



Erdős Bootcamp CoverMyMeds Project Statement

In this challenge we are presenting you with a set of (SIMULATED!) pharmacy claims data-billing claims that were run from a pharmacy to a third-party payer who covers some portion of the prescription drug price on behalf of a patient. Most commonly these third-party payers are the patient insurance plans, but other third parties such as manufacturer discount drug programs can also be run in the same manner as insurance claims to reimburse the pharmacy some portion of the cost. Ultimately, 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.

[As part of this project, we would ask you to build a method of predicting the copayments required of patients ahead of time using this claim billing data.](#) The ultimate initial goal of which would be to provide doctors information about the costs that their patient would expect to see in the pharmacy in case affordability of the medication was an issue. A secondary goal for the project would be to also provide information about the potential formulary status of the medication on each insurance plan. And lastly- a stretch goal would be to develop a method of grouping similar medications together so that all the options under a patient's insurance could be compared together based on their relative formulary statuses and copayment requirements.

In this dataset we have provided a list of pharmacy transactions from a handful of pharmacies taken over the course of a year. These transactions have been designed to incorporate several real effects seen in live pharmacy data- such as deductible phases beginning near the start of the year, lower copays for lower formulary tier drugs, higher rejection rates for the higher tier drugs on the formulary, and different categories of insurance plans. A brief description of the identifiers included in this dataset:

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