictive algorithm was constructed by statistically learning the correlations between the 6-month total cost and the preceding 12-month demographic and clinical data, and was validated prospectively. Potential economic impacts of our model on resource utilization were discussed.

II. Methods

A. Ethics statements

This work was done under a business arrangement between HealthInfoNet (HIN) and HBI Solutions, Inc. and the data use is governed by the business agreement (BAA) between HIN and HBI. No protected health information was released for the purpose of research. Instead, HBI implemented their application that was the foundation for our agreement and then reported on the findings resulting from applying this model to the products/services that HIN is now deploying in the field.

B. Population

The study targeted to cover patients visiting any HIN connected facility from January 1, 2009 through December 31, 2013. To be qualified for the study, all the patients involved were alive and in Maine. Totally there were 1,273,114 patients in the retrospective cohort, and 1,358,153 patients in the prospective cohort.

978-1-4673-6799-8/15/$31.00 ©2015 IEEE

C. *Cost assignment*

In this study, the cost values used for analysis were estimated by the encounter types (Outpatient, Emergency Department and Inpatient visit) for each patient on a monthly basis, according to investigations on the historical data of encounter-based healthcare resource utilization in past few years [8,9].

D. *Cohort construction*

The statistical learning consisted of two phases: retrospective modeling and prospective validation. A respective cohort of 1,273,114 patients, with the clinical information between January 1, 2012 and December 31, 2012, was assembled to develop the model to predict the risk of the resource utilization between January 1, 2013 and June 30, 2013. This model was validated by a prospective cohort of 1,358,153 patients, carrying the clinical information between July 1, 2012 and June 30, 2013 to predict the resource utilization risks between July 1, 2013 to December 31, 2013.

E. *Model development and validation*

The retrospective modeling phase consisted of three steps: (1) training, (2) calibrating, and (3) blind testing. The retrospective cohort samples were randomly partitioned into sub-cohorts. Random forest methodology [10,11] was applied to construct decision trees to estimate the future 6-month resource utilization risks upon preceding 12-month clinical history. Specifically, a group of trees was grown using randomly selected samples and predictors (clinical features) of the training cohort. At each node, trees were split by choosing a split predictor value producing the maximum node separation [11]. Predicted risk scores were calculated by averaging the decisions of each tree. The total population was stratified into low, intermediate, and high risk levels according to the output of the predictive algorithm. In the calibration process, estimated maximum costs were assigned to each risk group, with a confidence level of 0.600. After calibration, the model's performance was blind tested, and then prospectively validated.

III. RESULTS

A. *Prospective performance of the model*

The proposed algorithm stratified the total population into 3 distinctive risk levels based on the output of the tree-based predictive algorithm. A one-to-one mapping between the risk level and the estimated maximum cost values in future 6 months was developed, enabling an estimation of the ranges that the future 6-month cost would probably fall in. The prospective results of sensitivity and confidence level as well as the estimated maximum costs in future 6 months at each risk level were summarized in Table I. The confidence levels remained at a fairly high level, fluctuating between 0.858 and 0.937 for all three risk levels. It illustrated that the predicted cost ranges had an acceptable accuracy for individual patients regardless of their risk levels. Given the estimated maximum cost values assigned to each risk level, the algorithm correctly identified 80.0%, 24.0%, and 5.7% of patients at low, intermediate, and high risk, respectively. Such impressive percentages of patients found by the algorithm demonstrated that our cost prediction model is capable of stratifying patients based on the healthcare services delivered to them in the future 6 months.

B. *Potential economic impacts*

By using the risk strategy in combination with patients having the chronic conditions, we examined the economic values of our model based on a quantity measure of potential cost savings. To evaluate the total savings caused by our model, we assumed that medical interventions would be delivered to those high-risk patients having chronic diagnoses in past 12 months, and that these medical interventions would reduce the future inpatient and emergency department admissions, which would result in cost savings. Penetration rate defined as the reachable proportion of patients, and effectiveness rate defined as the proportion of admissions that were effectively reduced by the intervention, were two variables determining the final saving values. In the prospective cohort, there were totally 74,538 patients with chronic disease history predicted as high risk. These patients totally had 19,522 inpatient visits and 39,067 emergency department visits in future 6 month. Based on historical data analysis in Maine [8,9] the average cost was $8000 per inpatient visit and $925 per emergency department visit, generating a total cost of $192.3 million. Assuming the penetration rate and the effectiveness rate were 0.60 and 0.25, the total savings per 6 month by using our model would be $28.8 million, giving $21.2 savings per person. The total savings per 6 month as a function of the penetration rate and the effectiveness rate is shown in Fig. 1.

The potential cost savings brought by the model were also investigated based on the categories of chronic diseases. Table II shows the prospective results of the future 6-month costs and savings as well as the patient and admission counts of 10 chronic diseases that had the largest saving amounts, with a penetration rate of 0.60 and an effectiveness rate of 0.25. The total savings among these 10 diseases ranged from $5.42 million to $15.5 million, while the total costs of the inpatient, outpatient and emergency department admissions were

TABLE I. PROSPECTIVE RESULTS OF OUR RISK MODEL PREDICTIVE OF FUTURE 6-MONTH COSTS

Prospective (July 1, 2012 – June 30, 2013)			
Performance	***Risk level***		
	Low	Intermediate	High
% of total population	69.2	24.7	6.0
Estimated maximum cost ($)	340	1,870	13,301
Confidence level[a]	0.858	0.901	0.937
Sensitivity[b]	0.800	0.240	0.057

[a]Confidence level is defined as the proportion of samples with future 6-month costs less than the estimated maximum cost.

[b]Sensitivity is defined as the ratio of the samples in a particular risk group to the total samples, all with future 6-month costs less than the estimated maximum cost.

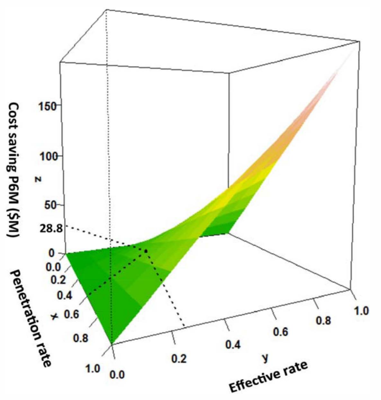

Fig. 1. Potential cost savings per 6 months in relation to penetration and effectiveness rates, resulting from medical intervention delivered to high-risk patients with chronic diseases. $28.8 million saving was achieved with 0.60 penetration rate and 0.25 effectiveness rate, as marked on the plot.

between $63.2 million and $177 million. Essential hypertension, disorders of lipid metabolism, and diabetes mellitus without complication are three chronic conditions having the top savings, due to the largest high-risk patient counts and considerable admission counts they accounted for.

IV. Discussion

Prospective experimental results demonstrated that our model has potentials of resource utilization forecasting. This feature highlights opportunities of measuring the expected healthcare spending in different scales using our model, which can offer economic benefits to care consumers, payers and providers altogether. Patients can get knowledge of their future expense ranges that offers guidance to their financial management in the light of the individual-based cost prediction. In the meanwhile, estimations of the future cost trends can improve the budgetary information for hospitals, clinics and insurance programmers, and give incentives to these healthcare providers and payers of developing effective business plans at a high level. More importantly, a potentially large amount of cost savings resulting from reduced care admissions is achievable, by applying our predictive model. Identification of high-risk population enables timely medical treatment to be delivered to a specific patient group that tends to demand for large care resources, which can improve those patients' health condition and thereby drop down their future hospital or clinical visits, resulting in savings of healthcare expense. Based on the clinical records in our database, around $28.8 million will be saved per 6 month among more than 1 million patients, with each person saving $21.2 on average. Given that all the data used in our study were collected from hospitals in Maine, even higher cost savings are expected by a national-wide application of our risk strategy algorithm. All these positive impacts on economics will make contributions to addressing the national issue of the increasingly high proportion of healthcare in federal budgets.

V. Conclusion

The total healthcare expenditures in the US are undergoing a dramatic increase and has placed a significant burden in the national economics, which requires an in-depth understanding of cost strategy. Our study derived a prospectively-validated statistic model to predict the resource utilization risks of future 6-month costs for each of more than 1 million patients in Maine. This model can provide a better understanding in the healthcare expense trends, and benefit the resource allocation management and targeted intervention delivery. It is an essential step of suppressing preventable costs while maintaining the healthcare quality.

TABLE II. ECONOMIC IMPACT OF THE PREDICTIVE MODEL ON FUTURE 6-MONTH CHRONIC-DISEASE-BASED HEALTHCARE UTILIZATION OF HIGH-RISK PATIENTS.

Chronic disease	Patient count	Economic impact on Inpatient care			Economic impact on emergency department care			Total cost[a] ($M)	Total savings ($M)
		Visit count	Cost ($M[b])	Saving ($M)	Visit count	Cost ($M)	Saving ($M)		
Essential hypertension	35,216	11,027	88.2	13.2	16,147	14.9	2.24	177	15.5
Disorders of lipid metabolism	32,454	9,744	78.0	11.7	13,077	12.1	1.81	155	13.5
Diabetes mellitus without complication	19,233	6,807	54.5	8.17	9,768	9.04	1.36	108	9.52
Esophageal disorders	17,269	6,116	48.9	7.34	10,902	10.1	1.51	93.1	8.85
Other nervous system disorders	15,859	5,834	46.7	7.00	12,016	11.1	1.67	92.2	8.67
Other nutritional, endocrine, and metabolic disorders	16,497	5,588	44.7	6.71	8,970	8.30	1.24	87.5	7.95
Cardiac dysrhythmias	13,173	5,505	44.0	6.61	6,047	5.59	0.839	83.8	7.45
Coronary atherosclerosis and other heart disease	13,088	5,899	47.2	7.08	6,416	5.93	0.890	83.0	7.97
Chronic obstructive pulmonary disease and bronchiectasis	11,007	5,372	43.0	6.45	6,795	6.29	0.943	71.2	7.39
Thyroid disorders	13,648	3,805	30.4	4.57	6,134	5.67	0.851	63.2	5.42

[a]Total cost includes costs on inpatient, outpatient and emergency department care.
[b]$M: million dollar

ACKNOWLEDGMENT

We thank and express our gratitude to the hospitals, medical practices, physicians and nurses participating in Maine's HIE. We also thank the biostatistics colleagues at the Department of Health Research and Policy of Stanford University for critical discussions.

REFERENCES

[1] (2009, 9). The unsustainable cost of health care. Social Security Advisory Board. [Online]. Available: http://www.ssab.gov/documents/TheUnsustainableCostofHealthCare_graphics.pdf

[2] L. Virta, P. Joranger, J. I. Brox, and R. Eriksson, "Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden," BMC Musculoskelet Disord, vol. 13, p. 17, 2012.

[3] S. L. Brilleman, S. Purdy, C. Salisbury, F. Windmeijer, H. Gravelle, and S. Hollinghurst, "Implications of comorbidity for primary care costs in the UK: a retrospective observational study," Br J Gen Pract, vol. 63, pp. e274-82, Apr 2013.

[4] G. C. Pope, J. Kautter, R. P. Ellis, A. S. Ash, J. Z. Ayanian, L. I. Lezzoni, et al., "Risk adjustment of Medicare capitation payments using the CMS-HCC model," Health Care Financ Rev, vol. 25, pp. 119-41, Summer 2004.

[5] A. S. Ash and R. P. Ellis, "Risk-adjusted payment and performance assessment for primary care," Med Care, vol. 50, pp. 643-53, Aug 2012.

[6] A. S. Ash, R. P. Ellis, G. C. Pope, J. Z. Ayanian, D. W. Bates, H. Burstin, et al., "Using diagnoses to describe populations and predict costs," Health Care Financ Rev., vol. 21, pp. 7-28, 2000.

[7] N. Mittmann, N. Liu, J. Porter, S. J. Seung, P. K. Isogai, R. Saskin, et al., "Utilization and costs of home care for patients with colorectal cancer: a population-based study," CMAJ OPEN, vol. 2, pp. E11-E17, 2014.

[8] A. Pfuntner, L. M. Wier, and C. Steiner. (2013, 1). Costs for hospital stays in the United States, 2010. [Online]. Available: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb146.pdf

[9] S. Machlin and S. Chowdhury. (2011, 3). Expenses and characteristics of physician visits in different ambulatory care settings, 2008. [Online]. Available: http://meps.ahrq.gov/data_files/publications/st318/stat318.shtml

[10] L. Breiman, "Bagging predictors," Machine Learning vol. 24, pp. 123-140, 1996.

[11] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, "Random survival forests," Ann. Appl. Statist, vol. 2, pp. 841-860, 2008.