Big Data for Reproductive Health (bd4rh)

https://sites.google.com/view/bd4rh/home

Students: Dennis Harrsch (Global Health, Computer Science), Elizabeth Loschiavo (Sociology), Zhixue (Mary) Wang (Computer Science)
Team Leaders: Amy Finnegan, PhD; Megan Huchko, MD, MPH; Kelly Hunter

Basis of our Work

One-third of women who begin using a modern method of contraception in low-income countries discontinue within the first year, putting them at risk for unintended pregnancies as well as maternal morbidity and mortality.

Comprehensive data on discontinuation exist, but they are inaccessible to non-technical users. Our project aims to change that.

Previous Work

Past Data+ and Bass Connections teams have developed a web platform that curates data on reproductive health so users such as researchers, advocates, and policymakers, can develop insights around the factors that influence contraceptive use and discontinuation.

Data Sources

The Demographic and Health Surveys (DHS) Program conducts household surveys in numerous middle and low income countries. This includes the Contraceptive Calendar in the Woman's Questionnaire: a monthly record of the respondent's reproductive history over 5 years. The DHS Program also reports contraceptive prevalence and discontinuation data.

Additionally, we utilize publicly available reproductive health and social indicator data from the World Bank, World Health Organization, and World Population Prospects reported in 5 year intervals.

Tools

We used R Shiny to produce the four current applications, and they are hosted on a local Duke virtual machine. We employ Google Sites for our main bd4rh website, a central hub for our tools.

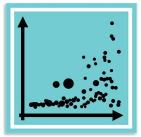
Objectives

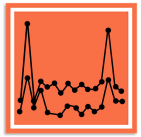
1. Create a machine learning tool that interactively clusters women's contraceptive trajectories and makes cluster membership predictions based on demographic indicators

2. Scale existing web applications by adding more country and survey data while meeting performance metrics

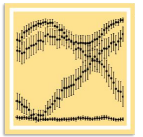
Future Work

The 2019-2020 Bass Connections team will continue to build upon the visualization platform and collect user feedback from key stakeholders via usability studies.


Correlation Plot


Line Graph

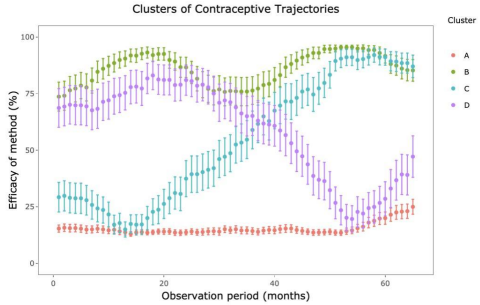

See the Switch

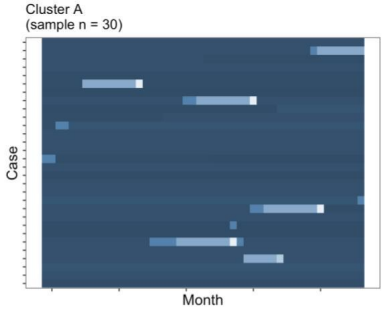

Cluster Prediction Model

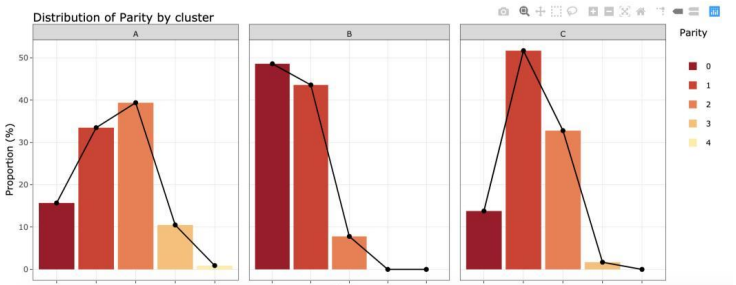
bd4rh
Big Data *for* Reproductive Health

Machine Learning

Step 1: Creating an interactive tool to dynamically apply predictive modeling to women's contraceptive efficacy trajectories.







START

Starting with **static** clusters on around 600 thirty-year-old women in Kenya, the results showed intuitive high, high-low, low-high, or low **efficacy trajectory** clusters.

To make this accessible, we created a public app to let users **interactively cluster** data of interest. We added visualizations of **sequences** and **cluster demographics** to aid user data exploration.

Next, we ran supervised machine learning, using k-nearest neighbors and logistic regression, to **predict dynamic cluster membership** based on demographic indicators.

We found class imbalance, with **deteriorating accuracy** as the number of classes increased. The dynamic model was **82%** accurate on 2 clusters and **50%** accurate on 4, even with oversampling.

MOVING FORWARD

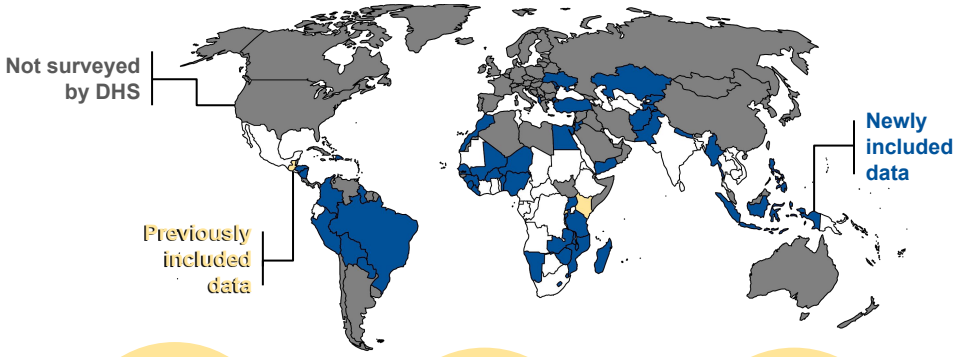
We want to **improve the model accuracy** via preprocessing techniques and other supervised learning algorithms and also run usability studies to determine utility for stakeholders.

bd4rh
Big Data for Reproductive Health

Scalability

Step 2: Incorporating more countries and surveys from the DHS Program dataset.

55 countries, **135** surveys, and over **1,800,000** women now included!



START

Originally, **only 3 countries,** each with one survey, were available per visualization. Access times ranged from **30 seconds** to **2 minutes**, potentially inhibiting use for stakeholders.

To improve latency, we **cleaned the data,** removing up to **18** unnecessary variables, regrouping them, and then organizing clean files by country and information available.

Next, we **rewrote existing code** to ensure that only the necessary data would be downloaded, reducing overhead performance costs and decreasing latency by up to **1 minute and 40 seconds.**

We then **shifted hosting** from an R Shiny-owned server to a temporary, local Duke-hosted **virtual machine,** further reducing latency and dependencies.

We also added **new functionalities** to better visualize specific patterns, including **postpartum** trajectories.

MOVING FORWARD

We want to **continue scaling** our apps by adding more countries and surveys and further reducing access times. We plan to incorporate **user feedback** to ensure our apps are useful for stakeholders.

Duke GLOBAL HEALTH INSTITUTE | Duke the center for GLOBAL REPRODUCTIVE HEALTH | Evidence Lab at the Duke Global Health Institute | Duke BASS CONNECTIONS | Data+