Patient Persistency on a Bone Density Treatment

A Report for Medical Providers and Pharmaceutical Executives

By the Data Glacier Data Dynamos Group:

Jen Turley, Queen Echerenwa and Ateeb Aqil

December 18, 2021

Insights into Drug Persistency

A classification model for drug persistency was created using a health database file of clinical and demographic data for patients on a bone density treatment; the ultimate goal of this project is to understand better the factors influencing drug persistency.

Utilizing the Microsoft Excel Spreadsheet file **healthcare_dataset.xlsx** which contains data for 3424 patients prescribed a bone density treatment, exploratory data analysis and machine learning classification models were applied.

The Persistency_Flag column within this data set indicates for each patient whether the patient continues taking the medication (is persistent) or if the patient has discontinued the prescribed treatment (is non-persistent).

When a patient fails to be persistent with a medical treatment, there may be adverse health consequences for the patient. When many patients fail to be persistent, there are adverse consequences for the medical providers and the pharmaceutical company that produces the drug. Therefore, it is in the interest of medical providers and pharmaceutical companies to understand the degree of drug persistency for a medication and what are the factors influencing drug persistency.

Data Exploration:

Exploration of the data revealed an imbalance between persistency and non-persistency: of 3424 patients total, only 1289 were persistent while 2135 were non-persistent.

Alarmingly, a majority of patients were non-persistent!

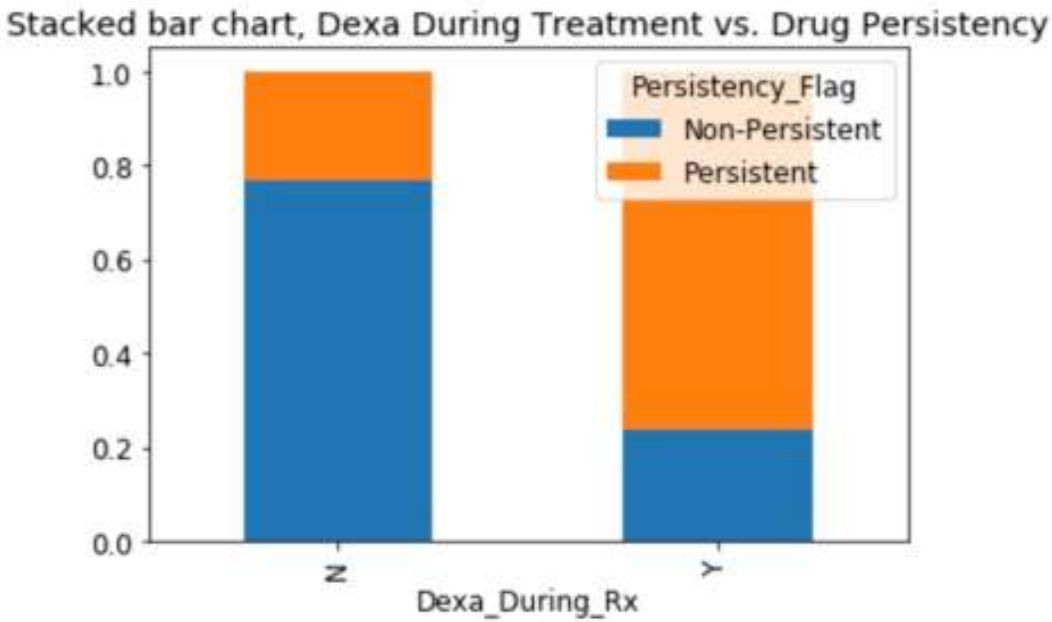
Can drug persistency be accurately predicted? If so, are there identifiable factors associated with non-persistence? Which of these factors might be modifiable by the medical provider?

Yes, Drug Persistency Can Be Predicted:

- Using machine learning methods for classification, patients' drug persistency could be predicted from the clinical and demographic data provided with greater than 80% accuracy.
- Which demographic or clinical factors are the most influential for drug persistency?

The top factor influencing Drug Persistency was whether the patient had a Dexa Scan while the treatment was ongoing.

Having a Dexa scan while bone density treatment was ongoing greatly increased the likelihood of Drug Persistency.



The next most influential factors

- If during the bone density treatment there were medical encounters such as
  - Screenings for malignacies
  - Visits relating to other long-term therapies
  - immunization
  - General medical checkups

  the likelihood of drug persistency was increased.

  The above list of visit types shows *in order of the degree of positive correlation* these helpful visit types associated with drug persistency.

  Visits with their medical providers make a difference for patients, increasing the likelihood of drug persistency.

Other factors influencing persistency, but more modestly:

- The medical specialty of the prescribing provider
- Whether the patient lives in the Midwest
- Whether the patient has certain other medical conditions including gastro-esophageal reflux disease, disorders of lipoprotein metabolism or other lipidemias, or certain joint disorders, among others.

Fortunately, the factors most strongly correlated with drug persistency are potentially modifiable by the treating medical provider.

Recommendations to prescribing medical providers:

- order Dexa scans while the patient's bone density treatment is ongoing

-encourage the patient to attend regularly for routine check-ups and any recommended immunizations

Technical information about this project:

- The data cleansing required on this dataset was minimal, and there were no missing values.
- One issue was the majority of data columns were of type categorical; these were transformed to numerical (flag) format, with some dummy variables added as required.
- Methods used for classification included Logistic Regression, Random Forest, Add a Boost, and Bagged Decision Trees With Random Undersampling (to address the class imbalance between persistent and non-persistent). These methods all achieved comparable classification accuracy at around 80%. The AUC ROC values achieved ranged from mid-70s to 90. There were no special challenges in applying the machine learning classifiers from the standard packages. Coding was done in Python with use of classifiers from SciKit-Learn and Statsmodels. Matplotlib was used to create the bar chart in this short summary report for business users.

Thank you!