Title: Advancing Public Health: Leveraging Machine Learning for Enhanced Infectious Disease Monitoring through Integration of Demographic Surveillance and Clinic Records Retrospectively in Low- and Middle-Income Countries

Tathagata Bhattacharjee[1], Emma Slaymaker[1], Chodziwadziwa Kabudula[2], Jim Todd[1]

[1] Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom
[2] Medical Research Council/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), South Africa

Corresponding author: *Tathagata Bhattacharjee (tathagata.bhattacharjee@lshtm.ac.uk), Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom*

ABSTRACT

Background: In low- and middle-income countries (LMICs), infectious diseases present a substantial public health challenge, impaired by resource limitations. Data collected from various points is often disjointed, hindering the integration of demographic surveillance and clinic records within the same geographic area. This fragmentation limits the comprehensive understanding of disease burden, transmission dynamics, and intervention efficacy. Thus, there is a critical need to explore innovative approaches to integrate retrospectively the disparate datasets for more effective monitoring in intervention areas.
Objective: This research aims to investigate the potential of leveraging machine learning (ML) techniques to improve infectious disease surveillance in LMICs by retrospectively integrating data from demographic surveillance and clinics.
Methods: Obtaining analysis-ready data from real settings is often difficult for quality concerns. The aim is to determine the most effective method of linking records necessitating the creation of a controlled data environment for testing and comparison. To achieve this, the initial step involved generating synthetic datasets that mimic real-world conditions. The approach included three levels: (a) error-free data for perfect matches, (b) minor errors with slight variations, and (c) major errors with significant changes in characteristics. The record linkage stage applied ML for efficient integration. Feature engineering extracted relevant information, followed by algorithmic comparison and matching. This approach facilitated linking datasets.
Results: The study encompassed three distinct synthetic datasets, each infused with varying degrees of errors. This approach aimed to investigate the efficacy of utilizing ML techniques on retrospective data for infectious disease monitoring within LMICs. Evaluations were conducted on ML algorithms, scrutinizing their adaptability, resilience, precision, and computational efficiency.
Conclusion: By utilizing ML techniques for retrospective data integration, this study represents a significant advancement in infectious disease surveillance capabilities in LMICs with the potential to inform targeted public health interventions to improve disease control efforts in resource-constrained settings.

BIOGRAPHY

Tathagata Bhattacharjee has extensive experience in data systems and analysis for Health and Demographic Surveillance Systems (HDSS). He pioneered the implementation of electronic field data capture systems for Vadu HDSS, transitioning from laptops to Android tablets, a pioneering effort in India. His global contributions include serving on the INDEPTH Network's Scientific Advisory Committee and leading the successful integration of data from 47 HDSS sites into a Common Data Model (CDM) through the iSHARE project. At LSHTM, he developed an innovative ETL pipeline for data standardization at ALPHA sites. Bhattacharjee also led efforts to migrate data from ALPHA specifications to the OMOP CDM for the INSPIRE Network, including harmonizing COVID-19 data from the IDSR African Region. He adeptly configured the INSPIRE Platform-as-a-Service (PaaS) on Microsoft Azure to streamline complex data processing tasks. His work linking population data with disease surveillance has garnered significant interest in Africa.

- Mobile Number*: +91 8975002968
- Category*: (Oral presentation/ Poster presentation): Oral Presentation
- Linked In:
- WhatsApp No: (for conference updates) +91 8975002968
- Research Interest*: Record Linkage for Population Studies and diseases, Data Standardization using OMOP CDM
- Fax No: