Machine Learning for Predicting Default from HIV Services in Mozambique

Background

In partnership with the Mozambique Ministry of Health (Ministério da Saúde), the U.S. Health Services and Resources Administration (HRSA), and the U.S. Centers for Disease Control and Prevention (CDC), ICAP at Columbia University worked with Dimagi to leverage the CommCare mobile health application to test the use of a predictive algorithm to identify HIV-positive patients at higher risk of default. CommCare, and its Mozambique-specific version, *Infómovel,* was developed by Dimagi to support community-based HIV and TB services in Mozambique. The main front-line users of the app are facility-based case managers (*gestor de casos*) and community health workers (*activistas*).

Objectives

The OpCon Mozambique project was designed to apply Machine Learning (ML) to predict the risk that individuals receiving HIV treatment will "default" - e.g., fail to return for a scheduled appointment - using Infómovel. Key project activities included: development of the ML algorithm; preliminary design of a user interface (UI) or "flag" within Infómovel; and rapid field testing to elicit user feedback. The project had three main activities:

- A Dimagi scoping visit in May 2018 to explore Infómovel utilization, activista workflow, and data management;
- Programming the ML algorithm and assessing its predictive capabilities by generating a ML dataset with data from ICAP-supported health facilities;
- A field-testing visit in September 2018 to obtain activista feedback on the idea of applying ML to generate an intuitive risk "flag" into the Infómovel user interface.

Scoping Visit Insights

The baseline scoping visit highlighted the fact that Infómovel can only accommodate the introduction of patients newly enrolled on ART who are receiving supportive home visits. As a consequence, health workers use OpenMRS to generate weekly lists of persons who miss appointments (including, but not limited to those in Infómovel). These OpenMRS lists are used to trigger outreach tracing activities, rather than the defaulter follow-up functionality in Infómovel, which is not currently in use. This limited the data available for the ML analysis. In response, the team used anonymized data on missed appointments from OpenMRS for the analysis.

Machine Learning Results

The ML algorithm was programmed to identify the 20% of patients at highest risk of default. Dimagi then used data from four ICAP-supported health facilities in Nampula and Zambézia provinces to develop a dataset containing 1,547 scheduled visits from 491 unique patients. Table 1 compares the actual rate of visits missed by 7, 10, 14 and 28 days with the rate of missed visits for the "high risk" patients flagged by the ML algorithm. In row 1, for example, the actual rate for all patients was 31.8%. In contrast, the rate for patients "flagged" by the ML algorithm was 48.3%. Thus, the algorithm effectively identified patients at higher than average risk.

Table 1

# days by which a visit was missed	Rate of missed appointments for all patients	Rate of missed appointments for patients in ML-predicted "riskiest" 20%
7	31.8%	48.26%
10	28.3%	43.01%
14	26.2%	39.98%
28	19.8%	34.77%

User Feedback Results

The Dimagi team discussed the feasibility and acceptability of the prediction app and its UI with activistas during the field-testing visit. Insights included:

- There is a high level of user acceptance for the existing Infómovel application.
- Many activistas are not familiar with the current Infómovel workflows for patients who miss appointments, since they use lists generated by OpenMRS, which allow them to identify individuals who miss appointments other than those in Infómovel

Conclusion

Despite the challenges of data management and utilization within Infomovel and MRS, the ML algorithm was able to identify a subset of patients at higher risk of default, providing proof-of-concept for the project. Users were open to the idea of a 'flag' within Infomovel, and interested in learning more about how it would be used.

A larger and cleaner dataset will improve the model, and development of "high-risk" workflows will enable tailored interventions to be provided to patients at high risk of default. These promising results were shared with the Ministry of Health, the U.S. Centers for Disease Control and Prevention (CDC), the U.S. Health Resources and Services Administration (HRSA), and the Dimagi teams implementing Infomovel in Mozambique.

Next Steps

The OpCon Mozambique pilot provided a foundation for future predictive work. Suggested next steps are described below; activities 1-3 are already included in the Infómovel workplan for the current year.

1. OpenMRS and Infómovel Integration

Currently OpenMRS is the only source of medication pick-up and scheduled pick-up dates, and therefore the source used to classify patients as 'active', 'faltoso' or 'abandonos'. Integrating OpenMRS and Infómovel, will enable a better integrated dataset for further analysis and improve data quality. Luckily, Infomovel-OpenMRS integration functionality is expected to be included as an update of the Infomovel app and will be available for CDC implementing partners in Mozambique during FY19; this will facilitate data collection across platforms to power an improved ML model.

2. Uploading Data on Missed Appointments to Infómovel

The critically important defaulter workflow is not being fully utilized in the Infómovel application. Uploading data on defaulting to Infómovel will help to keep the Busca Activa case list relevant and ensure that the activista users are sensitive to the differentiated workflows in the application. Working with activists on different patient statuses leading to different workflows will make the environment more receptive for future predictive risk flags.

3. Supervisory App

To better manage the app workflows and data management including tracking individuals who missed appointments, Dimagi is currently developing a supervisory app. This will help address consistency in following patients through the expected pathways in the application, and the data workflows on the back end of the application that would feed in to analyses.

4. Monitor Data Collection Stability to Re-Run ML Predictive Models

Once data collection and integration improve, new models can be tested to determine improvements in performance and to iterate on initial results. Following this, the aim would be to then design and field test a prototype UI for integration within the frontline health workflows, collaborating across organizational levels to determine requirements and test prototypes to determine optimal functionality.

This project is supported by the U.S. President's Emergency Plan for AIDS Relief, through the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under grant UJ7HA31180, Optimizing Momentum toward Sustainable Epidemic Control. This information or content and conclusions should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government.

PEPFAR HRSA Health Resources & Services Administration dimagi ICAP GLOBAL. HEALTH. ACTION. Columbia University Mailman School of Public Health