

# Patient Copayment Prediction

covermy  
meds  
empowering patients



Mandy Cheung, Craig Franze, Shirali Obul, Charles Ruggiero  
[https://github.com/ErdoS-Red-Eye/CMM\\_Patient\\_Expenses/](https://github.com/ErdoS-Red-Eye/CMM_Patient_Expenses/)

# Problem Statement

- The primary goal of this project is to predict a patient copay for a prescribed medication given basic information including drug name and insurance number.
- Recent research suggest that this is a difficult problem facing physicians and the lack of accurate information impedes informed conversations about financial trade-offs between physicians and patients.<sup>1</sup>

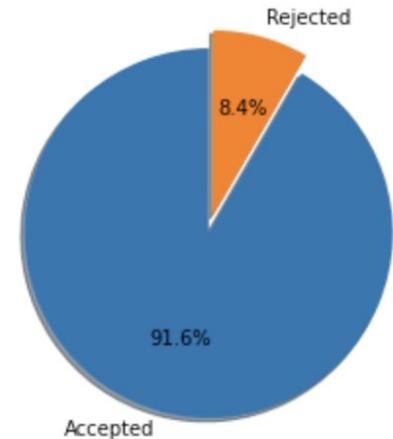
<sup>1</sup>Sloan C.E., Millo L., Gutterman S., BA, Accuracy of Physician Estimates of Out-of-Pocket Costs for Medication Filling, JAMA Netw Open. 2021;4(11):e2133188. doi:10.1001/jamanetworkopen.2021.33188

# Data

- The data set consists of approximately 14 million simulated pharmacy transactions for the year 2022.

(13910244, 9)

	tx_date	pharmacy	diagnosis	drug	bin	pcn	group	rejected	patient_pay
0	2022-01-02	Pharmacy #6	G99.93	branded tanoclolol	725700	1UQC	NaN	False	13.39
1	2022-01-02	Pharmacy #42	U60.52	branded oxasoted	664344	NaN	52H8KH0F83K	False	7.02
2	2022-01-02	Pharmacy #37	Q85.91	branded cupitelol	725700	1UQC	NaN	False	13.39
3	2022-01-02	Pharmacy #30	U60.52	generic oxasoted	571569	KB38N	6BYJBW	False	10.84
4	2022-01-02	Pharmacy #18	N55.01	branded mamate	664344	NaN	ZX2QUWR	False	47.00
...	...	...	...	...	...	...	...	...	...
13910239	2022-12-30	Pharmacy #42	U27.71	branded colifunene	322463	NaN	HO8HUGL	True	0.00
13910240	2022-12-30	Pharmacy #45	N59.44	generic tafistitrisin	664344	NaN	TFZOR5R49	False	6.28
13910241	2022-12-30	Pharmacy #54	W50.87	generic tanoclolol	691847	N098KI	6SP1DG	False	6.94
13910242	2022-12-30	Pharmacy #0	I68.27	branded prazinib	96934	S76J7V6	NaN	False	13.93
13910243	2022-12-30	Pharmacy #46	G99.93	branded bovirol	322463	3071UTS	NaN	False	12.22



# Data

*tx\_date* – The date on which the pharmacy transaction was attempted

*pharmacy* – The particular pharmacy where the transaction was attempted

*diagnosis* – The diagnosis of the patient associated with the transaction

*drug* – The drug that the patient was prescribed that the pharmacy is attempting to bill

*bin* – The broadest identifier of a patient’s insurance plan (banking identification number)

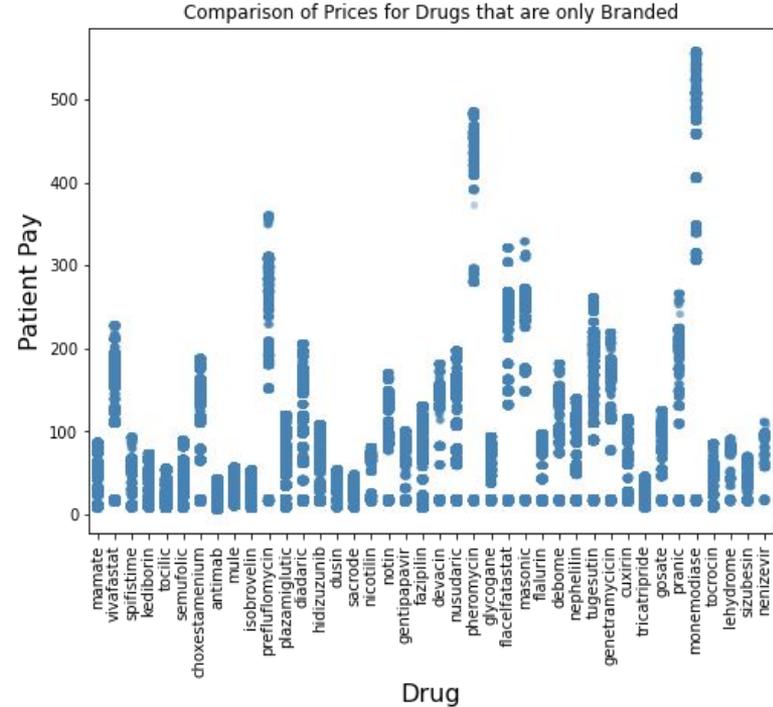
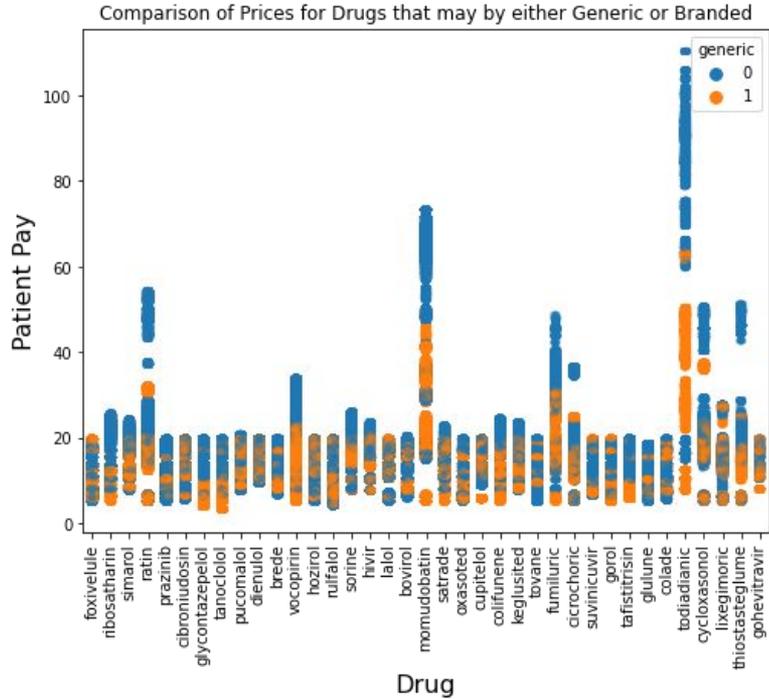
*pcn* – An identifier that more narrowly specifies a plan underneath the broader “bin”

*group* – Another identifier that more narrowly specifies a plan underneath the broader “bin”

*rejected* – Whether the billing transaction was rejected by the plan

*patient\_pay* – The amount of copayment for which the patient is responsible

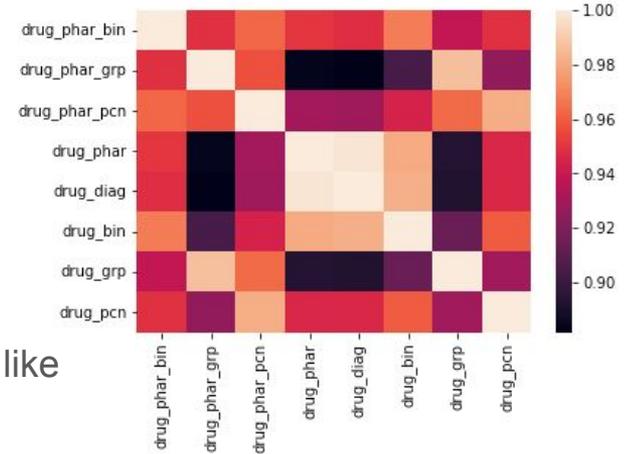
# Exploratory Data Analysis



# Challenge

The central challenge of this project is predicting a continuous variable (patient pay) with only categorical variables.

1. Feature engineering with *target encoding*
  - a. Engineer 8 features by aggregating over various columns.
  - b. Allows for impute claim entries when some of the features like drug, group, pcn or pharmacy are missing.
  - c. Creates a highly correlated feature set.
2. *Ordinal encoding* with regressors that natively support categorical variables.
  - a. Minimizes feature engineering
  - b. Limits the possible regressors that may be used.



## Models (CV Scores)

<b>Model</b>	<b>Categorical Variable Encoding</b>	<b>RMSE</b>	<b>Mean Absolute Percentage Error</b>
Gradient Boost	Feature aggregate with target encoding	11.989	3.917
Random Forest	Feature aggregate with target encoding	11.982	3.873
Histogram-based Gradient Boosting	Ordinal encoding ( <i>month, bin, pcn, group, drug</i> )	2.157	1.410

# Applications

A web application (prototype) allows for the prediction of drug prices.

CoverMyMeds Copayment Predictor

# covermymeds covermymeds

Prediction of patient copayment  
ahead of time by machine learning.  
How accurate can it be?

Using a set of (SIMULATED) pharmacy data-billing claims that were run from a pharmacy to a third-party payer (insurance plan) who covers some portion of the prescription drug price on behalf of a patient. As part of the claim process the amount that the payer reimburses the pharmacy and copayments required of the patient are set by complicated negotiations and contracts between the drug manufacturer, the payer, and the pharmacy. Those negotiations also often cover decisions on what drug claims will ultimately be approved (preferred / non-preferred / non-covered formulary status of each drug) based on the relative discounts that the payer can secure relative to other drugs in the same class that may treat the similar types of medical conditions. Therefore, using this claim billing data into machine learning algorithms to build a method of predicting the copayments required of patients ahead of time will be useful for doctors to prescribe drugs based on patient affordability.

Features in the model

Enter Bin Encoded

664344

Enter Pcn Encoded

KBOSN

Enter Group Encoded

DGLGRYP

Enter Drug Encoded

branded colifunene

Enter Month

1

Submit

Model Result

Predicted price

Filter Search in the dataset

Enter Month

1

Enter Pharmacy

Pharmacy #47

Enter Diagnosis

K32.86

Enter Drug

branded colifunene

Enter Bin

664344

Enter Pcn

KBOSN

Enter Group

DGLGRYP

Date	Month	Pharmacy	Diagnosis	Drug	Bin	Pcn	Group	Rejected	Patient_pay
1/3/22	1	Pharmacy #46	K32.86	generic cibriludodin	725700	9C5MOR3	S2QKZ0OFNWS6X	False	19.54
1/3/22	1	Pharmacy #18	K32.86	generic rulfalol	96934	S76J7V6		False	6.17
1/4/22	1	Pharmacy #10	K32.86	branded tanoclool	691847	N098KI	6SP1DG	False	15.55
1/5/22	1	Pharmacy #7	K32.86	generic lalol	539437		1CAHL	False	15.47
1/5/22	1	Pharmacy #22	K32.86	generic rulfalol	691847	N098KI	6SP1DG	False	6.94
1/6/22	1	Pharmacy #40	K32.86	branded lalol	725700	327CKV	IOEAN1DWVV3Y	False	11.15
1/6/22	1	Pharmacy #31	K32.86	branded lalol	664344		STGRDKRJ5RD	False	10.77
1/6/22	1	Pharmacy #19	K32.86	branded tafistitrisin	757349	RMQHB	SJVO3GXUURGO	False	10.9
1/6/22	1	Pharmacy	K32.86	branded	725700	327CKV	IOEAN1DWVV3Y	False	11.15

# Applications

Given a diagnosis, e.g. 'B45.03', along with the month, bin, pcn, and group numbers the models can be used to make a financial comparison of drugs that have been prescribed over the past year for the diagnosis.

*month= 5, bin=725700, pcn=1UQC, group=NaN*

```
1 option_cost('B45.03', 5, 725700, '1UQC', np.nan)
```

	<b>drugs</b>	<b>predicted_pay</b>
<b>2</b>	generic todadianic	49.365954
<b>4</b>	branded todadianic	92.863260
<b>0</b>	branded pranic	221.182812
<b>1</b>	branded masonic	271.147075
<b>3</b>	branded monemodiase	541.926317

## Future Work

- Build separate models for the generic and branded drugs.
- Examine the outliers where the models perform poorly and see if there is a particular commonality amongst them that could explain this.
- Combine engineered features, such as the ones used in the Random Forest model, or rejection rate for example, with the HistGradientBoosting model.
- Develop the app further by integrating these improvements to help doctors easily obtain predicted prices of medications ahead of time.