Please provide a draft title or explanation of your research area. You can change this later. *

Broad: Improving inference of pathogen transmissibility and effects of interventions during ~~epidemic~~ outbreaks.

Please describe your proposed project, including the core research question(s) you seek to address, and your research approaches/methodologies. Please provide as much detail as possible, and avoid generalisations. *

Project Overview

One of the main ways to characterise transmissibility and effectiveness of interventions is to estimate the time-dependent reproduction number, R_t.
The primary focus of my proposed project is to discuss the limitations of existing methods (which make use of Bayesian inference techniques and branching process theory) for R_t inference, and to investigate potential improvements to these methods.

R_t, the time reproductive number

The time-dependent reproduction number, R_t, is defined as the expected number of secondary cases generated by an infectious case once an epidemic is underway. This statistic indicates the magnitude of the intervention required to control the outbreak (e.g. the proportion of contacts that must be prevented for cases numbers to begin falling), for the given pathogen. Given perfect contact tracing information, R_t inference would be as simple as counting the average number of secondary cases that a primary case generates.

Inferring R_t

In reality, such information is not available and instead, R_t inference is estimated using two types of data. One data type is the incidence (number of new symptomatic cases), whilst the other concerns an epidemiological time distribution between all infector-infectee pairs. The second piece of data would ideally be the generation interval (the distribution of times from infection in a primary case to infection in a secondary case) and in which case the incidence data would be indexed to the date of infection of cases. In practice, a proxy for the generation interval is used (owing to the complexity and ambiguity of determining exactly when an infectee becomes infected). This is the so-called ~~the~~ 'serial interval' (the distribution of ~~times~~ between symptom onset in an infector-infectee pair.) To infer the time-dependent reproduction number accurately, we should then index the incidence data with date of symptom onset. Broadly speaking, there are two statistical methods ([5] and [6]) which a large number of studies base their R_t inferences on, both of which use Bayesian inference techniques to generate time-evolving confidence intervals and expectations for R_t.

Accurate and precise R_t estimation is of significance during an epidemic since it is the primary indicator of the necessary stringency of public health measures. Consequently, the lack of accurate or precise estimates ~~are likely contributions to the sub-optimal economic status or general health of nations worldwide.~~

Open research questions and my project

A very broad concern is whether or not the a) response to and/or b) methodology for generating R_t estimates can be improved.

Related to a) is measuring how public health measures correspond to the spread of a pathogen through a population. Detailed analysis of how these public health measures have impacted R_t can provide evidence for policy makers during future outbreaks about which interventions control infections most effectively.

Our initially proposed project would focus more heavily on matters concerning b). Specifically, there has only been preliminary research undertaken into whether the serial/generation interval varies with time. Typically, a serial interval is estimated at the beginning of a pandemic, but is not then updated throughout. This is despite the fact that since these initial calculations, there are numerous factors that will change both i) the interaction between the pathogen and the host (e.g. the pathogen mutates, triggering a different immune response from the host) and ii) the behaviour between hosts (e.g. a public health measure that reduces the probability of a primary case infecting a secondary case X days after the primary symptom onset.) Considering the second example, we can see that the 'window of infection opportunity' for the average infectious person will be reduced (they may be legally obliged to quarantine once they have tested positive.) This implies that the average infectious person will generate ~~less~~ infections and by definition this would reduce the time-dependent reproduction number.

Proposed Case Study

I plan to analyse a real world data-set (from the 2018-2020 ~~Kivu Ebola epidemic in Beni, DRC~~) to investigate what effect including/not including serial interval updates has on the inferred reproduction number. From this I hope to develop the current understanding of how accurate serial intervals need to be, in order to recover accurate R_t estimates. This analysis will involve Bayesian inference for the R_t inference using the data discussed, as well as standard analytical techniques to discuss theoretical inference.

The mechanism behind transmission of the Ebola virus suggests a non-negative serial distribution (which is ideal for the methodology proposed). A negative serial interval includes the possibility that an infector may display symptoms after the infectee, despite the infectee being infected after the infector. Although it is possible to extend our method in this direction, it is not straight forward. This in combination with using data from the second largest Ebola virus outbreak in history, lends itself favourably to an excellent case-study, where many insights can be inferred robustly.

Further work

Finally, there is scope within the time-frame of this PhD project to investigate more broadly and thoroughly how these approaches to R_t estimates can be extended to include greater epidemiological realism. For example, can the Bayesian inference methods be combined with spatial and/or age structures to develop a better picture of how an outbreak is spreading through a population? Can these methods be used effectively by multiple local constituencies? At what scale is this most effectively applied?

External partner provision *

The external partner for this project is Jonathan Polonsky (World Health Organisation). He has sourced the data from the 2018-2020 Kivu Ebola epidemic, which enables the standard calculation of R_t as well as the calculation of updated serial interval estimates through time.

Our collaborator plans to attend monthly meetings, providing a unique understanding and interpretation of the data, as well as local situational awareness and experience in dealing with outbreak response scenarios.

Please describe the relevance of the project to your external (non-university) partner organisation, and how this project provides a route to real-world impact *

My collaborator's organisation has the following mission statement:

"To advocate and catalyze global and country actions to resolve the human resources for health crisis, to support the achievement of the health-related millennium development goals and health for all."

Central to responding to health crises is the ability to effectively monitor the state of an epidemic. The time-dependent reproduction number is an essential statistic in this process, and as a result this project is highly relevant to our external partner's organisation.

With regards to a route to impacting the world, there are three obvious mediums that I can envisage.

1) Presenting my findings at conferences/talks, which other ~~academics~~ in the Mathematical Epidemiology field attend. This work can then make a meaningful contribution to the field and be implemented by other academics also working on improving R_t inference. Additionally, releasing user-friendly code to an open source platform such as GitHub will enable accessibility to the work I have done.
2) Currently several research groups (several universities, Public Health England and the Joint Biosecurity Centre) contribute to the estimate that the UK Health Security Agency provides [7]. ~~A~~n potentially immediate impact is that my work could ~~eventually~~ contribute to refining these published estimates, which could then influence policy change in the ~~UK.~~
3) Another route to real world impact is the extension of an existing online software interface, called EpiEstim App [8], which can be used readily by anyone. My work could extend this existing app by adding an option for those who wish to generate a non-static serial interval, which updates with time.

Throughout this process, we are also confident that the bi-directional communication with our collaborator at WHO will help inform the progress of the project to involve more beneficiaries/end-users e.g. partners of WHO who share an interest in R_t inference, the WHO themselves who also have specific interests.

(required)

The context of the research *

The threat that infectious diseases pose to plants, animals and humans is one of significant consequence globally [1]. Control of infectious diseases through public health measures is an intensely researched area (due to their effectiveness [2]), particularly during the outbreak stage of an epidemic.
The time-dependent reproduction number (R_t) is defined as the number of secondary infections generated by infected individual [3]. This statistic is very helpful when determining suitable strength of interventions to contain epidemic growth [4].

Broadly speaking, there are two main ways of estimating R_t ([5] and [6]). This project will involve building on the work developed in [5]. Since the turn of the century, continual tracking of R_t has increasingly become more helpful to guide how interventions should change through time.

Currently none of these estimates include non-static (time evolving) serial interval (the distribution of times between symptom onset in an infector-infectee pair) estimates. There is preliminary evidence to suggest that time evolving serial intervals may have a significant impact on R_t estimates.

The aims and objectives of the research *

Aims: To improve the techniques that generate R_t estimates and to develop the understanding (within the field of mathematical epidemiology) about the significance (if any) of time varying serial intervals on R_t inference.

Objectives:

- Develop a hypothesis on how characteristics of changing serial intervals will affect R_t inference.

- Investigate real world data (initially from the Kivu Ebola Epidemic in Beni, DRC), where we may infer the reproductive number (with and without updating serial intervals) to test my hypothesis.

-Extend existing theory on R_t inference to incorporate heterogeneities into the model framework, e.g. spatial/age models.

The novelty of the research methodology (if any) *

Currently there has not been a comprehensive study into the theory behind how different serial intervals affect R_t estimation (although there has been some preliminary analyses). There have been very few case studies of investigating the use of time-evolving serial intervals on R_t inference, although there have certainly been some, e.g. [9]

The potential impact, applications, and benefits *

The ~~ongoing~~ consensus is that public health measures do tend to shorten the serial interval during epidemic outbreaks [7] y case-study research could either support or contradict this assumption.

The mathematical analysis that I have proposed could serve as an explanation for why this phenomena is observed.

Supposing that my research supports the consensus, it could help endorse the requirement for certain data collection, i.e. contact tracing, or for the improvement of data collection if we discover that certain data resolution is critical for accurate R_t inference.

How the research relates to the EPSRC remit *

This project contains promising and innovative research methods within the realm of epidemiology. The centre that I am based at within the MathSys CDT at Warwick University is supported by some of the most experienced epidemiological modellers in the UK, many of whom are currently members of SPI-M (Scientific Pandemic Influenza Group on Modelling).

The EPSRC seeks to support research that creates a resilient nation, which it describes fully in the form of five different ambitions. My project is highly relevant to Ambition R3 ("Develop better solutions to acute threats: cyber, defence, financial and health") since it seeks to mitigate the threat posed by infectious diseases. The project also provides "data science and analytics to anticipate, understand and model threats and optimise solutions", whilst employing "complexity science and uncertainty qualification to understand interdependencies for better decision making".

I have spent the past year studying towards an MSc in Mathematics of Systems, where I have taken various modules in modelling epidemics and related mathematical topics, including working group project on 'Adaptive Management during an ongoing pandemic'. Most crucially, I have spent the last several months working on some of the background analyses, which will naturally extend to this proposed project.
I believe this makes me an apt and ~~competent~~ candidate for this PhD project.

[1]: Cohen, M. Changing patterns of infectious disease. *Nature* 406, 762–767 (2000). https://doi.org/10.1038/35021206

[2] Flaxman, S., Mishra, S., Gandy, A. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 584, 257–261 (2020).
µ
[3] https://doi.org/10.1016/j.annemergmed.2008.06.461

[4] https://plus.maths.org/content/epidemic-growth-rate

[5] Anne Cori, Neil M. Ferguson, Christophe Fraser, Simon Cauchemez, A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, *American Journal of Epidemiology*, Volume 178, Issue 9, 1 November 2013, Pages 1505–1512, https://doi.org/10.1093/aje/kwt133

[6] Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004 Sep 15;160(6):509-16. doi: 10.1093/aje/kwh255. PMID: 15353409; PMCID: PMC7110200.

[7] https://www.gov.uk/guidance/the-r-value-and-growth-rate

8 https://shiny.dide.imperial.ac.uk/epiestim/

[9] Ali, S. T., Wang, L., Lau, E. H., Xu, X., Du, Z., Wu, Y., Leung, G. M., & et al. (2020). Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions.. *Science (New York, N.Y.), 369* (6507), 1106-1109. https://doi.org/10.1126/science.abc9004