2023 Humana Mays Healthcare Analytics Case Competition Problem Prompt

Dustin James Harper—Senior Data Scientist, Humana Pharmacy Analytics and Consulting

This document outlines the problem statement and available data for the 2023 Humana Mays Healthcare Analytics Case Competition. For information about registration, schedule, submission logistics, and the leaderboard, see <https://mays.tamu.edu/humana-tamu-analytics/>.

# Motivation and Opportunity

Oncology is a clinical area of focus that has seen significant advances in research and new therapies. Despite these advances, cancer remains a leading cause of death, with approximately 600,000 people dying from cancer each year in the US alone. While new treatments are coming to the market, many of them are associated with potentially significant side effects which are a barrier to people staying adherent to their life-saving medications.

One of these medications is Osimertinib, an oral tyrosine kinase inhibitor used for patients in early-stage lung cancer of the non-small cell variety, with a specific targetable mutation known as EGFR. Osimertinib is known to be a largely effective medication, with patients receiving it being twice as likely to survive compared to those who do not take the medication. Additionally, patients taking the medication as prescribed are 80% less likely to have a recurrence of their cancer.

As with many other oncology drugs, tolerance of Osimertinib can be difficult due to the side effects associated with the medication—specifically nausea, fatigue, pain, high blood glucose, and constipation. Many of these side effects are manageable with proper counseling and avoidance techniques, but many patients may opt to discontinue their treatment rather than seeking guidance on how to manage them. Approximately one quarter of Humana members taking Osimertinib have side effects and discontinue their Osimertinib therapy within the first 6 months.

To address this problem, we need to leverage our data and analytics to target members at risk, encourage medication adherence, and allow our oncology patients to live longer, fulfilling lives.

# Predictive Modeling Target

As explained previously, taking Osimertinib as prescribed can help people live longer, but treatable side-effects might cause people to end their treatment prematurely. To this end, your assignment is to build a model to predict which therapies will end prematurely after a reported side effect, also known as an adverse drug event or ADE.

It is in everyone’s best interest for a patient to keep taking their Osimertinib therapy for as long as possible. In this case, a successful therapy is six months (180 days) of continuous Osimertinib therapy. Conversely, an unsuccessful therapy is any therapy that ends before 180 days. However, to specifically target members who may be discontinuing due to experiencing an ADE, we’ve defined the target more specifically to include an ADE at some point during the unsuccessful therapy. In the training data, this target is recorded in the column labeled `tgt\_ade\_dc\_ind`.

## Unsuccessful Therapy: tgt\_ade\_dc\_ind==1

The target is defined as a therapy that ends prematurely (before 180 days) and has an ADE reported at some time during the therapy. The target definition has already been done for you and is available in target\_train.csv as the column ‘tgt\_ade\_dc\_ind’.

## All Other Therapies: tgt\_ade\_dc\_ind==0

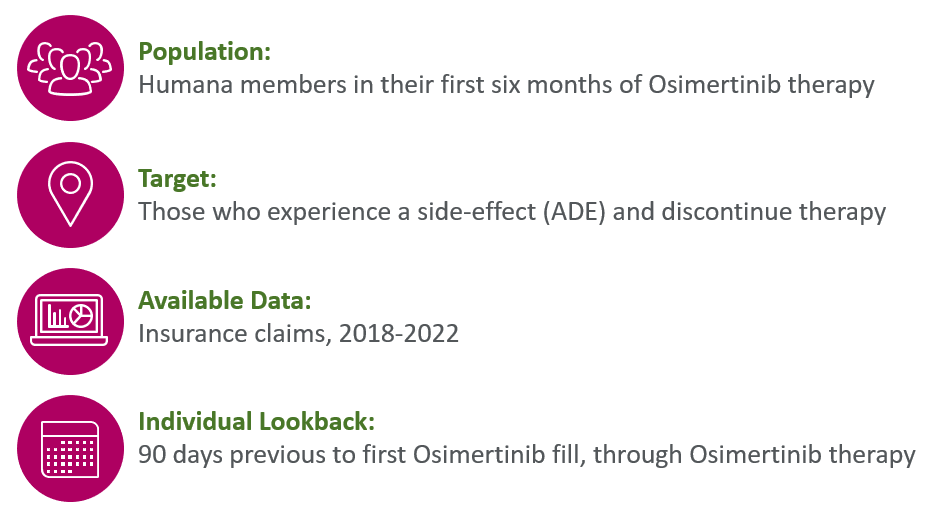
Since the target definition is so specific, there are several other types of therapies not included in the target group. The following are a few examples:

* Successful therapies with 180 days of continuous treatment
* Therapies that end prematurely with no reported ADEs
* Therapies where the member changes to another insurance plan or dies before 180 days

# Available Data

Since we’re trying to predict if a therapy will end prematurely after an ADE, our data is organized based on a specific therapy with one member, a start date and end date.

Each category of data is separated into a train and holdout set. Use the train set to train your model and submit your predictions on the holdout set for scoring.



## File Descriptions

The following sections provide a brief overview of the available data. Detailed descriptions of all fields are available in data\_dictionary.csv.

With the exception of the target\_holdout, the target and holdout sets contain the exact same data columns, but for different sets of individuals.

### Target: target\_train (1232 records), target\_holdout (420 records)

Unique on the person identifier and therapy identifier. Contains information about the therapy start and end dates, the target identifier, and protected attributes for the individual (sex, race, age, etc.)

*Important Note*: When you submit your results for Round 1, you will need to submit an ID, score and rank for each individual ID in the target\_holdout file. The ID will come directly from target\_holdout.csv, and the score and rank will come from your predictive model. You will notice that there is no target identifier or therapy end date included in target\_holdout.csv.

Sum of tgt\_ade\_dc\_ind in target\_train.csv: 117

### Medical Claims: medclms\_train (100159 records), medclms\_holdout (23232 records)

Unique on claim identifier. Contains simplified information about all medical claims for an individual during the time 90 days before their Osimertinib therapy and through the end of therapy. This data includes visit and process dates, diagnosis codes and indicators for diagnosis codes of interest. E.g. since nausea is a known side-effect of Osimertinib, we added an indicator column for a diagnosis code related to nausea.

Sums from medclms\_train:

|  |  |  |
| --- | --- | --- |
| ade\_diagnosis | seizure\_diagnosis | pain\_diagnosis |
| 6848 | 333 | 63 |

Sums from medclms\_holdout:

|  |  |  |
| --- | --- | --- |
| ade\_diagnosis | seizure\_diagnosis | pain\_diagnosis |
| 1841 | 37 | 24 |

### Pharmacy Claims: rxclms\_train (32133 records), rxclms\_holdout (6670 records)

Unique on claim identifier. Contains simplified information about all pharmacy claims for an individual during the time 90 days before their Osimertinib therapy and through the end of therapy. This data includes service and process dates, drug identifier codes (NDC) and indicators for drug codes of interest. E.g. since anticoagulants are a known to have drug interactions with Osimertinib, we added an indicator column for a drug code for an anticoagulant.

Average rx\_cost in rxclms\_train: 2463.950

Average rx\_cost in rxclms\_holdout: 2159.679

### Data Dictionary: data\_dictionary.csv (49 records)

Contains a description for each data column available in the claims datasets.

### Race Code Descriptions: race\_cd\_desc.csv (7 records)

Contains definitions for the coded race codes in the target files.