|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-B-022 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | N/A | Bangalore, 08-11-2018 | |
| **DOCUMENT** | | | | |
| **Source:** | | Swasti | | |
| **Title:** | | Proposal: Leveraging the power of computing to deliver effective, efficient and affordable healthcare to the most marginalized in HIV settings | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Shama Karkal  Swasti  India | | Tel: +91 - 98456 24994 Email: [shama@swasti.org](mailto:shama@swasti.org) |
| **Contact:** | | Prachi Patel  Swasti  India | | Tel: +91 - 9538444433  Email: [prachi@cms-india.org](mailto:prachi@cms-india.org) |

|  |  |
| --- | --- |
| **Abstract:** | We would like to submit a proposal for Leveraging the power of computing to deliver effective, efficient and affordable healthcare to the most marginalized in HIV settings. Swasti along with its sister organizations Vrutti and Catalyst Management Services implemented Bill and Melinda Gates Foundation funded Avahan India AIDS Initiative Phase III. The programme was aimed at reduction of HIV prevalence in the country through engagement with vulnerable population including female sex workers, men who have sex with men and transgenders. As part of the programme, we have cross sectional data of 1,20,129 individuals in year 1, 14,320 individuals in year 2 and 14,707 individuals in year 3 and longitudinal data for 6,891 people at three points in time. Given this open dataset, we propose a task to predict individuals who need to be prioritized based on likelihood of them not testing for HIV or not adhering to ART treatment. This when integrated with government programmes across the world can be used for focussing on the most marginalized and hence increasing cost effectiveness and efficiency of implementing large scale programmes. |

Organization Name

Swasti

Contact Name

Shama Karkal

Contact Email Address

[shama@swasti.org](mailto:shama@swasti.org)

Contact Phone Number

9845624994

Project Title

Leveraging the power of computing to deliver effective, efficient and affordable healthcare to the most marginalized

Overview

Swasti along with its sister organizations Vrutti and Catalyst Management Services implemented Bill and Melinda Gates Foundation funded *Avahan India* AIDS Initiative Phase III. The programme was focused on working with vulnerable populations including Female Sex Workers (FSW), Men who have sex with men (MSM) and Transgenders (TGs) to stabilize HIV prevalence rate in the country. The programme was implemented across five high prevalence states in India through 84 community organizations, reaching more than 1,36,000+ key population.

The focus of the programme was to identify and work on the determinants and influencers of key behavior patterns like testing for HIV, consistently using condom with clients and partners. This approach targets the root causes and hence is likely to have more impact in a cost effective way.

Impact

90-90-90. This is an ambitious treatment target adopted by all to help end the AIDS epidemic across the world by 2020. This means that by 2020, 90% of all people living with HIV will know their HIV status, 90% of all people with diagnosed HIV infection will receive sustained antiretroviral therapy and 90% of all people receiving antiretroviral therapy will have viral suppression[[1]](#footnote-1).

By the end of 2017, the world had achieved 75–79–81[[2]](#footnote-2). This progress is also not uniform across the countries with India at 77-56 and the progress against the third target is minimal because of the lack of viral load testing facilities across the country. Hence, a large gap exists in the targets and actual progress. With limited infrastructure, capacities and budget, it is imperative that novel cost-effective solutions are identified and implemented. In light of this, we believe that our approach of identifying and focusing on the individuals who are likely to drop out at any stage of the programme is likely to have far reaching effects in achieving the targets by 2020 and eliminating HIV by 2030.

Data Availability

Monitoring data has been the key source of data and information for Avahan to be implemented effectively. As part of monitoring, cohort data was designed, set up and used for 1,36,000+ people across five states in India. In total, we have cross sectional data collected for 1,20,129 individuals in year 1, 14,320 individuals in year 2 and 14,707 individuals in year 3 as part of yearly sampled outcome monitoring surveys conducted in each of the community organizations. As part of these surveys, we have longitudinal data for 6,891 people at three points in time. This data is currently available in formats supported by MS excel, SPSS, Stata and R. The first survey was conducted using paper based formats which were then entered into a MS Access software. This process was quality assured at multiple levels from data collection to entry and analysis. The sampled surveys were done using Computer Assisted Programming Interface (CAPI) built on an Open Data Kit (ODK) platform

Given this open dataset, we propose a task to predict individuals who need to be prioritized based on likelihood of them not testing for HIV or not adhering to ART treatment. The data can be split for training and testing as follows:

1. The data from the first two years will be provided as the train dataset
2. The last year’s data will be the test dataset, and will not be disclosed

The specific outcome variables are:

1. Use condom consistently with clients
2. Use condom consistently with partner/husband
3. Test for HIV as per the norms
4. Undergo ART treatment regularly

The areas of enquiry in the survey, which can be used as input features for prediction, include:

1. Demographic and social profile – This includes continuous variables like age, number of years in sex work etc, categorical variables like caste, religion, gender and ordinal variables like income range, no of years in migration among others
2. Access to services from the programme and the targeted interventions – This includes uptake of each of the services like STI testing, HIV testing, condoms as categorical variables
3. Social protection – This section includes variables on access to key civic identities and social protection schemes as categorical variables
4. Financial security – This section has categorical variables on financial literacy, access to financial products and continuous variables like quantum of loan taken, repaid
5. Gender based violence – This includes categorical variables on incidence of violence, prevention and redressal mechanisms and continuous variables on number of incidences of violence
6. Sexual behavior including testing and treatment behavior – The key areas of enquiry in this section include number of clients, condom usage behavior with clients and partner as categorical variables, HIV testing, STI symptoms, ART treatment as categorical variables
7. Reproductive and sexual health – Awareness on key reproductive health practices and cases of abortion is captured in this section using continuous, categorical and ordinal variables
8. Mental health – Each of variables are captured as categorical variables using the CESD scale for measuring depression
9. Self efficacy – These are categorical variables looking at behavior at individual level
10. Collective efficacy - These are categorical variables looking at behavior as group of members

Power calculations for sampling was done at the state level and further sampling was done for the members in all the community organizations. Hence, the source of data are individuals who have been part of the programme and have taken services. Detailed consent was taken from each of the respondents before this data was collected. Ethical clearance was also obtained from the Institutional Research Board at Sigma Research and consulting for the study. The study was done by external field investigators who were trained on the protocols, tools and ethics of data collection. The quality of data collected by these field investigators was assured through rigorous field processes like back checks, spot checks and random field visits.

As the programme implementers, we have access to the complete dataset of each of these surveys. As the programme was supported by Bill and Melinda Gates Foundation, the data is available to be used by researchers for analyzing and providing deep insights for the sector to learn and replicate. However, the identifiers for each of the individuals would be removed before sharing and the unique identifiers would be retained using an encryption algorithm for matching data across the datasets.

While Helsinki principles are applicable to studies involving human experimentation and the programme involved no experimentation, we have adhered to the key principles of:

1. Informed decision making
2. Assessment of risks and benefits
3. Right to withdraw

If additional data is needed from any of the individuals or geographies for doing this study, we will be able to facilitate the same. However, the process has financial implications and we would request support for the same.

Currently we do not foresee any conflict of interest with this exercise.

Benchmarking

Through the exercise supported by ITU and WHO, we would expect to be able to predict which individuals are likely to NOT do the following:

1. Use condom consistently with clients
2. Use condom consistently with partner/husband
3. Test for HIV as per the norms
4. Undergo ART treatment regularly

This when integrated with the technology solution being used on the field, will help the users/frontline field workers to identify the most at risk and hence provide increased focus and prioritization to avert the situation.

The benchmarking process will use the undisclosed test set. Each participant will be expected to create a model that takes as input features for an individual and outputs whether they are at risk or not, and the probability of being at risk.

For evaluation of the submissions, we propose two models of submission:

* + - 1. We will ideally ask people to submit code for the model that we will run on our server. We will input the test data to this model code and compare the accuracy of the predictions made to the actual outcome values. For security reasons, this code will be run in a sandbox environment so that it does not have permissions or access to rest of the server.
      2. Alternatively, we can disclose a random sample of only the input features for the test set and ask participants to submit a text file with the IDs of people who are at risk and provide probability of risk for each individual in the test set. Participants can upload this text file which will be evaluated for accuracy w.r.t the true undisclosed outcome values. This benchmarking process is similar to the process that Kaggle.com follows for its data science competitions.

In both cases, their submissions will be evaluated based on the accuracy of their predictions w.r.t. the true outcome values from the test set. Specifically, we will use precision, recall and AUC metric for comparing submissions. Since there are four outcomes, we can combine the four different evaluation metrics or consider each as a separate task. We chose these metrics because this is an unbalanced prediction task; we expect the fraction of people at risk to be different from the fraction of people not at risk.

Organizer Details

Swasti has been working in the field of HIV prevention and treatment since 2004. The focus of our interventions have been on evidence based programming and working with the communities on identifying strategies for prevention and not cure. Hence, our interventions would be highly efficient and effective in reaching the 90-90-90 if we can predict the members who are unlikely to undergo HIV testing or know their results or not undergo ART treatment. In order to help Swasti and the national government implementing targeted interventions across the country to achieve and replicate these outcomes, this project is important to us.

We have tried to replicate these models using regression models but due to lack of resources and partnerships to integrate artificial intelligence and data mining techniques into the current data, we have not been able to carry out a similar exercise before.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. http://www.unaids.org/en/resources/documents/2017/90-90-90 [↑](#footnote-ref-1)
2. http://www.unaids.org/en/resources/presscentre/featurestories/2018/july/90-90-90-targets-workshop [↑](#footnote-ref-2)