Title: Combining novel data streams to forecast acute febrile illness in the Dominican Republic

Background

Acute febrile illness (AFI) is characterised by fever of less than two weeks in duration and is a common symptom of many infectious and non-infectious diseases. In the Caribbean, AFI can be caused by a myriad of disease agents including protozoa, helminth, bacteria, rickettsia and viral. Despite its prevalence and contribution to overall morbidity and mortality, the ability to determine the aetiology of AFI is not yet commonly available at point of care facilities in the Caribbean. As a result, the true burden of each of the underlying aetiologies cannot be known, including their spatial and temporal distribution and their contribution to the total burden of acute febrile illness in the region. While a fever in the developing world contributes to morbidity in lost working days, in more vulnerable populations such as those found in the Caribbean, AFI can lead to considerable mortality.

The Dominican Republic (DR) is a small-island developing state (SIDS) situated within the Greater Antilles archipelago of the Caribbean on the island of Hispaniola. Similar to other SIDS, the Dominican Republic is vulnerable to natural disasters and other external shocks with an overdependence on international trade and the tourism industry. The epidemiology of the DR is diverse and both communicable and non-communicable disease contribute to the disease burden. The climate makes the country particularly susceptible to several arboviruses, most notably dengue. This project focuses on two viral diseases in particular that contributed to the burden of AFI; influenza and dengue.

Dengue is the most prevalent arbovirus globally. Transmitted by mosquitoes of the *Aedes* genus, dengue fever (and dengue haemorrhagic fever) incidence has been increasing globally along with the geographical range of the disease and in 2017 incidence was greater than 100 million[1]. In the DR, dengue incidence peaked in 2010 at 408,000 cases[1]. Although incidence has nearly halved since then, recent years have seen a stall in progress[1]. The relationship between environmental factors, such as temperature and rainfall, and mosquito and virus survival is well-established. However, until recently, the ability to effectively incorporate these variables into models that can forecast outbreaks and thus become an integral part of early warning systems was not available [2].

Influenza outbreaks are a result of the combination of antigenic shift and drift. Influenza in healthy individuals is usually a self-limiting disease however in more vulnerable populations, it can lead to serious complications. Estimating the burden of influenza is challenging due to the lack of information on morbidity and mortality, a result of few confirmatory tests and underreporting. Consequently, with current data sources, highlighting the dynamics of influenza in space and time has been particularly difficult.

Both dengue and influenza pose a threat to the health of the population in the Dominican Republic. Uniquely, this project offers the ability to analyse multiple novel data sources that may help to improve models that forecast AFI. The novel data sources available are: surveillance data, community serological surveys, large-scale human mobility data and environmental data. An outbreak of influenza, once in a human population, is largely determined by human interactions. Large-scale data on human mobility will allow us to better track the spatial and temporal dynamics of influenza within the population and thus highlight key agents and locations to interrupt transmission and contain an outbreak. Both dengue and influenza outbreaks will be driven by particular serotypes of the virus. Community serological surveys will provide details of the serotypes circulating within outbreaks and will provide better estimates of the burden of disease, as well as age specific patterns of infections that could not be gleaned regular community surveys of disease prevalence. Combining these novel data sources within a mathematical model, Kucharski *et al*.[3] were able to quantify the relative contribution of herd immunity, climate and control measures to

the outbreak dynamics. Ultimately leading to a greater understanding of the transmission and control of dengue in Fiji.

This project aims to investigate the role of influenza and dengue in driving acute febrile illness outbreaks in the DR. By combining novel data sources including human mobility and serological surveys provided by the collaboration with the General Directorate of Epidemiology (DIGEPI), we will better understand the transmission dynamics in our population. The data may provide insights into age-specific pattern of infection and prior immunity in the population that would not otherwise be available. Utilising climate data within geospatial temporal models we will be able to incorporate this information into an early warning system that can forecast acute febrile illness in the DR.

Objectives I hope to achieve during the PhD:

1. Understand the methods used to forecast dengue, influenza and AFI through a systematic review.
2. Maintain and build the collaboration with DIGEPI in DR through data sharing and analysis updates.
3. Develop a mathematical and statistical model within a Bayesian framework that can predict dengue and influenza outbreaks
4. Robustly evaluate model performance by comparing results against a null model.

Methodology

Firstly, this project will require a thorough exploratory analysis of the available data sets (human mobility data, serological surveys, surveillance data and environmental datasets). Exploring trends such as age specific patterns to infection and temporal patterns in serotypes. These findings may generate epidemiological insights of use to later analyses and more-widely to other transmission models of dengue and influenza.

Secondly, it is important to understand the recent history of the epidemiology of dengue and influenza in the DR and the methods that have been used to forecast outbreaks of dengue, influenza or AFI. A systematic review of these methods will be completed. From the understanding developed through initial exploratory analyses and systematic review, model development for both mathematical and statistical models will then begin. Model development will start by describing how each system operates for example; we know influenza is transmitted from human to human after initial exposure whereas the dengue transmission must incorporate the mosquito life cycle and the viral development inside the mosquito. The ability to incorporate priors is a key benefit of implementing these models within a Bayesian framework and informative priors will be used for processes where biological data of parameters already exist, for example, the intrinsic latent period and mosquito lifespan. Other processes we know less about will have minimally informative priors.

To describe the transmission dynamics of these virus', a mathematical model similar to that of Kucharski *et al.*[3] which followed a susceptible-exposed-infected-removed structure would be a good starting point. We know both infections have a latent period where you are exposed and that both infections provide immunity to the serotype that caused infection upon recovery so they would not return to the susceptible state and are removed. A Bayesian hierarchical model that estimates model parameters using Integrated Nested Laplace Approximation (INLA) could be used to integrate the climate data information that has been found to improve outbreak forecasts[4].

Many factors that influence transmission could be included within the mathematical model I will develop, including control measures, seasonality of transmission and environmental information. Model selection will be used to select which of these factors to include to

maximise predictive performance. Within a statistical model, selection will also be needed for covariate selection. Lastly, all model outputs will need to be evaluated. For example, we will likely compare the output of AFI incidence using root mean square error. When forecasting AFI, model performance could be evaluated by comparing the results to those from a null model [5]. This model will exclude some data sources such as serology and human mobility to assess if these new models combining novel data sources are predicting changes different from those that would have been found without the addition of novel data sources.

Practical implementation of this project

The work from this project could be used by national control programmes or organisations, such as DIGEPI in DR with whom data collection and analysis is in collaboration with, but also at larger scales by Regional Climate Centres to generate disease forecasts. Forecasting methods with more rigorous real-time evaluation are needed for their integration into the public health sector. Once in place, they can help to prevent excess morbidity and mortality by providing public health agencies with the information and time to plan and implement interventions that could reduce the risk of disease and prevent outbreaks. The models developed will estimate monthly AFI incidence for both dengue and influenza, allowing health agencies to know when and where an outbreak starts. This information can be used to plan resource allocation efficiently to ensure interventions and control strategies are implemented at the right time and place.

My suitability to this project

I believe this project aligns well with my previous experience and interests. My previous research experience has focussed on using large data sets within geospatial analyses of mosquitoes as vectors of disease and for malaria burden estimates. Within my role as [removed] with [removed] I gained experience maintaining collaborations and this will be important to ensure the data availability long term with the collaborators in this project, DIGEPI. I was also fortunate enough to be a part of two projects from initial conception through to the final stages of analysis. The first project sought combine novel anecdotal evidence into the consensus for plasmodium falciparum malaria seasonality and the other utilised a novel catchment model to improve our estimates for the clinical burden of malaria in Madagascar. I am currently writing up both projects and am hoping to submit these to very journal in the new year.

I recently attended a course in ‘Bayesian methods for Epidemiologists’ and I believe this along with my experience with R, which I use on a daily basis for data manipulation, as well as my familiarity with common statistical R packages for modelling, such as JAGS, INLA and TMB, has set me up well to begin this PhD project.

Within this PhD, I hope to broaden and deepen my skillset in geospatial statistical analysis and gain additional training in mathematical modelling to develop a well-rounded ability to model infectious disease outbreaks in the hope that we can prevent excess morbidity and mortality. In particular, I am excited by the prospect of collaborations with organisations within the DR and within the MRC cohort of students, which will contain people with a diverse and varied background that I am excited to learn from.

References

1. IHME, 2017, *Global Burden of Disease Results Tool*, IHME, viewed 2/11/2019, <http://ghdx.healthdata.org/gbd-results-tool>.

2. Lowe, R., Stewart-Ibarra, M., Petrova, D., Garcia-Diez, M., Borbor-Cordova, M., Mejia, R., Regato, M., Rodo, X. (2017) Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador, *Lancet Planet Health*, 1: e142-51.

3. Kucharski, A., Kama, M., Watson, C. *et al.,* (2018) Using paired serology and surveillance data to quantify dengue transmission and control during a large outbreak in Fiji, *eLife*, 7:e34848. DOI: 10.7554/eLife.34848.

4. Lowe, R., Gasparrini, A., Van Meerbeeck, C., Lippi, C., Mahon, R., Trotman, A., Rollock, L., Hinds, A., Ryan, S. and Stewart-Ibarra (2018) Nonlinear and delayed impacts of climate change on Dengue risk in Barbados: A modelling study, PLoS Medicine, 15(7): e1002613.

5. Lowe, R., Coelho, C., Barcellos, C., Carvalho, M., Catão, R., Coelho, G., Ramalho, W., Bailey, T., Stephenson, D., Rodó, X. (2016) Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil, *eLife*, 5: e11285.