VIMC Model Method

Our cholera burden simulation work is based on a spatial static model that considers temporal incidence rate trends, outbreak patterns, district-level vaccination campaign strategy, and indirect effects as well as the waning effectiveness of vaccines.

The first crucial information we need to simulate cholera burden is the spatial distribution of cholera infection risks. We use the 5 km by 5 km spatial incidence rate estimates from a published study (<https://doi.org/10.1016/S0140-6736(17)33050-7>) for the 35 countries in sub-Saharan Africa. As for the other 10 countries outside of sub-Saharan Africa involved in this study, the country-level annual cholera case reported by the WHO Weekly Epidemiological Record (WER) was adopted to estimate the incidence rates, which are assumed to be spatially homogeneous across the country. The spatial incidence rate in Bangladesh is based on an internal dataset and its population-weighted mean incidence rate is used for India. For each scenario, 50 random layers are drawn for each country to facilitate the stochasticity of the process.

Since the underlying incidence rate is unlikely to remain constant, we implement the incidence rate trend. We assume a deterministic log-linear relation between time (in years) and country-level annual incidence rates and assume the relation remains constant over the whole simulation period, then calculate and project their annual incidence rates from 2000 to 2100. To use the year-dependent incidence rate trend for each country on top of their underlying spatial incidence rate, we assume the yearly incidence rate trend and spatial incidence rates converge in the year 2014—for the year 2014, the original spatial incidence rate suffices; for any year other than 2014, we calculate a multiplicative multiplier for the projected yearly incidence rate trend relative to its 2014 projected estimate and multiply it to the spatial incidence rate to make a new spatial estimate for that year.

We also implement the district-level stochastic cholera outbreak spatial multiplier to the spatial incidence rate data. We adopt a dataset from a published paper (<https://doi.org/10.1101/2021.10.25.21265347>) that has the level-2 administration outbreak data in multiple countries from 2010 to 2020 and for countries and districts without outbreak data, we assume that there is no time-varying outbreak pattern. This dataset informs us of certain scalars to apply to different districts across outbreak times and non-outbreak times and the probabilities of outbreaks in each year. For each level-2 administration district, the spatial outbreak multipliers are generated over the first decade in the period from 2000 to 2100 and repeatedly used for the rest of the years following the decennial pattern, e.g., the same outbreak multiplier is applied to 2001, 2011, 2021, etc. Given there are 50 sets of stochastic simulations to implement outbreak patterns into, different stochastic draws of the outbreak multipliers are generated in correspondence to each simulation individually.

The vaccination campaign is hypothesized to be undertaken on the district level in each country. First, the level-2 administration districts are ranked based on the size of population affected by cholera. The affected population size is calculated by multiplying the 5 km by 5 km spatial population data in 2020 with the mean spatial incidence rate in each district, which is from the published study (<https://doi.org/10.1016/S0140-6736(17)33050-7>) as well. Second, the vaccines are given to the highest-ranking districts to fully cover their population until no vaccine is left. Due to the incomplete coverage of the vaccination campaign intended to fully vaccinate the overall population in a district, we assumed arbitrarily that only 80% of the population will receive the vaccines in all supposedly fully vaccinated districts. Additionally, if the vaccination campaigns are undertaken in adjacent years, we will not vaccinate a vaccinated (fully or partially) district again during the following three years.

To facilitate the mechanism of vaccination campaign bringing indirect protection to the unvaccinated, we assume that a certain vaccination coverage level will reduce the underlying spatial cholera incidence rate among susceptible people in each level-2 administration district. We select two published studies (10.1093/cid/cit009; <https://doi.org/10.1093/inthealth/ihy085>) and used their data to develop a logistic regression model based on the data, which will inform us of the degree of reduction in the spatial incidence rate given a coverage value from any districts. As for the waning effectiveness of the cholera vaccine, we use the data from a published meta-analysis (<https://doi.org/10.1016/S1473-3099(17)30359-6>) to inform the remaining proportion of the protected population a certain number of years after the vaccination campaign. We assume the protection from vaccines doesn’t last for more than 5 years.

The simulation is done across two counterfactual scenarios: the vaccination campaign scenario and no vaccination scenario, each with 50 stochastic runs. The stochastic simulations are facilitated by multi-layer spatial data, including the spatial incidence rate and the spatial outbreak-related incidence rate multiplier. Each layer is used for one stochastic run. We also consider four different settings, each of which has two scenarios and 100 stochastic simulations in total (50 simulations for each scenario). These four settings are different in terms of whether the outbreak-related incidence rate multiplier and the incidence rate temporal trend scalar are applied and the drawn set of underlying spatial incidence rate data. Additionally, the same set of outbreak-related incidence rate multipliers is used for the two settings that have the outbreak-related incidence rate multiplier applied for the purpose of efficiency.

Our code and relevant data to this study can be found at GitHub (<https://github.com/HopkinsIDD/gavi_vimc_cholera.git>).