

ABSTRACT

Introduction: Acute respiratory infections (ARIs) are communicable diseases of public health concern due to high morbidity and mortality burden especially in low and middle income countries (LMICs). Accurate and precise estimates of disease burden inform policy and resource allocation hence contributing to the realisation of the sustainable development goals aimed at achieving universal health coverage and reducing preventable deaths. However, such estimates at sub-national levels are mostly imprecise due to small sample sizes or unavailable due to lack of sampling at such levels. Such limitations are overcome by methods such as small area estimation (SAE).

Methods: This study utilised data from a previously conducted study at 9 government owned outpatient healthcare facilities at the Kilifi health and demographic surveillance system (KHDSS) at the Kenyan coast. Each health facility was considered to represent a location out of the 15 locations in the KHDSS. ARI was defined as presence of at least one of cough, sneezing, nasal congestion, difficulty breathing or increased respiratory rate for age. Patients presenting to the clinic with any of the listed ARI symptoms were eligible for recruitment. A total of 15 samples per week per health facility were targeted for collection from the eligible patients over the entire study period. Samples were tested using multiplex polymerase chain reaction (PCR) assay for respiratory syncytial virus (RSV A and B), human rhinovirus (HRV), human coronaviruses (HCoV-OC43, NL63 and E229), influenza (A, B and C), parainfluenza viruses (PIV 1-4), adenovirus (ADV) and human metapneumovirus (hMPV). Data analysis was conducted on the number of cases that turned positive for each of the 7 virus pathogens. Hierarchical Bayesian SAE models were used to estimate and predict the prevalence and incidence rates of the viral pathogens stratified by age group (all ages, 0-59 months and 60 months and above) and controlling for auxiliary variables (mean age of participants, number of health facilities in a location and mean monthly rainfall). For each virus, 5 models were fitted each assuming a different underlying probability distribution; Poisson, zero-inflated Poisson, binomial, zero-inflated binomial and logit-normal transformation of the observed proportions of positive cases. The model with best fit for each virus pathogen by age category was selected with low values of the deviance information criteria (DIC) and number of effective parameters (pD). Data analysis was conducted in R and Bayesian model specification done using R-INLA.

Results: For a period of 18 months starting January 2016, a total of 7985 ARI cases were recruited at 9 outpatient health facilities within the KHDSS. Generally, virus infections were observed throughout the study period, though in relatively low numbers in the last 6 months of the study. Crude estimates showed that the highest prevalence and incidence were borne by children aged 0-59 months (p -values < 0.05) for all viral pathogens except FLU. In contrast, model-based estimates revealed variations of ARI prevalence across age groups while incidence rates were highest in the under 5s age band for all viruses. Model-based estimates over the 18 months of study yielded prevalence estimates for all ages at 4% for RSV, 14% HRV, 9% HCoV, 4% FLU, 4% PIV, 8% ADV and $<1\%$ hMPV. Among the under 5s the most prevalent infections were HRV (12%) and FLU (5.6%). In this age band, incidence rates per 100000 children were 502 for RSV, 3249 HRV, 330 HCoV, 1006 FLU, 910 ADV and 477 hMPV. Spatio-temporal variations of both prevalence and incidence rates were observed across all age bands. In general, the lowest prevalence ($<5\%$) and incidence (<100 cases per 100000 population) rates for RSV, PIV and ADV were observed in the second half of 2016 while the first halves of both years (2016 and 2017) had relatively higher estimates across the 15 locations. The remaining viral pathogens displayed variations of the time interval with the lowest or highest burden of disease by location while for pathogens such as FLU these variations could not be differentially identified. The logit normal model generally gave the best fit for most virus targets when estimating the prevalence while for estimation of incidence rates the Poisson and binomial models were mostly the best performing across age categories.

Conclusion:

Our results have shown that estimates of ARI prevalence and incidence obtained by crude/ unadjusted methods differed from those obtained via SAE model-based techniques. SAE methods provided the advantage of quantifying disease burden within the KHDSS while adjusting for space and time variations as well as other auxiliary variables such as number of health facilities in a location rainfall patterns. This provided better measures, within a Bayesian framework, that were otherwise either over or under estimated by direct methods. We observe that there exists a considerable burden of viral ARI in Kilifi with a substantial part of the incidental burden borne by children aged under 5 years. This has implications on future interventions such as vaccination, localised therapies, surveillance and planning.