

Electronic Data Capture for Large Scale Typhoid Surveillance, Household Contact Tracing, and Health Utilization Survey: Strategic Typhoid Alliance across Africa and Asia

Deus Thindwa^{1*}, Yama G Farooq^{2*}, Mila Shakya³, Nirod Saha⁴, Susan Tonks², Yaw Anokwa⁵, Carl Hartung⁵, Andrew Pollard², Robert S. Heyderman^{1,6} on behalf of STRATAA Study Consortium

¹Malawi Liverpool Wellcome Trust Clinical Research Programme (MLW), Malawi; ²Oxford Vaccine Group, Centre for Clinical Vaccinology and Tropical Medicine, Department of Paediatrics, University of Oxford and the National Institute for Health, United Kingdom; ³Oxford University Clinical Research Unit-Patan Academy of Health Sciences, Nepal; ⁴International Centre for Diarrhoeal Diseases Research, Dhaka, Bangladesh; ⁵Nafundi, Seattle, WA; ⁶Division of infection and immunity, University College London, United Kingdom

Background: Paper based systems for data collection and management are labour intensive and can result in slow, poor quality data collection, and increased data cleaning time. In contrast, electronic data capture systems have potential to automate validation to pre-emptively correct data errors in real time. In STRATAA, we have used Open Data Kit (ODK) and Structured Query Language (SQL) routines to generate timely, high volume and quality data.

Methods: STRATAA is a comprehensive programme assessing the population dynamics and epidemiology of typhoid fever in Malawi, Bangladesh and Nepal to inform design of vaccine and public health interventions. Census and survey data collection forms were developed through a structured iterative process and then implemented using ODK, with customizations by Nafundi, on Android-based tablets. Data are uploaded onto MySQL databases, where SQL routines are run nightly to enforce data cleaning on critical variables beyond ODK's validation routines. Daily anonymized data are backed up from 3 sites centrally. Database reports and descriptive analyses are generated weekly. To assess efficiency and quality of data capture, volume, accuracy and time of census data collection were quantified.

Results: We recorded demographics of 311,204 individuals from 74,475 households in three countries within average of 14.7 weeks range (13-16) using 20, 25, and 37 enumerators from Malawi, Bangladesh and Nepal respectively. Overall, 28.4 errors per 10,000 data points were found; 3.4, 7.6, 17.4 errors per 10,000 data points for Malawi, Bangladesh, Nepal respectively. These values meet acceptable quality threshold of 50 errors per 10,000 data points established by the Society for Clinical Data Management.

Conclusions: This robust, easy to use system allowed for high volume data to be collected over short time periods. Errors were low and varied moderately by country. Access to data is in real time, facilitating quality checking and decision making.