Integrating Data With Epidemic Simulators to Improve Pandemic Preparedness: Chikungunya in Colombia

Wilbert Van Panhuis, University of Pittsburgh; *Guido Camargo, University of Notre Dame Department of Biological Sciences; Fernando de la Hoz, Universidad Nacional de Colombia Department of Public Health; Donald Burke, University of Pittsburgh Graduate School of Public Health; Michael Wagner, University of Pittsburgh School of Medicine*

Many existing datasets that could be used to counter epidemic threats are not used because standardizing and integrating datasets is time consuming and overwhelming. An epidemic is a complex system comprising pathogen and host populations, mosquitoes, and environmental factors. Figuring out the relationships between these factors is a bottleneck that constrains data integration –and the use of Big Data– for pandemic preparedness. Epidemic simulation models, developed by academic research, can be used to solve the data integration problem because they are blueprints of the epidemic system that mathematically define relationships between all system components. Our innovation is to use a new, ontology-based, data model derived from epidemic simulators to represent datasets in a machine-readable format and to integrate these datasets by simulating a real epidemic. As a use-case, we standardized data about the chikungunya epidemics in Latin America and integrated disease and climate data into an agent-based simulation of the 2014-2016 chikungunya (CHIKV) epidemic in Colombia. This model included 45 million agents that represented the population of Colombia distributed across 10 million locations (households, schools, and workplaces) in ~800,000 grid cells. CHIKV and Zika virus emerged recently in Latin America that has already gained decades of experience with the similar dengue virus (DENV). We found that mosquito-control interventions could have prevented 656,209 clinical cases of CHIKV if implemented by all 770 municipalities with vector populations. Information about previous DENV outbreaks could be used to reduce the scale of the intervention by 60% to 301 municipalities while the impact was reduced by only 25%. This information can help countries with limited resources to target their interventions to the most important hotspots of transmission. Our standard data model can catalyze the integration of existing datasets, which are currently unused, through epidemic simulation into cohesive scenarios for better preparedness against epidemic threats.