Prediction and Prevention of ADE-related Dropouts from Tagrisso Targeted Therapy

Humana-Mays Healthcare Analytics
2023 Case Competition

Table of Contents

1. Executive Summary

1.1 Study Proposal

Osimertinib (sold under the brand name Tagrisso) is a targeted therapy drug used to treat non-small-cell lung cancer (NSCLC) in cases where the cancer cells have specific abnormal EGFR genes. It has been proven to be effective as part of a variety of cancer care treatments, doubling survival rate in some cases. However, a quarter of therapy members experience Adverse Drug Events (ADEs) and drop out within the first 6 months of therapy start. This non-adherence leads to reduced disease-free survival rates among dropouts.

The objective of our study is to apply the latest data mining and machine learning techniques to derive insights from the medical data on past therapy members provided to us, and combine this with our creative problem solving and research to devise strategies that will help CenterWell Specialty Pharmacy and Humana reduce the dropout rate from its Tagrisso program.

1.2 Modeling and Analysis

Our analysis focused on predicting premature therapy discontinuations due to adverse drug effects using the 'tgt_ade_dc_ind' variable. Using a merged training dataset, the XGBoost model was selected for its standout performance. To counter data imbalance, we applied the Synthetic Minority Over-sampling Technique (SMOTE) and refined the model with Bayesian optimization. Post-training, the model was tested on a holdout dataset, and key influencing factors were identified through feature importance and SHAP value analysis. This approach produced a predictive tool that provides actionable insights for stakeholders.

1.3 Results and Recommendations

Based on our model's insights, we identified 4 strategies for CenterWell and Humana to adopt:

1. Predicting ADE-driven dropouts
2. Improving the effectiveness of Tagrisso therapy and its administration
3. Improving patients' quality of care
4. Bridging fundamental gaps in research

We believe these data-driven, research-backed recommendations will help CenterWell make more informed decisions regarding the care of therapy members, leading to a lower dropout rate.

2. Case Background

Cancer, the second leading cause of death in the United States following heart disease and responsible for more than 600,000 deaths annually, is one of the foremost global health challenges, posing a significant obstacle for healthcare providers and patients. Encouragingly, the past few decades have witnessed remarkable advancements in cancer treatments, and significantly improving survival rates. However, in the evolving healthcare landscape, medication adherence has become critical to improving patient outcomes and reducing the burden of chronic disease. Medication compliance, especially in complex conditions such as oncology, is a key determinant of treatment success.

Past studies consistently underscore the profound importance of adherence, emphasizing how non-compliance can compromise the effectiveness of therapeutic interventions and, in turn, contribute to preventable complications. These complications necessitate hospitalization, which subsequently translates into increased medical costs. From a business perspective, it is imperative that we address this challenge within the realm of oncology, exploring innovative solutions that ensure patients receive the full benefits of advanced cancer therapies while minimizing adverse outcomes. By doing so, health insurance and pharmacy companies can better serve their clients, boost patient satisfaction, and enhance operational efficiency.

2.1 Business Problem

The aim of this analysis is to address major business problems faced by CenterWell Specialty Pharmacy and Humana - **The high dropout rate for Tagrisso treatment**. Despite having high effectiveness at improving patients' disease-free survival and decreasing relapse rates, Tagrisso exhibits undesirable side effects (ADEs) in about a quarter of users. This has led to a ~10% dropout rate from CenterWell's Tagrisso program, leading to poorer care for customers and a loss of revenue for CenterWell.

The objective of our analysis is two-fold:

- Identify risk markers that will help us predict the dropout potential for patients
- Use insights mined from the data and modeling process, and combine them with medical and business research to devise actionable strategies to reduce dropout rate.

2.2 Key Performance Indicator for this problem:

A. AUC-ROC Curve:

The AUC-ROC (Area Under the Receiver Operating Characteristic) Curve is a graphical representation that illustrates the diagnostic ability of our classification prediction model in distinguishing between patients who drop out from the Tagrisso treatment and those who don't.

The curve plots the True Positive Rate against the False Positive Rate at various decision thresholds.
The AUC, or the area under this curve, quantifies the overall ability of the model to discriminate between the two classes. In our project, a higher AUC indicates that the model is better at predicting actual dropouts as dropouts and those who continue the therapy as non-dropouts. Given the serious implications of incorrectly predicting a dropout, we aim for a high AUC to ensure that our model has strong discriminatory power. In this case, we want to achieve an AUC value closer to 1 while preventing the model from overfitting.

B. AUC for Precision-Recall Curve:

Given the imbalanced nature of our merged training dataset, where only a small fraction of patients drop out from the therapy, the Precision-Recall (PR) curve becomes especially pertinent. This curve plots Precision (the fraction of predicted dropouts that are actual dropouts) against Recall (the fraction of actual dropouts that are correctly predicted by the model). Since our primary focus is on the minority class – the patients dropping out, the PR curve provides a more informative picture of our model's performance in these critical instances. The area under the PR curve gives us a single value summarizing the model's performance across all levels of precision and recall, allowing us to optimize for the best balance, especially given the high cost of false negatives in this context. Ideally, we want to have an AUC value close to 1.

C. Confusion Matrix:

The confusion matrix offers a comprehensive view of how our classification model's predictions compare to the actual outcomes. It categorizes predictions into four groups: True Positives, True Negatives, False Positives, and False Negatives. For CenterWell Specialty Pharmacy and Humana, understanding this breakdown is crucial. True Positives represent patients correctly predicted to drop out, allowing timely interventions. False Negatives represent patients incorrectly predicted to continue therapy but who actually drop out, a scenario we aim to minimize. The matrix not only provides insights into the model's accuracy but also highlights areas where the model might need further refinement to better serve the patients and address the high dropout rate effectively. Ideal Scenarios are as follows:

- True Positives: As high as possible. Represents correctly predicted dropouts.
- True Negatives: As high as possible. Represents correctly predicted continuations.
- False Positives: As low as possible. Represents false alarms where non-dropouts are incorrectly identified.
- False Negatives: Especially critical for representing missed dropouts. This should be minimized given the high costs associated.

D. Precision, Recall, and F1 Score: Given the business problem, a high recall (sensitivity) might be prioritized to identify as many actual dropouts as possible. Precision ensures that the identified dropouts are actual dropouts. The F1 score balances both.

3. Data Preparation

3.1. Data Description

There are two datasets provided by Humana this year, the training dataset, the one used to train the model, and the holdout dataset, the one used to make predictions.

- The training dataset includes 1,232 Humana members' information (10 columns), from 2018-2022, and their medical (100,159 records, 27 columns) and pharmacy claims (32,133 records, 24 columns) during the time 90 days before their Osimertinib therapy and through the end of therapy, also includes whether this member meets the target criteria (*tgt_ade_dc_ind*) and the end date (*therapy_end_date*).
- The holdout dataset includes 420 Humana members' information (8 columns), from 2018-2022, and their medical (23,232 records27 columns) and pharmacy claims (6,670 records, 24 columns) during the time 90 days before their Osimertinib therapy and through the end of therapy.

3.2 Data Exploration[1]

3.2.1. General Data Overview (Target Analysis)

In this section, we present a visual analysis of the target dataframe, with the aim of providing a comprehensive overview of the data. The following charts illustrate key aspects of the dataset.

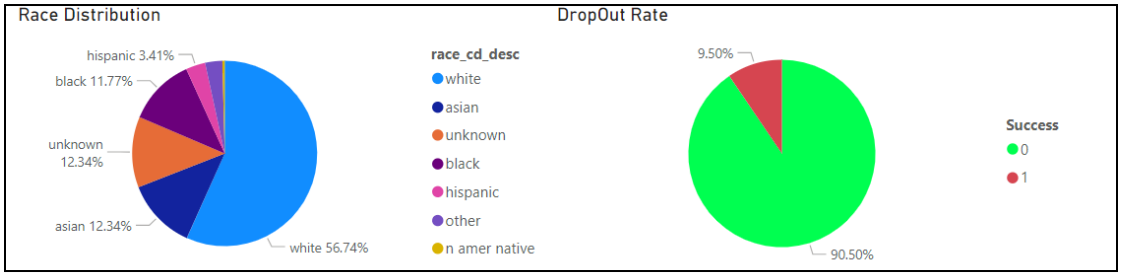

Figure 3.2.1.1 Race Distribution and Dropout Rate

- **Race Distribution**: This pie chart displays the distribution of races in the dataset, enabling us to identify the predominant racial groups. It reveals that the majority comprises white individuals, constituting 56.74%, while Hispanic individuals make up the minority at 3.41%.
- **Dropout Rate**: This pie chart explains the success rate (Dropout Rate) of individuals undergoing this treatment. In this chart, 0 (90.50%) represents successful treatments, while 1 (9.50%) represents treatments that were discontinued. In the subsequent section, we will delve deeper into this indicator.

[1] Power BI Dashboard for Data Exploration: Click me!

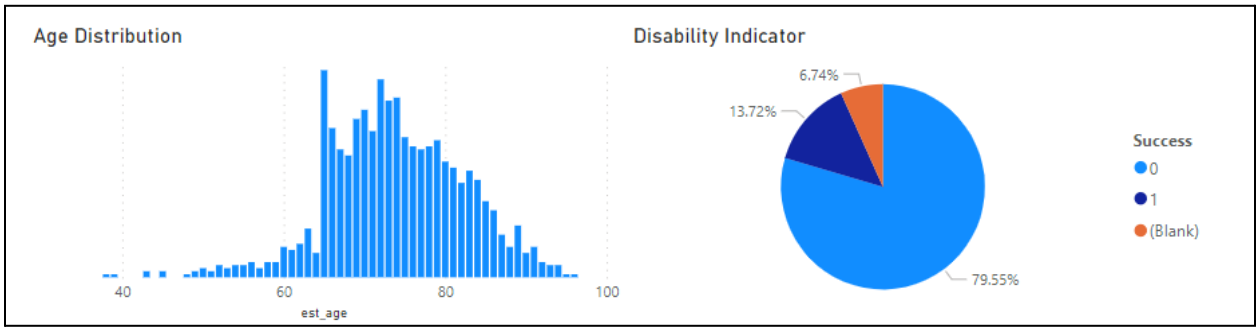

Figure 3.2.1.2 Age Distribution and Disability Indicator

- **Age Distribution**: This histogram illustrates the distribution of ages in the dataset, helping us identify the age groups that are most prevalent. The majority falls within the age range of 65 to 85 years, indicating a higher likelihood of receiving this treatment among older and elderly individuals.
- **Disability Indicator**: This pie chart illustrates the presence or absence of disabilities among the study population, providing insights into the prevalence of disabilities within the dataset.

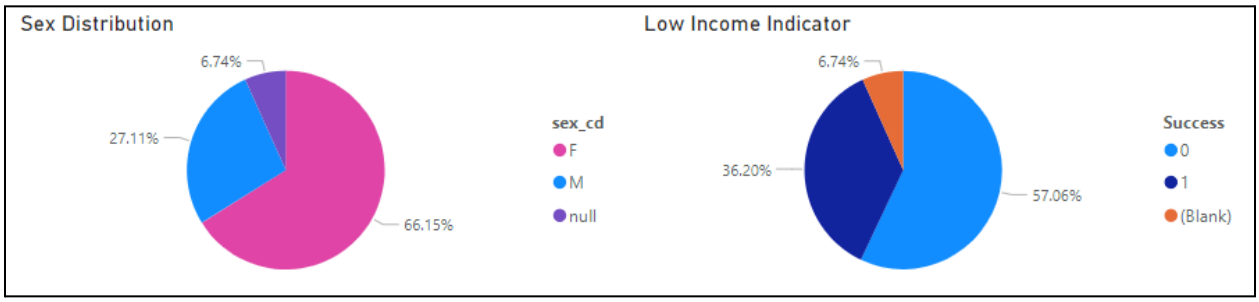

Figure 3.2.1.3 Sex Distribution and Low Income Indicator

- **Sex Distribution**: This pie chart offers insights into the sex distribution within the dataset, indicating that the majority of customers are females, constituting 66.15%.
- **Low-Income Indicator**: This pie chart summarizes the economic status of the study population by categorizing individuals into two groups: 1 ("Low-Income") and 0 ("Non-Low-Income").

3.2.2. DropOut Rate per Race (Target Analysis)

Delving deeper into the Target analysis, it is valuable to incorporate dropout rate analysis based on sex and race variables. Here, 1 represents individuals who discontinued the treatment, while 0 represents those who completed it.

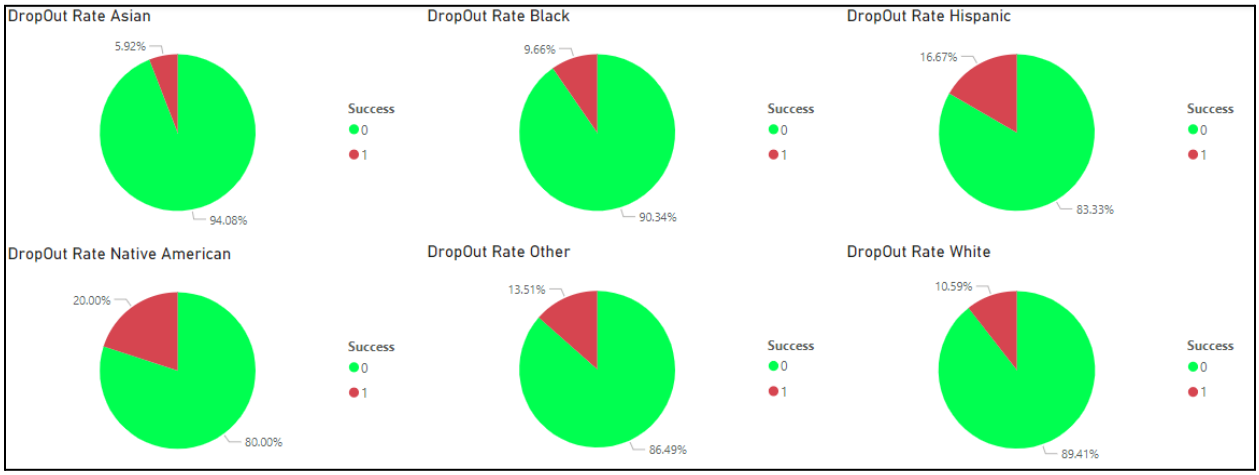

Figure 3.2.2.1 DropOut Rate per Race

Some key findings that emerge from these analyses include:

- Native Americans and Hispanic individuals exhibit the highest dropout rates, at 20.00% and 16.67%, respectively.
- Asians have the lowest dropout rate, standing at 5.92% compared to other racial groups.

3.2.3. DropOut Rate per Sex (Target Analysis)

In the pie charts below, we present data categorized into two groups: females and males. This categorization is done based on the assumption that success rates may vary between genders.

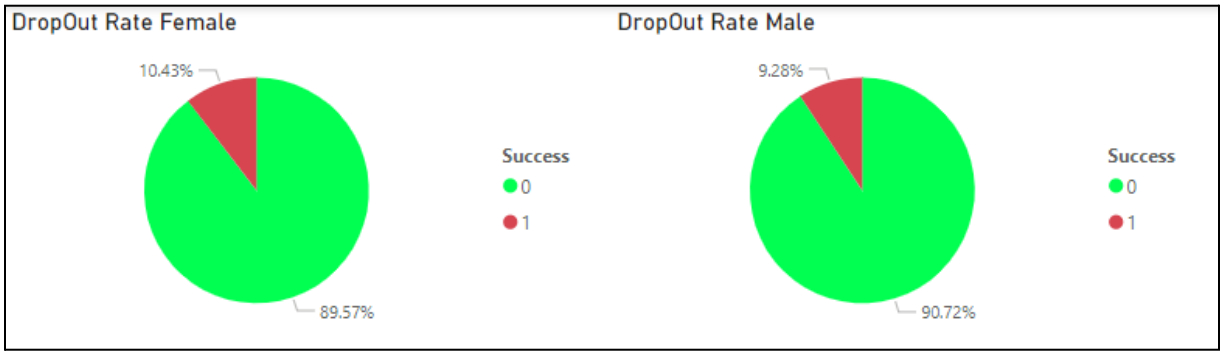

Figure 3.2.3.1 DropOut Rate per Sex

Observing the data, we note that the dropout rates are nearly equal for males and females, leading to the conclusion that the treatment is effective for both genders, with a slight advantage for males, as their dropout rate is lower by 1.15%.

3.2.4. Conclusion

These visualizations play a crucial role in understanding our dataset and form the basis for further analysis in this research.

3.3 Data Cleaning

In the initial phase of data cleaning, we identified relatively irrelevant variables with the following criteria and removed them from the datasets.

1. Variables that are duplicates of other variables.
2. Variables with low variance.
3. Variables that are not related to our analysis.
4. Variables that has too many missing values (the percentage of missing value > 50%)

This process helped us clean the dataset, leaving only the important and relevant data for further examination. Finally, we have a dataset containing 1149 rows × 190 columns.

3.4 Data Transformation & Feature Engineering

After removing some columns, we prepared the data for analysis by transforming it as needed, which involved the following steps.

1. Generate new columns to illustrate the count of ambulance usage, ER visits, inpatient visits, and outpatient visits by consolidating data from the *pot*, *util_cat*, and *hedis_pot* columns.
2. Generate a new column for disease score that depict the number of diseases based on the International Classification of Diseases (ICD-10) code by using the 1st, 2nd, and 3rd digits from the *primary_diag_cd*, *diag_cd2* , *diag_cd3*, *diag_cd4*, *diag_cd5*, *diag_cd6*, *diag_cd7*, *diag_cd8*, and *diag_cd9* columns.
3. Generate a new column for disease-specific drugs to depict the count of distinct drug types utilized from the *hum_drug_class_desc* column.

Second, the raw datasets encompassed all medical claims and pharmacy claims related to an individual, spanning a critical time frame from 90 days before the Osimertinib therapy and continuing through the end of the therapy. To maintain the utmost precision in our analysis, we meticulously removed any medical claims and pharmacy claims that fell outside the specified follow-up period. Furthermore, to enhance the granularity and clarity of our analysis, we transformed and aggregated multiple medical claims and pharmacy claims at the individual level. This method enabled us to construct a holistic and representative portrayal of everyone's healthcare utilization. By combining and summarizing these claims data at the individual level, we created a more manageable and insightful dataset for our subsequent analyses and evaluations.

Third, as part of our analysis, we explored whether changes observed before and after starting Osimertinib therapy played a role in patients' decisions to discontinue treatment prematurely. To investigate this, we divided our dataset into two periods: "pre-therapy" and "post-therapy." This division was based on medical visit dates (*visit_date*) and pharmacy service dates (*service_date*) relative to the start of Osimertinib therapy (*therapy_start_date*). This approach enabled us to examine potential differences and trends in patient outcomes before and after starting therapy and find out factors that might influence early treatment discontinuation.

Last, in the final stages of preparing our dataset for machine learning, we undertake a critical step known as one-hot encoding. We converted categorical variables into a binary format, 0 and 1, making them understandable to machine learning algorithms. By this process, each unique category becomes an independent binary feature, enhancing the compatibility and performance of the machine learning model.

Newly Generated Feature	Definition
pre_visit_days/ post_visit_days	The duration in days from the first visit date to the therapy start date/ The duration in days from the therapy start date to the last visit date
pre_service_days/ post_service_days	The duration in days from the first service date to the therapy start date/ The duration in days from the therapy start date to the last service date
pre_no_visit_avg/ post_no_visit_avg/ no_visit_avg_dif	Average visit frequency before/ after the therapy start date, and the difference between *post_no_visit_avg* and *pre_no_visit_avg*
pre_er_avg/ post_er_avg/ care_er_dif	Average emergency room visit frequency before/ after the therapy start date and the difference between *post_er_avg* and *pre_er_avg*
pre_ambulance_avg/ post_ambulance_avg/ care_ambulance _dif	Average ambulance usage frequency before / after the therapy start date and the difference between *post_ambulance_avg* and *pre_ambulance_avg*
pre_ip_acute_avg/ post_ip_acute_avg/ care_ip_acute_dif	Average inpatient acute care frequency before/ after the therapy start date and the difference between *post_ip_acute_avg* and *pre_ip_acute_avg*
pre_ip_mhsa_avg/ post_ip_mhsa_avg/ care_ip_mhsa _dif	Average inpatient MHSA care frequency before/ after the therapy start date and the difference between *post_ip_mhsa_avg* and *pre_ip_mhsa_avg*
pre_ip_rehab_avg/ post_ip_rehab_avg/ care_ ip_rehab _dif	Average inpatient rehab care frequency before/ after the therapy start date and the difference between *post_ip_rehab_avg* and *pre_ip_rehab_avg*
pre_ip_snf_avg/ post_ip_snf_avg/ care_ip_snf _dif	Average inpatient SNF care frequency before/ after the therapy start date and the difference between *post_ip_snf_avg* and *pre_ip_sf_avg*
pre_opt_avg/ post_opt_avg/ care_opt_dif	Average outpatient care visit frequency before/ after the therapy start date and the difference between *post_opt_avg* and *pre_opt_avg*
pre_physician_office_avg/ post_physician_office_avg/ care_physician_office _dif	Average physician office visit frequency before/ after the therapy start date and the difference between *post_physician_office_avg* and *pre_physician_office_avg*
pre_urgent_care_avg/ post_urgent_care_avg/ care_urgent_care _dif	Average urgent care frequency before/ after the therapy start date and the difference between *post_urgent_care_avg* and *pre_urgent_care_avg*
normalized_pre_therapy_score/ normalized_post_therapy_score/ score_diff	The normalized mean of no. of unique disease diagnosis before/after the therapy date for each unique therapy_id (The two normalized feature were derived by first mapping *primary_diag_cd* and *diag_cd2* through *diag_cd9* in the medical claims datasets using the ICD-10 codes provided by the World Health Organization (WHO). Then, the number of occurrences for each unique disease was calculated and defined as the SCORE. Lastly, for each unique therapy id, the mean SCORE from visit dates before/after the therapy date was computed and normalized) and the difference between *normalized_post_therapy_score* and *normalized_pre_therapy_score*

pre_ade_diagnosis_avg/ post_ade_diagnosis_avg/ med_ade_dif	Average ADE report frequency before/ after the therapy start date and the difference between *post_ade_diagnosis_avg* and *pre_ade_diagnosis_avg*
pre_seizure_diagnosis_avg/ post _seizure_diagnosis_avg/ med_seizure_dif	Average seizure report frequency before/ after the therapy start date and the difference between *post_seizure_diagnosis_avg* and *pre_seizure_diagnosis_avg*
pre_pain_diagnosis_avg/ post _pain_diagnosis_avg/ med_pain_dif	Average pain report frequency before/ after the therapy start date and the difference between *post_pain_diagnosis_avg* and *pre_pain_diagnosis_avg*
pre_fatigue_diagnosis_avg/ post _fatigue_diagnosis_avg/ med_fatigue_dif	Average fatigue report frequency before/ after the therapy start date and the difference between *post_fatigue_diagnosis_avg* and *pre_fatigue_diagnosis_avg*
pre_nausea_diagnosis_avg/ post _nausea_diagnosis_avg/ med_nausea_dif	Average nausea report frequency before/ after the therapy start date and the difference between *post_nausea_diagnosis_avg* and *pre_nausea_diagnosis_avg*
pre_hyperglycemia_diagnosis_avg/ post _hyperglycemia_diagnosis_avg/ med_hyperglycemia_dif	Average hyperglycemia report before/ after the therapy start date and the difference between *post_hypergycemia_diagnosis_avg* and *pre_hyperglycemia_diagnosis_avg*
pre_constipation_diagnosis_avg/ post _constipation_diagnosis_avg/ med_constipation_dif	Average constipation report frequency before/ after the therapy start date and the difference between *post_constipation_diagnosis_avg* and *pre_constipation_diagnosis_avg*
pre_diarrhea_diagnosis_avg/ post _diarrhea_diagnosis_avg/ med_diarrhea_dif	Average diarrhea report frequency before/ after the therapy start date and the difference between *post_diarrhea_diagnosis_avg* and *pre_diarrhea_diagnosis_avg*
pre_rx_cost_avg/ post_ rx_cost_avg/ cost_dif	Average per day cost of prescription before/ after the therapy start date and the difference between *post_rx_cost_avg* and *pre_rx_cost_avg*
pre_spcl_avg/ post_spcl_avg/ rx_spcl _dif	Average frequency of specialty drug used before/ after the therapy start date and the difference between *post_spcl_avg* and *pre_spcl_avg*
pre_ddi_ind_avg/ post_ddi_ind_avg/ rx_ddi_dif	Average frequency of drug usage with a known interaction with Tagrisso before/ after the therapy start date and the difference between *post_ddi_ind_avg* and *pre_ddi_ind_avg*
pre_anticoag_ind_avg/ post_anticoag_ind_avg/ rx_anticoag _dif	Average frequency of anticoagulant before/ after the therapy start date and the difference between *post_anticoag_ind_avg* and *pre_anticoag_ind_avg*
pre_diarrhea_treat_ind_avg/ post_diarrhea_treat_ind_avg/ rx_diarrhea _dif	Average frequency of drug used to treat diarrhea before/ after the therapy start date and the difference between *post_diarrhea_ind_avg* and *pre_diarrhea_ind_avg*
pre_nausea_treat_ind_avg/ post_nausea_treat_ind_avg/ rx_nausea _dif	Average frequency of drug used to treat nausea before/ after the therapy start date and the difference between *post_nausea_ind_avg* and *pre_nausea_ind_avg*
pre_seizure_treat_ind_avg/ post_seizure_treat_ind_avg/ rx_seizure _dif	Average frequency of drug used to treat seizure before/ after the therapy start date and the difference between *post_seizure_ind_avg* and *pre_seizure_ind_avg*
pre_drug_class_avg_1-30/ post_drug_class_avg_1-30/ rx_drug_class_avg_dif_1- 30	Average frequency of drug used by the Humana Drug Class Description before/ after the therapy start date and the difference between *post_drug_class_avg* and *pre_drug_class_avg*

Table 4.3.1 Description of Newly Generated Features

4. Statistical Analysis and Modeling

4.1. Model Selection

In the pursuit of predicting the ‘tgt_ade_dc_ind’ variable, an array of advanced machine learning models was employed to ensure a comprehensive assessment of potential predictors. The dataset was subjected to a train-test partitioning strategy, wherein 70% was allocated for training and the remaining 30% for validation. This approach ensured that the models had ample data for learning while also reserving an untouched subset for the evaluation of predictive performance.

Among the models explored, Logistic Regression augmented with Integrated Nested Laplace Approximation (INLA) implemented in R stood out, achieving an Area Under the Curve (AUC) of 0.89. In parallel, gradient boosting models were also employed. LightGBM demonstrated an AUC of 0.92, while CatBoost and XGBoost yielded AUCs of 0.93 and 0.95, respectively. These models, renowned for their prowess in handling complex non-linear relationships and classification problems, provided insights into the variable's intricate predictive patterns.

To bolster the reliability and robustness of the predictive models, k-fold cross-validation was performed. This rigorous technique repeatedly partitions the training data into multiple subsets, systematically using each for both training and validation. By averaging performance across these different partitions, k-fold cross-validation mitigates the risk of overfitting and offers a more generalized perspective on model efficacy.

After meticulous evaluation and comparison of the various models' performance metrics, the XGBoost model emerged as the most promising. With the highest AUC of 0.95, it outperformed the other contenders, suggesting superior discriminative ability in predicting the ‘tgt_ade_dc_ind’ variable. XGBoost's renowned capability for handling large datasets, dealing with missing values, and its efficient implementation of gradient boosting further solidified its position as the optimal choice. Thus, for the subsequent phases of this study, XGBoost was selected as the primary predictive model.

For a dataset with feature vector X and binary target variable Y (‘tgt_ade_dc_ind'), the prediction after m iterations can be articulated as:

$$E_m(X) = E_{m-1}(X) + \alpha_m \times h_m(X, r_{m-1}),$$

where $E_{m-1}(X)$ is the prediction up to the $(m-1)^{th}$ iteration. The term $\alpha_m h_m(X, r_{m-1})$ is the contribution from the m^{th} tree, and h_m is a function trained to predict the residuals r_{m-1} using the feature vector X, and α_m is the corresponding weight. The learning process in XGBoost

revolves around minimizing a loss function, which measures the difference between the actual target values Y ('tgt_ade_dc_ind') and the model's predictions. This is represented as:

$$\alpha_m, r_{m-1} = argmin \sum_{i=1}^{m} L(Y, E_{m-1}(X) + \alpha_m \times h_m(X, r_{m-1})),$$

where L is the differentiable loss function, and the optimization seeks the parameters α_m and r_{m-1} that reduces this loss. XGBoost continually minimizes the residuals, aiming to boost the model's predictive accuracy with each step.

4.2. Strategy for dealing with imbalanced dataset

The analysis of the target variable, 'tgt_ade_dc_ind', revealed a pronounced class imbalance, with 1,115 instances labeled as '0' and a mere 117 instances labeled as '1'. Such an imbalance poses challenges in machine learning, as models tend to be biased towards the majority class, often leading to suboptimal predictive performance for the minority class. Recognizing the criticality of this issue, especially in applications where the minority class might be of paramount importance, it was deemed essential to address this imbalance to ensure the robustness and reliability of the predictive models.

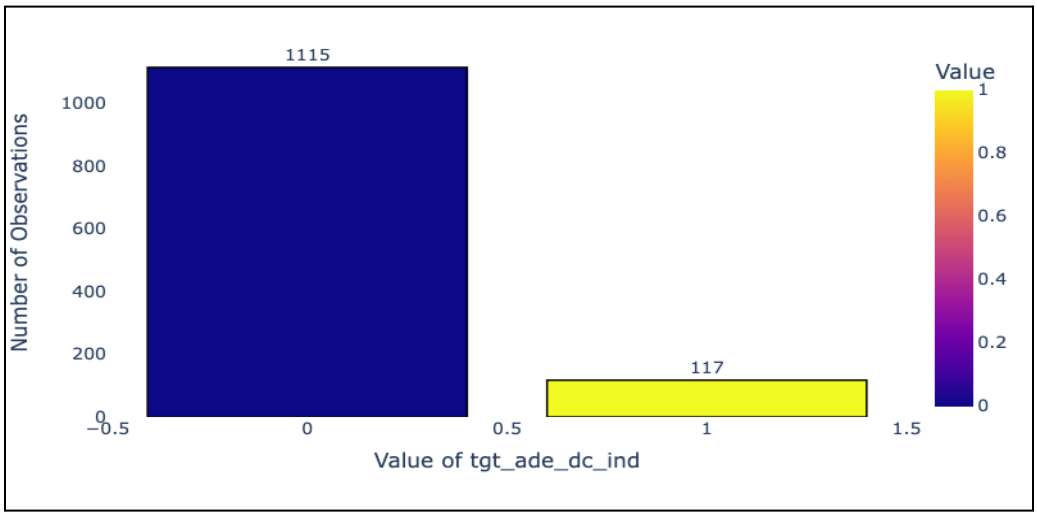

Figure 4.2.1 Distribution of target variable

To counteract this imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. SMOTE operates by creating synthetic samples in the feature space. For each observation in the minority class, SMOTE selects *k* nearest neighbors from the minority class. It then takes the difference between the feature vector of the observation under consideration and its nearest neighbor. A random number between 0 and 1 is multiplied with this difference, and the result is added to the feature vector of the observation. This process generates a new, synthetic data point that lies between the observation and its neighbor. By iteratively applying this method, SMOTE effectively augments the minority class, making it comparable in size to the majority class, thus rectifying the imbalance.

Our application of SMOTE ensures a more balanced representation of the two classes, enabling machine learning models to learn and generalize better. By synthesizing new examples, rather than merely replicating existing ones, SMOTE provides a richer and more diverse dataset, paving the way for models to capture the underlying complexities and nuances of the minority class. This technique is particularly pivotal in scenarios where the cost of misclassifying the minority class is high.

4.3. Hyperparameter Optimization

After implementing SMOTE to deal with the imbalance dataset, we are now focusing on dealing with model improvement. Traditionally, for a binary classification problem, we need to use random forest or LASSO regression to do feature selection. However, since we are using XGBoost as our predictive model, it has its own built-in method for calculating feature importance. Hence, we skip the feature selection part to avoid redundancy in ensemble models and reduce computational cost.

For the very beginning of the modeling process, we trained the XGBoost model and had a decent model performance relative to other models like LightGBM and CatBoost. In this section, we need to do model tuning and try to improve our model performance and prevent overfitting further. To tune the model, we utilized Grid Search and Bayesian Optimization to get the optimized combination of hyperparameters of the XGBoost model.

Grid search is an exhaustive method that systematically evaluates every possible combination of hyperparameter values within a predefined space. It's a straightforward and deterministic approach, ensuring that every specified combination is evaluated. While it's thorough, it can be computationally intensive, especially when dealing with high-dimensional hyperparameter spaces. Grid search is best suited for situations where the hyperparameter space is relatively small or when there's an abundance of computational resources available. Bayesian Optimization is a probabilistic model-based optimization technique. Instead of exhaustively evaluating all possible hyperparameter combinations, Bayesian optimization constructs a probabilistic model of the objective function and then intelligently selects the next set of hyperparameters to evaluate. Technically, in the very first step, we select the initial set of hyperparameters, which is known in 5.1, to evaluate them in the true function f. Secondly, we fit a Gaussian process[2] on the observed points. What's more, we use the acquisition function to find the next point to evaluate. We use the Expected Improvement (EI) acquisition function, which is defined as:

$$\alpha_{UCB}(x) = E[max(f(x) - f(x^{+}), 0)]$$

[2] A Gaussian process is defined by a mean function m(x) and a covariance function k(x,x′). Given any set of input values, the corresponding outputs are distributed as a multivariate Gaussian.

To Choose the next point based on the expected improvement over the current best value. After doing this, Bayesian Optimization will evaluate this point in the true function f, and update the Gaussian process with the new data point. We make Bayesian Optimization process repeat until our convergence criterion[3] is met. In general, This selection is based on prior evaluations and aims to strike a balance between exploring new regions of the hyperparameter space and exploiting areas believed to be optimal.

Key Parameters	Original XGBoost Model	After Grid Search	After Bayesian Optimization
learning_rate	0.100	0.050	0.027
colsample_bytree	None	0.900	0.884
gamma	None	0	0.383
max_depth	None	5	5
n_estimators	100	50	259
subsample	0.800	0.900	0.876
min_child_weight	None	None	4
lambda	None	None	0.239
alpha	None	None	0.057

Note: **'learning_rate'** determines the step size at each iteration while moving towards a minimum of the loss function. It shrinks the feature weights to make the boosting process more conservative. **'Colsample_bytree'** is the fraction of features that can be selected for building the tree. A value of 0.8 means 80% of the features are used. **'Gamma'** denotes minimum loss reduction required to make a further partition on a leaf node. It specifies a regularization term to the objective function. **'Max_depth'** is the maximum depth of a tree, determining how deep each tree can grow during any boosting round. **'n_estimators'** is the number of boosting rounds or trees to build. It's important to tune it with the learning rate since a small learning rate requires more boosting rounds. **'subsample'** is the fraction of the training data that can be randomly sampled to train each tree. **'Min_child_weight'** is the minimum sum of instance weight (hessian) needed in a child. It's used to control over-fitting. Higher values make the algorithm more conservative by setting a constraint on leaf nodes. **'Lambda'** indicates L2 regularization term on weights, and it's used to avoid over-fitting. **'Alpha'** is L1 regularization term on weights; it's also used to avoid over-fitting.

Table 5.3.1 Hyperparameter Optimization Results

According to table 5.3.1, we could observe that Bayesian Optimization yields better performance with more restrictive hyperparameters, and the optimization process is compared using the same evaluation metrics (accuracy) and the same validation strategy (k-fold cross-validation). Finally, we decided to use the hyperparameters generated by Bayesian Optimization since it not only generates more restriction for parameters to prevent overfitting but also has high efficiency and

[3] We defined a fixed number of trials (50 trials) for the Optuna optimization process on Python.

capacity to handle large search spaces. In the end, the newly trained model performance (evaluated as the AUC) is approximately 0.96.

4.4. Model Results

To gain a comprehensive understanding of our XGBoost model's performance and the importance of individual features, we've visualized the ROC-AUC curve, precision-recall curve, confusion matrix and assessed feature importance.

- **Feature Importance:** The feature importance plot of XGBoost quantitatively displays the significance of each feature in influencing the model's outcomes. Notably, features such as *'post_ade_diagnosis_avg',* *'post_fatigue_diagnosis_avg',* and *'post _service_days'* emerge as top contributors, underlining their paramount influence on the model. Conversely, features like *'care_opt_dif'* and *'pre_service_days'* have lesser importance. Depending on the model's objectives and interpretability needs, it might be beneficial to delve deeper into these features or even contemplate their omission in subsequent versions.

- **Performance Metrics**:
    - **ROC-AUC**: The model boasts an impressive ROC AUC score of 0.96, indicating its ability to differentiate between positive and negative classes effectively.
    - **PR Curve AUC**: The AUC for the Precision-Recall Curve stands at 0.92, reflecting the model's strong capability in precision and recall balance, especially when the positive class is of interest.
    - **Confusion Matrix:**
      True Positives: 25
      False Positives: 3
      True Negatives: 307
      False Negatives: 10

The confusion matrix reveals that the model performs particularly well in identifying negative cases (with 307 true negatives) and has a low false-positive rate (only 3 instances). However, the presence of 10 false negatives suggests some room for improvement in capturing all positive cases.

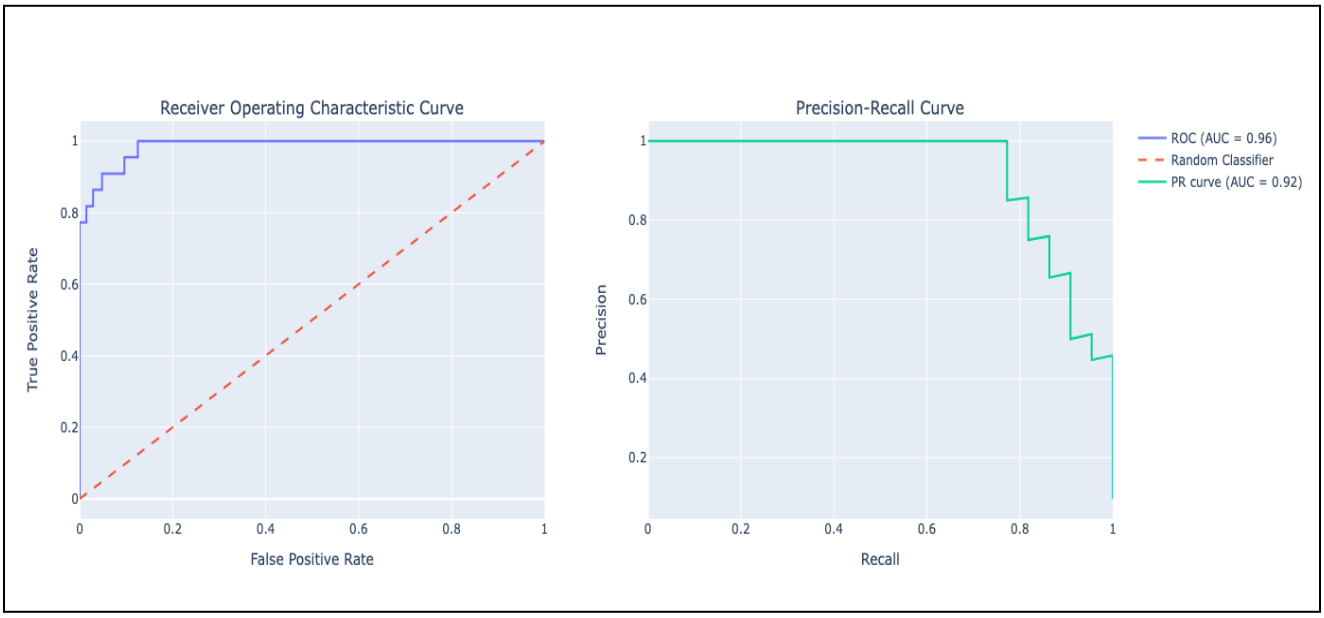

Figure 4.4.1 ROC curve & Precision-Recall Curve

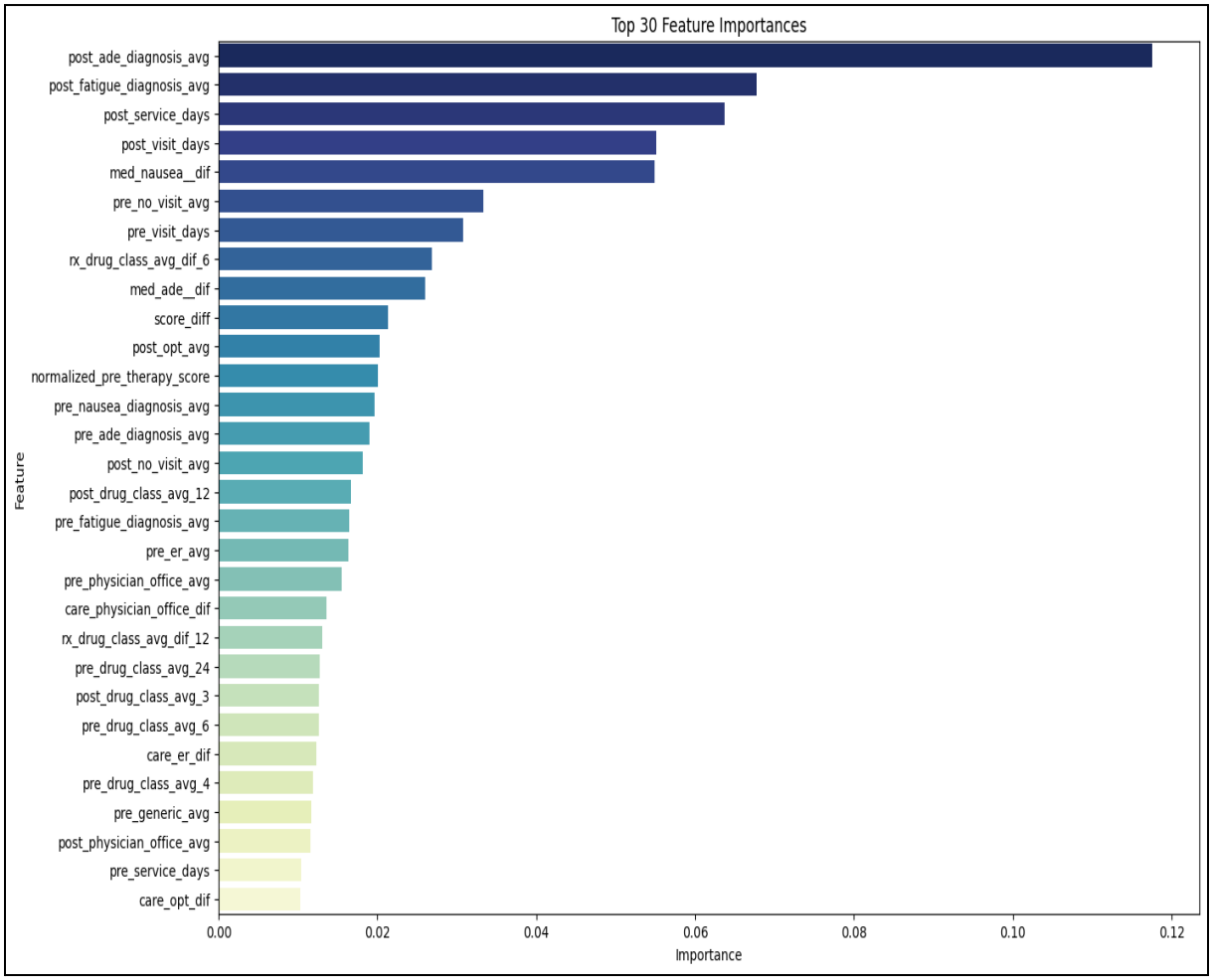

Figure 4.4.2 Top 30 features

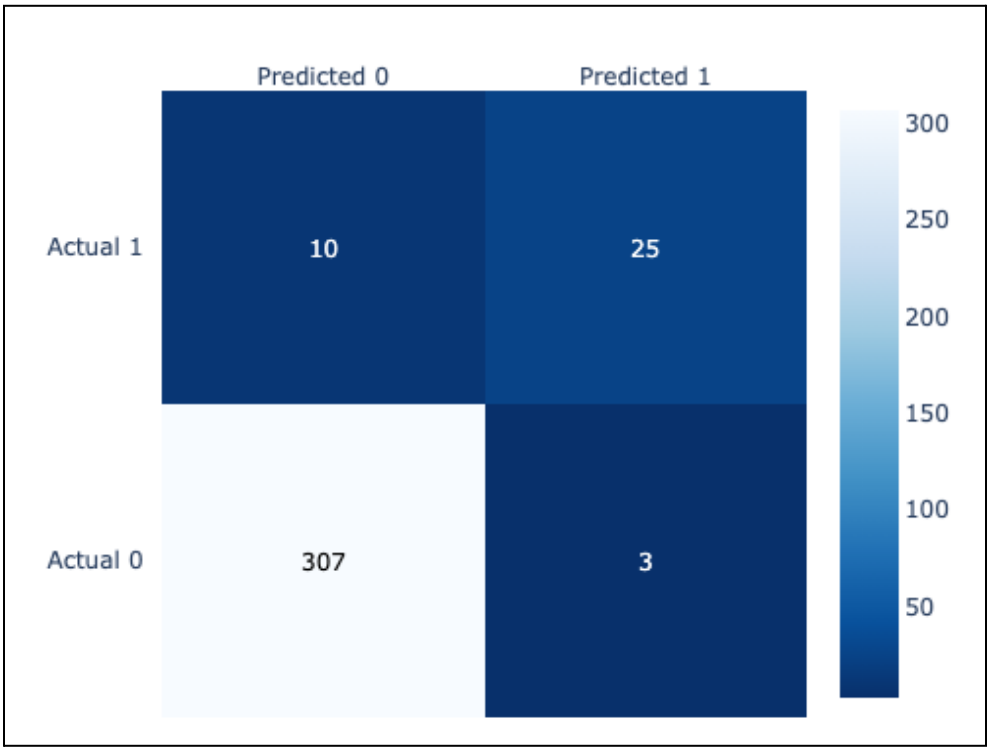

Figure 4.4.3 Confusion Matrix Plot

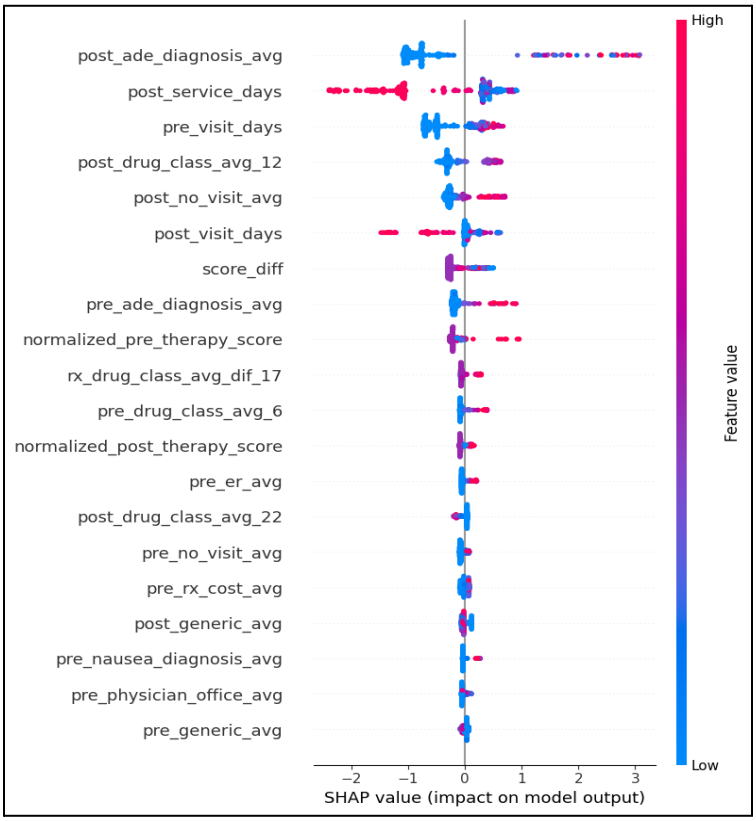

Figure 4.4.4 SHAP value of features

Overview of SHAP Value Graph:

- The SHAP value graph presented elucidates the impact of the top 20 features on the model's predictions. This breakdown enhances the transparency of the model's decision-making process.
- Prominent Features:
  'post_ade_diagnosis_avg', *'post_service_days'*, and *'pre_visit_days'* display the most substantial variation in SHAP values, underscoring their critical roles. These features exhibit significant fluctuation in their influence on the model's predictions across different data points.

Insights from Color Gradient:

- Cooler shades for *'post_ade_diagnosis_avg'*, representing lower values, largely align with the negative SHAP axis. This suggests that patients with fewer average ADE reports after therapy are less likely to discontinue the therapy.
- In contrast, warmer shades for *'pre_ade_diagnosis_avg'* lean towards the positive SHAP axis. This indicates that higher frequencies of average ADE reports before therapy commencement often elevate the prediction.

Less Influential Features:

- The SHAP value graph corroborates the significance of features like *'post_ade_diagnosis_avg'* and *'post_service_days'*, which were also spotlighted by the XGBoost feature importance.

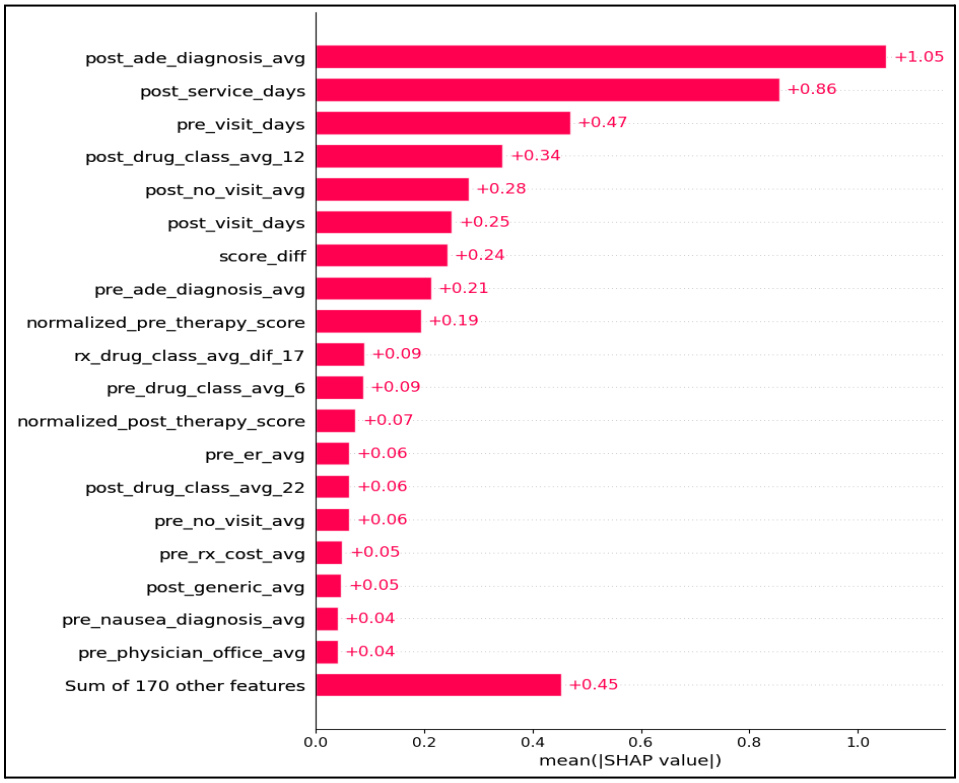

Figure 4.4.5 SHAP bar plot

Overview of mean absolute SHAP Value Graph:

- **The Paramount Importance of Focusing on Magnitude:**
    - When we discuss the magnitude of these SHAP values, we're essentially tapping into the core essence of how much weight each feature carries. By emphasizing this aspect, we can discern the overall influence that a specific feature imparts on the model's predictions. It's pivotal to note that this is irrespective of whether the impact is positive or negative, making the magnitude a neutral and unbiased metric of influence.

- **Features that Stand Out in their Influence:**
    - Among the myriad features, a few distinctly stand out. The top tier includes: '*post_ade_diagnosis_avg*', '*post_service_days*', '*pre_visit_days*', '*post_drug_class_avg_12*', and 'post_no_vist_avg'. These particular features, with their towering mean SHAP values, undeniably play a central role in the predictive outcomes. Their dominance in the graph underscores their indispensable role in the model's decision-making matrix.
    - Not to be overshadowed, the secondary influential features include: '*score_dif*' and '*normalized_pre_therapy_score*'. While they might not match the absolute influence of the top-tier features, they undeniably play a considerable, influential role in the model's predictive architecture.

- **Features with a More Subdued Influence:**
    - On the contrary, there are features that, while contributing, don't resonate with the same level of influence. These encompass '*pre_physician_office_avg', 'post_generic_avg',* and '*pre_nausea_diagnosis_avg'*. Their positions, characterized by relatively lower mean SHAP values, suggest a more restrained impact on the predictions. This nuanced understanding suggests they may warrant additional investigation, particularly if the overarching objective leans towards a more streamlined model.

5. Model Interpretation

Clearly, both the feature importance chart and the SHAP value figure highlight some standout features that are important. When we look at all the variables, these features can be grouped into the following categories. This grouping helps us understand how these important aspects affect the model's predictions and why they matter in our analysis. Recognizing these influential factors helps us make better decisions and improve our strategies for more accurate results. The following are the five categories of importance groups:

1. Healthcare Visits

'pre_no_visit_avg' and *'post_no_visit_avg'* emerged as highly important variables in both figures, displaying a positive correlation with the target variable, *'tgt_ade_dc_ind'*. This suggests that individuals who had fewer healthcare visits before starting therapy or those who had more healthcare visits were more likely to report adverse drug events (ADE) or discontinue the therapy. Among all the healthcare visits, *'pre_er_avg', 'pre_opt_avg'* and *'pre_physician_office_avg'* were also significant variables. The more visits to the emergency room (ER), or the fewer outpatient visits, the higher the likelihood of individuals meeting the criteria. Furthermore, *'post_service_days*,' *'pre_visit_days*,' and *'post_visit_days'* are also highly important variables in the model. Individuals with longer periods of medical and pharmacy claims after therapy start date are less likely to report adverse drug events (ADE) or discontinue therapy prematurely. These variables highlight the significance of the duration of medical and pharmacy-related interactions in influencing therapy outcomes.

2. Medical Condition Severity (Comorbidities)

'normalized_pre_therapy_score', *'normalized_post_therapy_score'*, and *'score_diff'* were also important variables, as they provided insights into the person's health condition. A higher score on these variables suggests the presence of a broader spectrum of health issues, indicating that the individual may have a greater susceptibility to various types of diseases. Consequently, when assessing the prediction outcome, which focused on the likelihood of dropping out of the therapy program, these health indicators played a crucial role in understanding the complex interplay between health status and therapy adherence.

3. Medical Condition Reporting

The variables '*pre_ade_diagnosis_avg*', '*post_ade_diagnosis_avg*', and '*med_ade_dif*' reflect the frequency of adverse drug event (ADE) reports, consistently ranking among the top 10 in both figures. These variables demonstrate a positive correlation with the treatment outcome. The more frequently adverse drug events are reported, the higher the likelihood of premature therapy discontinuation before the full treatment course is completed. Additionally, '*post_fatigue_diagnosis_avg*' stands out as the second most important feature, and frequently nausea reports, *'pre_nausea_diagnosis_avg'* and '*med_nausea_dif*' are also one of the most important features. This, in turn, influences whether the individual meets the criteria for reporting ADE or discontinuing therapy.

4. Pharmaceutical Usage

When considering the usage of drugs for different conditions, we assessed the frequency of each drug category, which includes a total of 30 categories based on the humana drug class description. After running the model, we identified several drug categories that proved to be crucial to the model's performance. These include drugs related to managing cholesterol (*'pre_drug_class_avg_6'* and *'rx_drug_class_dif_avg_6'*), addressing gastrointestinal disease (*'post_drug_class_avg_12'* and *'rx_drug_class_dif_avg_12'*), pain management (*'post_drug_class_avg_22'*), and addressing mental health (*'rx_drug_class_dif_avg_17'*). These drug categories played a significant role in influencing the model's predictions and outcomes.

5. Expenditure:

The expenditure for drugs, specifically *'pre_rx_cost_avg'* and *'cost_dif*,' also plays a crucial role in predicting the target outcome. A higher existing prescription expenditure is associated with a higher likelihood of early termination of therapy. Additionally, a greater difference in expenditure after starting therapy, indicating increased average spending post-therapy initiation, is correlated with a higher probability of discontinuing the therapy prematurely. These findings underscore the financial aspect's impact on therapy adherence and outcome prediction.

6. Business Implications and Recommendations

Our strategies to predict and reduce dropouts are grouped under four key ***findings:***

6.1. There are ways to predict ADE-driven Dropout

At-risk Patient + Lack of timely intervention = Premature dropout

- Our model has identified several "markers" that are indicative of dropout, based on existing health conditions, drug usage and level of medical care.
- We have also devised an aggregate "Score" metric that quantifies a patient's level of health and comorbidities before and after therapy commencement, which has proven to be a significant indicator of dropout.
- With a recall of ~70% of dropouts and nearly perfect precision, our model is able to "screen" for patients with a high likelihood of facing ADEs and dropping out prematurely.
- We recommend that CenterWell flag patients fitting the criteria we've identified and place them on a higher level of monitoring to allow for more timely intervention and support from their end.

<u>Our suggested Risk Markers include metrics such as:</u>

Medication Usage: Chemotherapy, Gastrointestinal, Pain Management, Anticoagulants etc.
Symptoms: ADE Diagnosis, Fatigue, Nausea, Diarrhea, Pain, Seizures etc.
Medical visits: Pre/Post Therapy Average Visits, Outpatient visits, ER Visits
Therapy: Days in Therapy

(please refer to previous chapter for full list)

Also, **major differences** between the pre-therapy and post-therapy values for the above factors are a significant signal for dropout.

6.2. There are ways to improve therapy effectiveness & reduce dropouts

6.2.1. Promote Adjuvant Chemotherapy via Comprehensive Geriatric Assessments

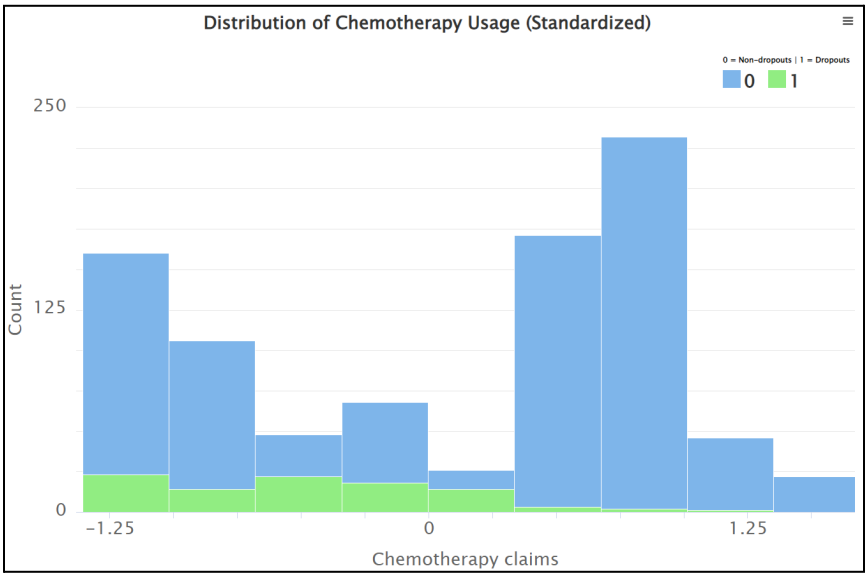

Figure 6.2.1.1 Distribution Plot of Chemotherapy Usage
(0 = Non-dropouts, 1 = dropouts)

- Tagrisso works best when coupled with adjuvant chemotherapy[4]
    - The FLAURA2 trial showed that when adding chemotherapy to Tagrisso, the risk of disease progression or death is reduced by 38% when compared to Tagrisso alone.
    - However, 80% of treatment dropouts are on less intensive chemotherapy plans than the cohort average.
- A primary reason for reduced chemotherapy allotments is the age of the patient - the median age of the training set is 73 years. As older adults are more prone to the toxic effects of chemotherapy, they are often put on more conservative treatment plans.
- However, not all people age the same way - some can bear more than their age suggests, and vice versa. Recent studies[5] have proven that "**Comprehensive Geriatric Assessments**", which gauge elderly patients' vulnerability to the toxic effects of chemotherapy, are effective tools for allowing physicians to customize treatment plans for older cancer patients. Such assessments have also proven to help in reducing the occurrence of side-effects faced by chemotherapy patients, at no adverse effect on survival rates.[6]

[4] Tagrisso plus chemotherapy demonstrated strong improvement in progression-free survival for patients with EGFR-mutated advanced lung cancer in FLAURA2 Phase III trial
[5] Use of a Comprehensive Geriatric Assessment for the Management of Elderly Patients With Advanced Non-Small-Cell Lung Cancer: The Phase III Randomized ESOGIA-GFPC-GECP 08-02 Study
[6] For Older Adults, Geriatric Assessment Reduces Cancer Treatment Side Effects

- We recommend CenterWell incorporate such assessments into the onboarding process, to ensure that prospective Tagrisso users receive the best possible chemotherapy care for their condition before starting targeted treatment. This will maximize Tagrisso's effectiveness and minimize toxic side-effects of treatment.
- Such assessments are now even easier to implement with a clinically tested tool[7] that lets physicians perform the assessment in less than 10 minutes. By coordinating with patients' primary physicians, we can help ensure better care is provided on-site.

6.2.2. Monitor First 90 Days of Treatment

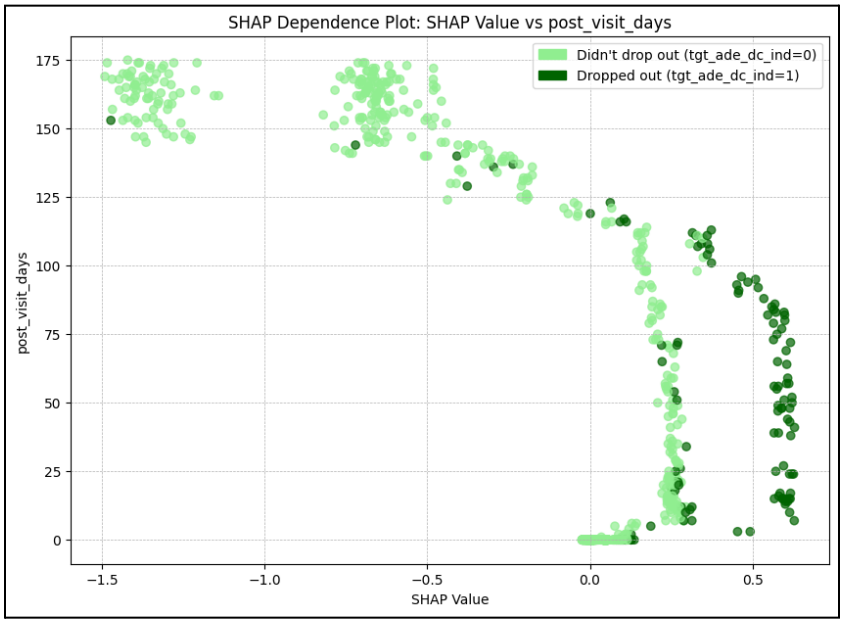

Figure 6.2.2.1 SHAP Dependence Plot for *post_visit_days*

- One-third of dropouts occur in the first month of therapy and 80% of all dropouts take place within the first 90 days. "Days in therapy" is a significant predictor of dropout.
- By connecting at-risk patients with one-on-one care from CenterWell's Cancer Center of Excellence, and training care providers to monitor for extreme changes in identified risk markers, we believe it is possible to reduce dropouts in this critical period.
- Major changes in symptoms and usage of medical services are key risk markers.

6.2.3 Prepare at-risk patients for possible side effects of taking Tagrisso

- Studies show that patients who report higher levels of communication with their care providers (physicians and nurses) are significantly less likely to consider dropping out of cancer treatment. They also report feeling better informed and more trusting of care providers than their peers.[8]

[7] Implementing the cancer and aging research group (CARG) tool in the ambulatory oncology setting to drive informed treatment selection.
[8] Cancer clinical trial patient-participants' perceptions about provider communication and dropout intentions - PMC

- We believe that the first step to a successful treatment is the establishment of a two-fold communication prior to therapy start:
    - Informational communication (risks and prognosis)
    - Relational communication (trust-building)
- Through this, we can simultaneously inform patients of the possibility of ADEs and build an interpersonal relationship between the CenterWell representative and the patient. We believe we can better prepare prospective Tagrisso users for what lies ahead of them and reduce the "shock" that they may face using Tagrisso.

6.2.4. Train and Communicate with on-site caregivers & physicians

- Studies show that older cancer patients consistently receive sub-par cancer care compared to their younger peers.[9]
- Regardless of whether therapy members live alone, with family or in nursing facilities, their caregivers should be aware of the risks of Tagrisso and best practices for treatment so as to remedy this deficiency when it comes to Tagrisso usage.

6.3. There are methods of providing better patient care and support

6.3.1. Encourage physical therapy as part of normal routine

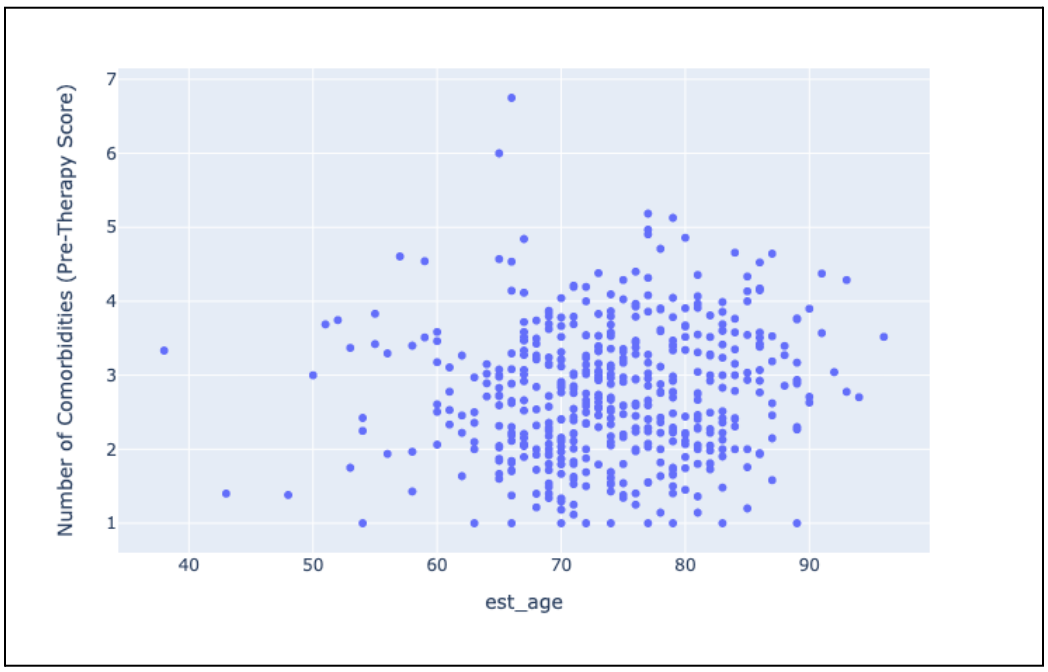

Figure 6.3.1.1 Scatter Plot of Estimated Age versus No. of Comorbidities

This plot shows the number of comorbidities each therapy member has prior to therapy start.

[9] Influence of Age on Guideline-Concordant Cancer Care for Elderly Patients in the United States - ScienceDirect

- Physical therapy with progressive muscle relaxation (PT&R) is a proven intervention for improved cancer care, both physically and mentally. Until recently, such measures were deemed impractical to administer to older patients with multiple conditions outside of therapy centers.[10]
- A pilot test conducted by The Ohio State University's Cancer Center showed that it was feasible to conduct **virtual physical therapy and relaxation** treatments among the older age demographic.[11]
    - In spite of comorbidities and other limiting factors, 90% of participants reported they "strongly liked" the program
- Providing therapy members access to such therapy can help create resilience among lung cancer patients by increasing physical fitness and reducing anxiety and depression. We recommend CenterWell incorporate virtual PT&R as part of the regular treatment plan.

6.3.2. Encourage regular counseling visits

- Nearly 3,000 claims for mental health medication were found in the claims data shared with us.
- Psycho-oncological studies have shown that psychological intervention leads to better treatment outcomes, quality of life and medical cost offsets. The inability to handle treatment-related stress is a leading cause of dropout in cancer care.[12]
- We recommend that counselor/psychologist visits be made a regular part of the therapy member's health checkup, as it is proven to be an effective strategy for improving the quality of patient care and offsetting medical costs.

6.3.3. Connect patients with Support Groups

- As part of the holistic approach, it's important for the elderly to receive emotional support outside of their treatment. This may help them cope with the unavoidable side effects of cancer treatment.
- The American Cancer Society and National Cancer Institutes provide resources to cancer patients who are looking to find support, counseling and peer groups for cancer patients. By connecting therapy members to such resources and groups, CenterWell can help alleviate the mental and social pressures of cancer treatment.[13]

---

[10] A Model-Based Cost-Effectiveness Analysis of an Exercise Program for Lung Cancer Survivors Following Curative-Intent Treatment - PMC
[11] Resiliency among Older Adults Receiving Lung Cancer Treatment (ROAR-LCT): A Novel Pilot Supportive Care Intervention Study
[12] Emotional distress: the sixth vital sign--future directions in cancer care
[13] Full article: Both group peer counselling and individual counselling reduce anxiety and depression, and increase self-esteem and overall life satisfaction in palliative cancer care

6.4. There are fundamental knowledge gaps in Geriatric Cancer Treatment

- An analysis of FDA-approved clinical trials has revealed that, despite accounting for 56% of cancer cases, 65+ years olds account for only 40% of participants in cancer trials.[14] This imbalance makes treatment of older cancer patients more difficult, owing to the lack of information on how new cancer drugs affect people of that demographic.

- Being a market leader in healthcare, Humana can benefit from investment in Geriatric Cancer Clinical Trials & Geriatric Cancer Care research. Researchers predict cancer incidences to continue to skew towards older demographics, making such investments essential for Humana's long-term success.

- By investing in geriatric cancer research, Humana can equip its subsidiaries operating in elderly health care with the medical knowledge and tools they need to make optimal business and medical decisions.

[14] FDA analysis of enrollment of older adults in clinical trials for cancer drug registration: A 10-year experience by the U.S. Food and Drug Administration.

6.5. Cost Analysis

6.5.1. Assumptions:

1. Therapy Cohort - 2000 members
    a. Members at risk of dropout - 200 (10%)
2. Therapy duration for non-dropout = 12 months
3. Therapy duration for dropout = 3 months
    a. 9 months revenue lost due to dropout
4. Cost of Tagrisso = $16,999 per 30 days
5. Recommendations are implemented only for at-risk members
    a. Recommended strategies can cut dropouts to 100 (5%)

6.5.2. Cost Estimates:

	Cost Estimates	
Recommendation	**Cost per Member**	**Total Cost (per year)**
Dropout Prediction System[15]	—	$50,000-$200,000
Comprehensive Geriatric Assessments[16]	$5000-10000 (once per member)	$1-2 Million
Virtual Physical Therapy Sessions[17]	$200-$400 per month	$500,000-$1,000,000
Investment in Geriatric Cancer Research[18]	—	$3-5 Million
Additional Monitoring of at-Risk patients[19]	$800 per month	$1.92 Million
Total Cost of Recommendations		**$6.5-10.2 Million**

6.5.3. Revenue recovered

200 dropouts * 50% retention * 9 months revenue recovered * $16,999 pm = ~$15,300,000 revenue recovered

The recommendations are financially feasible based on our assumptions and market research.

[15] What is the Cost to Deploy and Maintain a Machine Learning Model? | phData
[16] Is comprehensive geriatric assessment hospital at home a cost-effective alternative to hospital admission for older people?
[17] Telehealth Physical Therapy: The Guide to Virtual PT - GoodRx
[18] How Much Does an Oncology Clinical Trial Cost?
[19] Registered Nurse Salaries in the United States for CenterWell | Indeed.com

7. Conclusions

Our analysis of the data has revealed a category of individuals that are prone to dropping out of Tagrisso therapy due to ADEs and other related factors. We have identified the common markers that distinguish dropouts from their peers, and have developed an predictive model effective at identifying dropouts from a given cohort of therapy members. In addition, we have combined our data insights with the latest medical research, cancer care strategies and business understanding to craft strategies to reduce dropout rates. Our research has shown that these are proven methods of reducing dropout rates and improving the quality of care received by cancer patients. We believe CenterWell and Humana will benefit from implementing these recommendations into their Tagrisso program and operations.

However, our analysis revealed that not all Humana members had complete drug data, and even fewer had medical records, with only one-third having such records available. Therefore, to enhance the performance of future analyses, it is crucial to include more comprehensive data. Gathering complete data, especially medical records, can provide a more holistic view of patient health and enable a more accurate assessment of the effectiveness of the therapy programs.

Also, some variables with a considerable amount of missing data were not included in our current model. However, we believe that these variables may provide valuable information. In the future, if further research can develop improved missing value imputation strategies, we recommend including these variables to make the model's explanation even more powerful and comprehensive. This would further enhance the insights and outcomes of the analysis.

Additionally, in future analyses, we suggest that researchers conduct more detailed investigations into the events that occur in the 30 days preceding a patient's decision to discontinue the therapy program. For instance, it would be valuable to examine whether there is a frequent occurrence of adverse drug events (ADEs), reports of unpleasant health conditions, or significant changes in medical expenditures that may have become unaffordable for the patients. By gaining a deeper insight into the reasons behind a patient's decision to drop out of the program, we can empower CenterWell and Humana to make more informed decisions and offer timely support to encourage patients to remain in therapy. This not only benefits the healthcare providers but also contributes to the well-being and longevity of oncology patients, creating a win-win situation for all parties involved.

8. References

Power BI Dashboard for Data Exploration:
https://app.powerbi.com/view?r=eyJrIjoiM2EyYWZlZTgtNTM4My00MzZmLWI2MmYtNzhhMzIwNDkxNWRhIiwidCI6IjAxZjNkMGExLTNkYzgtNDBkNy04YjUyLTI0YjM4YmEzM2JjYiIsImMiOjl9

Adrian Kemp. (2023, May 17) Tagrisso plus chemotherapy demonstrated strong improvement in progression-free survival for patients with EGFR-mutated advanced lung cancer in FLAURA2 Phase III trial. Retrieved from https://www.astrazeneca.com/media-centre/press-releases/2023/tagrisso-plus-chemo-improved-pfs-in-lung-cancer.html

Bultz, B. D., & Carlson, L. E. (2006). Emotional distress: the sixth vital sign--future directions in cancer care. *Psycho-oncology*, *15*(2), 93–95. https://doi.org/10.1002/pon.1022

CenterWell Specialty Pharmacy. (n.d.). Retrieved from https://www.centerwellspecialtypharmacy.com/center-of-excellence/cancer

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Coop, R. (2023). What is the Cost to Deploy and Maintain a Machine Learning Model? phData. Retrieved from https://www.phdata.io/blog/what-is-the-cost-to-deploy-and-maintain-a-machine-learning-model

Corre, R., Greillier, L., et al. (2016). Use of a Comprehensive Geriatric Assessment for the Management of Elderly Patients With Advanced Non-Small-Cell Lung Cancer: The Phase III Randomized ESOGIA-GFPC-GECP 08-02 Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *34*(13), 1476–1483. https://doi.org/10.1200/JCO.2015.63.5839

Fang, P., He, W., Gomez, D. R., Hoffman, K. E., Smith, B. D., Giordano, S. H., . . . Smith, G. L. (2017). Influence of Age on Guideline-Concordant Cancer Care for Elderly patients in the United States. *International Journal of Radiation Oncology Biology Physics*, *98*(4), 748–757. https://doi.org/10.1016/j.ijrobp.2017.01.228

Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Ha, D., Kerr, J., Ries, A. L., Fuster, M. M., Lippman, S. M., & Murphy, J. D. (2020). A Model-Based Cost-Effectiveness Analysis of an Exercise Program for Lung Cancer Survivors After Curative-Intent Treatment. *American journal of physical medicine & rehabilitation*, *99*(3), 233–240. https://doi.org/10.1097/PHM.0000000000001281 Ohio State University, Columbus/OH/USA. (n.d.). *Resiliency among Older Adults Receiving Lung Cancer Treatment*

(ROAR-LCT): A Novel Pilot Supportive Care Intervention Study [Research Poster]. Retrieved from https://cattendee.abstractsonline.com/meeting/10925/presentation/2473

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Mbewe, A., Pike, P., Lewis, R., Kortmansky, J., Chiang, A. C., & Kanowitz, J. (2021). Implementing the cancer and aging research group (CARG) tool in the ambulatory oncology setting to drive informed treatment selection. *Journal of Clinical Oncology*, *39*(28_suppl), 209. https://doi.org/10.1200/jco.2020.39.28_suppl.209

Ledesma, P. (n.d.). How much does an oncology clinical trial cost? www.linkedin.com. Retrieved from https://www.linkedin.com/pulse/how-much-does-oncology-clinical-trial-cost-patricio-ledesma

Lundberg, S. M., & Lee , S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Ramsay, K., Ramsay, J., & Main, D. C. J. (2007). Both group peer counselling and individual counselling reduce anxiety and depression, and increase self-esteem and overall life satisfaction in palliative cancer care. *Counselling Psychology Quarterly*, *20*(2), 157–167. https://doi.org/10.1080/09515070701240071

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *71*(2), 319-392.

Shailaja, K., Seetharamulu, B., **& Jabbar, M. A**. (2018, March). Machine learning in healthcare: A review. In *2018 Second international conference on electronics, communication and aerospace technology (ICECA)* (pp. 910-914). IEEE.

Sharon Reynolds, Geriatric assessment reduces cancer treatment side effects. (2021, December 15). Retrieved from https://www.cancer.gov/news-events/cancer-currents-blog/2021/cancer-geriatric-assessment-fewer-side-effects

Singh, S., Gray, A., et al. (2022). Is comprehensive geriatric assessment hospital at home a cost-effective alternative to hospital admission for older people?. Age and ageing, 51(1), afab220. https://doi.org/10.1093/ageing/afab220

Singh, H., Kanapuru, B., Smith, C., Fashoyin-Aje, L. A., Myers, A., Kim, G., & Pazdur, R. (2017). FDA analysis of enrollment of older adults in clinical trials for cancer drug registration:

A 10-year experience by the U.S. Food and Drug Administration. *Journal of Clinical Oncology*, *35*(15_suppl), 10009. https://doi.org/10.1200/jco.2017.35.15_suppl.10009

Zhou, Q., Ratcliffe, S. J., Grady, C., Wang, T., Mao, J. J., & Ulrich, C. M. (2019). Cancer Clinical Trial Patient-Participants' Perceptions about Provider Communication and Dropout Intentions. *AJOB empirical bioethics*, *10*(3), 190–200. https://doi.org/10.1080/23294515.2019.1618417