Healthcare Costs and Utilization By Continuity of Care During the COVID-19 Pandemic: Exploring How Model Choices and Fit Impact Interpretation

EE491

Authors: Cory Pack[1], Jennifer Cheng[1], Taylor Marlin[1], Jessamine Winer-Jones[1], Natalia Coenen[1], Alina Bogdanov[1], Lee Kallenbach[1],
Affiliations: 1 Veradigm, Chicago, IL, USA

Introduction

- Healthcare costs can be difficult to accurately model because the data is highly skewed and a large proportion of values are zero.[1]
- One option is to use a two-part cost model that accounts for likelihood of non-zero costs and the non-gaussian distribution of costs.[2,3]
- Regardless of model choice, selection of appropriate covariates is critical to model stability and validity.[4]

Objective

- To explore methods for improving the fit of multivariable cost modeling using the healthcare costs of patients with different utilization patterns before, during, and after the early COVID-19 pandemic.

Methods

Data Source

- The Veradigm Network EHR linked to healthcare claims data spanning March 1, 2019–February 28, 2022

Time Periods

- This study used 3 time periods to identify patients
  - Y2019: March 1, 2019–February 29, 2020
  - Y2020: March 1, 2020–February 28, 2021
  - Y2021: March 1, 2021–February 28, 2022

Figure 1: Patient Selection

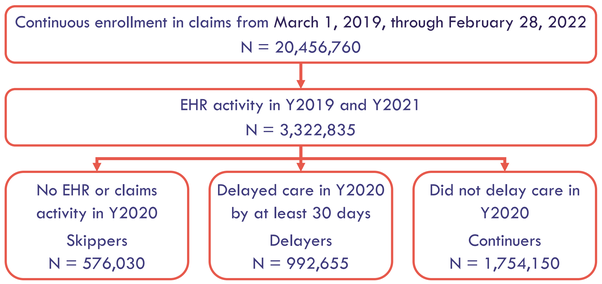
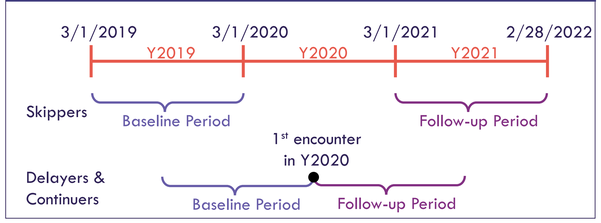


Figure 2: Study Design



Covariates

- Demographics: Age, sex, race, ethnicity, geographic region,
  - Age was measured on 3/1/2020 for skippers and at 1st encounter for delayers and continuers
- Clinical: Charlson Comorbidity Index (CCI) in baseline and follow-up and immunocompromised status
- Baseline healthcare costs

Outcome

- Healthcare costs in the follow-up period

Table 1: Model Covariates

	1	2	3	4	5	6
Utilization Cohort	Y	Y	Y	Y	Y	Y
Demographics†	Y	Y	Y	Y	Y	Y
Baseline Costs (categorical)	Y	Y	Y	Y	Y	Y
Follow-up CCI (numeric)	Y	Y	Y	Y	Y	Y
ΔCCI (numeric)	Y	–	–	–	–	–
ΔCCI (categorical)	–	Y	Y	Y	Y	Y
Cohort*Age	–	–	Y	–	Y	Y
Cohort*Baseline Costs	–	–	–	Y	Y	Y
Immunocompromised	–	–	–	–	–	Y

†Age, region, sex, race, and ethnicity

Table 2: Cohort Characteristics

	Skippers	Delayers	Continuers
	N = 576,030	N = 992,655	N = 1,754,150
Age, (Mean, SD)	35.5 (22.5)	38.1 (24.0)	42.2 (23.3)
Sex (N,%)			
Female	334,566 (58.1%)	593,421 (59.8%)	1,058,668 (60.4%)
Male	241,428 (41.9%)	399,140 (40.2%)	695,361 (39.6%)
Unknown/Not Reported	36 (0.0%)	94 (0.0%)	121 (0.0%)
Race* (N,%)			
White	282,442 (49.0%)	524,249 (52.8%)	965,823 (55.1%)
Black	72,691 (12.6%)	117,535 (11.8%)	212,483 (12.1%)
Asian	32,726 (5.7%)	58,722 (5.9%)	94,952 (5.4%)
Other	101,469 (17.6%)	170,247 (17.2%)	305,853 (17.4%)
Unknown	86,702 (15.1%)	121,902 (12.3%)	175,039 (10.0%)
Ethnicity (N,%)			
Hispanic	49,868 (8.7%)	92,337 (9.3%)	155,077 (8.8%)
Non-Hispanic/Not Reported	526,162 (91.3%)	900,318 (90.7%)	1,599,073 (91.2%)
Geographic Region (N,%)			
Northeast	78,813 (13.7%)	147,784 (14.9%)	248,607 (14.2%)
Midwest	134,223 (23.3%)	213,009 (21.5%)	359,312 (20.5%)
South	164,504 (28.6%)	300,535 (30.3%)	541,350 (30.9%)
West	181,024 (31.4%)	303,285 (30.6%)	557,418 (31.8%)
Other/Unknown	17,466 (3.0%)	28,042 (2.8%)	47,463 (2.7%)
Baseline CCI (Mean, SD)	0.55 (1.20)	0.62 (1.28)	0.87 (1.47)
Change in CCI (Mean, SD)	0.04 (1.06)	0.09 (0.98)	0.01 (1.07)
Change in CCI			
Increase	72,992 (12.7%)	139,243 (14.0%)	240,580 (13.7%)
Same	434,220 (75.4%)	757,086 (76.3%)	1,275,269 (72.7%)
Decrease	68,818 (11.9%)	96,326 (9.7%)	238,301 (13.6%)
Immunocompromised (N, %)	29,191 (5.1%)	63,714 (6.4%)	150,295 (8.6%)

Methods, Cont.

Data Cleaning

- Individuals with the top 1% of costs in the baseline period, those with the top 1% of costs in the follow-up, and those with missing age (N = 6) were excluded from the models

Statistical Methods

- We evaluated twelve models of healthcare costs in the follow-up period.
  - The 6 one-part models used only a generalized linear model (GLM) with a log-link and gamma-distribution of follow-up costs
  - The 6 two-part models included a logistic regression (LogR) model of having non-zero follow-up costs and a GLM of follow-up costs
- Primary predictor: utilization cohort (skipper, delayer, continuer)
- Covariates (Table 3)
  - All models included age, region, sex, race, ethnicity, baseline costs, and CCI in the follow-up period
    - All variables were categorical except CCI in the follow-up period
    - Age categories: 0 – 18, 19 – 34, 35 – 44, 45 – 54, 55 – 64, 65+
    - Cost categories: 0 – 9.9th percentile ($0 costs), 9.9th – 25th, 25th – 50th , 50th – 75th , 75th – 99th , 99th – 100th
  - Models varied on incorporation of additional covariates including change in CCI (continuous or categorical), immunocompromised status, and interaction terms
  - ΔCCI categories: increase, decrease, stay the same
- Approach
  - Models were trained on 70% of the data and tested on the remaining 30% of the data
  - Model performance was assessed with Akaike information criterion (AIC) and root mean squared error (RMSE)
  - Reported average marginal effect (AME) relative to continuers

Figure 3: Unadjusted Healthcare Costs

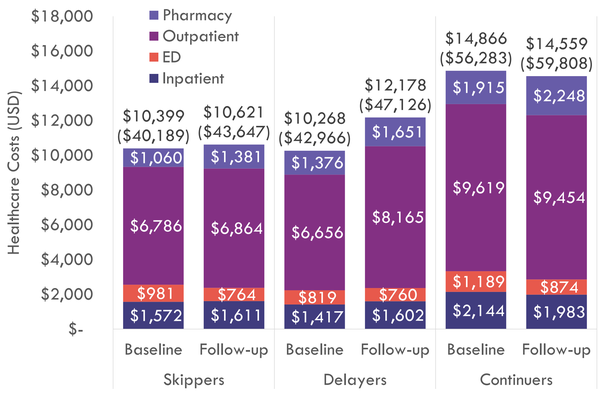
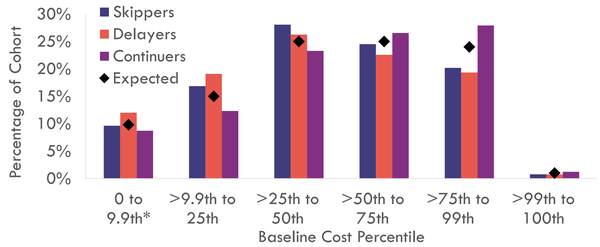


Figure 4: Observed versus Expected Cost Distributions by Cohort

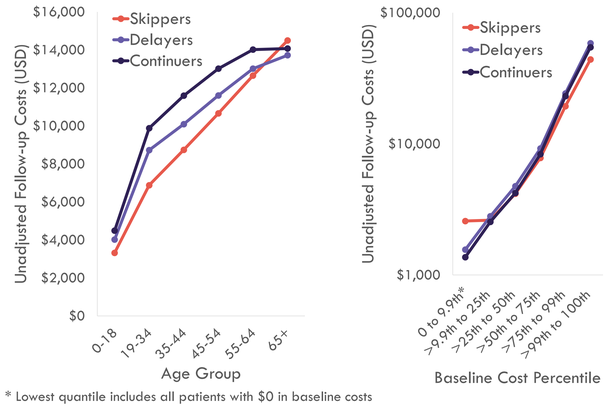


* Lowest quantile includes all patients with $0 in baseline costs

Descriptive Results

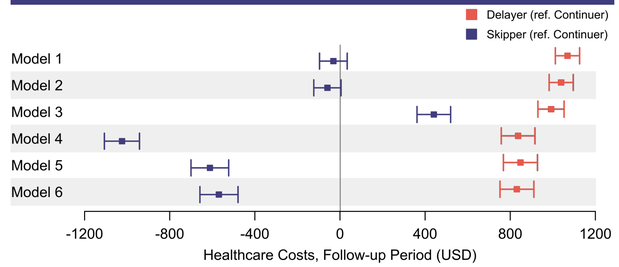
- The initial study population included 576,030 individuals who skipped care in the first year of the COVID-19 pandemic, 992,655 who delayed care, and 1,754,150 who continued care as expected (Figure 1).
- Skippers were younger and had a lower CCI than delayers or continuers (Table 2).
- Mean (SD) unadjusted healthcare costs at baseline were lower among skippers and delayers compared to continuers (Figure 3).
  - Between baseline and follow-up, mean unadjusted costs increased $222 among skippers and $1,910 among delayers, while costs decreased by $307 among continuers.
- After excluding those with the top 1% of baseline costs, a higher-than-expected proportion of continuers had costs in the 75th to 99th percentile (Figure 4).
- Plots of unadjusted follow-up costs suggested potential interactions between utilization cohort and age and between utilization cohort and baseline costs.

Figure 5: Key Covariate Interaction Plots



* Lowest quantile includes all patients with $0 in baseline costs

Figure 6: Model Predicted Marginal Costs



Modeling Results

- The two-part model always outperformed the one-part model in RMSE (Table 3)
- Treating ΔCCI as a categorical variable improved model performance
- Adding additional covariates generally improved model performance except for the addition of immunocompromised status
- After adjusting for covariates, delayers had higher costs than continuers in all models with the AME ranging from $831 to $1,069 (Figure 6)
- Estimates of the AME of skipping ranged from -$1,024 to $441 (Figure 6)
  - Two models estimated no effect on costs, one model estimated that skippers had higher costs than continuers, and the three best fitting models estimated that skippers had lower costs than continuers.
- In the best performing model (model 5) mean adjusted costs of delayers was $848 ($768 to $928) higher than continuers and mean adjusted costs of skippers was -$611 (-$700 to -$523).

Table 3 : Model Performance*

	AIC		RMSE	
	LogR	GLM	1 Part	2 Part
Model 1	1,319,650	39,250,565	24,518	17,234
Model 2	1,316,682	39,250,301	23,118	15,978
Model 3	1,315,398	39,248,709	23,128	15,987
Model 4	1,315,575	39,233,802	23,165	15,973
Model 5	1,311,881	39,244,986	23,094	15,941
Model 6	1,311,711	39,231,018	23,326	16,055

* For each column orange indicates worse performance and green indicates better performance.

Conclusions

- Use of a two-part model consistently improved overall model fit
- Size and direction of effect size estimates depended highly on choice of covariates
- Models consistently predicted that patients who delayed care during the COVID-19 pandemic had higher adjusted costs than those who continued care as usual.

References

1. Berk ML, Monheit AC. Health Affairs. 2001;20(2):9-18.
2. Deb P, Norton EC. Annu Rev Public Health. 2018;39:489-505.
3. Belotti F, et al. Stata J. 2015;15(1):3-20
4. Heinze G, et al. Biom J. 2018;60(3):431-449.

Disclosures

This study was funded by Veradigm.

All authors are employees of Veradigm which provided the data used in the execution of this study