Patient Identification for Telehealth Programs

Martha Ganser*, Sauptik Dhar[†1], Unmesh Kurup[†], Carlos Cunha[††], and Aca Gacic*.

*Robert Bosch Healthcare Systems, Inc., Palo Alto, CA
[†]Robert Bosch Research and Technology Center, Palo Alto, CA
[††]Robert Bosch Data Mining Services, Palo Alto, CA
[1] sauptik.dhar@us.bosch.com

Abstract— **Telehealth provides an opportunity to reduce healthcare costs through remote patient monitoring, but is not appropriate for all individuals. Our goal was to identify the patients for whom telehealth has the greatest impact. Challenges included the high variability of medical costs and the effect of selection bias on the cost difference between intervention patients and controls. Using Medicare claims data, we computed cost savings by comparing each telehealth patient to a group of control patients who had similar healthcare resource utilization. These estimates were then used to train a predictive model using logistic regression. Filtering the patients based on the model resulted in an average cost savings of $10K, an improvement over the current expected loss of $2K (without filtering).**

Keywords -- healthcare; telehealth; logistic regression.

I. Introduction

Telehealth can prevent costly healthcare interventions through continuous care with remote monitoring [1]. Patients engage daily by taking vital signs and learning disease self-management skills. Meanwhile, a care team monitors the patients' status daily, assessing risk and providing intervention when necessary. The data generated is stored in a data warehouse where it can be further linked to medical insurance claims that contain information on patients' healthcare utilization, diagnoses, procedures, and costs. Although telehealth enabled care management has proven financial and healthcare benefits, there are costs associated with distributing the monitoring devices and operating the care management staff. In order to maximize the benefits, patients may be prioritized according to the expected impact on health outcomes and costs.

Currently, patients are selected for telehealth programs using clinical groupers and statistical models. Clinical groupers provide methods of categorizing patients by the level of healthcare resource use and morbidity [2]. Statistical models have also been developed to predict patients who are high cost or at high risk of hospitalization [3]. Although both methods are useful for identifying high risk patients, not all high risk patients are a good fit for telehealth. For example, patients with end-stage renal disease (ESRD) are high risk but difficult to impact in a way that would reduce the cost for a payer. Conversely, patients with lower risk scores may benefit from preventative measures. Intervention-specific models will allow us to select the most appropriate patients for telehealth.

Our goal was to develop predictive models which support prioritization of patients for enrollment in the Health Buddy System (HBS), a telehealth program developed by Robert Bosch Healthcare. In an effort to capture the effect of the telehealth program directly, we were interested in predicting the individual cost savings for patients who were enrolled in HBS.

Evaluating cost savings from a non-randomized healthcare program requires consideration of *selection bias*, wherein program participants are different from non-participants with respect to some features, measureable or immeasurable, due to the nature of program enrollment[4]. In healthcare economics, cost savings are often evaluated on a population level using a difference-in-differences approach, propensity scores, instrumental variables, or a combination of these methods [4], which specifically address analytical challenges such as selection bias. However, most such studies lack a more detailed analysis of the cost-savings at an individual level. At the individual level, prediction is especially challenging due to the distribution of medical costs. Costs are skewed right, with high costs driven largely by inpatient hospitalizations[5]. Hospitalizations are relatively rare and can be an important indicator of disease acuity; however, they may also occur as a result of non condition related events such as physical injuries, adding noise to the model [5]. Most *first principle* based approaches, typically adopted in healthcare research, are not well-tailored to handle such complex dependencies [6]. This motivates the need to adopt a data-driven approach towards estimating a generalizable cost-savings model for patients at an individual-level.

Many machine learning methods have been applied previously to healthcare data, including the prediction of healthcare costs [7]. However, we found no literature pertaining to prediction of individual-level cost savings for a health intervention.

In this paper we estimate a data-driven cost-savings model for patients at an individual level, and use this model to identify the patients likely to save through the HBS telehealth program. The rest of the paper is organized as follows. Section II describes the data preparation and the experimental set-up. The modeling results of the cost-savings model are provided in section III. Finally the conclusions are presented in section IV.

II. Data Preparation and Modelling

A. Health Buddy Demonstration Study

All data came from the Care Management for High Cost Beneficiaries (CMHCB) demonstration study conducted by the Centers for Medicare and Medicaid Services (CMS) in conjunction with Robert Bosch Healthcare[8, 9]. The study consisted of 11,570 Medicare patients who were offered the Health Buddy to manage their chronic conditions over a period of three years (see http://innovation.cms.gov/Files/reports/CMHCB-HealthBuddyMontefiore.pdf). For this work we utilized only two years; *baseline year* as the year immediately preceding the start of the intervention, and *demonstration year* as the first year of intervention. Further, we removed patients with insufficient data, ESRD, and those whose death falls within six months from the end of the study. The current analysis includes 2805 cases who used the Health Buddy device at least once in the study period and 5092 controls who did not receive a Health Buddy device.

B. *Data Preprocessing*

Data came from two sources:

– *Telehealth Utilization Data* was used to determine program participation and compliance.
– *Administrative Claims Data* contained information on demographics, medical expenditures, and claim counts for baseline and demonstration year.

Claims also included medical diagnosis and procedure codes. We used the Johns Hopkins ACG® system [10] to process claims into meaningful features for medical conditions and healthcare resource utilization (HRU). Of particular importance was the patient grouping information based on morbidity and related health resource utilization patterns, called the resource utilization bands (RUBs), which ranged from healthy users with low HRU (RUB 0) to users with complex conditions with very high HRU (RUB 5).

We define *cost savings* as the difference between actual expenditures incurred by patients on HB program and expected expenditures of the same patients had they not been on the HB program. Actual expenditures of telehealth patients can be calculated directly from the claims data; however, expected expenditures need to be estimated from patients who had not participated in the telehealth program (control group), and who are similar to cases with respect to selected features (measured and not measured). As discussed before, the computation of cost savings presented two main challenges: (1) selection bias between cases and controls, and (2) high variability of cost among similar patients caused mostly by variability in hospitalizations and emergency room admissions.

We addressed these issues by comparing the costs of each HBS patient with the average cost within the group of controls who share the same RUB. We then built separate cost savings models for each RUB, viz., RUB 5, RUB 4 and RUB 3 (shown in Fig. 1). The assumption in this approach is that the selection bias is addressed by comparing patients with similar HRU needs and disease conditions. Furthermore, by aggregating the control patients we reduce the variability in the computation of the cost savings and in turn, the variability in model estimation.

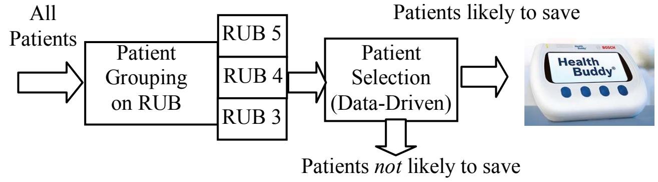


Fig.1. Schematic representation of the cost-savings' modeling workflow.

Equation (1) provides the difference-in-differences computation we used to estimate cost savings. For example, in the case of a Health Buddy patient in RUB 5 we obtain the cost saving for the i[th] patient as:

$$Cost\_Saving_i = \begin{cases} \text{yes} \ ; \ \text{if } (\overline{y}_2^{Control} - y_{i,2}^{HB}) - (\overline{y}_1^{Control} - y_{i,1}^{HB}) \geq 0 \\ \text{no} \ ; \ \text{else} \end{cases} \qquad (1)$$

where $\overline{y}_2^{Control}$, $\overline{y}_1^{Control}$ is the mean cost of control patients in RUB5 during the demonstration and the baseline year, respectively, and $y_{i,2}^{HB}$, $y_{i,1}^{HB}$ is the cost incurred by the i[th] HBS patient in RUB 5 during the demonstration and the baseline year, respectively.

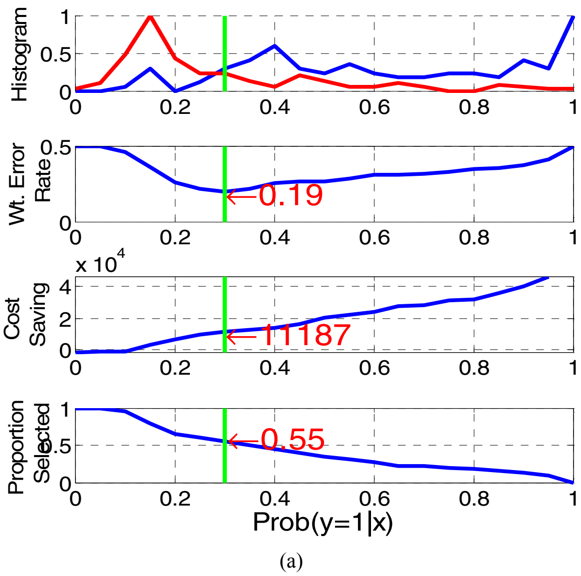
We build binary classifiers of "+1" and "-1" classes (i.e., the cost-savings model) separately for each RUB to identify the patients who are likely to save (see Fig.1). For the sake of brevity, the analysis in this paper is limited to the highest severity RUB 5. Additional results for RUB 3 and 4 will be provided in an extended version of the paper. After discussions with application domain experts, we selected 40 variables (listed in Table 1) for modeling. These variables capture the demographics, clinical, and historical claims information for each patient. Next, the data was uniformly scaled in the range of [0, 1]. The final dataset used for modeling is the result of the selection and preprocessing steps detailed above and contains the following:

– Number of patients who save (class '+1') = 476 (41%),
– Number of patients with loss (class '-1') = 679 (59%),
– Dimension of each sample = 40

C. *Modeling*

To build this classifier we tried several approaches including Decision Trees, L2-regularized Hinge Loss Support Vector Machine (SVM), L1-regularized L2- Loss SVM, L1-regularized Logistic Regression, and L2-regularized Logistic Regression (see [6] for more details). Since all of the methods provided similar classification accuracy, we settled on the L2-regularized Logistic Regression because it provides two main advantages:

– L2-regularization controls the model complexity and results in more generalizable models (i.e. models which can avoid over fitting). In addition, solving the L2-regularized Logistic Regression is more tractable in comparison to the other L1 based approaches.
– The logit loss provides a probabilistic output, which results in a more interpretable model in comparison to the SVM based approaches.



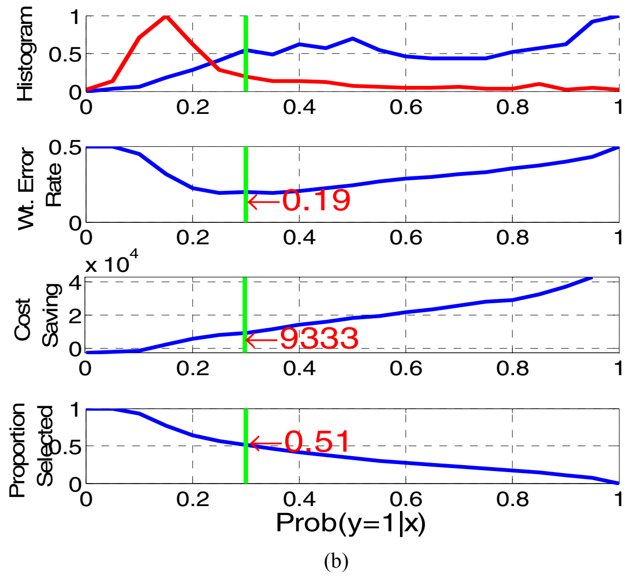


Fig.2. Model Performance with C=10, θ=0.3 for one random partition of the 5-fold split. (a) training data. (b) test data.

The L2-regularized Logistic Regression formulation is defined next [6]. Given input training data $(\mathbf{x}_i, y_i)_{i=1}^N$ with $\mathbf{x} \in \Re^D$ and $y \in \{-1,+1\}$, solve:

$$\min_{\mathbf{w},b} \quad \underbrace{\frac{1}{2}\|\mathbf{w}\|_2^2}_{\text{L2-Regularizer}} \quad + \quad \underbrace{\frac{C}{N}\sum_{i=1}^{N}\log(1+e^{-y_i(\mathbf{w}^T\mathbf{x}_i+\mathrm{b})})}_{\text{logit loss}} \quad (2)$$

where N represents the total number of training samples, D is the dimension of input samples, $C \geq 0$ and θ are user-defined parameters selected on the basis of application domain knowledge. The output is a probabilistic model with the final decision rule given as:

$$D(\mathbf{x}) = \begin{cases} +1, \text{ if } P(y=+1\,|\,\mathbf{x}) = \dfrac{1}{(1+e^{-(\mathbf{w}^T\mathbf{x}_i+b)})} \geq \theta \\ -1, \text{ else} \end{cases}$$

III. RESULTS

We provide the experimental results for (5,5) double resampling. Double resampling is a machine learning technique typically used to evaluate the predictive power of an algorithm (see [6] for more details). This approach involves a two-level partitioning of the data. The inner partition is used to perform model selection (i.e. selection of the optimal model parameters), in this case $C \geq 0$ and θ. The outer partition is used to test the predictive power of the selected optimal model. The resampling is done multiple times (while maintaining the same prior probabilities for each partition), and the average weighted error rate[6] over the several partitionings is reported in Table II together with the standard deviations. Fig. 2 provides the performance of the model for the parameters $C=10$ and θ =0.3 for one outer-partitioning of the resampling technique. The weighted error rate is similar between training and test sets, indicating a good model fit. Further, the estimated model with $C=10$ using all the data is provided in Table I. As seen from Table II (and also Fig. 2), through selection of patients on the basis of the estimated model with threshold θ=0.3, we can save ~\$10K by retaining at least 50% of the HB patients; in comparison to the current loss of ~\$2K.

Based on our analyses we can infer that such a data-driven filtering approach can prove useful in selecting at least 50% of the patients with reasonably high cost-savings. Finally, feature selection could shed more light on the effective variables to build a better predictive model. For example, we see that features such as inpatient payment, race, and carrier payment are important for selecting the patients who are likely to save, while features like number of home health agency claims, number of hospice claims, and hospice payments are the least predictive for building such a cost savings model. As a word of caution, it is not recommended to associate a causal behavior between cost-savings and the above features based on the estimated model. For such steps appropriate experiments proving causal relationship need to be set up.

IV. SUMMARY

The problem of estimating individual-level cost savings for health interventions is difficult due to the identification of an appropriate comparison group and

TABLE I. FEATURE WEIGHTS

FEATURES	w	FEATURES	w
Inpatient payments	5.53	Race = White	-0.4
Race = Asian	-4.93	East Phase II	-0.39
Carrier payment	4.58	CHF Diagnosis	0.38
Outpatient payment	2.79	Age over 65	-0.36
Race = N. Am. Native	-2.13	Elixhauser comorbidity	-0.33
HHA payment	1.98	Diabetes claim count	-0.32
DME count	1.95	Age under 65 (disabled)	-0.21
DME payments	-1.82	Female	0.17
Inpatient payments	1.65	Age	-0.12
Baseline year cost	1.53	West Phase I	-0.08
Total no. of claims	-1.34	Outpatient claim count	0.07
Race = Other	-1.16	AHRQ cormorbidity	-0.07
CHF claim count	-1.03	Diabetes diagnosis	-0.05
Charlson comorbidity	-0.96	West Phase I	-0.04
COPD claim count	-0.96	SNF payments	-0.04
costY1Diff	-0.72	Carrier claim count	0.034
Race = Black	-0.56	COPD diagnosis	0.03
SNF claim count	-0.53	HHA claim count	0
Race = Hispanic	-0.42	Hospice claim count	0
East Phase I	-0.42	Hospice payments	0

TABLE II. PERFORMANCE OF THE ESTIMATED MODEL OVER (5,5) DOUBLE RESAMPLING

Performance Metric	Test	Training
Wt. Error Rate (%)	20.87(3.04)	18.90(0.78)
Proportion of HB patients selected (in %)	51.91(2.68)	51.56(0.69)
Cost Saving (after selection)	\$10,200 (\$2,037)	\$10,359 (\$748)
Cost Saving (before selection)	-\$1,886 (\$851)	-\$2,324 (\$211)

adjustment for selection bias. Furthermore, the variabiliy of medical costs, driven largely by hospitalizations, added additional challenges. We computed ground truth by comparing Health Buddy patients with an aggregated group of controls with similar resource utilization and estimated separate logistic regression models for each RUB. Filtering the patients based on the predictive model resulted in an average cost saving ~$10K in comparison to the current incurred loss of ~$2K (without filtering). Our model used linear parameterization, which is easy to interpret and can be tuned to control the interplay between selected patient population size and targeted cost savings per customer's expectations.

It is important to note that associating causal relationship through the estimated predictive models should be done with caution [11]. This paper does not provide a causal-effect analysis relating the patient's demographics, clinical and historical claims information to the cost-savings. However, when this approach is complemented with domain knowledge, insights regarding the features necessary/unnecessary to prediction of cost savings can still be derived. For example, inpatient payments, which exemplify regression to the mean, could create an impression of higher cost-savings. Future research could exclude such features. Moreover, the current model does not incorporate the cost-values into the loss function of the model. Incorporating the cost-values could further help to identify patients with high vs. low cost savings. This could yield a better patient identification model which additionally weighs the patients in terms of the amount they save.

Future efforts may benefit from estimating cost savings for groups of patients rather than individuals, similar to the actuarial cell approach used by clinical groupers to predict resource use. This would further mitigate the negative effect of cost variation on individual patients' estimated cost savings. In addition, using a comparison group that was not offered the Health Buddy would further reduce the selection bias.

REFERENCES

[1] L. C. Baker, S. J. Johnson, D. Macaulay and H. Birnbaum, "Integrated Telehealth and Care Management Program for Medicare Beneficiaries with Chronic Disease Linked to Savings," *Health Affairs,* vol. 30, no. 9, pp. 1689-1697, 2011.

[2] S. Weir, G. Aweh and R. E. Clark, "Case Selection for a Medicaid Chronic Care Management Program," *Health Care Financing Review,* vol. 30, no. 1, pp. 61-74, 2008.

[3] 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.

[4] S. R. Khandker, G. B. Koolwal and H. A. Samad, Handbook on Impact Evaluation, Washington, D.C.: The World Bank, 2010.

[5] P. Diehr, D. Yanez, A. Ash, M. Hornbrook and D. Y. Lin, "Methods For Analyzing Health Care Utilization And Costs," *Annual Review of Public Health,* vol. 20, pp. 125-44, 1999.

[6] V. Cherkassky, Predictive Learning., VCtextbook , 2013.

[7] D. Bertsimas, M. V. Bjarnadottir, M. A. Kane, J. C. Kryder, R. Pandey, S. Vempala and G. Wang, "Algorithmic Prediction of Health-Care Costs," *Operations Research,* vol. 56, no. 6, pp. 1382-1392, 2008.

[8] C. Urato, N. McCall, J. Cromwell, N. Lenfestey, K. Smith and D. Raeder, "Evaluation of the Extended Medicare Care Management for High Cost Beneficiaries (CMHCB) Demonstration: Health Buddy Program at Montefiore," RTI International, Research Triangle Park, NC, 2013.

[9] C. Urato, N. McCall, J. Cromwell, N. Lenfestey, K. Smith and D. Raeder, "Evaluation of the Extended Medicare Care Management for High Cost Beneficiaries (CMHCB) Demonstration: Health Buddy West Program," RTI International, Research Triangle Park, NC, 2013.

[10] J. P. Weiner and C. Abrams, Eds., The Johns Hopkins ACG System Technical Reference Guide, Version 10.0, Baltimore, Maryland: Johns Hopkins Bloomberg School of Public Health, 2011.

[11] V. Cherkassky and S. Dhar, "Interpretation of Black-Box Predictive Models," in *Measures of Complexity: Festschrift for Alexey Chervonenkis*, 2016.