UMEÅ UNIVERSITET

Umeå University Medical Dissertations, New Series No 2305

Navigating Epidemics

By leveraging data science and data-driven modelling

Zia Farooq

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie/medicine doktorsexamen framläggs till offentligt försvar i Sal B, Våning 9, Norrlands universitetssjukhus Måndag, 03 June 2024, kl. 09:00.
Avhandlingen kommer att försvaras på engelska.

För att delta digitalt via zoom:
https://umu.zoom.us/j/62878331943
Passcode: 112233

Fakultetsopponent: Professor Ian Hall,
The University of Manchester, Manchester, United Kingdom.

Department of Public Health and Clinical Medicine
Section of Sustainable Health

Organization
Umeå University
Department of Public Health
and Clinical Medicine
Section of Sustainable Health

Document type
Doctoral thesis

Date of publication
13 May 2024

Author: Zia Farooq

Title: Navigating Epidemics: By leveraging data science and data-driven modelling

Abstract

Ours is an era of global change—including climate change, land-use change, urbanization, increased mobility of humans, species and goods, and environmental shifts. Concurrently, we are witnessing a tangible increase in the rate of (re-)emerging infectious diseases, mostly driven by global change factors. This complex landscape of infectious diseases necessitates strategies underpinned by computational tools such as data-driven models to enhance our understanding, response, and to anticipate the risk of epidemics.

In this thesis, I leveraged data science algorithms and developed data-driven models that extend beyond specific pathogens, providing insights to prepare for future epidemics, with a focus on Europe. I delved into three temporal contexts: 1) retrospective analyses to understand the contribution of global change factors—specifically climate change and human mobility—fuelling the disease outbreaks and expansion (papers I & IV), 2) develop model to improve disease severity estimation during an outbreak for immediate response (paper III), and 3) future disease transmission risk trajectories under various projected scenarios of global change (paper II)—each playing a crucial role in proactive public health planning and response.

In paper I, we assessed the predictive ability and the influence of eco-climatic factors on West Nile virus (WNV) expansion and transmission—a climate-sensitive, multi-hosts and multi-vector zoonosis of public health concern with in Europe. Utilizing an advanced machine learning classifier XGBoost, trained on a diverse dataset encompassing eco-climatic, sociodemographic predictors to the WNV occurrence data, the model accurately predicted the WNV risk a season ahead. Furthermore, by employing an explainable AI algorithm, we uncovered both local and European-level drivers of WNV transmission. Higher temperatures in summer and spring, along with drier winters, were pivotal in the escalated frequency of WNV outbreaks in Europe from 2010 to 2019.

In paper II, we projected WNV transmission under climate change and socioeconomics scenarios by augmenting the outputs of climate ensemble models into machine learning algorithms. We projected WNV risk trends and maps at local, national, regional, and European scale. We predicted a three-to-five-fold increase in WNV transmission risk during the next few decades (2040-60) compared to 2000-2020 under extreme climate change scenarios. The proportion of disease-reported European land areas could increase from 15% to 23-30%, putting 161 to 244 million people at risk. Western Europe remains at largest relative risk of WNV increase under all scenarios, and Northern Europe under extreme scenarios. With the current rate of spread and in the absence of intervention or vaccines the virus will have sustained suitability even under low carbon emission scenarios in currently endemic European regions.

In paper III, we developed a method to quantify an important epidemiological parameter-case fatality ratio (CFR)—commonly used measure to assess the disease severity during novel outbreaks. In our model, we accounted for the time lags between the reporting of a cases and that of the case fatalities and the probability distribution of time lags and derived the CFR and distribution parameters using an optimization algorithm. The method provided more accurate CFR estimations earlier than the widely used estimators under various simulation scenarios. The method also performed well on empirical COVID-19 data from 34 countries.

In paper IV, we modelled annual dengue importations in Europe and the United States driven by human mobility and climate. Travel rates were modelled using a radiation model based on population density, geographic distance, and travel volumes. Dengue viraemic travellers were computed considering local mosquito bite risk, travel-associated bite probability, and visit duration. A dynamic vector life-stage model quantified the climatic suitability of transmission-permissive local areas. Dengue importations linearly increased in Europe and the U.S. from 2015-2019, rising by 588% and 390%, respectively, compared to 1996-2000 estimates, driven by increased travel volumes (373%) and dengue incidence rates (30%) from endemic countries. Transmission seasons lengthened by 53% and 15% in Europe and the U.S., respectively, indicating increasingly permissive climates for local outbreaks. These findings apply to other diseases such as chikungunya, Zika, and yellow fever, sharing common intermediate host vectors, namely *Aedes* mosquitoes.

This thesis highlights Europe's increasing vulnerability to infectious diseases due to global change factors, putting millions at risk. It emphasizes the significance of advanced modelling and innovative data streams in anticipating epidemic risks. Developing digital early warning systems to track disease drivers and taking urgent climate change mitigation and adaptation measures are crucial to anticipate and reduce future epidemic risks. The outcomes of this research can be used to develop technology-driven decision support tools to aid public health authorities and policymakers in making evidence-based decisions during and inter-epidemic periods.

Keywords
Epidemics, Data science, West Nile virus, Europe, case fatality ratio, human mobility, AI, XGBoost, XAI, SHAP, climate change, dengue, data-driven model, CFR, climate adaptation, climate mitigation, early warning systems

Language
English

ISBN
print: 978-91-8070-385-7
PDF: 978-91-8070-386-4

ISSN
0346-6612

Number of pages
47 + 4 papers