Humana-Mays Healthcare Analytics 2023 Case Competition

"Overcoming Medication Adherence Challenges"

Enhancing Patient Outcomes Amidst Adverse Drug Effects

TABLE OF CONTENT

EXECUTIVE SUMMARY

Humana provides coverage for patients taking Osimertinib that results in improved patient outcomes when successfully completed. The question is how we can improve the rate of patients who complete the trial, as many are not able to handle the side effects. In addition, we need solutions that are equitable with respect to race and gender. Our research determines that Humana may suffer $100-$300 million in additional costs due to non-compliance.

First, we provide a robust and repeatable process to train accurate models that are equitable in outcomes. This ensures our approach will work as more data is added (no re-engineering required) and can be applied to the unmet need of handling this re-occurring modeling issue demonstrated by the previous years' competition.

Next, we use a technique known as causal analysis to analyze the data. This allows us to quantify on a per-patient basis "what could have been" (i.e., the counter-factual) if Humana intervened in their care. This indicates that only 4.5% of patients are likely to be receptive to an intervention based on the given data, with a 2-7% decrease in likelihood of discontinuity. This could then reduce Humana's costs by $4.5-$13.5 million.

This analysis suggests that Humana may benefit most by developing new avenues of intervention beyond what currently exists. Thus, we develop a business strategy around expanding the Humana app to become a resource portal and community for patients. Leveraging Humana's existing infrastructure, personalized recommendations to dietary, activity, and other life-changes can be made based on clinical evidence of improved drug compliance, while minimizing development and intervention costs. This strategy is repeatable in nature allowing Humana to expand to other high-risk protocols for other cancers and diseases. The research suggests adherence rates could be improved by up to 20-25% by our business plan, resulting in an enhanced $20-$75 million in savings for Humana.

The remainder of this document is organized in the following manner. We will review the medical context, and the total value at risk to Humana, in the Case Background. This establishes the $100-300 million at risk to the firm. The Data Analysis section will then review how we processed and organized the data to enable our analysis. In Modeling we invent a new loss function to jointly satisfy the accuracy and fairness goals of Humana in a repeatable way, satisfying the need demonstrated by this and previous year's competition that have used this metric. The results of the model are analyzed in the Analysis section that establishes the 4.5% of intervenable patients. The implications of this analysis to the business are determined in Business Implications, where we identify other avenues that could increase the intervention rate by up to 25%. This informs our solution to the larger problem of patient adherence in Business Solution by leveraging Humana's existing app infrastructure and expertise for a repeatable process. Applied to just the current scope of oncology patients this could increase the savings to $75 million, and greater value is obtainable by re-applying the solution to other high-cost patient populations.

1. CASE BACKGROUND

1.1. Context

Globally, lung cancer is the second most commonly diagnosed cancer, and in the United States, Non-Small Cell Lung Cancer (NSCLC) constitutes 81% of all lung cancer diagnoses. In 2023[1], an estimated 238,340 adults will receive a lung cancer diagnosis in the U.S., with the worldwide count reaching 2,206,771 in 2020, encompassing both small cell lung cancer and NSCLC cases. Lung cancer risk is closely tied to age, with the majority of cases diagnosed in individuals aged 65 or older. Approximately 53% of patients with lung cancer are aged 70 or above, and men typically receive NSCLC diagnoses between 80 and 84, while women are most commonly diagnosed between 75 and 79.

In the past two decades, there has been a notable surge in the utilization of oral anticancer agents in cancer treatment, which now encompasses genetic mutations and traditional therapies. The use of oral forms exhibited a growth of 58.8% from 2008 to 2020[2]. Moreover, the expenses associated with anticancer therapies surged significantly, with overall costs escalating by 96.5% during this period. Particularly, oral forms saw a substantial 221.6% rise in costs, primarily attributable to the introduction of new and more expensive therapies. The cost of Osimertinib[3], a key treatment for lung cancer, is approximately $530 per dosage, posing a significant financial burden on patients and healthcare systems. The FDA estimates that more than 10,000 patients nationwide may be eligible for Osimertinib each year after undergoing tumor removal[4]. Humana, one of the leading Medicare advantage providers, holds a substantial 17.4% market share[5]. Assuming a similar market share in lung cancer treatments and Humana covering a remarkable 99% of the cost of treatment, the estimated value at risk for Humana in this context amounts to upwards of $300 million annually.

Humana Part D Prescription Plans	**Monthly Premium**	**Deductible**	**Annual Pay (i)**	**Humana Pays ((D)-(i))**	**Percentage (A)**
Humana Walmart Value Rx Plan PDP	$73.90	$545.00	$1,431.80	$192,018.20	99.26%
Humana Basic Rx Plan PDP	$85.60	$545.00	$1,572.20	$191,877.80	99.19%
Humana Premier Rx Plan PDP	$135.40	$0	$1,624.80	$191,825.20	99.16%
FDA estimates for Osimertinib Patients (annual) (B)	10,000		**Osimertinib Cost per Pill**		$530
Humana's Medicare Market Share (C)	17.40%		**Days of Consumption**		365
Value at Risk (A*B*C*D)	**$336,603,000**		**Cost Per Year (D)**		**$193,450**
4.5% Counter-Factual			**$15,147,135**		

[1] https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/statistics
[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9399396/
[3] https://data.cms.gov/summary-statistics-on-use-and-payments/medicare-medicaid-spending-by-drug/medicare-part-d-spending-by-drug/
[4] https://www.uclahealth.org/news/new-treatment-for-early-stage-lung-cancer-has-jami-way-living-the-normal-life-she-wants
[5] https://www.nytimes.com/2022/10/08/upshot/medicare-advantage-fraud-allegations.html

NSCLC with an epidermal growth factor receptors (EGFR) positive case is a pervasive issue with evolving survival rates, and the financial aspects of treatment are a crucial consideration for both patients and healthcare and insurance providers, such as Humana. Ensuring medication adherence is paramount in achieving favorable patient outcomes, with its impact often just as influential as the treatment itself. It directly affects the quality and duration of a patient's life, overall health results, and the economic burden on healthcare systems. Nonadherence to prescribed medications is a major contributor to treatment failures, leading to a staggering 125,000 deaths and up to 25% of hospitalizations in the United States annually[6]. Achieving adherence rates of 80% or higher is crucial for optimal therapeutic effectiveness, yet the reality paints a stark contrast with adherence to chronic medications hovering at around 50%[7]. Addressing medication adherence not only promises to enhance patient well-being but could also alleviate a substantial $100 to $300 billion in healthcare costs[8] each year, highlighting the financial and clinical significance of this issue.

1.2. Problem Statement

Patient outcomes have traditionally encompassed key metrics such as symptom recurrence, emergency visits, hospitalization rates, morbidity, and mortality. Healthcare providers like Humana can influence these factors by optimizing the timeliness of diagnosis, disease management, and treatment costs, along with facilitating patient access to healthcare resources and bolstering medication adherence. However, the persistent problem of medication nonadherence has resulted in deteriorating health conditions, inflated care costs, and increased healthcare expenditures. This analysis presents innovative strategies to tackle this challenge, which we address in this paper. By improving medication adherence, we can pave the way for enhanced patient outcomes, ultimately benefiting individuals and healthcare systems alike.

2. DATA ANALYSIS

2.1. Feature Engineering

Our model uses a minimal amount of feature engineering, which we will detail in this section. Due to the small amount of data, it was exceptionally easy to over-fit the data distribution and obtain low test-set performance. For this reason, a number of engineered features were not included due to their negative impact on model performance. These forgone features will be noted when we believe they would elevate performance given a larger collection of data.

[6] https://www.uspharmacist.com/article/medication-adherence-the-elephant-in-the-room
[7] Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc. 2011;86(4):304-314.
[8] DiMatteo MR. Variation in patients' adherence to medical recommendations. Medical Care. 2004; 42(3); 200-209.

From the target csv files, we take the "race_cd", "sex_cd", "est_age", "cms_disabled_ind", and "cms_low_income_ind" features directly with no modification. The features we use and derive from the other csv files are detailed in the follow medclms, rx, and date sections.

2.1.1. Medclms Based Features

These are the features we used based on the medclms csv files. We use the 8 variables from the claims file as-is: ade_diagnosis, seizure_diagnosis, pain_diagnosis, fatigue_diagnosis, nausea_diagnosis, hyperglycemia_diagnosis, constipation_diagnosis, diarrhea_diagnosis. The value of the diagnosis is taken as the last claim's set of diagnosis. The logic behind this decision is that we want an indication of the patient's most current status and ailments.

To augment these 8 features, we include a "history of" version of each feature as well. If a patient is ever diagnosed with any of the conditions, they receive a positive marker for the "history" variant of the feature identifier. In this manner we can distinguish between types of conditions.

2.1.2. Rx Based Features

Similar to mdclms, these are feature we used based on the rx csv files, and a similar strategy is used. The features ddi_ind, anticoag_ind, diarrhea_treat_ind, nausea_treat_ind, seizure_treat_ind were used as provided, with the last rx date used as the indicator. Again, a "history of" variant for each feature was included that contained a positive mark if any of the features were ever used in the patient's history.

An additional derived feature is 'total_cost' which we computed as the sum of 'rx_cost' + 'tot_drug_cost_accum_amt'. We found that keeping both feature separate made it easy to perfectly identify all patients using just these two features and promoted significant overfitting. Similarly, we found many cases where rx_cost was large while the accumulation was small, giving misleading indications on expense. Adding them together alleviated this issue. While this feature was still highly identifiable, we tackled the remaining risk via model design, which we will discuss in the next major section.

2.1.3. Date Based Features

While the current number of days is known for the training data, it is not known for the holdout data. To mitigate this, we devised an approximate estimate of the days based on the information contained in the medclms and rx files. Both contain, for each event, the date of the claim or rx event. The most recent of the dates was used to calculate a total number of days by comparing against the therapy start date in target_train/holdout.csv. Both training and testing were processed in this same way to ensure that the features are comparable between train/test time.

Notably, it is not the case that every medclms event has a corresponding rx event, and vice versa. For this reason, we include a "medclms_stale" and "rx_stale" feature that is the number of

days since a medclms or rx event respectively against the most recent known date. In this way the lack of a recent prescription or medical evaluation is encoded into the model for prediction. We expect the predictive power of this feature to be considerably greater in an operational environment where these can be calculated against the current date rather than estimated as was necessary in this competition.

2.2. Future Feature Engineering for More Data

Two feature engineering options in our code were not used because of overfitting. However, we believe they would provide improvements in accuracy if more data was available but are not viable with the limited number of samples given today are outlined below.

First was that the hum_drug_class_desc categories seemed useful, but too numerous to model with limited data, and so a smaller set of classes may work instead. We attempted to coalesce the different drug categories in hum_drug_class_desc into a smaller set by using the Levenshtein edit distance between descriptions into a configurable number of categories. However, the number of drug types was still too numerous and enabled tight patient identification, and so was removed.

A second item we considered is a more intelligent encoding of the patient history in medclms and rx events. The mdeclms encoding currently used does not reflect that the underlying data is a sequence of events over time, and instead appears as a single snapshot irrespective of how the events transpired. We devised a feature engineering approach that would rectify this limitation.

In particular if h_t represents the current history, and d_t the current date, we can update the history to the next date d_{t+1} using the new event features x_t:

$$\mathrm{h_{t+1}} = \mathrm{h_t} \cdot \gamma^{(d_{t+1}-d_t)/\rho} + x_{t+1}\,(1 - \gamma^{(d_{t+1}-d_t)/\rho})$$

Where $\gamma \in (0, 1)$ is a decay constant (that we would expect to set to around 0.3) and $\rho \geq 1$ would be a scale parameter that controls how many days it takes before the history is "forgotten". Our anticipation is setting $\rho = 30$ to 60 days would correspond to a hypothesis that after a few months the claim history is no longer relevant to the current prediction or condition.

This exponential decaying weighted average allows us to create a feature representation that captures the current and historical claims/prescriptions that allows the model to select a feature based on recency. Any value that is ever used will always be > 0, allowing the same kind of "any history" check as is done with our current simplified history indicator. However, large values near 1 like 0.6 would indicate a very recent event, while small values like 0.05 would dictate an event several months ago. When we used this feature, we found that simple models were able to obtain near 100% training accuracy and fairness with low validation-step performance, indicating that the representation was too informative for the amount of training data available due to overfitting.

3. MODELING

Our modeling approach is to provide a long-term solution to Humana's recurring problem of needing to balance accuracy and disparity concurrently. Existing approaches are unsatisfying as they require manual inspection of which models happened to have the best tradeoff, and don't allow using industry standard tools like grid-search to optimize the target goal. We solve this by designing a new loss function that handles both goals simultaneously, allowing us to train and optimize one model directly for the task at hand, and should continue to work when more data becomes available in a production setting, or applied to other predictive tasks entirely.

In developing our model, we take into consideration two primary factors that inform the design. First is the small amount of data available to train from, only 1,232 rows. This presents a significant risk of over-fitting the training distribution. Second is that the model must respect the disparity in true-positive-rates (TPR) as set by the competition criteria.

For both issues we will use the XGBoost library[9] as a unifying framework. However, we will not use the boosting feature of the library due to the risk of overfitting. Our testing indicated that Boosting would be unlikely to generalize to the test distribution and exhibited high variance in AUC.

Instead, we use XGBoost to implement a Random Forest[10] style model as a means of mitigating the risk of overfitting. A total of 350 trees were built independently and assembled together via averaging. A standard 50% feature selection was used on a per-tree, per-level, and per-node induction step. Contrary to a normal Random Forest that uses bagging[11], we subsampled the data for each tree induction to just 10% of the total data. By comparison, bagging would have on average $\approx 67\%$ of the original data represented at each induction. Our aggressive sub-sampling was done to exploit the bias-variance tradeoff[12]:

$$\underbrace{\mathrm{E}_{x,y,D}[(f_D(x)-y)^2]}_{\text{Expected Test Error}} = \underbrace{\mathrm{E}_x\left[\left(\bar{f}(x)-\bar{y}(x)\right)^2\right]}_{\text{Bias}^2} + \underbrace{\mathrm{E}_{x,D}\left[\left(f_D(x)-\bar{f}(x)\right)^2\right]}_{\text{Variance}} + \underbrace{\mathrm{E}_{x,y}[(\bar{y}(x)-y)^2]}_{\text{Noise}}$$

Because of the limited amount of data, the risk of error due to the bias term is especially high. The data in our testing is also easy to over-fit, which subtly reduces the variance when trained on most of the dataset and necessitated us removing a number of aforementioned engineered features. By increasing the subsampling to an extreme degree, we trade the individual tree induction to favor a low bias for higher variance in the error. Per the normal theory of Random

[9] https://xgboost.readthedocs.io/en/stable/index.html
[10] Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
[11] Breiman, L. Bagging predictors. *Mach Learn* **24**, 123–140 (1996). https://doi.org/10.1007/BF00058655
[12] https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote12.html

Forests, averaging over many trees then reduces the error attributable to the variance term, and thus provides us a strategy to mitigate the overfitting risk in this case.

Second is the issue of fairness in the model's predictions. An XGBoost implementation is used because it provides the facilities for defining a customized loss function to incorporate this into the tree induction process, providing a principled method of handling the disparity in treatment.

Let $y \in \{0,1\}$ represent the true training label denoted by TGT_ADE_DC_IND in the target_train.csv file. Our model's predictions $\hat{y}$ will be continuous values in the range of [0, 1]. Standard XGBoost will train with the binary cross entropy loss:

$$\ell_{BCE}(y, \hat{y}) = y \log \hat{y} + (1 - y) \log(1 - \hat{y})$$

Our desire is for a holistic process that satisfies both criteria and does not require manual intervention or redesign if more data becomes available. To quantify the equitable TPR between races and sexes the given formula is used to compute a raw disparity penalty via:

$$DisparityScore = \frac{1}{|Sexes| + |Races|}\left(\sum_{s \in Sexes} \frac{TPR(s)}{TPR(Male)} + \sum_{r \in Races} \frac{TPR(r)}{TPR(White)}\right)$$

However, this calculation is non-differentiable, meaning modern machine learning methods cannot use it as a target to learn from. We choose to incorporate this calculation directly into a new loss function via a continuous relaxation of the TPR calculation. We can define the smooth TP approximation as $sTP_g = y \cdot \sigma\left(2\left(\hat{y} - \frac{1}{2}\right)\right) \cdot g$ where g is used to indicate a 1/0 value for a specific group identity (race or sex). The sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ is used to squash a value into the range [0,1], and so we shift $\hat{y}$ by 0.5 to make the 50% value correspond to 0.5 as the output of the sigmoid. The multiplication by 2 makes the sigmoid sharper, closer approximating the indicator function (which would return the true TP). Similarly, the smooth false positives $sFP_g = y \cdot \left(1 - \left(2\left(\hat{y} - \frac{1}{2}\right)\right) \cdot g\right)$.

Combined, we compute the smoothed disparity score as an approximation of the true objective, where g' indicates the majority population corresponding to group g, returning:

$$DS = \sum_{g \in \{Sexes, Races\}} \frac{\frac{sTP_g}{sTP_g + sFP_g}}{\frac{sTP_{g'}}{sTP_{g'} + sFP_{g'}}}$$

This above calculation is numerically unstable and results in "Not a Number" issues if attempted during training. So we must instead compute the log(DS):

$$logDS = \sum_{g \in \{Sexes, Races\}} LogSumExp\left(\left(\log sTP_g - \log\left(sTP_g + sFP_g\right)\right) - \left(\log sTP_{g'} - \log\left(sTP_{g'} + sFP_{g'}\right)\right)\right)$$

Where LogSumExp is the "Log summation exponentiation" trick commonly used to evaluate such equations in a numerically sable way. In brief, it works by noting that $\log \sum_i e^{x_i} = \max_j x_j - \log \sum_i e^{x_i - \max_j x_j}$, but avoids numerical overflow in all calculations. This allows us to compute a smooth and differentiable approximation of the Disparity Score.

This is then combined to form our final training loss function, where $\lambda = 1.05$ is added as a regularizer to the disparity calculation:

$$\ell(y^d, \hat{y}) = \frac{\ell_{BCE}(y^d, \hat{y})}{e^{-logDS - \log \lambda}}$$

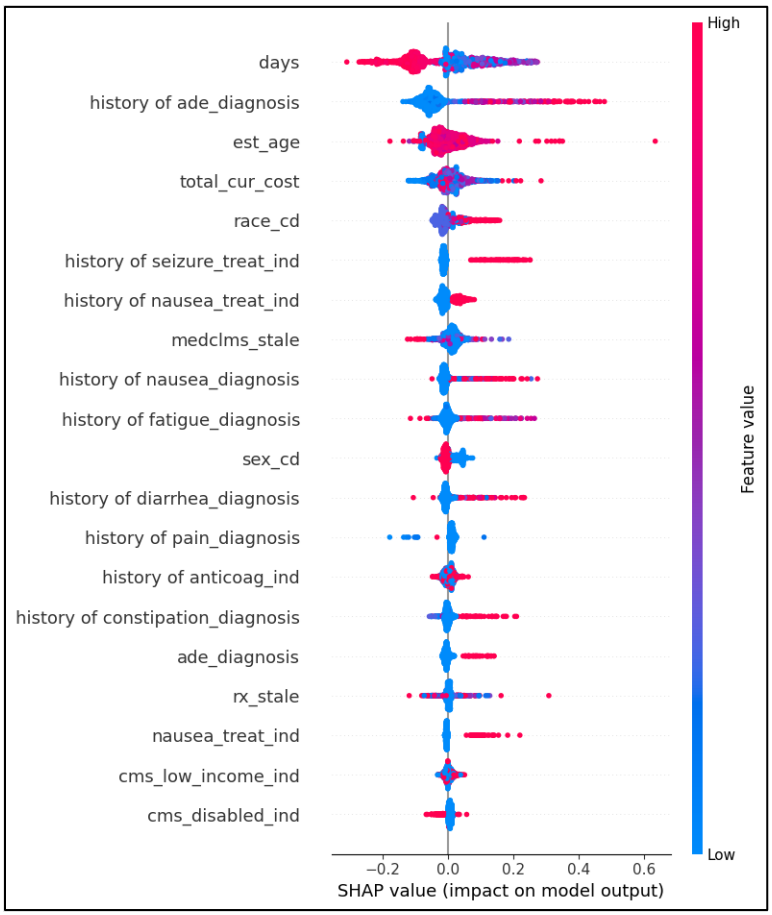
This calculation will encourage the model to be accurate and obtain good rankings for improved AUC, while also inflating the loss when the disparity between groups is higher. This allows us to train XGBoost directly to tackle the objective, without having to attempt less principled post-hoc corrections. The λ term further accounts for the fact that we do not know what the test-time disparity is, and so we penalize the model toward over-compensating on disparity due to its multiplicative depressive effect on the AUC score. $\frac{\partial}{\partial x}\ell(y^d, \hat{y})$ is computed via automatic differentiation using the JAX framework[13] for the gradient posting algorithm. XGBoost also requires the diagonal of the hessian, which we instead approximate by the constant 1 due to the $O(N^2)$ quadratic complexity of forming the hessian. Since this controls only the step size of the boosting algorithm, the worst possible impact is we require more iterations of boosting to obtain our model.

4. ANALYSIS

Our analysis begins with an inspection of the features' level of importance within the model based on the data. We use Shapley Additive explanations (SHAP) for two reasons. First, SHAP scores provide a theoretically grounded and useful way of interpreting feature importance and impacts, where the score indicates the marginal improvement in outcome averaged over all scenarios. Second, the "averaging over all scenarios" usually makes it computationally intractable, but the SHAP library has algorithms specific to decision trees that can compute SHAP exactly in reasonable time.

[13] https://jax.readthedocs.io/en/latest/index.html

The high-level analysis of the SHAP feature importance's is shown in the below figure. Note that in each figure one dot corresponds to one data point, as SHAP tells us the influence of each individual's value. A dot with a red color means the feature's value was high compared to all values in the dataset, and blue indicates that the feature had a low value. A dot occurring to the right indicates it increased the likelihood of a person having an ADE, and a dot to the left indicates a reducing in ADE likelihood. To rank the features, they are sorted from highest average magnitude to lowest.



We can immediately see a number of informative population-level trends. First is that the number of days the patient has successfully been on the treatment is the highest value predictor, which is intuitive. A patient who has made it 179 days has intrinsically been tolerant of the medication and side effects for a longer amount of time, and thus likely to continue to be tolerant of the treatment. A similar unsurprising result is the importance of patient age, with the oldest patients having a dramatic increase in risk of ADE discontinuation compared to the larger population.

Also of note is that the race and sex of the patient are of relatively high importance. This is by construction of our design, as the algorithm is incentivized to use these features to ensure a low disparity in treatment across races and sexes. In creating our model, we considered the legal

and ethical concerns of using protected attributes. No legal framework currently exists that directly controls how protected attributes are allowed to be used in the model. In building the model, we followed Humana's AI policy[14] covering protected attributes. We decided to explicitly use race and sex in our model to mitigate bias and eliminate the potential for a disparate impact among the protected attributes. Race and sex were both significant features in our model, so we decided it was important to explicitly include them rather than omit the features.

A broad theme of our results is that the "history of" features have a higher importance than the current features. We remind the reader from our Feature Engineering section that every "history of" feature has a corresponding current counterpart based on the most recent visit and diagnosis. This would indicate that the likelihood of discontinuation is a function of latent patient health factors, rather than acute diagnosis. This suggests more investigation into other factors could yield improved understanding of outcomes, and that once any diagnosis of an additional condition occurs Humana should temper its prognosis for that individual.

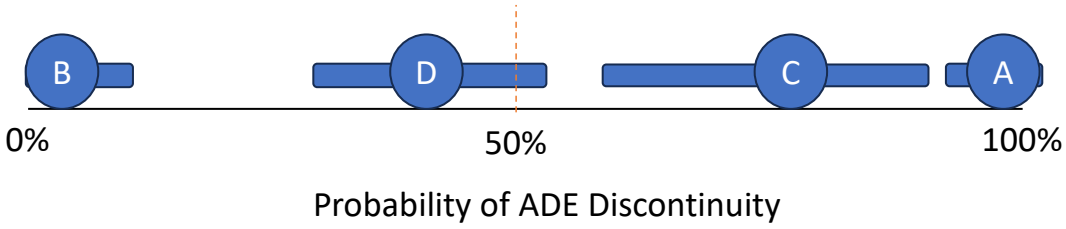
4.1. Intervention Analysis

The ultimate goal of this study is to determine if it is possible for some intervention to benefit a treatment patient, which would result in better outcomes for that individual and lower expenses for Humana. This requires careful analysis as a causal question. That is to say, asking the statistical question "what is correlated with successful/unsuccessful patients" may sound intuitive but does not inform how we can intervene to change the results. In the below diagram, this would correspond to considering patients "A" and "B" as the most important/informative.

Instead, we a method for causal analysis called do-calculus to answer the causal question: "which patients may have better outcomes if we intervened in the course of their treatment". To understand the importance of this perspective, consider the below diagram. Here the points have been augmented with a bar indicating the range of causal impact. That is to say, *how much could the probability for this specific individual be altered by an intervention*? Answering this is a counter-factual question of what an individual's risk of ADE termination in a different world "counter to the facts" that occurred.

[14] https://humana.gcs-web.com/static-files/8060a8b7-214d-471b-b2a2-c5d3f495a065

The below figure shows how we use do-calculus to compute these counter-factuals that directly answer Humana's question. The width of the bar indicates the range of possible counter-factual scores for each datapoint, indicating what could have been if some intervention had occurred. The naïve analysis would have preferred studying patients A and B, but little could have been done to alter their outcomes.



The most informative are actually patients D and C, which have wider counter-factual possibilities. This means Humana could have a large impact on the probability of D and C discontinuing the treatment, and thus increase patient survival and profits. Toward Humana's interest, there are two possible ways to "rank" the data that may inform interventions.

a) One could rank each individual by their counter-factual risk of crossing the 50% probability threshold. This focuses interventions on patients where Humana could change the outcome from unlikely-to-likely, and patients at risk of the opposite transition.
b) One could rank each individual by the magnitude of the counter-factual difference from their current prediction. This focuses interventions on patients where Humana has the greatest risk in outcome, and where an intervention will have the largest absolute impact on reducing that risk.

4.2. Counter-Factual Intervention Results

To obtain our counter-factual estimates, we must choose a set of features that Humana could intervene upon. Most of the features available in the data do not qualify as potential interventions, e.g., Humana cannot change a patient's diagnosis of a condition, or somehow force them to not develop side effects. We have identified four such features that are intervenable, meaning Humana could meaningfully impact them or their severity:

- *Low Income (1):* encoded by cms_low_income_ind. The basis for its selection is that Humana can reduce prices, increase coverage, waive premiums, and other financial disbursements to mitigate the individual's current limited income. If these disbursements are less than the expected gain, then it is a viable intervention.
- *Disabled individuals (2):* cms_disabled_ind. The basis for its selection is that, while Humana cannot remove someone's disability, it can provide coverage for caretakers, travel nurses, and

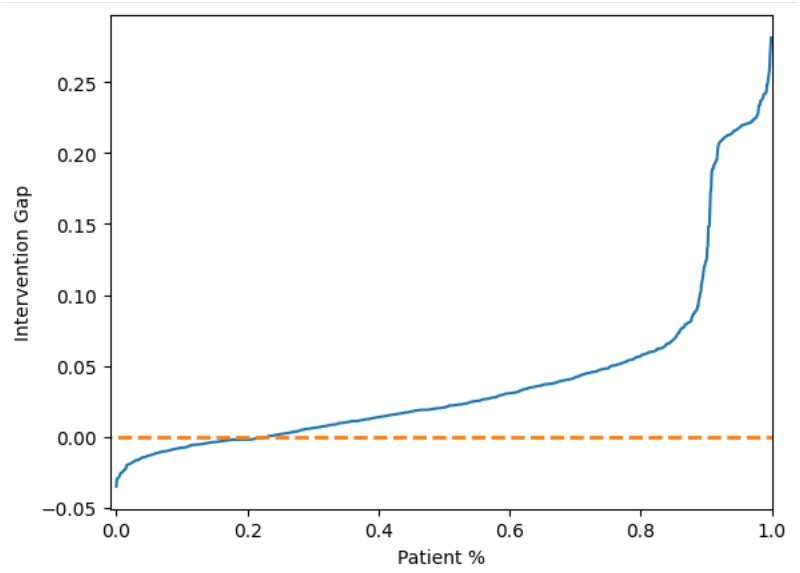
alternative/streamlines claim filing, to reduce the impact that the individual's disability has on their lives while in treatment.

- *Delayed treatment (3 & 4):* our derived features medclms_stale and rx_stale. The basis for these two variables is the same. If an individual's risk is high due to not having seen a doctor or pharmacist recently, traveling technicians, tele-health, and pharmaceutical mailing are all interventions within Humana's capabilities to intervene and bring the patient up to a non-stale outcome.

To estimate the causal effect, we use SHAP scores to compute E[f(X) | do(X_i=intervention)] for each of the four interventions listed above individually, and then take the sum of absolute magnitudes as the estimate of the counter-factual range of impact. In this notation $f(X)$ is the output of our model, do(X_i=intervention) indicates the evaluation of the counter-factual world where the i'th feature has been set to the given intervention.

4.3. Intervention for Outcome

We first look at the scenario where we rank everyone by the counter-factual ability to change their likely outcome (i.e., cross the 50% threshold). This is computed as $abs(f(X) - 50\%) - (f(X) - abs(\text{E[f(X) | do}(X_i\text{=intervention)]}))$. This gives us a score that is negative when the intervention could push the score above/below 50%. The results are shown in the below figure, where the dashed line indicates the threshold of success.

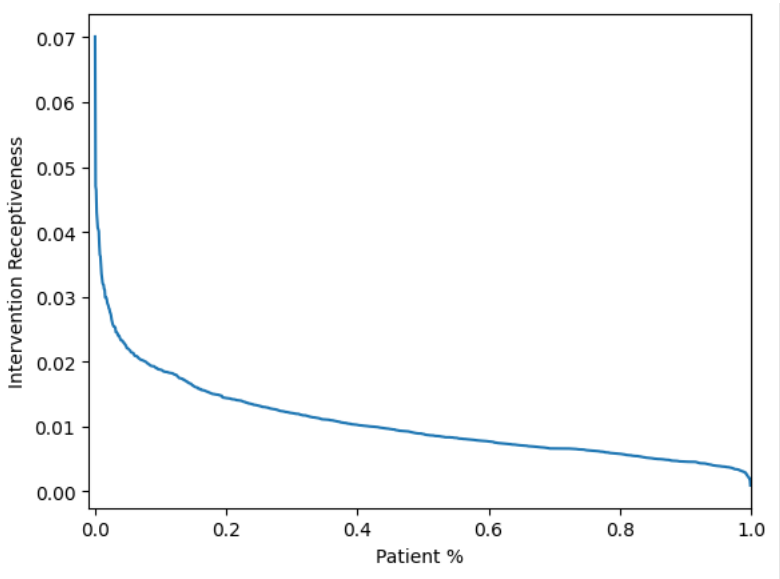


In this analysis we can see that 18% of the patient population could have their probability of being more or less likely to have an ADE that causes them to stop treatment reduced. Note that the right tail of the distribution goes up to 25% on the y-axis. This indicates that, even with

intervention, these patients are still an absolute 25 points away from the 50% mark. Unfortunately, such patients are, according to the model, unlikely to have their individual outcomes changed.

4.4. Intervention for Impact

Now we consider the selection of patients not by their probable change in outcome, but the impact in their change in probability. This is a simpler calculation of abs(E[f(X) | do(X_i=intervention)]). The results are plotted in the below figure.



Note that there is a small population of patients for whom a large interventional impact is possible, altering their outcome by up to 7%. Combined with the previous result, this also indicates that many patients are already near the 50% threshold.

4.5. Recommended Trial of Results

Based on this analysis, we can propose the following recommendation to trial the impact of interventions. The absolute magnitude of intervention success is muted according to the model, and a non-trivial number of patients are in a "could go either way" situation.

Our recommendation is to trial interventions on the union of the two intervention strategies: target patients who are near 50% change of discontinuity, and who are maximally receptive of an intervention. This can be computed on an individual basis allowing for customized intervention protocols to maximize the probability of successful outcomes.

If we define 2% as the minimum for a "large impact" of intervention, we find that only 95 patients in the data satisfy this criteria, and 286 patients satisfy the 50%-impact criteria. Only 39 patients satisfy both criteria, about 4.5% of the original patient population. This represents a small

enough trial population to manage overhead, while also providing the group most likely to result in a measurable outcome and individual benefit to intervention.

4.6. Development Considerations

The evaluation score can be succinctly summarized as computing the AUC of the predictions and multiplying them by the disparity in true-positive rates. Because AUC is a quantification of the quality of a ranking of predictions (i.e., are all successful therapy patients predicted higher than all unsuccessful patients), it makes intuitive sense from operationally intervening only on the cases where you are most likely to have an impact. The disparity score then balances this desire with that of inequitable treatment.

However, the chosen quantification of the operational objective may lead to inadvertent and undesirable outcomes that forge fairness entirely. Because disparity is calculated by TPR, it is possible to force all predictions to be negative, resulting in no apparent disparity. The AUC metric is invariant to monotonic transformations, meaning we can calculate: $\widehat{y'} = \frac{\hat{y}}{2} - 0.5$ to produce a set of predictions that maintain AUC while subverting the desired disparity concern.

While we have not, and do not, suggest using this approach in modeling, it points to a potential weakness in the chosen criteria. Data scientists operating in good faith attempting to maximize the objective may inadvertently discover strategies that pursue this metric, resulting in undesirable outcomes.

The basis of this issue stems from the combination of a relative score (AUC) and an absolute score (Disparity via true positive rates) into one combined value. This can be remediated by changing both to an absolute or a relative score. Because Humana's deployment of these models is likely to be used as a prioritization scheme, the AUC metric is well aligned with operational concerns. For this reason, we suggest altering disparity to also use a relative measure. One simple option would be to compute the per-protected group AUC instead of TPR and perform the calculation in the same manner as before.

5. BUSINESS IMPLICATIONS

In this section we will revie multiple methods of increasing patient adherence to medication in oncological situations. A consistent result across the cited medical literature is that adherence rates can be improved by 20-25% in the majority of cases. Based on these studies we believe it is reasonable to assume that the 4.5% identified by our analysis could be improved to the same 20-25% rate. This would lift the net impact potential savings to $20-$75 million.

5.1. Lack of Drug Adherence Harmful to Patients and Humana

With the rise of oral chemotherapy as opposed to the traditional intravenous chemotherapy, which was administered directly by healthcare personnel, patients are becoming increasingly responsible for the administration of their own treatment. While the flexibility this offers can be positive, the implications of this shift on the successful adherence to and completion of the prescribed course of treatment are significant. Negative ramifications of non-adherence most often include an "increase in physician visits, increased hospitalization rates, longer hospital stays, decreased patient satisfaction, poor patient-provider relationships, [and] compromised disease outcomes, such as decreased time to relapse and decreased survival."[15] Proper adherence is absolutely essential for the patient, and if the occurrence of adverse drug events (ADE) leads to a disruption or discontinuance, their consequences are quite severe to the point of impacting their survival. The increased responsibility of self-monitoring of side effects and strict adherence to the medication schedule raises the issue that "… patients are known to experience severe symptoms and miss as many as one-third of the prescribed medication doses."[16]

Lack of adherence leads to negative consequences for Humana as well. Studies show that "poor adherence has been linked to successive hospitalization, increased need for medical interventions, morbidity, and mortality and results in increased healthcare costs, with North America having estimates of approximately $100 billion being spent annually and $2000 spent per patient per year for additional visits to the physician."[17] The cost implications of sub-optimal medication adherence make this a valuable area of study. Our recommended interventions below can help mitigate the risk of non-adherence, improving not only patient survival rates but the financial impact to Humana as well.

5.2. Current Humana Resources for Adherence and Cancer Support

Humana currently offers several services to support cancer patients, including the Centerwell Specialty Pharmacy, Cancer Center of Excellence, and the Centerwell Pharmacy App. These resources collectively provide cancer patients with tools for managing side effects and maintaining medication adherence. The Centerwell Specialty Pharmacy and Cancer Center of Excellence offer clinical support, such as medication management, educational materials, and financial assistance. They also provide 24/7 access to pharmacists and a dedicated nurse who establishes a personal connection with patients through regular check-ins and phone consultations. The Centerwell Pharmacy Mobile App enables patients to conveniently refill prescriptions, set

[15] https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote12.html

[16] Greer, J.A., Amoyal, N., Nisotel, L., Fishbein, J.N., MacDonald, J., Stagl, J., . . . Pirl, W.F. (2016). A systematic review of adherence to oral antineoplastic therapies. Oncologist, 21(3), 354–376. https://doi.org/10.1634/theoncologist.2015-0405

[17] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6858102/

medication reminders, and manage their prescription and doctor information.[18] Additionally, Humana offers meal delivery with a cancer support diet option.

5.3. Current Resources used in the Market for Increased Oral Therapy Adherence

Many of the typical adherence strategies and interventions are currently being used by Humana to help patients with their treatments. The common techniques include an emphasis on patient education and communication, counseling regarding common side effects, and reminders to take medication.[19] The industry is currently seeing a shift in using more technology to assist with adherence strategies. The use of mobile applications is becoming more popular for symptom management and medication adherence. A previous worry revolved around the limited mobile phone ownership coined the "digital divide", which posed a challenge for tech-focused solutions. These worries have subsided, "According to the Pew Research Center (2021), 97% of Americans own mobile phones, 85% of which are smartphones."[20]

The growing trend of smartphone usage has led to an increasing number of studies looking at whether mobile applications can improve medication adherence among patients with cancer. The studies are showing positive results and are helpful to demonstrate which functionalities Humana should implement in their current mobile app offering.[21] The mobile applications are being used for more timely assessment and management of symptoms.[22] Studies have found the following functionalities to be beneficial: (1) personalized medication dosing schedule with reminder option, (2) educational resources for symptom management including information on common and urgent symptoms and recommendations for management of the symptoms, (3) symptom reporting module, (4) push (app) notifications with reminders to take medication and report side effect symptoms, (5) automated phone call and text reminders, and (6) Q&A section covering treatment of symptoms and other questions regarding treatment.[23] These functionalities were used in different studies in differing styles, but all yielded promising results around symptom management and medication adherence. In addition to the beneficial features used in the applications, it is important to note the differing approaches the app developers chose to increase engagement and communication.

5.4. Onboarding, Engagement, and Retention

Through proper onboarding and education regarding usage of the mobile app, any person regardless of age or tech literacy will be able to utilize the application. This process begins when

---

[18] https://www.centerwellspecialtypharmacy.com/center-of-excellence/cancer
[19] https://ashpublications.org/ashclinicalnews/news/2634/Top-Ten-Tips-and-Tricks-for-Treatment-Adherence
[20] https://www.pewresearch.org/internet/fact-sheet/mobile/
[21] Cazeau N. Mobile Health Interventions: Examining Medication Adherence Outcomes Among Patients With Cancer. Clin J Oncol Nurs. 2021 Aug 1;25(4):431-438. doi: 10.1188/21.CJON.431-438. PMID: 34269338; PMCID: PMC9642910.
[22] Karaaslan-Eşer A, Ayaz-Alkaya S. The effect of a mobile application on treatment adherence and symptom management in patients using oral anticancer agents: A randomized controlled trial. Eur J Oncol Nurs. 2021 Jun;52:101969. doi: 10.1016/j.ejon.2021.101969. Epub 2021 May 4. PMID: 33991868.
[23] Cazeau N. Mobile Health Interventions: Examining Medication Adherence Outcomes Among Patients With Cancer. Clin J Oncol Nurs. 2021 Aug 1;25(4):431-438. doi: 10.1188/21.CJON.431-438. PMID: 34269338; PMCID: PMC9642910.

treatment first starts with the patient. Staff at Humana can be trained and help the patient understand how to get started and use the application. It is important that patients are engaged with the application. Studies have shown that "… greater engagement with the mobile app was associated with improved adherence and fewer ED visits leading to hospitalization."[24] The ability to keep patients engaged is very important so that they receive the most potential benefit from the application. Additional studies showed that interactive exercises and "…use of tailored logic, educational games (i.e., gamification), adaptive skill-building, and a record of personal progress through the app content, have shown greater than 85% usability and retention rates."[25]

Furthermore, other studies examined mobile gaming apps that "included personal avatars, social networking, information on symptom management, and psychological support."[26] The study found that the gaming functionality allowed for better medication adherence in the tested group. The social networking function of the mobile gaming app also allowed for the patients to connect with other individuals with cancer to connect on an emotional level with others going through the same circumstances. The selection of functionality in the mobile application is crucial to the retention and engagement of the patient.

5.5. Communication and Accessibility

The use of a mobile application for medication adherence and symptom management has proven beneficial in regard to communication and accessibility between the patient and healthcare team. The use of mobile applications, especially the symptom reporting function allows for the patient to notify their healthcare team instantly and provide an opportunity to be monitored around the clock. "Such standardization for remote monitoring and clinician response to patient reports would ideally enhance engagement with mobile health interventions and sustain adherence over the long term."[27] Enabling patients to share their symptoms and receive prompt feedback via the app could mitigate avoidable delays and foster better adherence and communication between the patient and the healthcare team. Many patients avoid seeking assistance with side effects because they anticipate delays in contacting the healthcare team and receiving guidance. Enhanced communication and improved accessibility can be achieved through the utilization of mobile applications for managing symptoms and adhering to medication.

---

[24] Greer, J.A., Jacobs, J.M., Pensak, N., Nisotel, L.E., Fishbein, J., MacDonald, J., Ream, M., Fuh, C., Buzaglo, J.S., Muzikansky, A., Lennes, I.T., Safren, S., Pirl, W.F., Temel, J.S., 2017. Randomized trial of a smartphone mobile app for adherence to oral chemotherapy. J. Clin. Oncol. 20 https://doi.org/10.1200/JCO.2017.35.15_ suppl.10055, 35 (15_Suppl. l), 10055–10055.

[25] Payne, H. E., Lister, C., West, J. H., and Bernhardt, J. M. (2015). Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 3:e20. doi: 10.2196/mhealth.3335

[26] Cazeau N. Mobile Health Interventions: Examining Medication Adherence Outcomes Among Patients With Cancer. Clin J Oncol Nurs. 2021 Aug 1;25(4):431-438. doi: 10.1188/21.CJON.431-438. PMID: 34269338; PMCID: PMC9642910.

[27] Karaaslan-Eşer A, Ayaz-Alkaya S. The effect of a mobile application on treatment adherence and symptom management in patients using oral anticancer agents: A randomized controlled trial. Eur J Oncol Nurs. 2021 Jun;52:101969. doi: 10.1016/j.ejon.2021.101969. Epub 2021 May 4. PMID: 33991868.

6. BUSINESS SOLUTION

Our recommended solution is that Humana concentrate on enhancing and expanding the features of their mobile applications. Our recommendation is supported by the results of our model and the growing medical literature exploring this topic as described above. The four features we chose to intervene upon based on the model, (1) low income, (2) disabilities, (3) delay in seeing a doctor, and (4) pharmacy delays, will be addressed by making the changes outlined below to Humana's mobile application. Additionally, while we cannot directly intervene regarding a patient's history of a diagnosed side effect, their history does greatly impact their treatment. Therefore, we have included recommendations to mitigate this risk as well.

As part of our recommendation, we believe Humana can improve patient outcomes for medication adherence and ADE symptom management by making their current resources available through a mobile application. Currently, Humana has four mobile applications, (1) MyHumana, (2) Go365, (3) Go365 for Humana Healthy Horizons, and (4) CenterWell Pharmacy. There appears to be a barrier between all the resources available to patients and how they can access them. Most of the resources Humana currently offers to cancer patients listed above are not accessible through any of their application offerings. Most of the resources are listed on the website but could be improved by providing access via the mobile application.

We suggest either repurposing one of Humana's current applications or developing a new one that consolidates all of Humana's current adherence and support resources. It's crucial to ensure that this application is accessible to all Humana patients, not limited to those exclusively using CenterWell Pharmacy (i.e., CenterWell Pharmacy app may not be the best option for delivery of these resources if not all Humana patients are utilizing the CenterWell pharmacy.) Below we will describe which features Humana should include from their current resources, but also new features that should be included based on our intervened features and the medical studies.

6.1. Existing Humana Resources that should be Included in the Application

6.1.1. Medication Reminders and Dosing Schedule

The current medication reminder provided by Humana in the CenterWell pharmacy app should be retained. However, we believe there is room for improvement to enhance its effectiveness. Rather than requiring patients to manually input the medication reminder, healthcare providers should preconfigure the reminder to align with the relevant medication dosing schedule. The existing push notification format, which appears as a pop-up message on smartphones, should be maintained. Furthermore, we also recommend introducing additional delivery methods for reminders, such as automatic telephone calls and text messages.

6.1.2. Medication Refill and Delivery

Humana should continue utilizing their medication refill and delivery services, which are already available through the CenterWell Pharmacy App. Facilitating easy medication refills and home delivery is vital for improving accessibility to these essential medications. Increased accessibility has a beneficial effect on the pharmacy delays feature we used in our model. Additionally, the low-income feature in our model comes into play here, as Humana and CenterWell Pharmacy can reduce costs associated with utilizing their services for medication refill and delivery.

6.1.3. Telehealth

Humana currently has a Telehealth program available to their members, and we recommend building it into the application to make it more easily accessible and seamless with the rest of the features. Their account can be linked to their medical provider so that they can easily make virtual appointments, which may be especially helpful for those who are experiencing fatigue from their treatment.[28] The model demonstrates that a delay in seeing a doctor significantly impacts a patient's probability of experiencing an ADE that leads to them not adhering to the treatment plan, so this is an important issue to address that could be alleviated with a built-in telehealth function that would foster improved communication with the patient. Lastly, the telehealth function will serve as a beneficial intervention on the feature focused on delays in seeing a doctor used in the model. Patients will now be able to connect with a healthcare provider when needed 24/7, therefore decreasing usual delays.

6.1.4. Humana Well Dine®

Humana currently provides a meal delivery service to eligible members during recovery and for those with chronic conditions. They offer medically tailored diet plans, including one designed for cancer support.[29] This is a valuable program and including it as a feature within the application would make it more easily accessible to patients and help increase engagement with the app. It would also help reduce the risk of malnutrition in patients, which can impact their quality of life and survival rate. Patients could use the app to sign up for the meal service, view information about their provided diet plan, and track deliveries. Having a central, streamlined location for patients to view the services available to them would make these services more accessible and easier to manage for people who are already juggling many different aspects of their treatment plans. While this program would be helpful to anyone suffering with a chronic illness, it would have an even greater impact based on the low-income feature in our model. Humana's intervention here in providing meals to low-income patients who suffer from food insecurity could prove significant in helping them complete their course of treatment.

[28] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7489872/
[29] https://www.humana.com/home-care/well-dine

6.2. New Features we Recommend be Included in the Application

6.2.1. Option to Report Symptoms

Within the application, we recommend a feature providing patients with a prompt to report any side effects or symptoms they may be experiencing. This will facilitate tracking their symptoms over time, which not only would assist their doctors in having a more detailed understanding of their health but would also provide Humana with valuable data for the continuous improvement of this course of treatment. If the patient selects certain symptoms, or a significant severity or combination of symptoms, the app could raise a flag and notify the health care professional, giving them the opportunity to intervene with the goal of treating the symptoms and increasing the patient's probability of continued adherence to the treatment.

6.2.2. Standing Prescription Orders

As our model demonstrates, a patient having previously experienced an adverse drug event (ADE) strongly increases their probability of having another while taking this medication. This has a high impact on their likelihood of properly adhering to the treatment through completion. Our model also indicates that delays in communicating with a doctor and with receiving prescriptions from the pharmacy can negatively impact a patient's chances of successfully completing the treatment. Based on this knowledge, we recommend building a feature in the Humana application to include standing prescription orders for patients with an identified predisposition for common side effects. For example, if a patient has a documented history of nausea with similar treatments, their doctor can pre-emptively list the appropriate medication for this symptom in the patient's medication portal in the application when they begin treatment. If the patient begins experiencing this side effect as the treatment progresses, they can easily select the medication from their app menu, immediately notifying the health care provider's office who can seamlessly approve it or call the patient with any questions.

6.2.3. Side Effect Guide and Nutrition Plan

Patient education is critical to proper oral chemotherapy adherence, and the information given out during doctor's appointments while the patient may already be feeling overwhelmed or fatigued is often insufficient. It is essential that the patient have easy access to all relevant instructions regarding their treatment, such as dosage, what to do if a dose is missed, and any food or beverage interactions that may impact the treatment's efficacy and their side effects.[30] We recommend that an informational feature with treatment instructions and frequently asked questions be made available in the application, along with a guide explaining common side effects

[30] https://www.accc-cancer.org/docs/Documents/oncology-issues/articles/2003-2016/2008/JA08/ja08-improving-patient-adherence-with-oral-chemotherapy

and options for alleviating them. The impact of diet and nutrition is especially significant in symptom management, especially for ADE's like nausea, diarrhea, and constipation.[31]

Our model shows that these side effects increase a patient's likelihood of nonadherence or discontinuation of treatment, and providing patients with the information they need can empower them to take steps to manage these symptoms. We recommend that Humana work with their nutritionists to provide a nutrition guide tailored to small cell lung cancer, as well as how certain foods and beverages have the potential to alleviate symptoms for some patients. "Web-based nutrition" has been shown to help patients manage their side effects.[32] Providing access to this information, along with the option to speak directly with a nutritionist in-person or via telehealth would be a valuable addition to the Humana application.

6.2.4. Game and Reward Feature

We propose that Humana develops the mobile application with an interactive gaming format. Research indicates that mobile apps designed as interactive games boost engagement, resulting in improved medication adherence.[33] Furthermore, the game should incorporate a social networking feature, enabling patients to connect with others undergoing the same cancer treatment. This social aspect has been shown to provide valuable psychological support for patients in managing their symptoms and sticking to their medication regimen. Finally, as a complement to the interactive gaming format, Humana should introduce a reward system to encourage active participation. Humana can replicate their existing Go365 reward programs to create a reward system that incentivizes the patients to report their symptoms and adhere to their medication treatment plan. The combination of the interactive gaming format and the reward system will create a mobile application that promotes engagement and retention of the patient.

7. CONCLUSION

Addressing the critical issue of medication adherence and symptom management among cancer patients is paramount for both the well-being of the patients and the financial stability of Humana. The shift from traditional intravenous chemotherapy to oral treatments has placed greater responsibility on patients, making adherence and side effect management vital to successful outcomes. We propose implementing an XGBoost model to identify patients who take Osimertinib and who are at risk of discontinuing treatment due to the occurrence of ADE's. Our final model had an AUC of 0.832 on the official leaderboard, which indicates that our model performs well and can be implemented with confidence. One the model is implemented, we propose a solution focused on consolidating some of Humana's existing programs into one central application for

[31] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8781744/
[32] https://www.frontiersin.org/articles/10.3389/fnut.2023.1134793/full
[33] Cazeau N. Mobile Health Interventions: Examining Medication Adherence Outcomes Among Patients With Cancer. Clin J Oncol Nurs. 2021 Aug 1;25(4):431-438. doi: 10.1188/21.CJON.431-438. PMID: 34269338; PMCID: PMC9642910.

patients, as well as adding several new features that can further impact each patient's probability of successfully completing their course of treatment.

Through targeted interventions, Humana can enhance the well-being and survival rates of their patients while also mitigating the financial burden of treatment failures for patients, healthcare and insurance providers. By intervening on four key model features, we project that Humana has the potential to achieve savings of around $13.5 million. Furthermore, our suggested mobile application solution offers the opportunity to significantly enhance medication adherence and symptom management, with potential to reach $75 million in savings. Lastly, we envision that the proposed repeatable solution has the potential to yield even greater savings since it can be applied to a wide range of oral treatments, enhancing medication adherence and symptom management.