

# Improving HIV Patients' Appointment Adherence Using a Predictive Model

*TREO Talk Paper*

**Candis Riggs, MSHI**

University of Alabama at Birmingham  
cariggs@uab.edu

**James Willig, MD**

University of Alabama at Birmingham  
jwillig@uabmc.edu

**Navneet Baidwan, PhD**

University of Alabama at Birmingham  
baidwo02@uab.edu

**Bunyamin Ozaydin, MSEE, PhD**

University of Alabama at Birmingham  
bozaydin@uab.edu

## Abstract

Although much has been done to reduce the transmission of HIV, there are still 36.7 million individuals living with the virus worldwide, with 1.1 million of those living in the US<sup>1</sup>. At the University of Alabama at Birmingham (UAB) there is a specialty HIV clinic, the 1917 clinic, which has helped over 12,000 patients in their HIV healthcare journey since 1988. One way to reduce the prevalence of HIV is by engaging patients in their care. Patients that arrive at their clinic visits regularly have been shown to have improved medication adherence, reduced mortality, and improved health outcomes<sup>2</sup>. Our clinic experienced a 19% visit no show rate in fiscal year 2018. Identifying patients at risk for no show visits is important for clinic scheduling, staffing and resource allocation. Identifying those at risk is the first step to the implementation of strategies to decrease no shows.

The data used for this project consisted of appointment data for 4,950 HIV positive patients between October 2015 and July 2019 at UAB's HIV clinic. The data source utilized was UAB's scheduling software, IDX®. To ensure data accuracy, automate data pre-processing, and provide a list of patients to be contacted, a relational database was developed in MS SQL Server using IDX® data and an automated SSIS package was utilized for the extract, transfer, and load (ETL) processes. A predictive logistic regression model (predicting no show Yes vs. No) was developed using the data from the database. The data was split into training (80% or 64,354 records) and testing (20% or 21,450 records) sets prior to deploying the model. A probability score for each patient was generated based on the trained logistic regression model with the independent variables of: previous no show/total appointments ratio, age, race, lag days (the days between scheduled and appointment dates) and gender. A threshold value, above which the patient was more likely to no show, was selected on the basis of analysis of the receiver operating curve (ROC) which showed the distribution of probability scores. The model was then tested for precision, specificity, and the weighted average (F1).

Among the 21,450 testing records, the distribution of no-show probability scores ranged from 0.09 to 0.83. A threshold of 0.3 was chosen based on its comparatively lower false negative (FN) error of 1,549. The true positive (TP), or those patients predicted to no show and actually did no show, was 6,435. The model works with 44% precision, or specificity, 80% sensitivity, with an F1 of 57%. A cutoff point was chosen that emphasized sensitivity in our model as we preferred to bring the potential of an intervention to more patients than what was needed. By using predictive algorithms, such as ours, healthcare organizations will be better positioned to identify and have the opportunity to offer interventions to engage patients in longitudinal clinic care and the ensuing improved morbidity and mortality.

## References

1. del Rio C. The global HIV epidemic: What the pathologist needs to know. In Seminars in diagnostic pathology 2017 Jul 1 (Vol. 34, No. 4, pp. 314-317). WB Saunders.
2. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. Journal of general internal medicine. 2012;27(5):520-6.