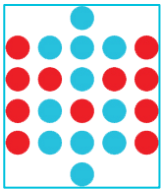


Predictive Analytics

Generating insights from data to improve health outcomes

THE OPPORTUNITY

More data, and more types of data, are collected than ever before. Beyond routine programmatic data, vast quantities of data from electronic medical records (EMR), geospatial and satellite datasets, text and audio data from media and search platforms, as well as myriad commercial and public datasets, create a vast font for analytics.

Machine learning (ML), artificial intelligence (AI), and predictive analytics more broadly identify patterns and relationships in large datasets, equipping decision makers with insight into what is likely to occur. Emergence of methods in ML and AI make it possible to harness these data to inform decision making and action. Importantly, improvements in technical infrastructure and computational power now make it possible to deploy these technologies at scale. These data, methods, and technologies offer the power to transform client care, identify and better characterize target populations, and accelerate achievements across the HIV clinical cascade.

THE DATA.FI SOLUTION

Predictive analytics can enhance the informational value of data in myriad ways. Data.FI works with the United States Agency for International Development (USAID) to customize analytical solutions to specific program challenges. Some examples include the following:

Predicting health outcomes at the client level

- Data.FI is applying ML to profile and predict clients at greatest risk of interrupting treatment and to identify the variables that drive predictions. Data.FI partners have used ML—combining test results, information collected via screening questions, and locational attributes—to predict with high accuracy which clients are likely to test positive for HIV. These methods can be used to answer additional questions across the cascade, such as: Who is most likely to test positive? Who should be targeted for pre-exposure prophylaxis (PrEP)? Who is at risk for viral non-suppression?

Improving data integrity

- Data.FI has developed client record deduplication guidance with USAID.[1] This work results in more complete and accurate clients records and gives health facility staff a better picture of treatment interruption. It also allows for client records to be augmented with additional data from other systems, such as laboratory or dispensing systems.

---

[1] Data.FI. (2020) Guidance for Deduplicating Client-Level Data. Washington, DC, USA: Data.FI, Palladium. Available at https://datafi.thepalladiumgroup.com/solutions/

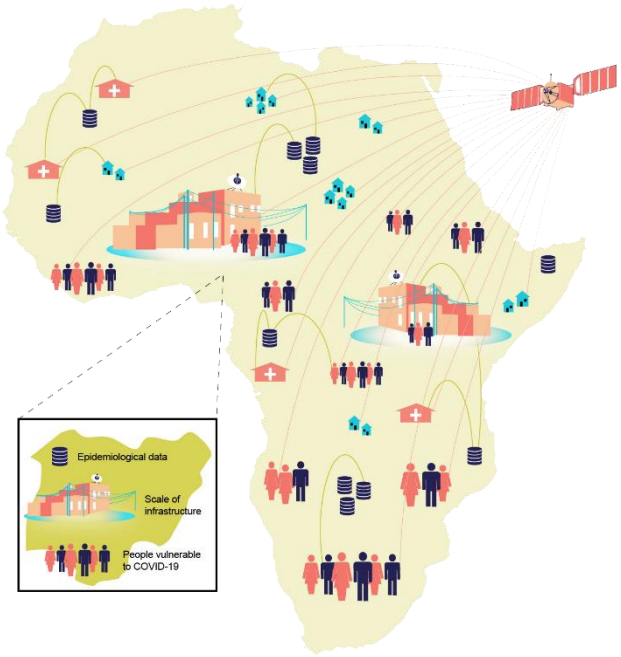
- Data.FI partners use anomaly detection algorithms to rapidly identify data quality and performance issues across large and varied datasets. Automated algorithms reduce the need for labor-intensive manual data reviews.

Predicting population-level indicators and size estimates

- Data.FI is using sophisticated geospatial analytics to identify and map at-risk populations, such as adolescent girls and young women most vulnerable to HIV. We apply ML and AI techniques to interpolate spatial data to identify where, at a hyperlocal level, populations at risk are concentrated and distributed geographically.
- Data.FI partners use time series forecasting techniques to predict probable population health trajectories based on recent trends and policies.
- Data.FI partners use social listening techniques to track opinions about health topics on media platforms, such as Twitter, to identify where and how often health topics are discussed, as well as network effects from such activity (how often content is shared, by whom, and with what audience.) Data.FI partners also use natural language processing techniques to capture the views and sentiments expressed, such as whether a topic is viewed favorably or unfavorably. Social listening helps decision makers improve client-centered care, identify underserved areas and populations, and pinpoint key influencers.

WHAT IS THE IMPACT?

Predicting health outcomes at the client level



In Mozambique and Nigeria, Data.FI applied a ML model to predict interruptions in treatment using de-identified EMR data, combined with AI-enhanced satellite imagery and other publicly available data. In Mozambique, the model identified the 20 percent of clients at greatest risk of treatment interruption. Of these, 75 percent interrupted treatment. We were able to derive the importance of each feature in the model's predictions. Two aspects of the client's past behavior—timeliness of attendance at past appointments and clinical history (e.g., past laboratory results)—were the most important categories of variables for the model's predictions. These types of insights can be used by service delivery partners to focus resources to retain clients who are more likely to drop out of treatment, moving countries closer to achieving the second HIV 95-95-95 goal.[2]

[2] United Nations Joint Programme on HIV/AIDS (UNAIDS). (2014). Fast Track: Ending the AIDS Epidemic by 2030. Geneva, Switzerland: UNAIDS. Available at https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf

Improving data integrity

Data.FI collaborated with the National Department of Health in South Africa to finalize and deploy a supervised ML-based patient matching model. The model first identified duplicate clients among over 20 million client records in the national tuberculosis/HIV information system and established a mechanism for other client records ingested in the national InfoHub to be analyzed and unified.

Predicting population-level indicators

In Eswatini, Haiti and Mozambique, Data.FI estimated the population size of vulnerable adolescent girls and young women, using publicly accessible survey data combined with ML/AI algorithms to build spatial interpolation models with high-resolution satellite imagery and other data products (e.g., earth observation data and gridded population information). We used this information to estimate the number of girls at risk at subnational levels and generate heat maps of girls at risk. Across countries, we found that there are substantial geographical variations in the distribution of at-risk adolescent girls and young women. These findings can be used by country operating units (OUs) and service delivery partners to monitor the current performance of Determined, Resilient, Empowered, AIDS-free, Mentored and Safe (DREAMS) programs, improve targeting of DREAMS interventions to geographic areas of greatest need, and to inform future program expansion to high-priority districts.

PUTTING THE SOLUTION INTO ACTION

Information generated from predictive analytics must be delivered to decision makers in the form they need it in, when they need it. Outputs of predictive analytic activities range from one-time analyses generating standalone information products to continuously deployed predictive algorithms processing data and generating predictions in real time. Data.FI partners are working to deploy solutions using a variety of platforms and have the needed expertise in information systems and software development to bring solutions where needed to decision makers.

SB-20-02

Data for Implementation (Data.FI) is a five-year cooperative agreement funded by the U.S. President's Emergency Plan for AIDS Relief through the U.S. Agency for International Development under Agreement No. 7200AA19CA0004, beginning April 15, 2019. It is implemented by Palladium, in partnership with JSI Research & Training Institute (JSI), Johns Hopkins University (JHU) Department of Epidemiology, Right to Care (RTC), Cooper/Smith, IMC Worldwide, Jembi Health Systems, and Macro-Eyes, and supported by expert local resource partners.

JANUARY 2021

FOR MORE INFORMATION

Contact Data.FI:

Emily Harris, Data.FI AOR
emharris@usaid.gov

Jenifer Chapman, Data.FI Project Director
datafiproject@thepalladiumgroup.com

https://datafi.thepalladiumgroup.com/