METHODOLOGY FOR PREDICTING SWITCHING BEHAVIORS IN PATIENTS

Brigham Hyde, PhD, President of Data and Analytics, EVERSANA™

Amish Dhanani, Associate Principal, Analytics, EVERSANA™ MANAGEMENT CONSULTING

EVERSANA™

eversana.com

EVERSANA™

Following our recent white paper, Patient Switching Behaviors Impact on Adherence and Engagement: A Predictive Analytics and Machine Learning Approach to Improving Hub Performance and Patient Outcomes, we committed to sharing our methodology.

Problem Statement (from the white paper):

Patient and hub services represent a significant and often ineffective spend surrounding overall patient support. How can we use data and analytics to build better patient services programs that predict next best actions and achieve improved patient engagement and outcomes across the treatment journey? As the industry continues its shift to value-based care, this challenge has never been timelier.

Here is the methodology (patient switch prediction):

Lifecycle of a patient prediction model

Following is the five-step process involved in setting up a successful patient switch prediction model:

1 DEFINE THE PATIENT UNIVERSE.

- Defining the right patient universe with significant heterogeneity is a key component of the prediction. There is an absolute need to have a bare minimum ("sufficient") data for prediction, where sufficiency is accounted for by the presence of relevant attributes and significant variation and diversity within the data.
- We assess the APLD and Health Records data to identify what is best suited data aggregation for an analysis on a case-by-case basis.

2 DEFINE THE TIME PERIOD OF THE ANALYSIS.

- Defining an ideal time period for the analysis is critical to ensure there is a sufficient look-forward and look-back period available to understand and interpret sufficient medical history and the therapy impact (side effects, comorbidities, etc.) on a patient before we begin the prediction window.

3 DEFINE THE PREDICTION DYNAMICS.

- Defining the right prediction dynamics is helpful to preempt and absorb the impact of the external factors (therapy area research, market dynamics, etc.) as well as internal factors (how soon a side effect can occur, etc.) while predicting an event.

4 IDENTIFY THE RELEVANT FEATURES (METRICS/KPIs).

- The relevance of features varies on a case-by-case basis and is driven by factors such as disease, therapy area, market landscape, economic burden, etc. Therefore, it is critical to set up an optimal feature selection framework that assists in selecting the optimal combination of features.

5 IDENTIFY THE RIGHT PREDICTION MODEL.

- The predictive model is selected by considering the business case, capability of fine-tuning the parameters, complexity of the computation and interpretability of the results.
- Select model from among various available options of classification, regression models and neural networks.

EVERSANA™

Lifecycle of a patient prediction model

Define the patient universe

- Assessing the data requirements (APLD and Health Records)
- Assessing the disease area and relevant disease and procedural codes
- Defining market and competition
- Assessing data pulse, identifying whether there is enough variation and diversity in the data to rationalize events

Define the time period of analysis

- What is the typical medical past of a patient in a therapy area of interest?
- What is a typical length of treatment to observe any side effects?
- What is the sufficient look-back and look-forward period to determine a patient behavior?
- How does changing the time period of analysis impact the results?

Define the prediction dynamics

- Defining the horizon of the prediction
- Defining the point of prediction (i.e., when do we want to predict – at the time of initiation or a quarter later, etc.)
- What an ideal prediction window should be, how fast a disease progresses, how soon market landscape changes, etc.

Identify the relevant features (metrics/KPIs)

- Identify key features that are most predictive for a business case
- Identify dominating hypothesis behind patient switch (i.e., higher drug cost, poor patient service, etc.)
- Identify the right mathematical transformation for the features
- Identify the optimal mode of feature generation

Identify the right prediction model

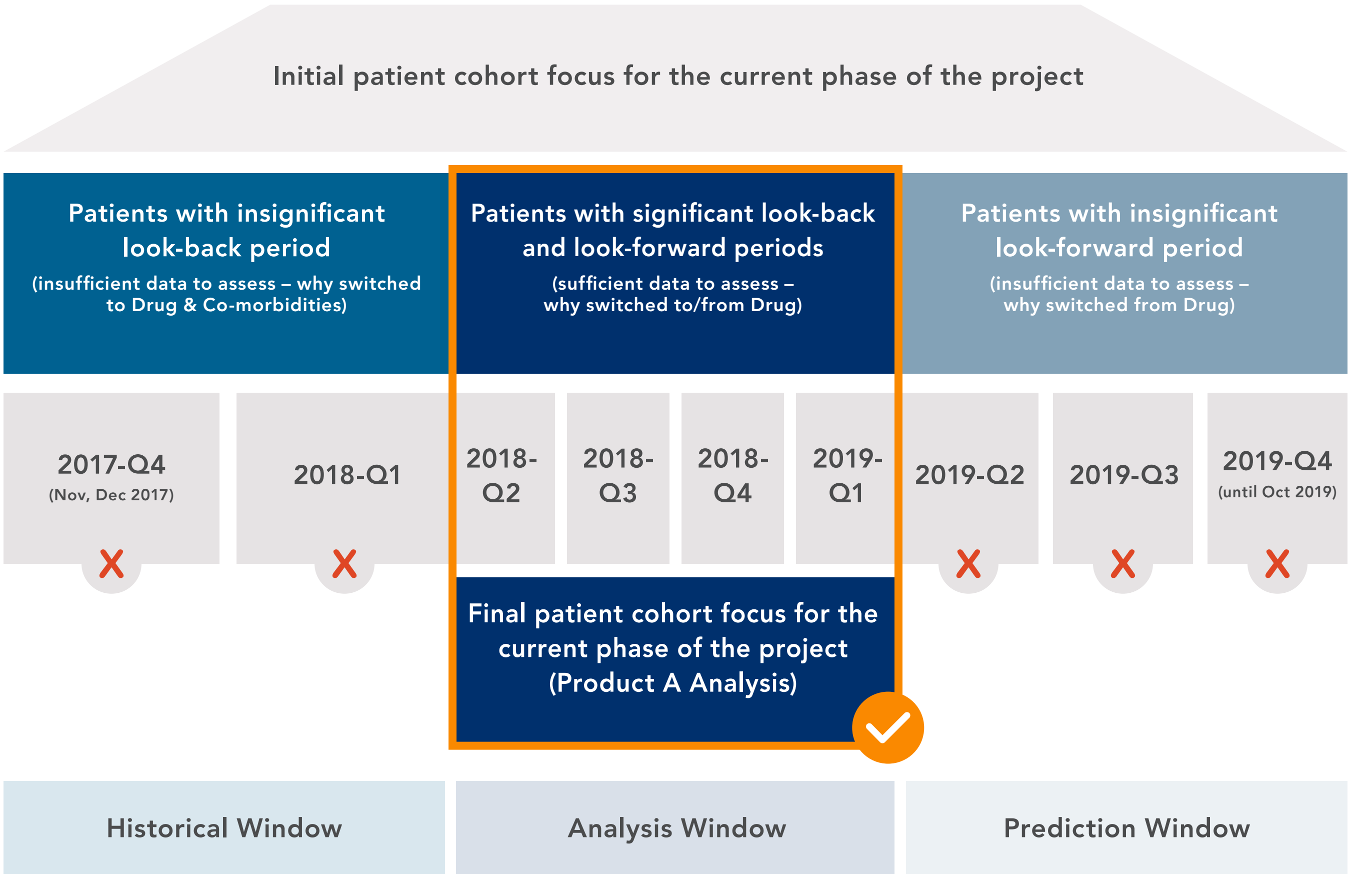
- What is the choice of statistical algorithm?
- How does the choice vary by use case (e.g., rare disease vs oncology vs cardiology)?
- What will be the defined values for associated fine-tuning parameters?
- What is the trade-off between the precision of the model versus interpretability of the model to action business insights?

Data Processing and Feature Selection

Leveraging the APLD and EHR data, the patient cohorts (switchers vs. non-switchers) comprising 83K patients were created based on an intuitive set of business rules. Subsequently, the feature variables were derived from available information such as Physician Characteristics, Patient Demographics, Treatment Characteristics, Insurance Characteristics and Medical/Clinical Characteristics. Features have been selected based on the therapy and market intelligence available with existing research and validated by consulting opinion leaders. Advanced feature selection framework is also leveraged to expedite the feature selection process.

Analysis Time Period and Prediction Window

The most recent possible period of – Q2, 2018 - Q1, 2019 – is chosen for analysis to provide sufficient look-forward and look-back periods.



Model Selection

The state-of-the-art classification model leveraging **gradient boosting** is used to identify heterogeneous cohorts comprising homogeneous groups of patients based on the values of selected features (components of classification). The training arrangement is a **30:70 test and train split** and is iterated using a tree ensemble approach. Hyperparameter tuning is performed for further optimization (optimizing tree depth, number of trees, regularization parameters, etc.).

Multiple iterations of the prediction model are run with permutations and combinations of the features to understand the optimal model accuracy, model balance as well as the impact of each individual feature. There will always be a trade-off between the model accuracy, model complexity and the resource availability, with an exponential cost component involved in improving an accuracy any further after attaining an intermediate equilibrium.

However, the rationale behind model selection should ensure that there is an absolute interpretability of the results even after multiple iterations over randomized test and train samples with significant statistical accuracy **(AUC >70%).**

EVERSANA™

Model Results Evaluation

Citing validation of the results over the test dataset described above, the results are evaluated based on the receiver operating characteristics (Accuracy, Sensitivity, Specificity, F1 Score (adj), Area under the ROC curve). AUC above 70% is considered good, and AUC above 90% is considered exceptional.

Also, the significance of the results is evaluated by probability scores, and the results are finalized only if a statistically acceptable significance **(Probability value < 0.005)** is achieved.

Key Notes

- Statistically, model contains significant predictive capacity
- More HCP switches indicate a likelihood of patient switch
- Previous anticoagulant therapy indicates likelihood of patient switch, provided there is other concerning situation (such as side effect, etc.)
- Side effects, especially bleeding, indicates likelihood of patient switch
- Higher average OOP costs indicate likelihood of patient switch

Evaluator	Value
Accuracy	93%
Precision	87%
Recall (Sensitivity)	74%
F1-Score	80%
R-Square value	60%

Actual / Predicted	Negative *(Patient Continues)*	Positive *(Patient Discontinues)*
Negative *(Patient Continues)*	42,912	1,184
Positive *(Patient Discontinues)*	2,859	8,210