Data.FI SOLUTIONS BRIEF

Predictive Analytics

Prospective insights from data to improve health outcomes

THE OPPORTUNITY

Over the past decade, public health has seen an explosion in the collection of data beyond routine programmatic data. From vast quantities of patients' electronic medical records (EMR) to a myriad of publicly available geospatial and satellite data sets, as well as text and audio data from media and search platforms, the opportunities to construct contextually informed models of public health challenges with data have likewise mushroomed. Improvements in technical infrastructure across low- and middle-income countries have made it possible to deploy these methods at scale.

Predictive analytics, specifically machine learning (ML) and artificial intelligence (AI), represent a shift from a retrospective to a prospective use of large data sets to predict patient outcomes, equipping decision makers with insights on the likeliest scenarios and obstacles.

These data, methods, and technologies offer the power to transform client care, identify and better characterize target populations, enhance the allocation and utilization of resources across the supply chain, and accelerate widespread improvements in health outcomes.

In short, predictive analytics offer an opportunity for projects and stakeholders to remain one step ahead of major global health problems.

THE DATA.FI SOLUTION

Predictive analytics can enhance the informational value of data in several ways. The Data for Implementation (Data.FI) project works with the United States Agency for International Development (USAID) to customize analytical solutions to specific program challenges. Some examples follow.

Predicting health outcomes at the client level

- Data.FI is applying ML to predict interruption in HIV treatment so that service providers can intervene to encourage retention in care before clients interrupt treatment. ML models use information on client demographics, clinical histories, and locational and temporal attributes of where and when clients seek care. Data.FI is also using ML to improve HIV testing yield by combining test results, data from screening questions, and contextual attributes to predict which clients are likely to test positive for HIV. Similarly, ML can be used to answer such questions as whom to target for pre-exposure prophylaxis and who is at risk for viral non-suppression.

Improving data integrity

- Data.FI 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 client records to be augmented with additional data from other systems, such as laboratory or dispensing systems. In addition, Data.FI developed a tool that algorithmically checks Monitoring, Evaluation and Reporting (MER) data for anomalies, equipping USAID partners with a sophisticated, quick, remote data quality tool to supplement more intensive data quality improvement efforts. Automated algorithms reduce the need for labor-intensive manual data reviews.

Optimizing healthcare supply chain

- Data.FI developed tools to optimize COVID-19 vaccine allocation based on supply and demand-side constraints. Our interactive R-Shiny app takes user inputs to generate vaccination plans consisting of the number of vaccines to allocate to each distribution point.

- Data.FI developed tools to manage and strengthen the supply chain at multiple levels, including demand forecasting, inventory management, and delivery planning. Using various methods, such as hierarchical forecasting to anticipate demand across a distribution network, transshipment optimization to transport and store medications efficiently at fulfillment centers, and route-planning heuristics for last-mile delivery, Data.FI partners provide solutions across the supply chain management spectrum.

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. Using time series forecasting techniques, Data.FI predicted probable population health trajectories based on recent trends and policies.

- Data.FI used 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, and network effects from such activity (i.e., how often content is shared, by whom, and with what audience). We also utilized 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.

---

[1] Data.FI. (2021). *Guidance for deduplicating client-level data*. Washington, DC, USA: Data.FI, Palladium. Retrieved from: https://datafi.thepalladiumgroup.com/wp-content/uploads/2021/01/Data.FI_Guidance-for-Deduplicating-Client-Level-Data_Solutions_Brief_SB-20-01.pdf.

Modeling scenarios for target achievement

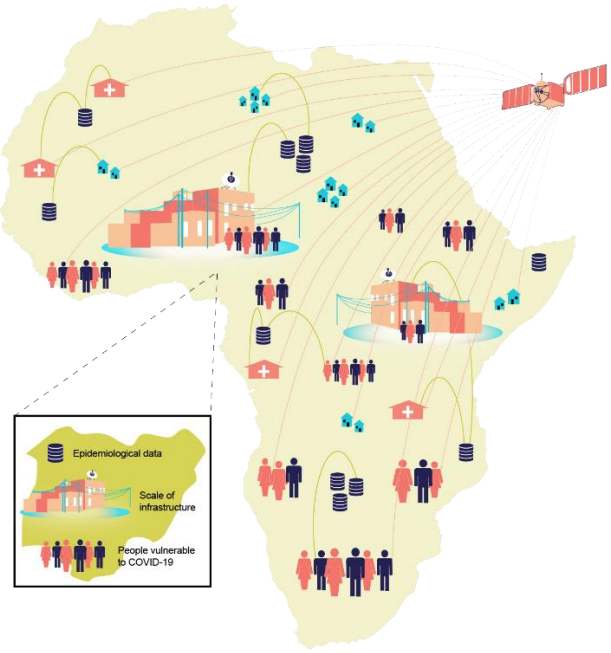
- In Nigeria, Data.FI developed a R-Shiny-based tool to generate annual targets quickly and reliably. This app takes in historical MER reporting data, user-defined parameters, and project targets at the site by integrating partner-level data within minutes.

- Data.FI's Human Resources for Health (HRH) Optimization Tool helps countries plan their HRH investments to best align HRH resources to meet programmatic needs. Combining site-level data on programmatic targets, current staffing distributions, and the Touch Foundation's service delivery mappings, the HRH Optimization Tool prescribes the optimal way to allocate HRH resources to maximize the achievement of programmatic targets in an equitable and efficient manner.

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 patient feature in the model's predictions. Two aspects of a 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. Service delivery partners can use these insights 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]



Improving data integrity

Data.FI collaborated with the National Department of Health in South Africa to finalize and deploy a supervised ML-based client matching model. The model first identified duplicate clients among more than 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.

---

[2] United Nations Joint Programme on HIV/AIDS (UNAIDS). (2015). Understanding fast track*: Accelerating action to end the AIDS epidemic by 2030. Geneva, Switzerland: UNAIDS* Retrieved from: https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf.

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 at-risk girls at subnational levels and generate heat maps to "map" them. 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 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 inform future program expansion to high-priority districts.

PUTTING THE SOLUTION INTO ACTION

Information generated from predictive analytics must be tailored for decision makers in the appropriate form and on demand. 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 is working to deploy solutions using a variety of platforms and has the needed expertise in information systems and software development to bring solutions to decision makers wherever needed.

In Mozambique, Data.FI developed and deployed a ML model that predicts risk of interruption in HIV treatment. Here, the Mozambique Ministry of Health participates in a meeting with Data.FI technical staff on HIV care and treatment.

September, 2022, Maputo. Data.FI/Mozambique

SB-23-02

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, DT Global, Jembi Health Systems, and Pendulum, and supported by expert local resource partners.

This publication was produced for review by the U.S. President's Emergency Plan for AIDS Relief through the United States Agency for International Development. It was prepared by Data.FI. The information provided is not official U.S. Government information and does not necessarily reflect the views or positions of the U.S. President's Emergency Plan for AIDS Relief, U.S. Agency for International Development, or the United States Government.

August 2023

FOR MORE INFORMATION

Contact Data.FI

Madeline Schneider, Data.FI AOR
mschneider@usaid.gov

Shreshth Mawandia,
Data.FI Project Director
datafiproject@thepalladiumgroup.com

https://datafi.thepalladiumgroup.com/