

Project Details	
Project Code	MRC23PHSBr Reyher
Title	Using machine learning to facilitate rapid and efficient responses to gastrointestinal disease outbreaks.
Research Theme	Population Health Sciences
Summary	Illness caused by the consumption of contaminated food is a major threat to public health and requires significant public resources to identify the source of infections. The successful candidate will work with the UK Health Security Agency to develop cutting-edge machine learning tools for the prediction of foodborne disease outbreaks. These tools will be used to support public health decision making and facilitate rapid response to future outbreaks.
Description	<p>Human infection by gastrointestinal pathogens is costly to both human health and the healthcare system. Unlike human-transmissible infectious disease, these infections are primarily associated with consumption of contaminated foodstuffs. However, due to the interconnected nature of the global food supply chain, identification of the source of infections is complex and labour-intensive. Spearheaded by the UK Health Security Agency (UKHSA), public health agencies have begun to routinely utilise whole genome sequencing (WGS) for pathogen surveillance. WGS provides high-resolution genetic relatedness between clinical disease isolates, allowing the confident identification of clusters of infections arising from a common source. As a result, deeper understanding into the infection dynamics of gastrointestinal disease has developed, with only a minority of cases leading to outbreaks requiring intervention. Distinguishing between random clusters of cases and an uptick in cases preceding a major outbreak, however, is non-trivial and requires significant expertise and oversight. In order to answer the research question ‘Can we use machine learning to predict future outbreaks of gastrointestinal pathogens and support public health decision making?’, this project will develop a reinforcement learning (RL)-based decision support tool to determine the optimal outbreak intervention window for gastrointestinal outbreaks. RL is an area of machine learning (ML) which has shown great promise in handling problems requiring multi-stage decisions and has been successfully used for outbreak management of a range of problem pathogens. An RL model will be trained using a combination of public health and pathogenomics data on clonal outbreaks of gastrointestinal infection provided by the UKHSA. The model will use variables associated with the infected individual (e.g. approximate geographical region, recent reported travel status, date, age, gender), as well as pathogen genetic information (e.g. lineage, presence of AMR and virulence genes), to predict an optimal intervention window. The cost of interventions, for example public health warnings and product recalls, and whether they are ‘true’ interventions or false alarms, will be used to solve the problem of when to initiate an epidemiological investigation using an RL framework. The project will achieve three primary objectives: a) Establish the costs of outbreaks and interventions The successful student will co-produce knowledge with UKHSA stakeholders to identify the costs associated with early, late and no intervention during outbreaks. This will be supplemented by interrogation of the literature to identify cryptic costs</p>

	<p>associated with other clinical factors, such as the presence of AMR and virulence genes. b) Build and train a RL-framework for intervention prediction The student will use ~55,000 clinical cases of Salmonella enterica and ~8000 Escherichia coli collected by the UKHSA from 2015-2022 to construct a model parameterized using the outcomes from (a) and trained on a combination of genetic and clinical metadata. The model will incorporate recent advances in ML to effectively generate dynamic intervention decisions. c) Generating human-readable and understandable outputs from ML The student will compare model outcomes from (b) to those used by Claire Jenkins’s team of experienced UKHSA epidemiologists to assess the impact of RL decision support on future outbreaks. The model will be packaged to produce human-readable outputs for use by non-experts using explainable ML methods. The project would suit an applicant with a strong first degree or masters involving bioinformatics or computational biology. The project will be tailored to the student; those with a mathematical background open to learning skills in bioinformatics and ML or those with a biological/biomedical background and experience in the areas of basic programming, data science or ML will also be considered.</p>
<b>Supervisory Team</b>	
<b>Lead Supervisor</b>	
Name	Professor Kristen Reyher
Affiliation	Bristol
College/Faculty	Faculty of Health Sciences
Department/School	Bristol Veterinary School
Email Address	kristen.reyher@bristol.ac.uk
<b>Co-Supervisor 1</b>	
Name	Dr Sion Bayliss
Affiliation	Bristol
College/Faculty	Faculty of Health Sciences
Department/School	Bristol Veterinary School
<b>Co-Supervisor 2</b>	
Name	Professor Andrew Dowsey
Affiliation	Bristol
College/Faculty	Faculty of Health Sciences
Department/School	Department of Population Health Sciences & Bristol Veterinary School
<b>Co-Supervisor 3</b>	
Name	Dr Lauren Cowley
Affiliation	Bath
College/Faculty	Milner Centre for Evolution
Department/School	Department of Biology and Biochemistry