

Research Article

Identifying Future High Cost Individuals within an Intermediate Cost Population

Juan Lu, MD, MPH, PhD

Department of Family Medicine and Population Health, Division of Epidemiology, Virginia Commonwealth University

Erin Britton, MPH

Department of Family Medicine and Population Health, Division of Epidemiology, Virginia Commonwealth University

Jacquelyn Ferrance, MPH

Department of Family Medicine and Population Health, Division of Epidemiology, Virginia Commonwealth University

Emily Rice

Department of Family Medicine and Population Health, Division of Epidemiology, Virginia Commonwealth University

Anton Kuzel, MD, MHPE

Department of Family Medicine and Population Health, Division of Epidemiology, Virginia Commonwealth University

Alan Dow, MD, MSHA

Department of Internal Medicine, Virginia Commonwealth University

ABSTRACT

Background: Improving health and controlling healthcare costs requires better tools for predicting future health needs across populations. We sought to identify factors associated with transitioning of enrollees in an indigent care program from an intermediate cost segment to a high cost segment of this population.

Methods: We analyzed data from 9,624 enrollees of the Virginia Coordinated Care program between 2010 and 2013. Each fiscal year included all enrollees who were classified in an intermediate cost segment in the preceding year and also enrolled in the program in the following year. Using information from the preceding year, we built logistic regression models to identify the individuals in the top 10% of expenditures in the following year. The effect of demographics, count of chronic conditions, presence of the prevalent chronic conditions, and utilization indicators were evaluated and compared. Models were compared via the Bayesian

information criterion and c-statistic.

Results: The count of chronic conditions, diagnosis of congestive heart failure, and numbers of total hospital visits and prescriptions were significantly and independently associated with being in the future high cost segment. Overall, the model that included demographics and utilization indicators had a reasonable discrimination ($c=0.67$).

Conclusions: A simple model including demographics and health utilization indicators predicted high future costs. The count of chronic conditions and certain medical diagnoses added additional predictive value. With further validation, the approach could be used to identify high-risk individuals and target interventions that decrease utilization and improve health.

Key words: Primary Care, Health Care Cost, Chronic Diseases, and Administrative Data Uses

Introduction

To achieve the triple aim of improving health, decreasing costs, and enhancing patient experience, the healthcare system must move toward population-focused models of organizing care¹. A population-focused approach, supported by adequate information systems to help understand the population, allows policy makers and planners to define the needs of specific segments of the populations and tailor care to the needs of specific individuals. In contrast, care in the current system is often poorly responsive to the needs of both individuals and communities and configured largely based on the supply of providers and other resources rather than by the demands of the population². In particular, because a small fraction of the population drives healthcare costs and presumably has the worst level of health³, controlling the costs and improving the health of the highest cost group is a priority.

Understanding populations in order to configure care is a

nascent field. The Bridges to Health model proposed dividing the population into eight groups based on disease burden with categories such as 'good health', 'limited, acute illness', 'advanced organ system failure' and 'near death'⁴. Each of the categories was linked to quality goals based on the Institute of Medicine's six domains of quality⁵. This theoretical approach to population segmentation has been further validated with analyses of populations of older adults^{6,7}.

For our community, our institution developed a care coordination program for indigent adults, the Virginia Coordinated Care program (VCC), in 2000⁸. The goals of the program were to improve the health of this population and reduce utilization of higher cost services such as emergency department visits and hospitalizations. The program enrolled individuals who were uninsured and under 200% of the federal poverty level (PFL). Enrollees were provided free primary care through an assigned primary care provider, access to free specialist care and testing, and low-cost medications. Overall,

the program saved costs and improved health⁹. Within this population, several segments were defined based on care utilization, ranging from episodic care to high cost, frequent care¹⁰. Yet, how leaders and planners can use this data to structure population-centered care to meet the future needs of the community most effectively remains to be defined.

The goal should not only be to target the highest cost group but also aim interventions at preventing individuals from entering into this group in the first place. Predicting healthcare costs is an option to shape healthcare delivery programs and potentially improve health and control costs. Health cost prediction models often incorporate demographic information and information on clinical conditions based on data from medical records or claims databases⁵.

Clinical conditions have been entered into prediction models as the diagnostic cost group (DCG), prevalent chronic conditions, or counts of chronic conditions. DCG models were originally developed to match HMO payments to the healthcare needs of enrollees. The system uses patients' age, gender and medical diagnosis profiles to predict healthcare expenditures¹¹⁻¹⁷. While these methods have been broadly used in Medicare databases, they are also applicable to private insurance and Medicaid databases¹⁸. However, the system is limited by its requirement of special resources (e.g., commercial software and expertise), dependence on a common classification structure within ICD-9 codes, and accuracy of these codes.

As an alternative, the presence of certain chronic conditions can predict health expenditures in various settings¹⁹⁻²². A third approach to predicting future costs is to use the count of chronic conditions. Several studies suggested that a simple count of chronic conditions can predict the length of hospital stay and mortality²³ and amount of health expenditures^{19,22}. The study by Farley et al¹⁹ showed that not only the simple count of diagnose clusters but also the counts of prescriptions and physician visits were better predictors of future costs than the comorbidity measures (i.e. Charlson and Elixhauser indexes) with a count of diagnose clusters being the best predictor among all measurements examined. Similarly, Fleishman and Cohen²², using data from the national Medical Expenses Survey, compared the ability of the DCG method, counts of chronic conditions, and the presence of ten prevalent chronic indicators to predict the top ten percent medical expenditure. The results showed that the count of chronic conditions significantly predicted future high-cost, controlled for DCG category, demographic characteristics and self-reported functional status.

To advance our understanding of how population management principles should guide the structure of care delivery, we sought to determine what factors correlate with transitioning to the highest cost segment of the population within the VCC program¹⁰. Using demographic data, diagnosis information, and medication utilization of healthcare utilization, we describe the factors that are associated with the transition of individuals from an intermediate cost segment to a high cost segment of the population.

Methods

Data sources and participants

The data from the VCC Program for enrollment, utilization and claims between 2010 and 2013 were utilized for the study, with an annual average of 26,974 adults enrolled in the program. Utilization data included all services with VCU Health System including prescription data and affiliated primary care practices. Based on annual data, the program stratified enrollees into subgroups of episodic, chronic, complex and specialty care, based upon diagnosis, utilization and prescriptions¹⁰. For the purpose of developing a prediction model on future healthcare expenditures, we focused on the subgroup of 9,624 enrollees with stable chronic conditions and intermediate healthcare cost. The criteria for classifying this subgroup include 1) annual hospital spending between \$7001 and \$19,999; 2) having a minimum of six and maximum of 12 emergency room (ED) visits; or 3) six plus prescriptions within a fiscal year (e.g., July 1, 2010 through June 30, 2011). Individuals who exceed the upper limit for any category were considered in the complex (highest utilization) group. Likewise, individuals only needed to exceed one lower limit of the criteria to be placed in the intermediate cost group.

Key chronic conditions and number of chronic conditions

Information on medical diagnoses was derived from the primary or secondary diagnoses for each encounter along with the ICD-9-CM codes. We created binary indicators for the ten most prevalent chronic conditions: mental health problems, hypertension, diabetes, chronic obstructive pulmonary disease (COPD), coronary artery disease, mild liver disease, cancer, heart disease, cerebrovascular disease and congestive heart failure. To increase the sample size, we further combined the diagnoses of asthma with COPD, moderate and severe liver disease with mild liver disease, and myocardial infarction (MI) with coronary artery disease. Further, we calculated the number of chronic conditions for each individual. Due to small numbers, individuals with six or more conditions were collapsed into a single group.

Health utilization variables

We used hospital visits and number of prescription drugs as measures of health utilization. The hospital visits is an aggregate measure of inpatient, outpatient and ED visits. The number of hospital visits was treated as continuous with the exception of individuals with 15 or more visits who were collapsed into one group. The number of prescription drugs was also collapsed into categories: 0, 1-2, 3-5, and 6 or more.

Analysis

We analyzed aggregate datasets through three panels between 2010 and 2013. Each panel included all enrollees who were classified as the intermediate cost segment in the preceding year and remained enrolled in the program the following year. The enrollees who were classified as the specialty care in their following program year (i.e. their care was driven by a single dominant disease state such as polytrauma or poisoning) were excluded.

For each panel, we identified the individuals in the top 10% of expenditures in the following year. Based on the information collected in the preceding year, we used logistic regression to estimate the association between the potential risk factors and entering the top 10% of expenditures for the following year. The

baseline model includes age, quadratic term of age, gender, race ethnicity categories, and fiscal year. Subsequently, we added number of chronic conditions, the presence of the ten most prevalent chronic conditions, total hospital visits, and number of prescriptions. Combinations of these risk adjuster sets were further examined. As a sensitivity analysis, we repeated analyses using the top 5% as cut off for high expenditure cases. We compared models using the Bayesian information criterion (BIC)²⁴, with lowest values indicating a better model fit.

The performances of the models were assessed with respect to calibration and discrimination. Calibration is the ability of a model to produce unbiased estimates of the outcome probabilities, while discrimination is the ability of a model to separate high from non-high expenditure cases. We utilized Hosmer-Lameshow (H-L) goodness of fit test²⁵, which assesses agreement between the observed and predicted risks over the full range of predicted probabilities, to evaluate calibration. The H-L test specifically identifies subgroups as the deciles of predicted risks to perform the test. Models for which expected and observed event rates in subgroups are similar are called well calibrated. We used c-statistics to measure the discrimination. C-statistics range from 0.5 to 1, where 0.5 corresponds to random chance and 1 corresponds to perfect discrimination. SAS software (Version 9.4, SAS Institute Inc., Cary, NC, USA) was used for all analyses.

Results

We analyzed the data from 9,624 VCC enrollees between 2010 and 2013, including 2410, 2798 and 4416 from the panels one, two and three respectively. The unadjusted mean expenditures and standard deviations for the panels 1, 2 and 3 were \$1062.18 (\$1692.22), \$1248.57 (\$4821.10) and \$1985 (\$6172.00), respectively.

Descriptive analyses

Table 1a shows the characteristics of all panels in the preceding year and unadjusted total hospital costs in the following year. The majority of VCC enrollees were adults between 18-65 years old, and 55% of enrollees were female. Approximately 65% of enrollees were Black Non-Hispanics. About 16% enrollees did not have any chronic condition; 17% did not have any hospital visits, and 28% did not use any prescription drugs.

Table 1b presents the ten most prevalent chronic conditions identified from all panels. Among the 9,624 enrollees, the conditions were mental health problems (38.32%), hypertension (30.91%), diabetes (28.53%), COPD (17.64%), coronary artery disease/MI (9.05%), liver disease (5.09%), cancer (4.53%), heart disease (3.77%), cerebrovascular disease (2.24%) and congestive heart failure (2.07%), respectively.

Univariate comparisons

Compared with younger age individuals and males, older age individuals and females were significantly associated with transitioning to the highest cost segment. When compared with White Non-Hispanics individuals, being Black Non-Hispanics and other Non-Hispanics individuals were significantly less likely to transition to the highest cost group. Increased numbers

of chronic conditions, total hospital visits and prescriptions were associated with transitioning to the highest cost group.

Multivariate analyses

Table 2 displays the estimated associations between the potential predictors in the preceding year and being in the top 10% of total hospital expenditures in the following year from several logistic regression models.

Among the demographic characteristics examined, age was a consistent predictor of entering the highest utilization segment across all models. Compared with males, females were more likely to be in the top 10% of spenders. This finding was independent of age, race ethnicity (Model 1), numbers of chronic conditions (Model 2), prevalent individual chronic conditions (Model 3) and combination of models 2 and 3 (Model 5). However, adding health utilization indicators to models 4, 6, 7 and 8 eliminated any significant gender disparity. Compared with White Non-Hispanics individuals, Black Non-Hispanics and other Non-Hispanics were less likely to be in the top 10% of spenders through all models.

The odds of being in the top 10% of expenditure in the following year significantly increased as the numbers of chronic conditions increased. The association was independent of age, gender, and race ethnicity and health utilization indicators. However, in models including prevalent chronic conditions (Models 5 and 8), the significant association weakened or disappeared.

Several individual chronic conditions predicted the odds of being in the top 10% of expenditure in the following year. In Model 3, hypertension, coronary artery disease/MI, liver disease, cancer, heart disease, and congestive heart failure each showed significant association with higher odds of being in the top 10% of expenditure in the following year. However, in models including the number of chronic conditions (Models 5, 7 and 8), the significant association between these individual conditions and the outcome disappeared, with the exception of congestive heart failure.

The odds of being in the top 10% of expenditures significantly increased as the numbers of hospital visits and prescription drugs increased. The association was independent of age, gender and race ethnicity (Model 4), number of chronic conditions (Model 6), individual chronic conditions (Model 7) and all components combined (Model 8). Noticeably, the increase was non-linear with a notable jump in the odds for the category of 6 or more prescriptions.

Table 3 describes the goodness of fit of the models. All models except for Model 1 fit well, as indicated by non-significant H-L goodness of fit test. The performances of the models were further evaluated using the BIC and C-statistic. Adding the number of chronic conditions (Model 2), individual chronic conditions (Model 3) or health utilization indicators (Model 4) each markedly improved baseline model, as indicated with lower BIC values and improved c-statistics. Among models 2-4, Model 4 has the best model fitness (lowest BIC) and highest c-statistic ($c=0.67$). Comparing models 5, 6, and 7 to their nested models indicates preference for the parsimonious models. Similarly, comparing model 8 to the best-performing

Table 1a: Baseline Characteristics and Total Hospital Costs (N=9624).

	Year 1 Characteristics		Year 2 Costs	
	Total N (%)	Top 10% N (%)	Mean (SD)	
Age				
Less than 18 years	11 (0.11)	0 (0.00)	498.84 (744.61)	
18-39 years	3359 (34.90)	259 (7.71)	1202.73 (3715.85)	
40-64 years	6083 (63.21)	688 (11.31)	1728.24 (5612.99)	
65 years or older	171 (1.78)	16 (9.36)	1527.22 (4308.67)	
Gender				
Males	4304 (44.72)	402 (9.34)	1503.79 (3986.96)	
Females	5320 (55.28)	561 (10.55)	1584.41 (6044.02)	
Race Ethnicities				
White Non-Hispanic	2552 (26.52)	290 (11.36)	1631.69 (4443.22)	
Black Non-Hispanic	6242 (64.86)	606 (9.71)	1535.96 (5422.35)	
Other Non-Hispanic	619 (6.43)	41 (6.62)	1231.66 (3157.11)	
Hispanic Any Race	211 (2.19)	26 (12.32)	1448.07 (2803.25)	
Panels				
1	2410 (25.04)	241 (10.00)	1062.18 (1692.22)	
2	2798 (29.07)	280 (10.01)	1248.57 (4821.10)	
3	4416 (45.89)	442 (10.01)	1985.08 (6172.00)	
Number of chronic conditions				
0	1550 (16.11)	117 (7.55)	1329.02 (3746.16)	
1	3610 (37.51)	294 (8.14)	1204.92 (3712.21)	
2	2580 (26.81)	287 (11.12)	1529.53 (4107.65)	
3	1200 (12.47)	152 (12.67)	2078.50 (7827.28)	
4	443 (4.60)	65 (14.67)	2496.79 (6883.74)	
5	172 (1.79)	24 (13.95)	2878.90 (11653.92)	
6 or more	69 (0.72)	24 (34.78)	5334.97 (9859.44)	
Number of total hospital visits				
0	1610 (16.73)	77 (4.78)	800.18 (3456.45)	
1	1790 (18.60)	127 (7.09)	919.79 (3339.05)	
2	1517 (15.76)	114 (7.51)	1184.83 (3925.12)	
3	1135 (11.82)	88 (7.73)	1469.09 (6536.43)	
4	915 (9.51)	94 (10.27)	1599.79 (5276.82)	
5	667 (6.93)	86 (12.89)	1738.52 (3946.39)	
6	437 (4.54)	59 (13.50)	1791.06 (3494.12)	
7	354 (3.68)	48 (13.56)	2013.23 (4130.18)	
8	278 (2.89)	48 (17.27)	2345.21 (3614.48)	
9	207 (2.15)	36 (17.39)	2872.59 (5282.33)	
10	164 (1.70)	31 (18.90)	3412.59 (8090.18)	
11	126 (1.31)	27 (21.43)	3934.79 (10745.47)	
12	94 (0.98)	22 (23.40)	3888.88 (8829.41)	
13	69 (0.72)	21 (30.43)	6428.59 (17734.13)	
14	49 (0.51)	10 (20.41)	3835.23 (7415.95)	
15 or more	209 (2.17)	75 (35.89)	4773.52 (6715.34)	
Number of prescription drugs				
0	2738 (28.45)	175 (6.36)	1141.36 (4046.03)	
1 – 2	1531 (15.91)	134 (8.75)	1361.45 (5630.15)	
3 – 5	2321 (24.12)	222 (9.56)	1438.85 (4446.01)	
6 or more	3034 (31.53)	432 (14.24)	2066.74 (5777.53)	

Table 1b: Ten Most Prevalent Chronic Conditions and Total Hospital Costs (N=9624)..

Year 1 Conditions		Year 2 Costs	
Conditions (ICD-9 Codes)	Total N (%)	Top 10% N (%)	Mean (SD)
Mental health problems (290-319)			
Yes	3688 (38.32)	397 (10.76)	1682.47 (5275.71)
No	5936 (61.68)	566 (9.54)	1451.24 (1451.24)
Hypertension (401)			
Yes	2795 (30.91)	372 (12.50)	1932.86 (6757.48)
No	6649 (69.09)	592 (8.89)	1364.00 (3979.60)
Diabetes (250)			
Yes	2746 (28.53)	267 (9.72)	1481.24 (5722.61)
No	6878 (71.47)	696 (10.12)	1563.24 (4698.99)
COPD/Asthma (490/496)			
Yes	1698 (17.64)	186 (15.04)	1787.24 (6806.19)
No	7926 (82.36)	832 (9.51)	1486.85 (4535.40)
Coronary artery disease/MI (414/410)			
Yes	871 (9.05)	131 (9.05)	8536.88 (2632.19)
No	8753 (90.95)	8753 (90.95)	1431.15 (4499.74)
Liver disease (140-239)			
Yes	490 (5.09)	79 (16.12)	2208.66 (5498.76)
No	9134 (94.91)	884 (9.68)	1503.97 (4982.59)
Cancer (571)			
Yes	436 (4.53)	76 (10.01)	2895.62 (6148.59)
No	9188 (95.47)	887 (9.65)	1475.51 (4943.02)
Heart disease (429.2)			
Yes	363 (3.77)	63 (17.36)	2532.24 (6707.99)
No	9261 (96.23)	900 (9.72)	1500.95 (4930.34)
Cerebrovascular disease (434.91)			
Yes	216 (2.24)	30 (13.89)	1989.90 (4656.24)
No	9408 (97.76)	933 (9.92)	1529.51 (5019.88)
Congestive heart failure (428.0)			
Yes	199 (2.07)	48 (24.12)	2847.29 (6894.70)
No	9425 (97.93)	915 (9.71)	1512.24 (4961.60)

nested model, i.e., model 6, indicated the more complex model was not preferable. Final model (Model 4) thus included demographic information and health utilization indicators.

Sensitivity analyses

The top 5% of total hospital costs was used as a cut-point to test for relative model performance. The results were consistent to those obtained using the top 10% of total hospital costs. However, the three most complex models had gross lack of fit, as indicated by significant H-L Chi-Square test results (Appendix Table 4), likely due to the limited data size for such complex models.

Discussion

Using data from an indigent care program, we demonstrated that integrating information on medical diagnoses, the count of chronic conditions, utilization indicators in the preceding year and demographic data was associated with high healthcare expenditures in the following year. We further evaluated whether combinations of the risk sets would improve prediction. Overall, the model that included demographics and utilization indicators (hospital visits and number of prescriptions) had the greatest association with transitioning to the highest cost segment ($c=0.67$). Identifying these complex patients may help target clinical interventions

Table 2: Odds Ratios and Associated 95% Confidence Intervals of Being the Top 10% of the Total Hospital Costs in the Following Year (n=9624)

Characteristics	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Age	1.02 (1.01-1.02)**	1.01 (1.01-1.02)**	1.01 (1.01-1.02)**	1.01 (1.01-1.02)**	1.01 (1.01-1.02)**	1.01 (1.00-1.02)**	1.01 (1.00-1.02)**	1.01 (1.00-1.02)**
Gender								
Males	-	-	-	-	-	-	-	-
Females	1.17 (1.02-1.34)*	1.17 (1.02-1.34)*	1.20 (1.05-1.38)**	0.98 (0.86-1.13)	1.20 (1.04-1.37)*	0.99 (0.86-1.15)	1.02 (0.88-1.17)	1.02 (0.88-1.17)
Race ethnicities								
White Non-Hispanic	-	-	-	-	-	-	-	-
Black Non-Hispanic	0.81 (0.70-0.95)**	0.79 (0.68-0.92)**	0.81 (0.69-0.94)**	0.83 (0.71-0.96)*	0.80 (0.69-0.94)**	0.82 (0.70-0.95)*	0.81 (0.69-0.95)*	0.81 (0.70-0.95)*
Other Non-Hispanic	0.54 (0.38-0.77)**	0.59 (0.42-0.84)**	0.60 (0.42-0.86)**	0.53 (0.37-0.75)**	0.60 (0.42-0.86)**	0.55 (0.39-0.78)**	0.55 (0.38-0.78)**	0.55 (0.38-0.78)**
Hispanic Any Race	1.10 (0.72-1.69)	1.14 (0.74-1.76)	1.15 (0.74-1.77)	0.94 (0.60-1.46)	1.14 (0.74-1.77)	0.95 (0.61-1.49)	0.95 (0.61-1.49)	0.95 (0.61-1.49)
Panels								
2012 - 2013	-	-	-	-	-	-	-	-
2011 - 2012	1.00 (0.85-1.19)	0.99 (0.83-1.16)	1.02 (0.86-1.21)	0.99 (0.84-1.18)	1.02 (0.86-1.21)	1.00 (0.84-1.18)	1.00 (0.84-1.18)	1.00 (0.84-1.18)
2010 - 2011	1.00 (0.85-1.17)	0.99 (0.85-1.17)	1.04 (0.89-1.22)	1.00 (0.85-1.18)	1.04 (0.89-1.22)	1.01 (0.86-1.19)	1.02 (0.87-1.20)	1.02 (0.87-1.20)
Number of chronic conditions (year 1)		1.22 (1.16-1.29)**			1.11 (0.93-1.33)	1.08 (1.02-1.14)**		0.98 (0.82-1.18)
Top 10 chronic conditions (year 1)								
Mental health problems			1.07 (0.93-1.24)		0.97 (0.77-1.22)		0.97 (0.84-1.13)	0.99 (0.79-1.25)
Hypertension			1.24 (1.07-1.44)**		1.12 (0.88-1.42)		1.04 (0.89-1.22)	1.07 (0.84-1.36)
Diabetes			0.95 (0.81-1.11)		0.86 (0.68-1.08)		0.98 (0.83-1.15)	0.99 (0.79-1.26)
COPD/ Asthma			1.11 (0.93-1.32)		0.94 (0.68-1.31)		1.03 (0.86-1.23)	1.06 (0.76-1.42)
Coronary artery disease/ MI			1.53 (1.25-1.89)**		1.38 (1.05-1.82)*		1.24 (1.00-1.53)	1.26 (0.96-1.66)
Liver disease			1.60 (1.24-1.89)**		1.42 (1.02-1.97)*		1.27 (0.98-1.66)	1.30 (0.93-1.82)

Cancer	1.77 (1.36-2.31)**	1.58 (1.14-2.19)**	1.17 (0.88-1.54)	1.19 (0.85-1.67)
Heart disease	1.42 (1.05-1.91)*	1.28 (0.90-1.81)	1.32 (0.98-1.79)	1.35 (0.95-1.66)
Cerebrovascular disease	1.10 (0.74-1.65)	0.98 (0.62-1.54)	0.92 (0.61-1.39)	0.94 (0.59-1.49)
Congestive heart failure	2.30 (1.63-3.26)**	2.07 (1.41-3.07)**	2.24 (1.56-3.20)**	2.28 (1.53-3.40)**
Total hospital visits (year 1)		1.14 (1.12-1.16)**	1.13 (1.11-1.15)**	1.13 (1.11-1.15)**
Number of prescription drugs (year1)				
0		-	-	-
1-2		1.29 (1.01-1.63)*	1.27 (1.00-1.62)*	1.32 (1.04-1.68)*
3-5		1.23 (0.99-1.52)	1.20 (0.97-1.49)	1.25 (1.01-1.55)*
6 or more		1.53 (1.25-1.86)**	1.51 (1.24-1.84)**	1.55 (1.27-1.90)**

* p<0.05; ** p<0.01

Model 1: age, age quadratic term, gender, race and panel indicator

Model 2: Model 1 + number of chronic conditions

Model 3: Model 1 + 10 most prevalent chronic conditions

Model 4: Model 1 + health utilization variables

Model 5: Model 1 + number of chronic conditions + 10 most prevalent chronic conditions

Model 6: Model 1 + number of chronic conditions + health utilization variables

Model 7: Model 1 + 10 most prevalent chronic conditions + health utilization variables

Model 8: Model 1 + number of chronic conditions + 10 most prevalent chronic conditions + health utilization variables

Table 3: Goodness of Fit Indications of the Logistic Regression Models.

Indicators	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
H-L test (χ^2 p-value)	21.35 (0.006)	11.96 (0.15)	5.07 (0.75)	9.53 (0.30)	6.08 (0.64)	6.60 (0.58)	8.24 (0.41)	8.68 (0.37)
C-statistic	0.58	0.61	0.62	0.67	0.63	0.68	0.68	0.68
BIC	6270.07	6223.53	6259.80	6014.46	6267.67	6016.29	6072.39	6081.51
Change in BIC	-	-46.54	-10.27	-255.61	+44.14	+1.83	+57.93	+65.22
Comparison Model	-	1	1	1	2	4	4	6

Model 1: age, age quadratic term, gender, race and panel indicator

Model 2: Model 1 + number of chronic conditions

Model 3: Model 1 + 10 most prevalent chronic conditions

Model 4: Model 1 + health utilization variables

Model 5: Model 1 + number of chronic conditions + 10 most prevalent chronic conditions

Model 6: Model 1 + number of chronic conditions + health utilization variables

Model 7: Model 1 + 10 most prevalent chronic conditions + health utilization variables

Model 8: Model 1 + number of chronic conditions + 10 most prevalent chronic conditions + health utilization variables

that improve health and reduce costs. Further inclusion of the count of chronic conditions or presence of ten prevalent chronic conditions in the model only slightly improved the discrimination aspect ($c=0.68$).

The criteria most predictive in these models could also help structure future interventions. With regard to individual risk factors, the count of chronic conditions and having a diagnosis of congestive heart failure were significantly associated with high future costs, regardless of demographics and utilization indicators. Identifying patients with multiple chronic conditions or heart failure before they develop a pattern of high healthcare utilization may be opportunities to intervene for the benefit of the patient and system. In addition, because both utilization indicators (hospital visits and number of prescriptions) were significantly and independently associated with high cost in the future, the complex patients with a higher risk for progression to a pattern of high utilization can be defined several ways. Leaders have some flexibility in their approach to identifying high-risk patients based on the data available within their information systems.

Our study, by systematically comparing the ability of the count of chronic conditions, the ten most prevalent chronic indicators, counts of prescriptions, and counts of total hospital visits to predict future high-costs, had results consistent with other studies^{19,22}. Our prediction models, however, combined clinical and administrative data without the complexity of DCG scores. Hence, the results derived from these models may be easier to implement. Moreover, our study focused on an uninsured adult population whose income was under 200% FPL. Our results may provide unique insights relevant to the newly insured populations under Medicaid expansion.

Noticeably, the model discrimination presented by this study is lower ($c=0.68$) than the results presented by other studies (c ranged from 0.81-0.85)^{22,26,27}. This difference may be a consequence of several factors. Our model included the counts of chronic conditions and presence of individual chronic conditions as the main predictors of high costs. These indicators are sensitive to information on severity of the conditions. For example, a study by Omachi and colleagues²⁸ showed that adding COPD severity measures significantly improved predictions of medical costs in the following year among a cohort of patients with COPD. Adding further information regarding the severity of each condition may increase the predictive power of the model. Additionally, the overall prevalence of chronic conditions were relative low in our dataset, likely due to the population being younger (mean age=43.3 years, standard deviation=12.6 years). Leaders and researchers should select the model that best fits the demographics of the population of interest.

This study has several limitations. Our dataset may not have contained all data on utilization or diagnoses for each patient. In addition, similar to the cost prediction results drawn from other administrative databases, we are unable to adjust the heterogeneities of disease conditions in the cost prediction model. Finally, the patient population included in this study is medically underserved and they are relatively younger and healthier, thus the results may not be generalizable to other patient population segments, such as those more affluent or individuals with better access to care, and those with older

age and sicker conditions. Further model validation studies are needed to confirm our findings.

In conclusion, the results from this study show that a simple model including demographics and health utilization indicators is associated with high future costs. Specific medical diagnoses, such as congestive heart failure, and the count of chronic conditions were also associated with higher future utilization. Further validation of the models is recommended to confirm the predictive capacity of this approach. If confirmed, this approach could be used to identify high-risk individuals and target interventions that decrease utilization and improve health.

Acknowledgement

The study was approved by the institutional review board and supported by CTSA award No. UL1TR000058 from the National Center for Advancing Translational Sciences. The authors report no declarations of interest.

References

- Berwick DM, Nolan TW and Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)* 2008; 27:759-769.
- Lillrank P, Groop PJ and Malmstrom TJ. Demand and supply-based operating modes--a framework for analyzing health care service production. *Milbank Q.* 2010; 88:595-615.
- Cohen S, Yu W. The concentration and persistence in the level of health expenditures over time: Estimates from the U. S. Population, 2008-2009. . 2012.
- Lynn J, Straube BM, Bell KM, et al. Using population segmentation to provide better health care for all: the "Bridges to Health" model. *Milbank Q.* 2007; 85:185-208; discussion 209-12.
- Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. 2001; Washington: National Academies Press.
- Eissens van der Laan MR, van Offenbeek MA, Broekhuis H and Slaets JP. A person-centred segmentation study in elderly care: towards efficient demand-driven care. *Soc Sci Med.* 2014; 113: 68-76.
- Zhou YY, Wong W, Li H. Improving care for older adults: a model to segment the senior population. *Perm J.* 2014; 18:18-21.
- Retchin SM, Garland SL and Anum EA. The transfer of uninsured patients from academic to community primary care settings. *Am J Manag Care.* 2009;15: 245-252.
- Bradley CJ, Gandhi SO, Neumark D, et al. Lessons for coverage expansion: a Virginia primary care program for the uninsured reduced utilization and cut costs. *Health Aff (Millwood).* 2012; 31: 350-359.
- Dow AW, Bohannon A, Garland S, et al. The effects of expanding primary care access for the uninsured: implications for the health care workforce under health reform. *Acad Med.* 2013; 88:1855-1861.
- Ash A, Porell F and Gruenberg L. An Analysis of Alternative

- AAPCC Models Using Data from the Continuous Medicare History Sample. Final Report to the Health Care Financing Administration. 1986. Boston: Health Policy Research Consortium. Brandeis/Boston Universities.
12. Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A. Adjusting Medicare capitation payments using prior hospitalization data. *Health Care Financ Rev.* 1989; 10:17-29.
 13. Ellis RP, Ash A. Refinements to the Diagnostic Cost Group (DCG) model. *Inquiry.* 1995; 32: 418-429.
 14. Ellis RP, Pope GC, Iezzoni LI. Diagnostic Cost Group (DCG) and Hierarchical Coexisting Conditions and Procedures (HCCP) Models for Medicare Risk Adjustment. Final Report to the Health Care Financing Administration. Contract No. 500-92-0020. . April 1996.
 15. Pope GC, Ellis RP and Liu CF. Revised Diagnostic Cost Group (DCG)/Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment. Final Report to the Health Care Financing Administration under Contract Number 500-95-048. February 1998.
 16. Pope GC, Liu CF, Ellis RP. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Final Report to the Health Care Financing Administration. . February 1999; 25:111-163.
 17. Pope GC, Ellis RP, Ash AS, et al. Principal inpatient diagnostic cost group model for Medicare risk adjustment. *Health Care Financ Rev.* 2000;21:93-118.
 18. Ash AS, Ellis RP, Pope GC, et al. Using diagnoses to describe populations and predict costs. *Health Care Financ Rev.* 2000; 21:7-28.
 19. Farley JF, Harley CR, Devine JW. A comparison of comorbidity measurements to predict healthcare expenditures. *Am J Manag Care* 2006;12:110-119.
 20. Baser O, Palmer L, Stephenson J. The estimation power of alternative comorbidity indices. *Value Health* 2008;11:946-955.
 21. Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. *J Clin Epidemiol* 2008; 61:1234-1240.
 22. Fleishman JA, Cohen JW. Using information on clinical conditions to predict high-cost patients. *Health Serv Res* 2010; 45:532-552.
 23. Melfi C, Holleman E, Arthur D, Katz B. Selecting a patient characteristics index for the prediction of medical outcomes using administrative claims data. *J Clin Epidemiol* 1995; 48:917-926.
 24. Raftery AE. Bayesian Model Selection in Social Research *Sociological Methodology* 1995; 25:111-163.
 25. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York: Wiley. ISBN 0-471-61553-6; 2000.
 26. Meenan RT, Goodman MJ, Fishman PA, Hornbrook MC, O'Keefe-Rosetti MC, Bachman DJ. Using risk-adjustment models to identify high-cost risks. *Med Care* 2003; 41:1301-1312.
 27. DeSalvo KB, Jones TM, Peabody J, et al. Health care expenditure prediction with a single item, self-rated health measure. *Med Care* 2009; 47:440-447.
 28. Omachi TA, Gregorich SE, Eisner MD, et al. Risk adjustment for health care financing in chronic disease: what are we missing by failing to account for disease severity? *Med Care* 2013; 51:740-747.

ADDRESS FOR CORRESPONDENCE

Juan Lu MD; MPH, PhD, Department of Family Medicine and Population Health, Division of Epidemiology, Virginia Commonwealth University, Richmond, VA, U.S.A., Tel: 804 8289786; Fax: 804 8289773; e-mail address: juan.lu@vcuhealth.org