

Executive Summary: Cover My Meds

A. Problem Definition:

For most patients, filling a prescription at a pharmacy is akin to visiting a roulette table. The patient's responsibility, copay, seems based more on chance than on logic. In actuality, copays are established through complicated negotiations and contracts between the drug manufacturers, third party payers, and the pharmacies. To rectify this problem, efforts were undertaken to develop models to first predict the amount copay and, secondly, to determine the drug's formulary status (preferred / non-preferred / non-covered) based on the simulated pharmacy claims data provided by CoverMyMeds.

B. Exploratory Data Analysis

An investigation was performed to determine which of the provided identifiers should be included in the modeling process, to assess balance, to discover missing fields, outliers, and other anomalies. The three included predictive variables were diagnosis, pcn, and drug. Four attributes were excluded: tx_date, although reporting higher copays January and February due to annual plan deductible limits, resulted in only small elevation in means compared to the remainder of the year; bin and group, since they were unique to pcn, an included predictor; pharmacy, which showed little variance. Patient-pay, the regression response variable was right skewed, ranging from \$0-\$558, with a mean of \$26 and a standard deviation of approximately \$40. Rejected, although not included in the response model since all copays were \$0, was used as a classifier for fitting the formulary status model.

C. Regression and Classification Models:

To predict copayments, we used linear regression, K-nearest neighbors, regression trees, and a random forest. Evaluating these models with the RMSE, we found the regression tree and random forest to perform best and roughly equivalently. To predict the formulary status, we used logistic regression, naive Bayes, decision trees, and random forests. The decision trees performed best according to accuracy and F1 measures.

D. Results and Implications

The benefits to patients and health care providers are immense. When multiple drugs are effective for a given diagnosis, the provider could prescribe the drug with the lowest copay. If patients could predict their copay, they could request the less expensive alternative drugs and avoid outright rejection. Informed prescribers and consumers would increase efficiency in the marketplace, and result in downward pressure on drug prices.

E. Future Exploration:

This three-week investigation did not do justice to this important question. With additional time, the analysis could be expanded to include drug retail prices by web scraping public sources such as GoodRx. Expanded sampling of point of sales (supermarkets, mail-order, large chain vs. small private pharmacies) would probably increase the significance of the pharmacy attribute, and incentivize consumers to seek the lowest cost POS. Finally, with time, our app could be improved with more detailed information on alternative drugs options for patients.