UiO : University of Oslo

PhD candidate: Solveig Engebretsen

Contributions to network science in public health

Thesis submitted for the degree of Philosophiae Doctor

Oslo Centre for Biostatistics and Epidemiology
Department of Biostatistics
Institute of Basic Medical Sciences
University of Oslo

Department of Infectious Disease Epidemiology and Modelling
Division for Infection Control and Environmental Health
Norwegian Institute of Public Health

2019

Acknowledgements

First and foremost, I would like to thank my team of supervisors for support throughout the three years, Birgitte Freiesleben de Blasio, Kenth Engø-Monsen and Arnoldo Frigessi. In particular, I want to thank my main supervisor Birgitte for her warmth and kindness, her sharing of knowledge, wisdom and advice and for always showing interest both in me and my project. Thanks to Kenth for interesting discussions, and in particular for discussions on the use of mobile phone data and for providing the data. I also want to thank Arnoldo for his unshakable optimism, positivity and enthusiasm, and for always insisting on high standards. I want to thank Ingrid K. Glad, for encouraging me to do a PhD, for her support, warmth and collaboration. I want to thank Nils Lid Hjort for showing interest in my work and inviting me to his interesting and stimulating FocuStat research kitchens. I want to thank Marijtje Van Duijn for warmly welcoming me to the University of Groningen and sharing her inputs and thoughts with me. I want to thank Veronica Vinciotti for her input and ideas, in particular on ideas for making a more sophisticated human mobility model. I also want to thank Gianpaolo Scalia Tomba for interesting discussions.

I want to thank my other co-authors Christopher S. Nielsen, Audun Stubhaug, Anne-Sofie Furberg, Emily S. Gurley and Mohammad Abdul Aleem. Special thanks to Christopher for his interesting and fun ideas, and especially for coming up with the idea for the first paper, at a time where I was mainly waiting for data, and for his support and advice.

My work has been partly funded by the Norwegian Research Council centre BigInsight (project number 237718) and by the Norwegian Institute of Public Health. I also acknowledge COSTNET (COST Action CA15109) for funding my visit to the University of Groningen, workshops and conferences.

During my PhD I have had two offices, one at the Norwegian Institute of Public Health (NIPH) and the other at the Oslo Centre for Biostatistics and Epidemiology (OCBE) at the Department of Biostatistics. I want to thank all my colleagues at both NIPH and OCBE for numerous lunches, tea and coffee breaks. I especially want to thank my office mates Lotta and Francesco at NIPH for advice, discussions and ice cream breaks. I also want to thank my office mate Andrea at OCBE for going to gym classes with me and for our chatty dinners. Thanks to Owen at OCBE for interesting discussions on whimsical language details and for volunteering to read my thesis. I am also eternally grateful to my friends, family and family in law for your love, support, dinners, holidays, joys, laughs, breaks, skiing, hiking and running trips. Finally, thank you to my superhero and rock Andreas for his comfort, encouragement and wonderful support and his invaluable comments, input and advice, also on the academic side.

List of papers

Paper I

Engebretsen, S., Frigessi, A., Engø-Monsen, K., Furberg, A. S., Stubhaug, A., de Blasio, B. F., & Nielsen, C. S. (2018). The peer effect on pain tolerance. *Scandinavian Journal of Pain*, 18(3), 467-477. `http://doi.org/10.1515/sjpain-2018-0060`

Paper II

Engebretsen, S., Engø-Monsen, K., Frigessi, A., & de Blasio, B. F. (2019) A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions. *PLOS Computational Biology* 15(3): e1006879. `http://doi.org/10.1371/journal.pcbi.1006879`

Paper III

Engebretsen, S., Engø-Monsen, K., Aleem, M. A., Gurley, E. S., Frigessi, A., & de Blasio, B. F. Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh. *[Manuscript]*

Contents

Chapter 1

Introduction

Network analysis is used to study relational phenomena between entities. For example, networks offer a mathematical representation of friendship or contact among individuals, mobility between locations or patient transfer among hospitals. Networks are thus an excellent framework for studying relational phenomena like, for instance, infectious disease transmission, peer influence or social support, with numerous applications in public health.

In many public health applications, the object of study is relational (Luke and Harris, 2007). Network analysis explicitly analyses the dependence structure, as opposed to most classical statistical applications, where the observations are assumed to be independent. While systematic network analysis date back to the 1930s (see for example Moreno et al. (1932); Moreno (1934)), it has seeds in as early as 1735, with Euler's solution of the famous Köningsberg bridge problem (Kolaczyk, 2009), published in 1741 (Euler, 1741). In recent years, there has been an "explosion of network data" and hence also methods for analysing the network data (Kolaczyk, 2009). Epidemic modelling has been an area of interest for applications of network models since the 1980s and is a rapidly growing field (Danon et al., 2011). In recent years, the whole field of epidemic modelling in networks has "greatly progressed in the understanding of the interplay between network properties and contagion properties" (Pastor-Satorras et al., 2015).

Infectious diseases that spread via person-to-person transmission is one of the most important applications of network analysis in public health. Networks allow us to study the potential transmission routes between individuals or locations, providing a framework to better understand the spreading pattern, predict future spread and study different intervention strategies (Danon et al., 2011). Individual contact networks have been used to model the spread of infectious diseases like sexually transmitted diseases (de Blasio et al., 2007; Klovdahl et al., 1994), SARS (Meyers et al., 2005; Small and Tse, 2005), influenza (Fox et al., 2017; Salathé et al., 2010) and smallpox (Eubank et al., 2004). Contact networks can also be exploited in contact tracing, to prevent spread of infectious diseases by tracking down contacts of infectious individuals. For instance, Centers for Disease Control and Prevention (CDC) exploits network analysis in contact tracing for tracking down tuberculosis cases (Luke and Harris, 2007; National Tuberculosis Controllers Association and Centers for Disease Control and Prevention, 2005). Contact tracing was also used during the swine flu pandemic (Danon et al., 2011). There are also many examples of networks used to model disease spread between different locations or populations. In such models, human mobility is often the main driver of the disease transmission. For example, the international network of passenger air flights provided by the International Air Transport Association has been used to model the global spread

of pandemic influenza (Colizza et al., 2007a) and SARS (Hufnagel et al., 2004).

There are numerous examples of social network analyses with applications in public health. Social network analysis has been used to study the interplay between social structure and relationships and health behaviour (Luke and Harris, 2007). Assortative mixing occurs when the ties in a network are between nodes with similar properties. Different lifestyle factors have been shown to be assortative in social networks, for instance obesity (Christakis and Fowler, 2007; Valente et al., 2009), smoking (Christakis and Fowler, 2008; Steglich et al., 2010) and alcohol consumption (Rosenquist et al., 2010; Steglich et al., 2010). Loneliness (Cacioppo et al., 2009) and happiness (Bollen et al., 2011) have also been shown to be associated with the social network. In social networks, the number of connections has been shown to be assortative (Newman, 2002), that is, highly connected individuals are connected to other highly connected individuals and vice versa. Assortativity in the number of connections is important for disease transmission. Outbreaks occur for less aggressive diseases and have a shorter duration in more assortatively mixed networks (Kiss et al., 2008). Negative vaccine sentiments also seem to be assortative in social networks – Salathé et al. (2013) investigate how health sentiments towards a new vaccine spread on the Twitter social network. They find that negative sentiments have a contagious effect, while they do not find a contagion effect in positive sentiments. There are mainly two processes which can explain assortativity of a trait in a social network – either contacts are selected based on similarity in the trait (homophily), or they influence each other, so that they become similar over time. These processes can also operate simultaneously. Christakis (2004) suggests that health interventions targeted towards an individual can have effects beyond the individual to its social network, if there is indeed a social influence effect of the trait. This suggests optimal treatment strategies with targeted interventions, just like for epidemic prevention in disease spreading.

Another benefit of network analysis in public health is the opportunity to study the role of social support, social capital and network position on personal health and health behaviour (Luke and Harris, 2007). Many studies have provided evidence that social support and capital are beneficial for health. Valente et al. (2009) found an association between weight status and network position – overweight adolescents had a higher out-degree and a lower in-degree than their normal weight peers. House et al. (1982) found that those who reported more social relationships and activities were significantly less likely to die during the follow-up period of their study. Other examples of studies finding a beneficial effect of social support can be found in, for instance, Kawachi et al. (1997, 1999); Ziersch (2005). Social support has also been shown to be beneficial for pain (Eisenberger et al., 2006; Master et al., 2009; Brown et al., 2003). In particular, a larger social network has been found to be associated with higher pain tolerance (Johnson and Dunbar, 2016).

This thesis develops new statistical methodologies with important applications to understand and predict various phenomena that rely on network relations. Paper I is a social network analysis applied to pain tolerance, where the interplay between pain tolerance and a friendship network is studied. We inves-

tigate whether pain tolerance is assortative in the social network, and whether network centrality is associated with pain tolerance. In both paper II and paper III, we use mobility networks of locations to develop mathematical models of spatial diffusion of an outbreak of an influenza-like illness. Paper II is a simulation study of the effect of urbanisation on the spread of an influenza-like illness, where locations in a grid model are connected through a mobility network. We explore the impact of travel restrictions and vaccination, when the urbanisation is varied. In paper III, we model the spatio-temporal spread of influenza in Bangladesh, using a dynamic mobility network informed by mobile phone data. We apply the model to the 2014-2017 influenza seasons in Bangladesh to obtain information about transmissibility. In the remaining parts of this introductory chapter, we provide a background for the topics explored in the three papers of this thesis.

1.1 Pain tolerance

Genetic factors have been shown to explain approximately half of the variance in pain tolerance (Nielsen et al., 2008; Trost et al., 2015; Angst et al., 2012), while the family environment has been shown to have a negligible effect (Nielsen et al., 2008; Trost et al., 2015). The social environment beyond the family is a candidate for explaining some of the variance in pain tolerance that cannot be explained by genetic factors. Reduced pain tolerance is a key feature of and a suspected risk factor for chronic pain (Edwards et al., 2005; Lötsch et al., 2017). Prevention strategies aimed at increasing pain tolerance could potentially reduce the number of people affected by chronic pain. Chronic pain has enormous cost for the society, with higher economic impact than most other health conditions (Phillips, 2009; Maniadakis and Gray, 2000; Fillingim, 2017), and is one of the major causes of work disability (Nielsen, 2013; Fillingim, 2017). Understanding the mechanisms behind pain tolerance is therefore of key importance, and it has been claimed that "we can only fully understand and treat pain, when we also acknowledge the interpersonal or social context in which pain occurs"(Vervoort et al., 2018).

The first part of the thesis (paper I) is a contribution to understanding the complex social nature of pain. The social communication model of pain is the dynamic interplay between biological, psychological and social determinants of pain (Craig, 2015, 2009). The social part of the social communication model is of key importance to clinicians and researchers, and has gotten relatively little attention (Craig, 2009; Mogil, 2018; Vervoort et al., 2018). Prkachin et al. (2018) request analyses of social networks to develop a comprehensive framework of social influence on pain.

Even though the social aspect of pain has received little focus, there are examples in the literature on studies of social features of pain. Social networks and interpersonal connections have been found to be relevant to the experience and communication of pain, for instance through stereotyping, social exclusion and social support (Prkachin et al., 2018). Social influence on pain can op-

erate through for instance cultural, economic, ethnic, healthcare and political aspects, as well as through facial expressions, social neuroscience, social norms and public standards (Craig, 2018). For instance, individuals learn to suppress the facial displays of pain following cultural or social norms (Kunz et al., 2018). Children report being less likely to reveal pain in front of a peer than a parent, because they expect peers to be less accepting of their pain (Kunz et al., 2018; Zeman and Garber, 1996). Martin et al. (2015) find that when familiar pairs are participating in experimental pain tests together, they report higher pain ratings and show increased pain behaviour, compared to being tested in isolation. This is the case for both mice and humans, and they speculate that this is due to emotional contagion. The effect was not found for stranger humans tested together. Block et al. (2018) test experimentally whether being in pain decreases popularity, whether adolescents prefer to interact with similar adolescents (in terms of pain) and whether adolescents become similar to their friends in terms of pain, by studying 17 adolescents on an arctic expedition. They do not find assortativity of pain in the social network, but they find that pain decreases popularity.

1.2 Influenza

Influenza is a classic example of a disease which spreads annually across the globe through human mobility (Barbosa et al., 2018). Seasonal influenza is an acute viral infection which easily transmits from an infectious individual to a susceptible individual. Influenza is airborne, so transmission can occur if an infected individual coughs and a susceptible individual breaths in droplets from the infected individual. It can also transmit through direct contact, or through for instance contaminated surfaces (World Health Organization, 2016).

There are four types of influenza viruses, and the seasonal influenza epidemics are caused by the two types A and B. Influenza A can be further broken down into subtypes, and currently circulating subtypes are A(H1N1) and A(H3N2) influenza viruses (World Health Organization, 2016). Influenza viruses are continuously changing. Occasionally, these changes are so large that there is effectively no prior immunity to the virus in the population, resulting in a virus with pandemic potential. For example, influenza strains that are usually circulating in animals can become infectious to humans, such as swine flu or avian flu (World Health Organization, 2019).

In this thesis work, we consider influenza spread in Bangladesh. There, influenza causes a substantial number of deaths each year and imposes a considerable economic burden (Bhuiyan et al., 2014; Ahmed et al., 2013, 2018). In temperate countries, influenza epidemics occur during winter. Studies of the epidemiology and seasonality of influenza strains in Bangladesh indicate peak activity during the rainy season, which is from June to October in Bangladesh (Zaman et al., 2009; Islam et al., 2013; Azziz-Baumgartner et al., 2011; Noedl et al., 2012). Bangladesh has been identified as a potential source for newly emerging avian pandemic influenza, due to the high population density and

farming system (Hill et al., 2017). Highly pathogenic avian A(H5N1) and avian influenza A(H7N9) are considered serious public health threats (Hill et al., 2017; Tan et al., 2015).

1.3 Influenza surveillance

Surveillance of infectious diseases is essential for effective detection, notification and implementation of optimal control strategies and policies (Abat et al., 2016). In addition, surveillance data are necessary to build and assess mathematical transmission models and to estimate key parameters and properties of the pathogen in question.

Different surveillance systems for influenza have different advantages and disadvantages, regarding for instance representability, completeness, signal, volume, privacy, cost and timeliness. Laboratory-based surveillance has a strong signal, with high sensitivity and specificity (Ciancio and Kramarz, 2014). Laboratory-confirmed surveillance data also provide additional information such as influenza type and subtype. However, outpatient surveillance of influenza-like illness (ILI) is timelier than laboratory-based surveillance, and in outbreak settings, rapid detection and timeliness is crucial to allow mobilisation of resources (Ciancio and Kramarz, 2014). ILI outpatient surveillance also normally has a higher volume than laboratory-confirmed data. For the ILI outpatient surveillance, the diagnosis is normally not confirmed by laboratory tests, and the data are thus subject to human error and noise. Mortality-based data have the advantage that they are legally required in most countries (Declich and Carter, 1994), and are thus a low-cost surveillance data source. However, mortality data are subject to long delays, and are less useful in outbreak settings (Declich and Carter, 1994). The accuracy is also varying, as cause of death is not always clearly defined (Declich and Carter, 1994), in particular since influenza is typically a contributing cause of death. In sentinel surveillance procedures, a selected sample of sentinel sites agree to report all influenza cases. These data are not complete, as they only contain reports for a sample of sites, but they are often more reliable and timely data sources, than surveillance data based on all sites (i.e. hospitals or general practitioner (GP) offices) (Declich and Carter, 1994). Other sources of data that can be used for influenza surveillance, which are not necessarily collected for surveillance purposes, are school and work absenteeism, hospital discharge data, drug sale data and news reports (Declich and Carter, 1994).

Influenza surveillance is complex and challenging due to the changing nature, rapid worldwide dissemination and non-specificity of clinical symptoms for influenza (Ciancio and Kramarz, 2014). Multiple sources and systems are therefore often used in influenza surveillance (Ciancio and Kramarz, 2014). For example, in Norway, the Norwegian Institute of Public Health uses ILI outpatient surveillance from all GP offices and emergency departments, for influenza surveillance (Norwegian Institute of Public Health, 2016). As many people do not seek healthcare for influenza, the ILI outpatient surveillance is subject to un-

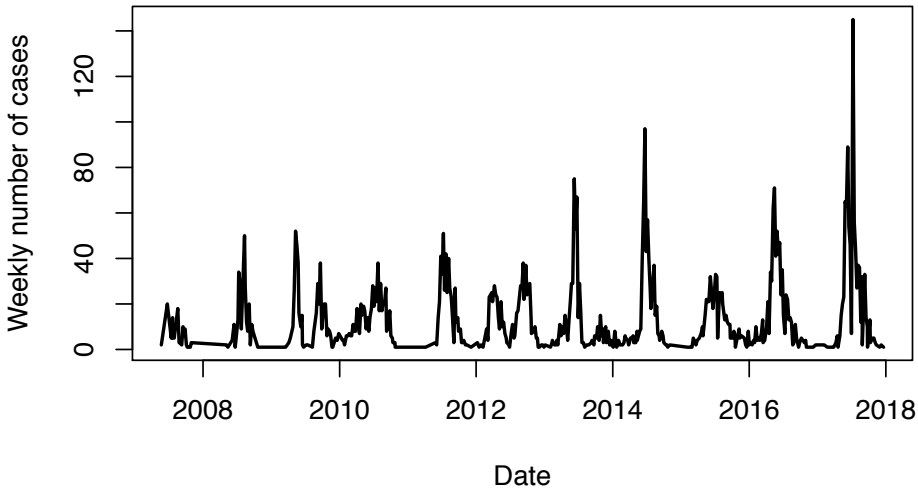

Figure 1.1: **Influenza cases in Bangladesh.** Weekly number of confirmed influenza cases in Bangladesh.

derreporting. In addition, laboratory-based surveillance of patients in hospitals and primary care is used for virological surveillance of the influenza virus. Death certificates are also used as a supplement to these main influenza surveillance systems in Norway.

Since April 2007, the Government of Bangladesh and the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) have been doing national laboratory-based influenza surveillance in Bangladesh (Azziz-Baumgartner et al., 2011). The surveillance data include data for 12 sentinel hospitals covering all seven divisions of Bangladesh. The inclusion criterion for testing is patients in the hospitals with severe acute respiratory infection, severe pneumonia for children under five (up to July 2017) or outpatients with ILI. The weekly number of cases show a clear periodic signal of one year, see Figure 1.1. Nevertheless, there are also some years with bimodality (2009, 2012 and 2013). This is particularly clear if we compare with the ILI cases from Norway in Figure 1.2. Here, the seasonal unimodal pattern is more pronounced, except from the bimodal 2009-2010 swine flu pandemic. The figures also show inherent noise and scarcity in the surveillance data from Bangladesh, compared to the Norwegian surveillance data.

In our times, we leave behind enormous amounts of digital traces. In recent years, these novel data streams have been increasingly used in disease surveillance and prediction (Simonsen et al., 2016; Bansal et al., 2016; Althouse et al., 2015). Though we do not use these types of data sources in this thesis, they could have been a valuable additional resource for our analysis, due to the low volume of the case data for Bangladesh. Novel data sources are particularly valuable in countries where disease surveillance is scarce (Simonsen et al., 2016). Traditional approaches used for disease surveillance typically have a bet-

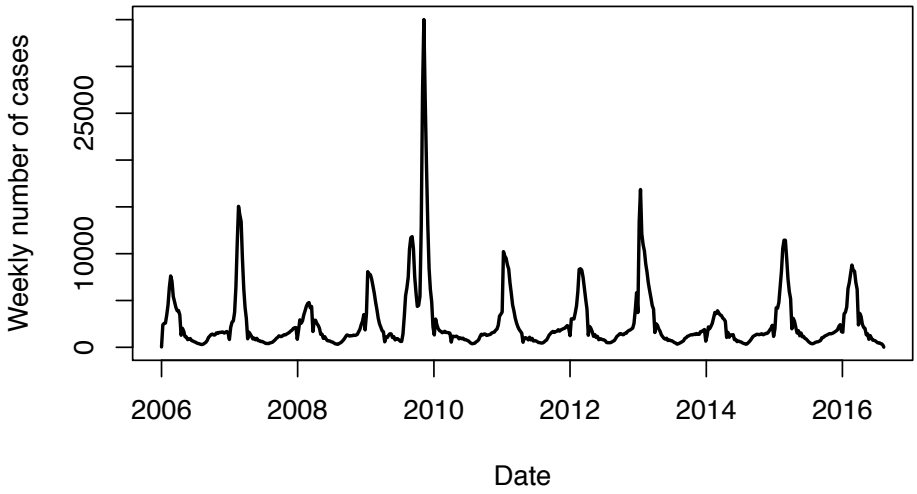

Figure 1.2: **ILI cases in Norway.** Weekly number of ILI cases in Norway.

ter signal, but lower volume potential than novel data sources like electronic health data and online sources like social media streams and search queries (Simonsen et al., 2016). Novel data sources can improve timeliness, spatial and temporal resolution of traditional infectious disease surveillance (Bansal et al., 2016; Althouse et al., 2015). CDC's prediction competition for the 2013-2014 seasonal influenza encouraged participants to use novel data sources in their predictions (Biggerstaff et al., 2016). Another famous example is Google flu trends (Ginsberg et al., 2009), which uses search query data to predict the disease activity in real-time (now-casting). Google flu trends performed well in the beginning, but later turned out to be a big disappointment (Lazer et al., 2014). Other examples using Google flu trends and/or Twitter posts for surveillance and forecasting can be found in for instance Signorini et al. (2011); Santillana et al. (2015); Nsoesie et al. (2013); Brooks et al. (2015).

1.4 Human mobility and infectious disease spread

Digital data sources contain timely information on disease and health dynamics around the world, and can be used to infer the dynamic links on everyday human individual movement (Salathe et al., 2012). Traditionally, these human mobility links have been estimated using stationary data like travel surveys or small-scale global positioning system studies, which cannot be used to understand the dynamical aspect of human mobility (Salathe et al., 2012).

Social mixing and human mobility patterns are key drivers for the spatial dissemination of infectious diseases. The airline transportation network, movement of individuals to and from work and the movement network of dollar bills (as a proxy for human mobility) are examples of movement networks that

have played important roles in understanding the spread of infectious diseases (Danon et al., 2011). The exploding use of mobile technologies enables collection of information on human mobility for "millions of individuals at once" (Pastor-Satorras et al., 2015). Mobile phone data can be used as a proxy for human mobility to understand and predict the spatial spread of an outbreak.

In low-income settings, where the burden of infectious disease is typically high, there has been a lack of appropriate data on human mobility, hindering prediction of spatial spread of disease outbreaks (Wesolowski et al., 2016). Classical models of human mobility like the gravity model (Zipf, 1946) and the radiation model (Simini et al., 2012) have primarily been developed and validated for countries in the western world (Finger et al., 2016). Generalisability to low-income settings is not obvious and straightforward, since the transportation networks and travel costs are not comparable. In this thesis we use mobile phone data to model spatial spread of influenza in Bangladesh and compare the performance to models using the radiation model and the gravity model.

Mobile phone data have previously been used in the context of infectious disease modelling for instance in studying malaria and rubella in Kenya (Wesolowski et al., 2012b, 2015a) and dengue fever in Pakistan (Wesolowski et al., 2015c). Tatem et al. (2009) use mobile phone data to estimate the travel patterns in Zanzibar for planning elimination of malaria. Bengtsson et al. (2015) use mobile phone data to predict the spatial spread of cholera in Haiti. Frias-Martinez et al. (2011) use call detail records to assess individual human mobility and social network information, to inform an agent-based model for the 2009 pandemic in Mexico. Finger et al. (2016) exploit mobile phone mobility data to evaluate the influence of a mass gathering, a religious pilgrimage, on the course of the 2005 cholera epidemic in Senegal. This heterogeneity in mobility and contact patterns would not be explained by classical and simplified human mobility models.

1.5 Mathematical models of influenza-like illness

The study of mathematical models for infectious disease dynamics in the literature date back to the work of Bernoulli in the 18th century (Bernoulli, 1760; Keeling and Rohani, 2011). The earliest mathematical model which can be used to model influenza transmission is the so-called SIR (susceptible-infectious-recovered) model, published in the 1920s (Kermack and McKendrick, 1927). The SIR epidemic model contains three disease states for individuals. They are either susceptible, infectious or recovered, and there is a system of differential equations governing the dynamics. The SIR model serves as a basis for subsequent, more sophisticated influenza models (Coburn et al., 2009). We consider extensions which include two additional disease states, an exposed compartment and an infectious asymptomatic compartment. The exposed class is the class of infected, but not yet infectious individuals. The SIR model assumes homogeneous populations, where all susceptible individuals have the same probability of encountering infection from an infectious individual in the population.

The basic reproductive number, R_0, is an important epidemiological parameter, defined as the number of expected new cases generated by one infectious individual, in a fully susceptible population. For instance, R_0 determines whether it is possible for a disease to successfully invade a population, the threshold level of vaccination required to obtain herd immunity and eradication, the final size of the epidemic and the speed of the spread, assuming a homogeneously mixing population (Keeling and Rohani, 2011; Biggerstaff et al., 2014a). Hence, R_0 has consequences both for the impact of the epidemic, and the probability of successful containment with different control strategies. As the population is normally not fully susceptible to the influenza virus, we often work with the effective reproductive number, R_e, which is the reproductive number in the population, taking the prior immunity into account. In a systematic review, Biggerstaff et al. (2014a) estimate an R_e for seasonal influenza of 1.28 (IQR 1.19 to 1.37). Most of the estimates of R_0 for seasonal influenza in the literature are for temperate countries. If R_e is below one, it is not possible for the disease to successfully invade the population. Hence, the ultimate goal for control measures is to decrease R_e below one, to ultimately control the outbreak.

In a simple SIR model, R_0 is given as the product of two model parameters – the transmission rate β, which depends on the contact rate and the probability of transmission upon a contact between an infectious and a susceptible individual, and the infectious period. The contact rate can be described by use of networks. Estimates of R_e can be found by fitting the model to incidence data, to obtain estimates of the parameters. This is how we obtain estimates of R_e in paper III, for the specific influenza transmission model we use, providing a basis for comparing different pathogens or subtypes of pathogens. Alternatively, one can estimate R_e from for instance the initial growth rate of the epidemic, the final size of the epidemic or using the average age at infection (Dietz, 1993).

Infectious disease spread is a complex process, and mathematical models offer valuable tools for understanding spreading patterns, developing and evaluating the effect of control measures for decision-making (Heesterbeek et al., 2015). Mathematical models of infectious disease also bring opportunity to predict infectious disease spread, which is essential to prevent and mitigate diseases (Pastor-Satorras et al., 2015). Quantitative forecasts from models can be used for preparedness planning, and epidemiological models are a powerful tool for providing guidelines for decision makers and public health practitioners (Keeling and Rohani, 2011). For example, CDC launched a prediction competition both for the 2013-2014 seasonal influenza and the following 2014-2015 influenza season (Biggerstaff et al., 2016, 2018).

Mathematical models can be used to provide understanding of characteristics of influenza spread and the relative importance and influence of different factors. In addition, mathematical models allow us to in silico investigate and compare effectiveness of different intervention strategies. Mathematical models provide opportunity to optimise the use of limited resources or targeted control measures, aiding in eradication and control of diseases (Keeling and Rohani, 2011). Disease outbreaks like a newly emerging pandemic influenza strain require urgent decisions to act upon. Mathematical models can aid in providing

timely guidelines to policymakers (Lofgren et al., 2014). Timely action can significantly reduce disease burden and cost (Brockmann and Helbing, 2013). A recent example is Kucharski et al. (2015), where the impact of the control measures used to fight Ebola in Sierra Leone are evaluated, and they investigate how many cases could have been averted if the control measures had been introduced earlier. Another example can be found in Chowell et al. (2009), who consider optimisation of age-targeted vaccination strategies to minimise severe morbidity and mortality for the 2009 pandemic influenza outbreak in Mexico.

The model environment allows for an ideal world where individual factors, like urbanisation, can be studied in isolation, without disturbance from other confounding factors. Mathematical modelling allows for in silico experiments that would be unfeasible, unethical and nearly impossible to perform in real life. For example, models allow us to answer questions like "How does the disease spread dynamics depend on seeding location?", "What are the most likely transmission routes?", "How effective is it to only vaccinate commuters, versus only vaccinating non-commuters?" or "What is the effect of population clustering on infectious disease spread?". In addition, stochastic models of infectious disease allow for studying scenario-based approaches, for instance worst case scenarios, best case scenarios and counterfactual scenarios. The insights gained from such modelling are often robust and generic, and can thus be applied to a wide variety of particular problems (Keeling and Rohani, 2011). Results found in a general setting are likely to be applicable to specific cases, whereas drawing general conclusions from specific country-wise case studies is questionable. In fully data-driven studies, we cannot easily control for confounding factors.

Models for infectious diseases range from simple, less realistic models assuming homogeneous mixing, to very detailed, realistic and computationally demanding agent-based models (Pastor-Satorras et al., 2015). Metapopulation models offer a compromise between simple homogeneous mixing models and agent-based models, including more structure and realism than the former, while being less detailed and computationally demanding than the latter. In metapopulation models, the nodes represent smaller patches or groups of individuals, and links represent contact between the groups. The groups can, for instance, be locations, households or age groups, or combinations of these, and can also be organised in a hierarchical structure. Household models are popular types of metapopulation models, where the individuals are partitioned into households (Neal and Roberts, 2004; Cauchemez et al., 2004; Clancy and O'Neill, 2007). High-quality data are often collected on household level (Ball et al., 2015). Household models and data were for instance used to estimate human-to-human transmissibility of avian A(H7N9) virus in China (Yang et al., 2015b) and to estimate the transmissibility of the 2009 H1N1 pandemic (Cauchemez et al., 2009; House et al., 2012; Yang et al., 2009).

As with many mathematical models, there is a trade-off between the transparency needed for understanding and the accuracy needed for prediction, in models for infectious disease spread (Keeling and Rohani, 2011). Transparency problems are for instance the main criticism for many artificial intelligence/machine learning methods, like deep neural networks. When the goal is good

predictions, there is no limit to how much heterogeneity, complexity and details that can in theory be included in the model. However, if the goal is to understand and describe infectious disease spread and the effect of different factors, like the urbanisation phenomenon in paper II, the model has to be simple enough so that one can isolate the phenomenon of interest. In addition, the amount of details included in the model must reflect the amount of data and information that is available. A complex model which is not properly informed by data sources can lead to a false sense of accuracy (Keeling and Rohani, 2011). Using a too complex model will affect the generalisability of the model to future outbreaks. Hence, the model complexity should reflect the research question. For example, Kucharski et al. (2014) find that models with a large number of age groups resulted in a worse fit than models with a moderate number of age groups, in their transmission model for pandemic 2009 influenza in Hong Kong, indicating overfitting, or that the data was not rich enough to defend the increased complexity. Another example is in a synthetic forecasting challenge for Ebola (Viboud et al., 2018), where they found that prediction accuracy was not driven by model complexity, and they denote this as their "perhaps most surprising finding" and "probably one of the most important lessons".

Moreover, which type of model to use depends on the specific setting and disease in question. When modelling spread of airborne pathogens, as we do in this thesis, homogeneous mixing, for instance within a location, can be a decent approximation. If spatial spreading patterns are of interest, as in paper II and paper III, spatial structure should be included in the model.

1.5.1 Spatio-temporal influenza spread models

Two frameworks have been used to model spatial spread of influenza in the literature: agent-based models and metapopulation models (Barbosa et al., 2018). An agent-based model can be viewed as a dynamic network, where the modelling unit is the individual, represented by a node. The individuals are kept track of and embedded in detailed contact networks, which are built to match socio-demographic data for the geographic area (typically country) studied (see for instance Ajelli et al. (2010)). Some examples of agent-based models that have been used to model influenza spread are Longini et al. (2005); Merler et al. (2011); Ferguson et al. (2006). In this thesis, the focus will be on metapopulation models where the nodes are locations, and the links represent travelling individuals. One of the earliest examples of a metapopulation model is Rvachev and Longini Jr (1985), where a metapopulation model is developed and applied to model the global spread of influenza. Another early example of a metapopulation model using detailed movement data is Sattenspiel and Herring (1998), who model the 1918 pandemic influenza in the Canadian Subarctic. Grais et al. (2004) update the metapopulation model in Rvachev and Longini Jr (1985), with more modern travel data (and volume). In Balcan et al. (2009b), a metapopulation model combined with early information on the 2009 influenza pandemic was used to predict the spread of the epidemic. Other examples of metapopulation models are for instance Tizzoni et al. (2014); Brockmann and Helbing

(2013); Meloni et al. (2011); Chao et al. (2010); Watts et al. (2005); Hufnagel et al. (2004); Colizza et al. (2007a,b).

1.5.2 Fitting infectious disease models to data

In order for the model results to be useful in specific settings, the model should be properly parametrised by empirical data. In addition, model assessment is not possible without comparing it to actual data, which requires tuning. In lack of data (and methods) to parametrise models, parameters are often assigned plausible values based on literature and expert opinion (O'Neill, 2010). Moreover, parameter estimates are often obtained by making assumptions, for instance on the infectious period, or through the use of ad hoc methods (Danon et al., 2011). However, in the last 20 years, there has been a lot of development in methods for inferring parameters of infectious disease models (de Angelis et al., 2015). Markov Chain Monte Carlo (MCMC) methods allow for parameter estimates without strong assumptions (Danon et al., 2011). Some of the earliest full Bayesian analyses using MCMC in models for infectious disease spread are O'Neill and Roberts (1999); Gibson and Renshaw (1998); O'Neill et al. (2000). Missing and partially observed epidemic data often make likelihood analysis complicated in infectious disease epidemiology, and hence data augmentation and imputation are often needed (McKinley et al., 2009). Full MCMC analyses seem to be mainly used in simpler cases with more detailed data, and not in large-scale outbreaks.

For large-scale epidemics, and for more complex and detailed models, MCMC breaks down (de Angelis et al., 2015). Evaluating or writing down the likelihood can be intractable or even impossible, especially under partially observed data (McKinley et al., 2009). In addition, there is a problem with computational power and time. However, there are many alternative methods which have been used to infer parameters of infectious disease models, which have been developed and used in complex models of large-scale epidemics. Examples of such methods are Sequential Monte Carlo (SMC), Approximate Bayesian Computation (ABC) and emulation (de Angelis et al., 2015).

Birrell et al. (2016a) use an SEIR model with age-structure in real-time monitoring of an emerging influenza epidemic using SMC to infer parameters. Farah et al. (2014) use Bayesian emulation to obtain Bayesian inference for the epidemic parameters in a model of the 2009 influenza pandemic. Cauchemez et al. (2008) use a method based on SMC to quantify the role of schools in influenza transmission and predict the effect of school closure.

Complex, large-scale computer simulation models require high quality data to improve parametrisation (Salathe et al., 2012). Often, multiple heterogeneous data sources have to be combined to inform the many parameters (de Angelis et al., 2015). Some examples of studies inferring parameters using multiple data sources are Birrell et al. (2016a,b); Jackson et al. (2015); Punyacharoensin et al. (2015); Birrell et al. (2011). There are also numerous other examples of applications of Monte Carlo methods or approximations in inferring epidemiological parameters of infectious disease models, for instance Funk et al. (2019);

Ho et al. (2018); Corbella et al. (2018); Pei et al. (2018); Probert et al. (2018); Birrell et al. (2017); Funk et al. (2017); Zimmer et al. (2017); Ajelli et al. (2016); Kucharski et al. (2016); Finger et al. (2016); Reis and Shaman (2016); Kucharski et al. (2015); Yang et al. (2015a); Schwartz et al. (2015); Poletto et al. (2014); Deardon et al. (2010); Kypraios et al. (2010); Jit et al. (2010); Ong et al. (2010); Balcan et al. (2009b); Kim et al. (2007); Salomon et al. (2002).

Another alternative is to use ABC, as first proposed by Tavaré et al. (1997). In ABC, simulations from the model with different parameters are compared to the empirical data, in order to assess which model parameters are in accordance with the observations. Simulations are thus used in order to avoid calculation of the likelihood. The parameters are accepted if the distance between the simulated data and the observed data is smaller than a specific tolerance. The result is thus an approximation to the posterior, unless the tolerance for the acceptable distance is 0, hence the *approximate* in "approximate Bayesian computation". Since ABC relies on simulations from the model, it can be used for all kinds of epidemic models where it is possible to obtain simulations (Blum and Tran, 2010). There is thus no limit to the complexity of the models which can be handled by ABC, except the computational cost of simulating multiple times from the model. For example, Irvine and Hollingsworth (2018) develop an ABC method for fitting complex agent-based models to data. In our application, we use a sequential version of ABC, ABC-SMC (Sisson et al., 2007; Toni et al., 2009), to infer the parameters of the model in paper III, due to the complexity of our infectious disease transmission model. There are many other examples of studies using ABC to infer parameters of infectious disease models, see for instance Brown et al. (2018); Meakin et al. (2019); Sun et al. (2015); Brooks-Pollock et al. (2014); Lu et al. (2013); Blum and Tran (2010); McKinley et al. (2009).

1.6 Urbanisation

From a United Nations report (United Nations, Department of Economic and Social Affairs, Population Division, 2018), we know that in 2018, 55% of the population was living in urban areas. They estimate that in 30 years, 68% of the population will live in urban areas. A key driver of the urbanisation process is migration from rural to urban areas, resulting in spatially expanding urban cities (McGranahan and Satterthwaite, 2014). Since people are continuously moving to the cities, and the cities are growing in size, we get clusters of areas with high population density and clusters of areas with low population density. Hence, as part of the urbanisation process, individuals are clustering in geographical areas (Rose et al., 2017). In this thesis, we focus on this population clustering when we study urbanisation, and will use the terms interchangeably. This is of course just a single aspect of urbanisation. Other characteristics are for example urban sprawl (Rose et al., 2017; McGranahan and Satterthwaite, 2014) and population growth (Cobbinah et al., 2015). By population clustering, we thus refer to the phenomenon that large cities are often surrounded by other

cities or suburbs with large population sizes. Rural areas also tend to appear in clusters (i.e. positioned closely together). This is found to be the case of for instance Australia, where the majority of the population is clustered around the coastal belt (Kapferer, 1990). Empirical evidence of population clustering has also been found in the Turku region of Finland (Vasanen, 2009).

Little focus has been given to the interplay between person-to-person transmission of infectious diseases and urbanisation. There have been some studies comparing health conditions in urban and rural areas, concluding that the health conditions are better in urban areas (for instance higher vaccination coverage and better access to healthcare services), and that overall health has improved with urbanisation (Alirol et al., 2011; Hay et al., 2005). High population density has been found to increase exposure to infectious diseases (Alirol et al., 2011). There is however one modelling study, Zhang and Atkinson (2008), which develops a model for the effect of urbanisation on the transmission of infectious diseases, focussing on population growth and land use development. However, the infectious disease spread model used is very simple, only allowing local transmission and not including human mobility. In addition, they do not consider the effect of interventions.

1.7 Travel restrictions

Travel restrictions aim to decrease the reproductive number by reducing the contact rate in the population. Historically, travel restrictions have a long tradition, dating back to at least the 14th century plague epidemics, where people were restricted from leaving or entering specific communities (Sattenspiel and Herring, 2003). Travel restrictions were also used during the 1918-1919 Spanish flu pandemic (Sattenspiel and Herring, 2003). More recently, internal travel restrictions were used during the Ebola outbreaks (Plan international, 2015), and for controlling the 2009 pandemic influenza virus in Mongolia, through interruption of provincial rail and road travel (Bolton et al., 2012). However, travel restrictions are very expensive, and their value in public health should be assessed in a model combined with health economic evaluations.

There have been many modelling studies investigating the effect of travel restrictions on controlling an infectious disease. Germann et al. (2006) and Ferguson et al. (2006) study the spread of a hypothetical pandemic influenza in the United States, and find that domestic travel restrictions do not have an effect on the final size of the epidemic, but slightly delay the epidemic time course. In a systematic review, Mateus et al. (2014) find that internal travel restrictions can delay influenza epidemics, and that extensive travel restrictions can reduce the final size. They also find that travel restrictions have minimal impact in urban centres with dense population and high mobility. Lee et al. (2012) found that internal travel restrictions in Korea delayed the peak timing and had a small reduction effect on the epidemic peak size. In addition, as a real life experiment, Brownstein et al. (2006) consider the influenza epidemic following the travel ban after 9/11 and find a delayed influenza season. They

speculate that this is due to the decrease in air traffic (however, note also the rebuttal in Viboud et al. (2006c)).

1.8 Vaccination

Vaccination reduces the effective reproductive number by reducing the proportion susceptible, by immunising susceptible individuals. Vaccination is probably the most important discovery in medicine in terms of saving lives. Vaccination is a more commonly used and more widely studied mitigation measure against infectious diseases than travel restrictions. Vaccination has been claimed to be the most effective way of preventing the public health impact of (pandemic) influenza (US Department of Health and Human Services, 2017). Vaccination can also be used as a control measure in an outbreak setting, as is the focus of this thesis. For example, Germann et al. (2006) and Ferguson et al. (2006) find that vaccination significantly reduces the final size of influenza epidemics, and in particular, is more effective than travel restrictions.

In case of limited vaccination supplies, one can consider different vaccination strategies to optimise the effect of vaccination. Spatially uniform vaccination strategies (pro rata) are usually the default vaccination guidelines (Matrajt et al., 2013). For instance, uniform vaccination guidelines were given during the 2009 influenza pandemic in for example Norway, Ireland and the United States (de Blasio et al., 2012; Cotter et al., 2010; Centers for Disease Control and Prevention, 2009; Rambhia et al., 2010). However, the optimal vaccination strategy measured in terms of minimising final size is not necessarily uniform (and most likely not). There are numerous examples of spatially targeted vaccination and antiviral strategies in the literature (Matrajt et al., 2013; Shrestha et al., 2016; Dimitrov et al., 2011; Wu et al., 2007; Keeling and White, 2011; Azman and Lessler, 2015; Duijzer et al., 2018b; Nguyen and Carlson, 2016; Keeling and Shattock, 2012; Araz et al., 2012; Teytelman and Larson, 2013).

Chapter 2

Aims for the thesis

The common denominator for this thesis is that all parts consist of developing new models and methodology with important applications, for systems of individuals which interact by forming networks. We use methods in network science to improve our understanding of infectious diseases and pain tolerance.

The aims of paper I are to:

- Provide better insight into the interplay between pain and social determinants
- Explore potential confounders
- Investigate how social capital and centrality are associated with pain tolerance
- Develop novel methodology for handling censoring in social network analysis

The aims of paper II are to:

- Investigate the consequences of urbanisation on infectious disease spread and effectiveness of prevention measures in a generic setting
- Develop novel methodology for generating spatial fields with controllable levels of population clustering

The aims of paper III are to:

- Develop a large-scale spatial transmission model informed by daily aggregated mobile phone data from Bangladesh
- Assess the performance of different detail levels for mobile phone mobility data in infectious disease modelling
- Assess how the gravity and radiation model capture the relevant mobility for infectious disease transmission in a low-income setting
- Increase the understanding of the spatial spreading patterns of influenza in Bangladesh
- Estimate the model parameters by using ABC informed by scarce influenza case data
- Obtain and compare estimates of R_e for different influenza seasons

Chapter 3

Data sources

3.1 Fit futures (paper I)

The study participants in paper I are from the Fit Futures study from 2010-2011, which is part of the Tromsø study in Norway. Fit futures is an extensive youth cohort study of adolescents in the first year of upper secondary school, normally aged 15-17 years. All the pupils in Tromsø and Balsfjord municipality were invited to the study, and the response rate was more than 90%. The pain tolerances of the participants were measured by a cold-pressor test, where the participants held their dominant hand in circulating cold water of temperature 3°C for as long as they could endure. The pain tolerance is censored at 105 seconds. After removing the participants with cognitive disabilities, the total number of adolescents in the study for whom we have pain tolerance information is 997.

The adolescents in the study filled out extensive questionnaires, were interviewed and participated in tests and medical examinations. The data contain sociodemographic variables, lifestyle information, health information, puberty information as well as information about their social network. The participants could name up to five friends whom they had had most contact with the last week, providing the social network. The nominated friends had to attend first year of upper secondary school in the Tromsø area.

3.2 Mobile phone mobility matrices (paper III)

The mobile phone data are provided by Telenor, through its subsidiary Grameenphone, the largest mobile operator in Bangladesh. The data contain information for more than 60 million customers. The mobility matrices date from 1. April 2017 to 30. September 2017. This coincides with the period with highest influenza activity. Call detail records from mobile phone data can be used to infer human mobility patterns. Every time a contact is made between a user and a receiver, their locations and the time point of the contact are registered. For each day, a user is assigned to their most frequent tower location. These individual time series are aggregated to upazila level (administrative level of Bangladesh, there are 544 upazilas in total in Bangladesh), by counting the number of individuals who have transitioned between upazilas from one day to the next. The mobile phone mobility data are anonymised. The transition matrices are also used to estimate the population size in each administrative unit, by scaling the mobility counts to the total population of Bangladesh, 163 million in 2016 (The World Bank, 2018). The daily mobility matrices are scaled up to the respective

estimated population size in each upazila. Moreover, we will use the average over the daily mobility matrices, denoted by the time-averaged mobility.

3.3 Influenza case data (paper III)

The influenza case data were plotted in Figure 1.1 and the surveillance system was described in Section 1.3. The data contain demographic information for the cases and information on the influenza type. We use the residence upazila of each case to estimate the observed case count in the different upazilas. We focus mainly on the 2017 seasonal influenza in this thesis, and the number of positive cases for 2017 was 890 out of 4229 tested.

Chapter 4

Models and Methods

The models and methods used in the three papers are introduced in this chapter. The methods used in paper I are described first. The social network is defined in Section 4.1. The peer effect in pain tolerance is analysed in two ways. We fit network autocorrelation models, which provide a neighbourhood correlation effect and measure how similar an individual is to the average of their friends. These models are introduced in Section 4.1.1. In addition, we measure pairwise correlation between friends' pain tolerance, introduced in Section 4.1.2.

The model structures used in paper II and III are similar. In Section 4.2, the equations for the local transmission process in the stochastic metapopulation model are described, common for paper II and III. The extension to vaccination is introduced in Section 4.3, which is used in paper II. The gravity model is introduced in Section 4.4 and the radiation model is introduced in Section 4.5. The clustering algorithm developed for paper II is presented in Section 4.6. The details of the infectious disease models used in paper II and paper III are specified in Section 4.7 and 4.8. Finally, the parameter fitting procedure for paper III is laid out in Section 4.9.

4.1 Social network analysis (paper I)

In order to assess the relationship between friendship ties and individual traits, social network analysis is used. In a social network, each individual is represented by a node, and there is a link between two nodes if they share a relationship. In a friendship network, the links represent friendships. It might be that individual i nominates individual j as a friend, but that the opposite does not occur, so that the friendship is not reciprocal. The friendship network is thus a directed network. If individual i nominates individual j as a friend, we have a directed link in our network from individual i to individual j. The node i is called the start node of the link, and the node j is the end node of the link. This gives us an adjacency matrix $\mathbf{W}$, which contains the friendship information. The adjacency matrix is defined by

$$W_{ij} = \begin{cases} 1, & \text{if } i \text{ nominates } j \text{ as a friend,} \\ 0, & \text{otherwise,} \end{cases}$$

where W_{ij} is the element in the ith row and the jth column of the matrix. Note that the diagonal of $\mathbf{W}$ consists of zeros.

4.1.1 Network autocorrelation model

In order to estimate the effect of friends' pain tolerance on the individual's pain tolerance, we fit a network autocorrelation model (O'Malley and Marsden, 2008) to the data. The model is given by

$$\mathbf{Y} = \rho \mathbf{W}_N \mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{Y}$ is the vector of pain tolerance for all the individuals, ρ is the autocorrelation between an individual's pain tolerance and the friends' pain tolerance, $\mathbf{W}_N$ is the adjacency matrix, normalised by the number of friends for each individual so that the rows sum to one, $\mathbf{X}$ is a matrix with other explanatory variables, $\boldsymbol{\beta}$ is the corresponding vector of coefficients and $\boldsymbol{\epsilon}$ is a vector of random noise, assumed to be normally distributed with variance σ^2. The autocorrelation coefficient ρ is a measure of the degree to which an individual has a similar pain tolerance to the overall pain tolerance of their friends, and can be interpreted as the increase of the individual's pain tolerance with an increase of the average pain tolerance of the individual's friends by one second.

4.1.1.1 Extension to censored data

We consider the problem of finding relationships between a right-censored variable and the social network. To do so, we develop a method for fitting a network autocorrelation model for the outcome, which is based on a pseudo likelihood approximation.

Since the full likelihood is intractable, we approximate the likelihood by the pseudo likelihood (Besag, 1975), where the observations are assumed to be conditionally independent. Due to the censoring, we do not observe the response variable (the pain tolerances), but we observe $Z_i = \min(Y_i,\ 105)$. Let δ_i be a censor indicator, so δ_i is one if observation i is censored and zero otherwise. The pseudo likelihood is given by

$$L(\mathbf{z}, \boldsymbol{\theta}, \mathbf{y}) = \prod_{i=1}^{N} \left(\frac{1}{\sigma} \phi \left(\frac{z_i - \mathbf{x}_i^T \boldsymbol{\beta} - \rho y_{-i}}{\sigma} \right) \right)^{1-\delta_i} \left(1 - \Phi \left(\frac{105 - \mathbf{x}_i^T \boldsymbol{\beta} - \rho y_{-i}}{\sigma} \right) \right)^{\delta_i},$$

where $y_{-i} = \sum w_{ij} y_j$, with w_{ij} being the jth element of the ith row of $\mathbf{W}_N$, so y_{-i} is the average pain tolerance of the friends of individual i, $\mathbf{x}_i$ is the vector of explanatory variables for individual i, N is the number of individuals, ϕ is the standard normal density, Φ is the standard normal cumulative probability function and $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}, \sigma)$. We use an EM algorithm to maximise the pseudo likelihood over the parameters, since the y_{-i} are unknown, due to the censored observations. The parameter estimates and the y_{-i} are updated cyclically. We start with an initial guess for the $\mathbf{y}$, $\hat{\mathbf{y}}^0$, where the observed pain tolerances are kept fixed, and the censored pain tolerances are random draws from a truncated normal on $[105, \infty)$. A first estimate for the model parameters, $\hat{\boldsymbol{\theta}}^1$, is then found by

$$\hat{\boldsymbol{\theta}}^1 = \text{argmax}_{\boldsymbol{\theta}} L(\mathbf{z}, \boldsymbol{\theta}, \hat{\mathbf{y}}^0).$$

With an estimate of $\boldsymbol{\theta}$, we can obtain an updated estimate for $\mathbf{y}$, $\hat{\mathbf{y}}^1$, by

$$\hat{y}_i^1 = \begin{cases} y_i, & \text{if } \delta_i = 0, \\ \text{mean of a trN}(\hat{\rho}^1 \mathbf{W}_{Ni} \hat{\mathbf{y}}^0 + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^1, \hat{\sigma}^1) \text{ on } [105, \infty) & \text{otherwise,} \end{cases}$$

where trN is the truncated normal distribution, $\hat{y}_i^1$ is the ith element of $\hat{\mathbf{y}}^1$ and $\mathbf{W}_{Ni}$ is the ith row of the matrix $\mathbf{W}_N$. We continue updating using $\hat{\mathbf{y}}^k$ to find $\hat{\boldsymbol{\theta}}^{k+1}$ by

$$\hat{\boldsymbol{\theta}}^{k+1} = \text{argmax}_{\boldsymbol{\theta}} L(\mathbf{z}, \boldsymbol{\theta}, \hat{\mathbf{y}}^k).$$

Then $\hat{\boldsymbol{\theta}}^{k+1}$ is used to find $\hat{\mathbf{y}}^{k+1}$ by

$$\hat{y}_i^{k+1} = \begin{cases} y_i, & \text{if } \delta_i = 0, \\ \text{mean of a trN}(\hat{\rho}^{k+1} \mathbf{W}_{Ni} \hat{\mathbf{y}}^k + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^{k+1}, \hat{\sigma}^{k+1}) \text{ on } [105, \infty) & \text{otherwise.} \end{cases}$$

The updating steps are repeated until we have reached convergence.

The variances of the parameter estimates are estimated by a jackknife procedure, with a normalisation factor specific for networks, since the observations are not independent (Snijders and Borgatti, 1999).

4.1.2 Pairwise correlations

Since the pain tolerance observations are right-censored, pairwise correlations are computed by Kendall's τ, which is a rank correlation coefficient. We compute the correlation between the pain tolerance of the start node and the end node of all the links. In order to assess the significance of the correlation, we perform a permutation test. This is done by randomly shuffling the observations for all the individuals, while keeping the network structure fixed.

4.2 Infectious disease modelling (paper II and III)

For paper II and III, we use a stochastic metapopulation model, where individuals can be in five different states depending on their disease status – susceptible to the disease, exposed, infectious with symptoms, asymptomatic infectious or recovered. The individuals are further divided into spatial locations, where homogeneous mixing is assumed within each spatial unit, that is, no heterogeneity between individuals within a spatial location. There is a separate set of stochastic differential equations governing the local disease dynamics in every location, and the processes are coupled through individuals who travel, according to the mobility network. The networks consist of nodes which are locations (upazilas

for Bangladesh, block/pixel units for the population clustering application) and edges which represent mobility between the locations. These edges are weighted according to the amount of travel. Our metapopulation model set-up is similar to the infectious disease models used in for instance Balcan et al. (2009a) and Colizza et al. (2007b), where they model the global spread of ILI. The local disease model is illustrated in Figure 4.1, and the stochastic differential equations for one time step, Δt, are

$$S^i(t+\Delta t) = S^i(t) - \text{Binom}\left(S^i(t),\ \beta \Delta t I^i(t)/N_i^t + r_\beta \beta \Delta t I_a^i(t)/N_i^t\right), \quad (4.1)$$

$$\begin{aligned} E^i(t+\Delta t) = E^i(t) + \text{Binom}\left(S^i(t),\ \beta \Delta t I^i(t)/N_i^t + r_\beta \beta \Delta t I_a^i(t)/N_i^t\right) \\ -\text{Multinom}\left(E^i(t),\ p_a \lambda \Delta t,\ (1-p_a)\lambda \Delta t\right), \end{aligned} \quad (4.2)$$

$$I^i(t+\Delta t) = I^i(t) + \text{Binom}\left(E^i(t),\ (1-p_a)\lambda \Delta t\right) - \text{Binom}\left(I^i(t),\ \gamma \Delta t\right), \quad (4.3)$$

$$I_a^i(t+\Delta t) = I_a^i(t) + \text{Binom}\left(E^i(t),\ p_a \lambda \Delta t\right) - \text{Binom}\left(I_a^i(t),\ \gamma \Delta t\right), \quad (4.4)$$

where β is the transmission probability per unit time, N_i^t is the population in location i on day t, $1/\lambda$ is the average latent period, p_a is the probability of being asymptomatic conditional on infection, r_β is the reduced infectiousness of the asymptomatic infectious, $1/\gamma$ is the average infectious period, $\text{Binom}(n,p)$ is the binomial distribution with n trials and success probability p and $\text{Multinom}(n,p_1,p_2)$ is the multinomial distribution with n trials and success probabilities p_1 and p_2. $S^i(t)$ is the number susceptible in location i at time t. Similarly, $E^i(t)$ is the number exposed, $I^i(t)$ is the number infectious and $I_a^i(t)$ is the number asymptomatic infectious. We assume that the population size is constant during the epidemic, so that $R = N - S - E - I - I_a$, where R is the number recovered. The basic reproductive number of the model is given by $R_0 = \frac{\beta}{\gamma}(1 + (r_\beta - 1)p_a)$ (Colizza et al., 2007b).

4.3 Implementing vaccination (paper II)

In paper II, we investigate different vaccination strategies. Including vaccination is a straightforward extension of the disease dynamics in Equations 4.1-4.4 to include five more compartments: individuals can be vaccinated and immune, vaccinated but susceptible, vaccinated exposed or vaccinated infectious with or without symptoms, with reduced infectiousness. The vaccinated and immune individuals behave similarly to the recovered class. We assume that individuals who are vaccinated but not immune behave similarly to the susceptible class. We have assumed that the vaccinated, susceptible individuals who get infected have the same disease progression as the unvaccinated, but they have reduced infectiousness. For simplicity, we only vaccinate the susceptible individuals. In practice, this means that we slightly overestimate the effective proportion vaccinated. Our vaccination model is of course a simplification, assuming that vaccinated individuals either gain complete or no immunity to the disease. Several more realistic aspects could have been included, see for instance the discussion

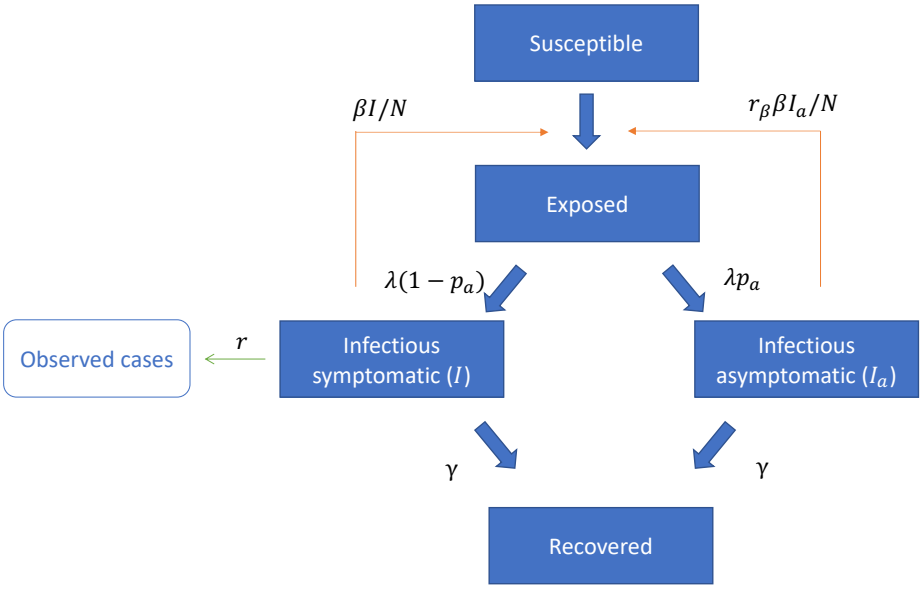

Figure 4.1: **Local infectious disease process.** The different compartments, transitions and transition rates for the infectious disease model. The susceptible can transition to the exposed class, with a rate which depends on the infectious symptomatic (I) and asymptomatic (I_a), the exposed can transition to the infectious classes, and the infectious can transition to the recovered class. β is the transmission rate, $1/\lambda$ is the average latent period, $1/\gamma$ is the average infectious period, p_a is the probability of being asymptomatic conditional on infection, r_β is the reduced infectiousness of the asymptomatic infectious and r is the probability of an infectious symptomatic case being observed (used in paper III).

in Britton (2010). However, we do not believe that these simplifications affect the relative performance of the different vaccination strategies.

We investigate three different vaccination strategies – vaccinating uniformly in space (pro rata), preferentially vaccinating rural locations and preferentially vaccinating urban locations. In all settings, we use a vaccination coverage of 40%, similar to the vaccination coverage in Norway during the 2009 influenza pandemic (de Blasio et al., 2012). We assume that the vaccines are introduced 75 days after the seeding day, and that the doses are distributed uniformly each day for six weeks. In the urban vaccination strategy, we only vaccinate the 50% of locations with the largest population size, and in the rural vaccination strategy, we leave out the 2% largest population size locations. The same total number of vaccines is used in all settings.

4.4 Gravity model (paper II and III)

The gravity model (Zipf, 1946) is a famous model for predicting population movement, and is a popular choice when modelling human mobility in infectious disease spread models (Truscott and Ferguson, 2012; Wesolowski et al., 2015b). For example, it has been found to successfully capture both spatio-temporal influenza (Viboud et al., 2006b) and measles (Xia et al., 2004) dynamics. The gravity model assumes that the number of individuals that move between locations i and j, T_{ij}, is given by

$$T_{ij} = \frac{m_i^\alpha n_j^\beta}{f(r_{ij})},$$

where m_i and n_j are the population sizes in location i and j, respectively, r_{ij} is the distance between the locations, α and β are free parameters and f is an increasing function, which can be selected to fit empirical data. A typical choice for f is $f(x) = x^\gamma$, where γ is a free parameter.

4.5 Radiation model (paper II and III)

The radiation model is a parameter-free alternative to the gravity model and claimed to be more universal (Simini et al., 2012). In the radiation model, the number of travellers between i and j is assumed to be dependent on the population density between i and j rather than the distance directly. The radiation model is given by

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})},$$

where s_{ij} denotes the total population in the circle with radius r_{ij}, centred at location i, and T_i is the total number of travellers from location i, often assumed proportional to the population size.

4.6 A theoretical single-parameter model for urbanisation (paper II)

In order to investigate how population clustering affects infectious disease spread, we need a series of countries where everything is fixed, apart from the population clustering, which varies continuously and in a controllable manner. For this purpose, we generate a synthetic country, and then apply a clustering algorithm to that country, which generates different versions of the country with different population clustering levels.

Our fictional country is a squared country, consisting of many smaller block units. Each block unit is assigned a population size. In our application, the population sizes were assigned randomly, as draws from a gamma distribution fitted to population sizes for Norwegian municipalities, and then rescaled to

match the desired total number of individuals. Then we apply the clustering algorithm to the country, by reshuffling the block units with the aim of monotonically increasing the population clustering in a continuous manner.

In order to induce positive correlations between the neighbouring locations, we use a geostatistical model to generate a random spatial field where we control the correlations. We use a model with a Matérn covariance function. In our application, we use a range parameter of 5.0 (measured in block units) and a process variance of 0.1, so that the covariance, C, between population sizes in two locations with distance d block units apart is (Cressie and Wikle, 2011, p. 126, section 4.1)

$$C(d,\kappa) = 0.1(2^{\kappa-1}\Gamma(\kappa))^{-1}(d/5.0)^{\kappa}K_{\kappa}(d/5.0),$$

where K_κ is a modified Bessel function of the second kind of order κ and the parameter κ is varied to control the population clustering. The block units are then mapped to the random spatial fields by maintaining the order, so that the block unit with the largest population size is mapped to the location of the largest number in the random spatial field, the block unit with the second largest population size is mapped to the location of the second largest number in the random spatial field and so on. The parameter κ controls the amount of population clustering. Hence, our model is a single-parameter model for urbanisation. The resulting versions of the country for some sampled values of κ are given in Fig 4.2. The version in the upper left of Fig 4.2 is the country without any clustering. Population clustering clearly increases with κ.

4.7 Infectious disease model specification for paper II

The infectious disease model is a metapopulation model as described in Section 4.2. We use a time step of $\Delta t = 12$ hours, in order to distinguish day time from night time. Every individual has a home location and a work location. During day time, individuals mix in their work location, while at night time they mix in their home location. The same individuals commute on the same links every day, as this has been shown to be an important realistic ingredient (Keeling et al., 2010). A gravity model fitted to data on commuting between Norwegian municipalities is used to model commuting. In addition to commuting travel, we include some non-deterministic travel. Every non-commuting individual has a fixed probability of travelling to a random location every morning. The destination location is random with probability proportional to population size. We let them travel for 24 hours (two time steps, day and night) before they return to their home location.

The number of people in location i during day time, $N_i^{\text{day},t}$, and night time, $N_i^{\text{night},t}$, are given by

$$N_i^{\text{day},t} = N_i^{\text{home}} + \sum_j w_{j,i} + \sum_j u_{j,i}^t - \sum_j u_{i,j}^t,$$

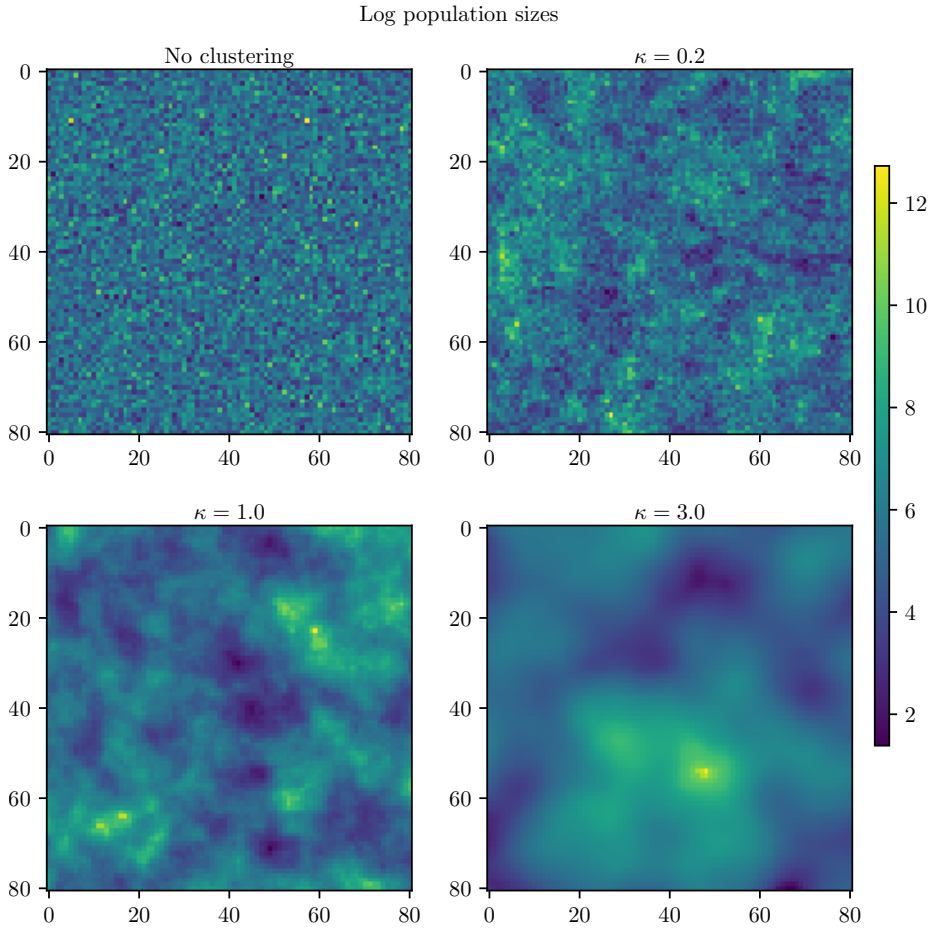

Figure 4.2: **Clustering levels.** Generated versions of the country for various κ. Upper left: No clustering. Upper right: $\kappa = 0.2$. Lower left: $\kappa = 1.0$. Lower right: $\kappa = 3.0$.

$$N_i^{\text{night},t} = N_i^{\text{home}} + \sum_j w_{i,j} + \sum_j u_{j,i}^t - \sum_j u_{i,j}^t,$$

where N_i^{home} is the number of people living in location i who do not commute, $w_{i,j}$ is the number commuting from i to j, $u_{i,j}^t$ is the number of non-commuters travelling from location i to location j on day t, and the t indicates that the number of people varies from day to day.

Hence, in the beginning of a new day, the commuters travel to their work location, and the non-commuting travellers to their destination. Then everyone mixes in their current location according to Equations 4.1-4.4 for 12 hours. Then the commuters are sent back to their home location, and everyone mixes in their current location for 12 hours. Then the non-commuting travellers are sent back home, and a new day begins where everything is repeated.

4.8 Infectious disease model for Bangladesh (paper III)

Also in paper III, the infectious disease model is a metapopulation model, as described in Section 4.2. We use a spatial resolution of upazila. As the mobile phone mobility matrices have a resolution of one day, we use a time step of $\Delta t = 24$ hours. The data do not contain individual identifiers, and we thus let those who travel every day be a random draw from the population, with uniform probability. Hence every day, we draw random individuals from the population who travel, according to the mobility estimates for that day.

The number of people in location i on day t, N_i^t, is given by

$$N_i^t = N_i + \sum_j w_{j,i,t} - \sum_j w_{i,j,t},$$

where N_i is the population size in location i and $w_{i,j,t}$ is the number of travellers from location i to location j on day t, estimated from the mobile phone data. We will also model the influenza epidemic for dates outside the range of the dates for which we have daily mobility matrices. For these days, we let $w_{i,j,t}$ be the time-averaged mobility.

Hence, in the beginning of a new day, individuals travel according to the mobility estimates. Then everyone mixes in their current location according to Equations 4.1-4.4 for one day. Then the travellers are sent back to their home location. Then the next day begins, where individuals travel according to the mobility estimates for the next day.

We assume a binomial observation model, so that the number of observed cases is binomial with parameters the actual symptomatic incidence and success probability r. Hence r is the probability that a symptomatic influenza case will be detected in the sentinel hospital case data.

4.9 Parameter estimation – Approximate Bayesian Computation (paper III)

We model the 2017 seasonal influenza epidemic in Bangladesh using the hospital case data to tune the parameters of our model. We estimate two parameters, the transmission parameter β as in Section 4.2, which determines R_e, and the reporting probability r. We assume that the mean latent period and the mean infectious period are fixed. In order to calibrate our model to the observed case data, we use ABC-SMC (Sisson et al., 2007; Toni et al., 2009). In an ABC-SMC, the posteriors are approximated sequentially, with a gradually decreasing tolerance level for the difference between the simulated and the observed data. The distance is measured in summary statistics. In our application, we use four summary statistics – the final size of the epidemic as the total number of observed cases, the duration of the epidemic, defined as the number of days between the day where 90% of the total number of observed cases have occurred and the day where 10% of the total number of observed cases have occurred, the peak number of observed infected individuals and the shape of the global observed incidence curve. The shape of the incidence curves in the observed data and the simulated observed data are compared by aligning the peaks.

As the date of the seeding event is unknown, we use the time-averaged mobility in the parameter fitting procedure, which is also more computationally efficient. After obtaining the parameter estimates, the seeding date is tuned to match the observed peak in the case data.

We assume prior distributions for the transmission parameter β and the reporting probability r. We fix the mean latent period to 1.9 days, in accordance with the fact sheet on seasonal influenza from WHO (World Health Organization, 2016), which states that the average latent period is two days. Similar values for the latent period for influenza can be found in Birrell et al. (2017); Corbella et al. (2018); Balcan et al. (2009a); Longini Jr et al. (2004). We fix the mean infectious period to three days, based on the infectious period used in previous models in the literature, for both seasonal and pandemic influenza (Germann et al., 2006; Balcan et al., 2009a; Colizza et al., 2007a; Longini Jr et al., 2004; Bajardi et al., 2011).

The reporting probability is likely to be very small, as we only have laboratory-confirmed hospital cases for 12 hospitals in Bangladesh. We assume a beta distribution as our prior distribution for r, with shape parameters 1 and 7, corresponding to a mean of $1/8$ and a variance of $7/(8^2 \cdot 9)$. This distribution is skewed towards 0. For the transmission parameter β, we assume a uniform prior on $(0.4, 0.7)$.

Chapter 5

Summary of the papers

5.1 The peer effect on pain tolerance (paper I)

The social network among the adolescents in the Fit Futures study is given in Figure 5.1, where the nodes are coloured by pain tolerance indicators. By visual inspection, we find some clustering of similar nodes in the network. Note that approximately half of the observations are censored. We find significant pairwise correlation in pain tolerance among friends, also when controlling for sex. We fit a network autocorrelation model, and find a significant peer effect, with an estimated autocorrelation coefficient of 0.44 (p-value $1.40 \cdot 10^{-12}$).

As sex is known to be assortative in social networks, and females have been found to have a lower pain tolerance than males, we control for sex. Controlling for sex decreases the estimated autocorrelation coefficient to 0.21 (p-value 0.0049). In agreement with the literature, we find that the boys have higher pain tolerance than the girls. We also control for school programme and age. Age does not appear to have an effect on pain tolerance, while those attending a general studies programme have significantly higher pain tolerance than those attending a vocational programme.

We further control for potential confounding variables, like the lifestyle factors smoking and physical activity, socioeconomic status and test sequence among the adolescents. Controlling for lifestyle factors and socioeconomic status does not have a large effect on the autocorrelation, but we find that lifestyle factors are associated with pain tolerance. The effects are in the expected directions – those who smoke have lower pain tolerance, and physically active individuals have higher pain tolerance. Test sequence appears to have a positive effect on pain tolerance, so that those who are tested later have a significantly higher pain tolerance. Controlling for test sequence slightly decreases the autocorrelation effect, but it is still significant. Though we have tried to adjust for possible confounders, the peer effect might be due to an unknown confounding variable. Stratifying on gender, we find that the peer effect is present only among males, and only through their male-male friendships. This suggests a possible masculinity effect or peer pressure effect among males. We do not find any significant effect of friendship size or popularity on pain tolerance, but the number of nominated friends (truncated at five) is borderline significant.

Our study is the first social network analysis of pain tolerance, and is in accordance with different experimental pain studies showing that social determinants are important factors for pain tolerance. Our findings are a contribution to understanding the social factors of pain and we hope to stimulate more research in the same direction.

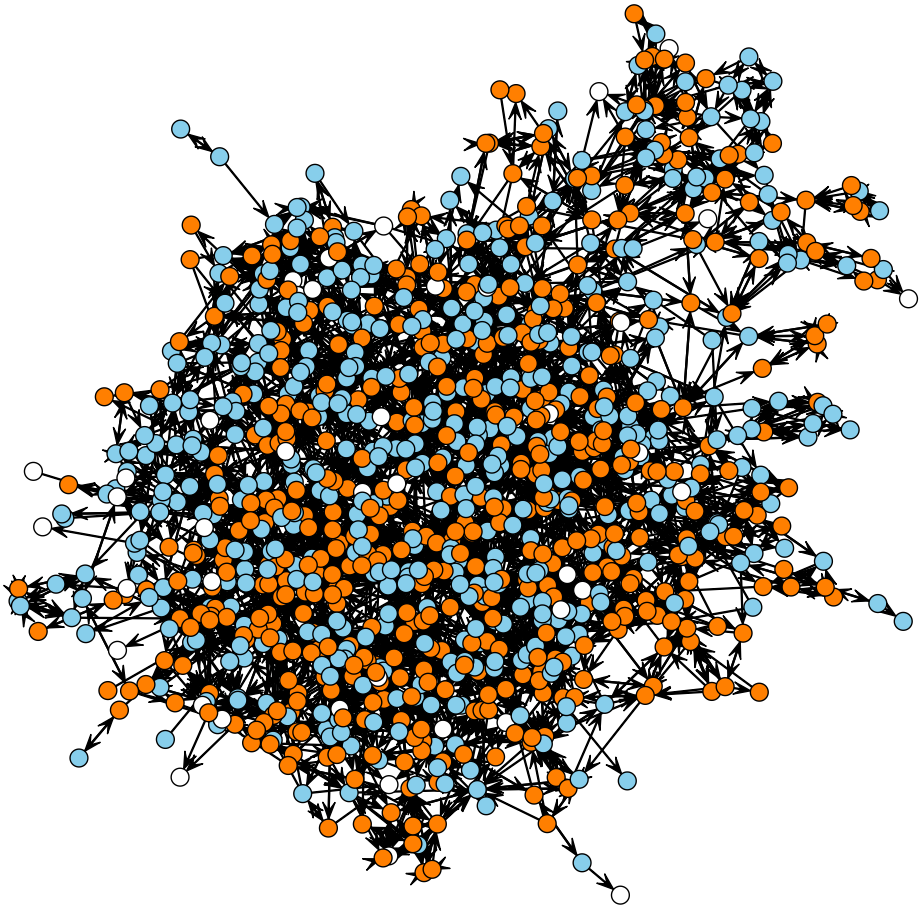

Figure 5.1: **Social network among the adolescents in the Fit Futures study.** The nodes are coloured by pain tolerance indicators. The orange individuals are the censored adolescents (those who held their hand in the water for more than 105 seconds), the blue individuals are those who did not hold their hand in the cold water for the full 105 seconds, and the white individuals are those for which the pain tolerance information is missing. The isolates are left out for visualisation purposes.

5.2 A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions (paper II)

In paper II, we investigate the effect of urbanisation on infectious disease spread in silico. Our purpose is not to describe single outbreaks in specific populations, or find the "best model" or strategy for a specific country, but rather study the phenomenon from a generic and principled point of view. This is done by generating a fictional country, with various levels of clustering. We use a plausible population size distribution based on a gamma distribution which is fitted to Norwegian municipality data. We also integrate plausible commuting patterns, by a gravity model fitted to data on commuting between Norwegian municipalities. Hence our study is a general and principled study of the urbanisation phenomenon represented in a very simplified and theoretical way, yet conserving important elements of realism.

We explore how urbanisation affects the disease dynamics and the effect of internal travel restrictions and vaccination, by performing simulations when our design parameter κ is varied. We hypothesise that the disease dynamics and effectiveness of control measures may vary with urbanisation, and the urbanisation can thus have direct important implications for the effectiveness and design of interventions. Our model parameter κ is a tuning parameter which controls the clustering, and the higher κ, the higher the clustering. We only consider restrictions on within-country non-commuting travel. The motivation behind this choice is to only postpone non-critical travel. We investigate three different vaccination strategies – a uniform vaccination strategy, an urban vaccination strategy and a rural vaccination strategy. In our application, we focus on an influenza-like illness with transmission potential corresponding to a pandemic influenza (basic reproductive number $R_0 = 1.5$). We choose outcome measures final size of the epidemic (total number infected), peak date (the day with the highest activity) and peak prevalence (the proportion infected on the peak date).

We find that in a baseline scenario with no control measures (no internal travel restrictions and no vaccination), urbanisation does not seem to have a large impact on the final size, timing or peak height of the epidemic. However, we find spatial clustering of peak dates in the locations, which is higher for the higher population clustering levels, causing more stress on the healthcare systems.

We find that population clustering has a stronger effect when we implement travel restrictions. Most importantly, we find a decrease in final size with increased population clustering. As expected, the travel restrictions also decrease the final size, and this effect is stronger for the higher clustering levels. Further investigating how this effect depends on how rural/urban the location is, we find that for the highest clustering levels, the effect of travel restrictions on reducing final size is largest for the more rural locations. Hence, according to our model, travel restrictions are most effective on reducing final size in the countries with high population clustering, and the effect is most visible in the less urban parts

of the country.

Travel restrictions appear to delay and reduce the peak. The peak is higher and earlier for higher κ. We find that the effect of travel restrictions on delaying the epidemic peak is larger for low κ values. The reduction in peak prevalence due to the travel restrictions is larger for lower clustering levels.

Though we do find that travel restrictions decrease the final size of the epidemic, the effect size was surprisingly low. For 99% travel restrictions, the final size was reduced by up to 19%. In practice, it may not be possible to achieve 99% travel restrictions, due to both compliance issues and economic costs. The vaccination strategies were more effective in reducing the final size. Hence, travel restrictions are more relevant in scenarios with no or few vaccines available.

Vaccination is slightly more effective in reducing the final size for low clustering levels than for high clustering levels. The urban vaccination strategy is found to be the most effective strategy for decreasing final size, however only slightly better than the uniform vaccination strategy. The urban vaccination strategy decreases the final size by 67% for the lowest clustering level and 63% for the highest clustering level. Preferentially vaccinating the more rural locations is clearly the most inferior vaccination strategy (42% reduction in final size for $\kappa = 3.0$ and 44% for the no clustering scenario). All vaccination strategies reduce the peak prevalence. As opposed to the travel restriction settings, the epidemic progression is unaffected by population clustering under all three vaccination strategies, as for the baseline scenario. This is possibly due to travel restrictions breaking down specific infection routes, as opposed to the vaccination.

5.3 Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh (paper III)

We develop a large-scale spatial transmission model informed by daily aggregated mobile phone data from Bangladesh, by extending the fine-scaled stochastic metapopulation model developed in paper II. We investigate which detail level is necessary to model infectious disease spread transmission, using influenza in Bangladesh as a case study. We perform a systematic comparison of the disease spread model informed by daily mobility (reference) with predictions informed by time-averaged mobility and mobility model approximations.

When we use the gravity model approximation to estimate mobility patterns in the transmission model, the synchrony and early spatial spread are overestimated, compared to the results with the daily mobility. The epidemic spreads quickly to the whole country, and the early spread is less restricted to the locations in proximity to the seeding location. For the radiation model approximation, the synchrony and early spatial spread are underestimated. The early spread is slower, and much more restricted to the locations close to the seeding location than for the daily mobility. Some of the effects we find can

likely be explained by the difference in density of the mobility networks. The radiation mobility network is much less dense than the time-averaged mobility network, while the gravity mobility network has a higher density. Note that we compare with the daily mobility data as the reference, however they are also an approximation, as they are aggregated and potentially biased. Nevertheless, they represent our only observations of the true, underlying human mobility.

We have daily mobility matrices for six months, but our model runs over a longer time. We thus have to use another estimate for those days where we do not have daily mobility information. We have chosen to use the time-averaged mobility. We investigate whether the time-averaged mobility is a good approximation to the daily mobility by comparing simulations using these two mobility estimates. The final size, peak prevalence, peak date and the spatial spread are very similar. In particular, the synchrony is neither over- nor underestimated, as with the gravity and radiation model approximations. We thus conclude that the time-averaged mobility is a good mobility approximation.

We fit the model to the 2017 influenza season, by seeding the model based on the early cases for the season and estimating the parameters using an ABC-SMC (cf. Section 4.9). The estimated parameters are provided in Table 5.3.

We use the posterior modes to simulate the 2017 seasonal influenza epidemic in Bangladesh, using the daily mobility. We find a good qualitative agreement between our estimated model and the case data. The correlation between the case data and our simulations is 0.63, which is high, especially given the spiky and noisy nature of the case data, likely due to some days with high testing activity, and some days with low testing activity, and/or varying healthcare seeking behaviour.

As we develop a model for infectious disease spread in Bangladesh informed by high-quality mobility data, we also seek to apply the model to increase our understanding of the spatial spreading patterns of influenza in Bangladesh, and to obtain estimates of R_e for different seasons. We investigate the spatial spread when we seed in a single upazila in the different divisions. A map of Bangladesh with the divisions is given in Figure 5.2. Regardless of seeding division, we find that the disease spreads early to Chittagong and Dhaka, while it takes longer before the epidemic reaches Rajshahi, Rangpur and Sylhet, except when we seed in these divisions. The early spread is to upazilas close to the seeding upazila(s), but also to larger cities around the country, in particular to Dhaka city (the capital). The spatial spread is more coherent when seeding in Dhaka and Chittagong, than when seeding in the other divisions. When we seed in Dhaka, the delay is four days between the epidemic sparks in the latest and earliest hit division. The corresponding delay is approximately one week when seeding in Chittagong, three weeks when seeding in Khulna, Barisal and Rajshahi, four weeks when seeding in Rangpur and six weeks when seeding in Sylhet.

Since the time-averaged mobility seems to be a good approximation to the daily mobility, this motivates a more general use. We use the time-averaged mobility for the other influenza seasons than 2017. We find a good qualitative agreement between the case data and the fitted models. The reporting probability is related to the severity of the epidemic, as one would expect more

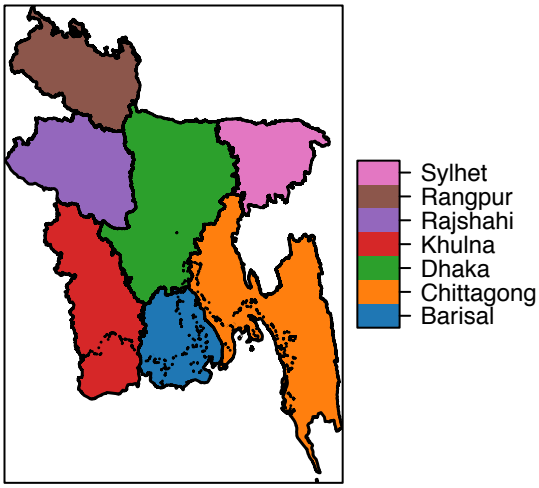

Figure 5.2: **Divisions of Bangladesh.** Map of Bangladesh with the seven divisions.

cases in the hospital sentinel influenza data for a more severe epidemic, while β is directly related to the effective reproductive number, reported in Table 5.3. We are able to estimate and distinguish the transmissibility for the different influenza seasons, as the credible intervals are narrow. The transmissibility for the 2014 influenza season is similar to the transmissibility for 2017. The 2015 and 2016 seasons have lower estimated transmissibilities, and 2015 seems to be the mildest influenza season. The 95% credible intervals for the reporting probabilities are overlapping, indicating that we do not have enough information in our data to estimate the reporting probability.

	R_e	r
2017	$1.220\ (1.201, 1.237)$	$3.52 \cdot 10^{-5}\ (5.02 \cdot 10^{-6}, 8.34 \cdot 10^{-5})$
2016	$1.196\ (1.180, 1.211)$	$2.15 \cdot 10^{-5}\ (2.75 \cdot 10^{-6}, 8.07 \cdot 10^{-5})$
2015	$1.152\ (1.141, 1.164)$	$2.49 \cdot 10^{-5}\ (3.58 \cdot 10^{-6}, 1.10 \cdot 10^{-4})$
2014	$1.213\ (1.198, 1.235)$	$2.89 \cdot 10^{-5}\ (3.65 \cdot 10^{-6}, 7.74 \cdot 10^{-5})$

Table 5.1: **Parameter estimates.** Posterior modes for the effective reproductive number R_e and the reporting probability r for different seasons. 95% posterior credibility intervals are given in parenthesis.

Chapter 6

Discussion

In this section, we will discuss the contributions of each paper, comment on the findings and relation to literature, discuss the statistical methods, discuss ethical considerations, sources of bias, other limitations and finally implications and future perspectives.

6.1 Contributions

6.1.1 Contributions of paper I

Paper I is a contribution to understanding the social nature of pain. It adds to explaining the social determinants of pain, which are of key importance to clinicians and researchers, and essential to the development of pain prevention and intervention (Craig, 2009; Mogil, 2018). We have performed the first in-depth social network analysis of pain, and hope to stimulate more research in this direction. As written in the editorial comment (Mogil, 2018) to our publication: "Studies like [...] Engebretsen et al. suggest we have just scratched the surface of an extremely interesting set of questions, the answers to which will contribute importantly to our emerging appreciation of social influences on pain". Our study is also a contribution to understanding gender-specific aspects of pain.

In addition to contributing to understanding social aspects of pain, our study has provided methodological contributions. Our study is the first, to our knowledge, to perform a social network analysis with a censored (response) variable of interest. We have proposed a framework for estimating a censored network autocorrelation model, based on a pseudo likelihood approximation.

6.1.2 Contributions of paper II

The main contribution of paper II is the proposal of a framework for generating spatial fields with controllable clustering of the population. This model is a single-parameter model, which captures key features of urbanisation. The applicability of the framework is wider than this, and it could for instance easily be extended to the problem of simultaneously investigating the population clustering and population growth, or in other settings where it is desirable to generate spatial fields with clustering of a quantity.

In addition, to our knowledge, our study is the first to systematically investigate the effect of urbanisation on infectious disease spread. We contribute to understanding how the effect of interventions can vary between different countries. It thus puts focus on a subject which has not been much studied, hopefully stimulating more research.

6.1.3 Contributions of paper III

Paper III improves the understanding of limitations of using mobility approximations instead of daily mobility when modelling infectious disease spread and the results are thus useful for guiding future studies. We find that the time-averaged mobility is a good proxy for the daily mobility, which motivates a wider application of time-averaged mobility, as it is not specific to certain dates. An additional advantage of the time-averaged mobility matrix is that it is subject to less privacy concerns than the daily mobility, as it contains less temporal information on individual movements.

In addition, our study improves the understanding of the spatial spread of influenza in Bangladesh. Influenza is a key public health challenge in Bangladesh. In addition, Bangladesh is a potential source for newly emerging avian influenza. The model can be used in preparedness planning in the event of a new pandemic and for influenza control.

Our model has been tuned to the 2017 seasonal influenza epidemic, and nicely fits the case data. We hope that our study contributes to popularising rigorous model calibration for infectious disease models, as applications of statistical methodologies to calibrate infectious disease models are limited (de Angelis et al., 2015). There has however been a lot of development in recent years (cf. Section 1.5.2).

We provide the first, to our knowledge, estimates of transmissibility of seasonal influenza locally in Bangladesh. R_e has consequences both for the magnitude of the epidemic, and for the mitigation strategies needed to control the outbreak (Biggerstaff et al., 2014b). Though there exist estimates of R_e for other countries, the values for the same seasonal outbreak are expected to vary between populations, due to underlying social, socio-demographic, climatic and geographical factors (Biggerstaff et al., 2014a). The specific value is important, because small variations might be crucial in determining whether or not a specific control measure will be successful in containing the outbreak. Moreover, estimates of R_e for tropical countries are needed to assess how the transmissibility depends on geographical and social factors (Biggerstaff et al., 2014a), as most estimates in the literature are for temperate countries.

Further, our study is a contribution to the literature on using novel data sources (mobile phone data) in infectious disease models, which is particularly useful for countries where census data are scarce.

6.2 Findings and relation to literature

6.2.1 The peer effect on pain tolerance

6.2.1.1 Peer effect on pain tolerance

In our study, we found a significant peer effect in pain tolerance. A related finding is that when people experienced pain together, their pain tolerances became similar, as cited in Prkachin et al. (2018). Hence, there was a form of experimental social influence in pain tolerance. Similar to our study, Block et al.

(2018) experimentally tested whether adolescents preferentially interact with adolescents with similar pain experience, or whether adolescents become similar to their interaction partners in terms of pain experience. The experimental setting was a three-week hiking expedition in the Arctic circle. Interestingly, unlike us, they did not find either of these effects in their study. However, their sample size was very small (17 adolescents), so the power was necessarily small, and individuals had a limited choice of interaction partners. This might explain the discrepancy between the studies. Females were found to be more likely to report pain than males in the arctic expedition, in agreement with the general finding that females have lower pain tolerance than males (Fillingim et al., 2009). In addition, sex is known to be assortative in social networks (McPherson et al., 2001), and this was also true for the social network among the adolescents on the arctic expedition. Since their assortativity measure does not control for sex, it is surprising that they did not find assortativity in pain.

6.2.1.2 Pain and social support

Social support has in general been shown to be beneficial for various aspects of health (Luke and Harris, 2007; House et al., 1982; Kawachi et al., 1997, 1999; Ziersch, 2005). Social support has also been shown to be beneficial for pain. Eisenberger et al. (2006) found a link between pain distress and social distress. Master et al. (2009) found that social support reduced pain, even when the support was in form of a picture of a romantic partner. Brown et al. (2003) found that individuals who were accompanied by another individual who provided active or passive support during a cold-pressor test experienced less pain. In addition, a larger social network has been found to be associated with higher pain tolerance (Johnson and Dunbar, 2016). In the arctic expedition experiment, they found that experiencing pain was linked to decreased popularity (Block et al., 2018). This effect was present in males only. In our analysis, we did not find a significant effect of any centrality measures (out-degree, in-degree, betweenness, closeness, eigenvector centrality) on pain tolerance, however the effect of out-degree was borderline significant (p-value 0.084). Hence, our results support, but do not confirm, the results from the literature finding that social support is beneficial for pain.

6.2.1.3 Gender aspects of peer effects in pain

We found that the peer effect in the social network was only present among males. In the arctic expedition (Block et al., 2018), they found decreasing popularity with decreased pain tolerance for males only. The more pain the males reported, the fewer individuals nominated them as friends. These results suggest a possible competition or macho effect in pain tolerance among males, and that the social component of pain is different for males and females. Friendship groups have a set of norms which members are expected to follow, and groups exert social pressure on members to uphold these norms (Valente, 2010). If being masculine or brave is part of the group's social norm, then that might cause

pressure to perform on the pain tolerance test. There are also other studies indicating that peer effects in pain are sex dependent. Edwards et al. (2017) found that experimental pain tolerance increased in the presence of another person. Male friends had the largest effect on males, in agreement with our finding. The authors speculate that the friendship effect is a sign of either group pressure or support effects, and that competition between men and men's wish for peers to view them favourably are possible explanations for why the effect for males is strongest for male friends. Keogh (2018) suggests that the society's gender-based expectations might have an impact on pain behaviour and management. Masculinity is associated with being stoic and aggressive, while femininity is associated with being more emotional and sensitive (Keogh, 2006). This might be why men are more likely to tolerate pain when accompanied by a male friend, as they want to live up to the expectations and not be seen as "weak" by their peers. Similarly, in Berke et al. (2017), men were randomly assigned to receive feedback that was either threatening or non-threatening to their masculinity. The men who received masculinity threatening feedback had a higher pain tolerance than those who did not. This might be related to our findings on test sequence, that those who were tested later had higher pain tolerance. Experience, bragging and competition among the adolescents might create the idea that masculinity is related to enduring the full pain test, and consequently might improve endurance for the boys who feel that their masculinity is threatened by the pain tolerance test. Hence, our peer effect findings on pain tolerance are in accordance with the literature, as we find that the peer effect that boys have on boys is the strongest peer effect.

6.2.2 A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions

6.2.2.1 Internal travel restrictions

In paper II, we found that internal travel restrictions reduced the final size of the epidemic, but that vaccination was more effective. This is in agreement with previous literature (Germann et al., 2006; Ferguson et al., 2006). However, these studies found no effect of internal travel restrictions on final size, while we found an effect, as in Mateus et al. (2014). We also found that the effect of travel restrictions on reducing final size was largest in the less urban locations, which is in agreement with Mateus et al. (2014), where they found no effect of travel restrictions in urban areas. Internal travel restrictions were also found to delay the epidemic, in agreement with Bolton et al. (2012); Germann et al. (2006); Ferguson et al. (2006); Mateus et al. (2014).

6.2.2.2 Vaccination

Preliminary results indicated that the rural locations were more protected from the epidemic than the urban locations. The idea behind the urban vaccination strategy is that since the rural locations are more isolated and protected from

the epidemic, they can be deprioritised when distributing vaccines. The idea behind the rural vaccination strategy is that the vaccination is more likely to be successful in eliminating the epidemic risk in the rural locations, for the same reasons.

We found that prioritising urban locations for vaccination was the most effective vaccination strategy, however there was almost no difference between the uniform and the urban vaccination strategy. Similarly, Araz et al. (2012) investigated, among other strategies, two vaccination strategies – a uniform vaccination strategy, and a sequential vaccination by population size-strategy, and found little difference between the two, in accordance with our result.

6.2.3 Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh

6.2.3.1 Mobility model approximations

We compared predictions from the disease spread model using daily mobility as a proxy for human mobility with predictions using the gravity model and the radiation model. We found that these model approximations were not accurate enough to capture the relevant spatio-temporal aspects of the disease spread in our study setting. Our results are in agreement with the findings in Finger et al. (2016), where a gravity model approximation was compared to mobile phone mobility data for modelling a cholera outbreak in Senegal. As in our setting, they found that the gravity model was not a good approximation. Similarly, Wesolowski et al. (2015c) compared a gravity model to mobile phone mobility for dengue transmission in Pakistan, and found that the mobile phone data predicted the spatial extent and epidemic timing of cases more accurately than the gravity model. Wesolowski et al. (2015b) compared gravity and radiation models to mobile phone mobility data in Kenya, and found that both models systematically failed to capture the mobile phone mobility. In agreement with our results, they found that the gravity model overestimated the spatial travel.

6.2.3.2 Time-averaged mobile phone mobility

Contrary to the gravity and radiation model approximations, the time-averaged mobility was found to be a good approximation to the daily mobility. There have been studies using individual mobile phone mobility trajectories over time to investigate characteristics of human mobility. Song et al. (2010b) found a 93% predictability in individual mobility, indicating a large degree of regularity in individual movements. Individuals have also been found to preferentially return to the same, previously visited locations (González et al., 2008; Song et al., 2010a; Pappalardo et al., 2016). González et al. (2008) found a high degree of both spatial and temporal regularity, where the individuals were characterised by time-independent characteristic travel distances. These findings are in agreement with our finding that the time-averaged mobility was good in capturing the essential mobility patterns of the daily mobility.

6.2.3.3 Estimated reproductive numbers

For the 2017 seasonal influenza in Bangladesh, we estimated an R_e of 1.22 with 95% credibility interval $(1.20, 1.24)$. Our estimate is similar to the estimates of R_e in the systematic review by Biggerstaff et al. (2014a), who found an estimate of 1.28 (IQR 1.19 to 1.37). These estimates are also similar to our estimates of R_e for other seasons – 1.20 for 2016, 1.15 for 2015 and 1.21 for 2014, with our estimates lying in the lower range. Note that most of these estimates from the literature are based on temperate countries in the Northern hemisphere.

The estimated effective reproductive number for the 2015 season was found to be significantly lower than the reproductive numbers for the other seasons considered. Previously, it has been found that seasons that are dominated by A(H1N1) and B are milder than influenza seasons dominated by A(H3) (Viboud et al., 2006b; Simonsen et al., 1997). Inspecting the subtypes of the cases for the different seasons, we found that the 2015 season had the lowest proportion of A(H3) cases, in accordance with this result.

6.2.3.4 Spatial spreading pattern

Regarding the spatial spreading pattern, our results are also in agreement with the sparse literature on spatial spread of influenza in Bangladesh. Similar to our result, Zaman et al. (2009) studied the 2007 seasonal influenza and found that the virus was likely to have started in Chittagong and Dhaka, before spreading to the rest of the country. We found that regardless of the seeding location, Chittagong and Dhaka experienced early initial dates compared to the rest of the country.

6.3 Methods

6.3.1 Choice of weight matrix (paper I)

In our network autocorrelation model, we have assumed that the individuals who can exert influence on the focal individual are the friends nominated by the individual, through our choice of the matrix $\mathbf{W}$ in the network autocorrelation model. Hence, the choice of weight matrix in the autocorrelation model reflects the underlying social influence (or homophilous selection) model. Leenders (2002) raises concern about how little focus there is in the literature on the choice of weight matrix. The parameter estimates and conclusions that can be drawn depend on the specific weight matrix used (Leenders, 2002). Many alternative constructions could have been tested. Using the outgoing links from an individual (like we do) assumes that people are more influenced by those they perceive to be their friends, than by those who nominate them as friends. Valente (2010) claims that this is a plausible assumption and thus probably the most appropriate way of modelling network influence, supporting our model choice. However, it could be appropriate to model the influence on the incoming links instead, as these individuals are the ones who could actively try to

influence the individual, and are the ones who would share information or opinions with the focal individual (Valente, 2010). In addition, we have chosen to use a row normalised weight matrix, but another option is column normalisation, assuming that the influence from individual j on individual i depends on the number of outgoing links from individual j. The idea behind column normalisation is that j distributes his/her influence among his/her friends (Leenders, 2002). Different covariates or network positions (for instance centrality) could also be taken into account when modelling influence, for instance assuming that the friends with highest centrality have the strongest influence effect (Valente, 2010). In our analysis, we have considered the effect of gender, but we could also have tested other covariates. Another option is to model the autocorrelation through the noise terms, assuming that the focal individual is influenced by (or select friends based on) their alters' tendency to deviate from their expected behaviours (Leenders, 2002). However, though there are many potential structures and covariates that could be tested, one quickly runs into the problem of testing too many models, and we have thus chosen the structure that we believe is most relevant.

6.3.2 Choice of commuting model (paper II)

In paper II, a gravity model was used to model commuting, with a power decay in distance. A power function of distance was also used in for instance Wesolowski et al. (2015c); Merler and Ajelli (2010); Ciofi degli Atti et al. (2008); Viboud et al. (2006b); Xia et al. (2004). We also fitted a gravity model with an exponential distance function, but chose to use the power distance function in our main analysis, due to a better fit.

Parts of why we chose to model commuting with the gravity model in paper II is because of its popularity in influenza and infectious disease spread models. However, we did find that using a gravity model approximation instead of the mobile phone mobility overestimated the spread and synchrony in paper III. This indicates that the results found in paper II may be less relevant in low-income settings, as the gravity model did not seem to capture the mobility in the setting of paper III well.

One of the limitations of this analysis is the fact that our gravity model was fitted to a municipality scale, which is different from our block unit scale, on which we applied the model. The population size distribution was also fitted to data on municipality scale. This scaling from municipalities to block units should be handled with care. Unfortunately, without data on finer scales, it is impossible to assess how well this population size distribution and modelled commuting pattern generalise to a finer spatial scale.

The fact that the radiation model is parameter-free is an attractive property in our application, since it does not have the same problem with parameters being fitted on one scale and used on another. However, the gravity model has been found to better fit empirical data, especially for finer scales (Masucci et al., 2013), which is what we have in our setting. Liang et al. (2013) also found that the radiation model did not accurately capture human mobility in urban areas.

Furthermore, the gravity model captured the Norwegian commuting data on the municipality scale better, with an $R^2 = 0.82$ compared to an $R^2 = 0.67$ for the radiation model. Hence, the fact that the radiation model does not have any parameters is also a drawback, since it makes the model less flexible and thus less likely to fit the data well. Barbosa et al. (2018) speculate that the fact that the radiation model does not have any parameters is the reason why its performance is not very robust to spatial scale.

There are some analytical inconsistencies in the gravity model, as pointed out in Simini et al. (2012). For instance, in the limit of the destination population going to infinity, the number of commuters also goes to infinity, and hence can exceed the origin population. Truscott and Ferguson (2012) point out that the interactions between large population centres are not accurately captured by the gravity model, and suggest an assortative model in order to better capture these features. There are also other alternative models for mobility flows, and other versions and extensions of the gravity model and the radiation model. A thorough review of mobility models is given in Barbosa et al. (2018). However, for simplicity, we have chosen to use the standard gravity model formulation.

Though the qualitative results were quite robust to the parameters of the gravity model, the shape of the distance function in the gravity model and to using a radiation model to model commuting instead of the gravity model, the effect sizes depended on the specific commuting model. This is not very surprising, since the commuting network is key in defining the spatial spreading pattern of the disease, and it depends on the underlying population clustering. For example, when commuting was modelled by a radiation model, the qualitative patterns were the same as for the gravity model, but the effect of travel restrictions was smaller. This is likely due to the construction of the radiation model, which directly depends on the population density between two locations instead of their distance, suppressing some of the clustering effects that are found with the gravity model.

6.3.3 Spatial scale (paper II)

When constructing the fictional country used in our analysis of urbanisation, we have chosen to use data on municipality scale in Norway to inform the population size distribution and gravity model of the country. The reason why we used the municipality scale is because it is the finest scale for which we had commuting data available, and we believe that it is important to use the same scale to inform the population size distribution and the commuting model. Commuting data on a finer scale is not available, due to both practical and privacy reasons. Census data are normally available on administrative scales, while an alternative would be to use mobile phone data, where one could have gridded locations on exactly the same scale as the block units, with corresponding commuting data inferred from the mobile phone data. However, we did not have such data available. We do not use the population sizes in the administrative units directly to assign population sizes to the block units, since a finer scale is necessary for the clustering algorithm to produce smooth transitions between

the clustering levels. In addition, the finer the resolution, the more reasonable is the homogeneous mixing assumption that is made in the infectious disease model. One can interpret the block units as a discretisation of administrative units, where multiple block units can be seen as making up one administrative unit. When selecting the number of block units, there is a trade-off between a high resolution (smoother clustering distributions) and computational feasibility, since the computational cost increases quadratically with the number of block units.

6.3.4 Observation process (paper III)

In our observation model for Bangladesh, we have used a binomial process to model how many of the actual cases are captured in the surveillance system, with a detection probability r, which is constant across both time and space. In reality, r might be time dependent, depending on for instance the influenza level and/or media attention. Most likely, the reporting rate is also space dependent, where different locations have different healthcare seeking behaviour due to for instance socio-economic status and availability of healthcare services. In particular, the probability of seeking healthcare in these 12 surveillance hospitals has been found to decrease with distance (Nikolay et al., 2017). Moreover, it may be that a beta-binomial process is more suitable for our setting, due to overdispersion. However, we have chosen to use the simpler binomial observation model and a stationary reporting probability, as we have very little data, and a beta-binomial distribution requires an additional parameter to be estimated. In addition, estimating fewer parameters is also computationally more efficient.

6.3.5 Time-averaged mobile phone mobility (paper III)

For the other seasons than 2017, we have chosen to use the time-averaged mobility. This is based on two assumptions: 1) that the time-averaged mobility captures the relevant daily mobility, and 2) that the essential mobility patterns are similar for different seasons. We found that the time-averaged mobility was a good approximation to the daily mobility, justifying the first assumption. In addition, our model with the time-averaged mobility was able to fit the case data well for the different influenza seasons. Hence, these results support the use of time-averaged mobility, and that the two assumptions are reasonable.

The time-averaged mobility represents a smoothed mobility network, as the connections that are in reality only open on a few specific calendar days will be open every day, but with a lower travel volume. There are many alternative ways to estimate the mobility for the days without daily mobility information. One alternative is to add noise to the time-averaged mobility network. However, this does not solve the problem with the connections being open on too many days. Another alternative is to include higher order effects like seasonal components, for instance weekends, separate matrices for each weekday, separate matrices for each month, holidays etc. In Finger et al. (2016), mobile phone data from 2013 were used to estimate the effect of a mass gathering on a cholera outbreak

in 2005, by considering the effect of a religious pilgrimage. They tested different ways of estimating the mobility matrices. They compared using time-averaged mobility over all days except from the days of the mass gathering, to different mobility matrices with various seasonal components. They found that the model using the time-averaged mobility for all days except the mass gathering was the best in terms of maximising the likelihood of the cholera case data. This further supports the use of time-averaged mobility over all days, instead of including more temporal structure.

6.3.6 Mobility assumptions (paper II)

In paper II, we have chosen a very simplified model for non-commuting travel, where individuals are assumed to return after 24 hours. In reality, the travel durations have been shown to have heterogeneous distributions (Poletto et al., 2013, 2012). This realistic ingredient has been shown to affect the epidemic spread and conditions for disease invasion (Poletto et al., 2013, 2012). We perform sensitivity analysis to this assumption by letting the travellers have heterogeneous length of stay, and find that the main conclusions are robust to this assumption.

6.3.7 Mobility assumptions (paper III)

When modelling mobility in paper III, we have chosen to send individuals to their destination locations in the morning according to the mobility estimates, and then send them back home after 24 hours. We have done this because we cannot track individuals in our mobile phone data, as they have no identifiers. In reality, it might however be that they stayed at their destination location for multiple days, or that they travelled to a third location. However, human mobility is characterised by preferential return and reproducible patterns (González et al., 2008; Song et al., 2010a; Pappalardo et al., 2016). Moreover, humans have a tendency to return home on a daily basis (Song et al., 2010a). Hence, we have chosen to send them back, in order not to overestimate the spatial mixing. By modelling this way, the population sizes in each upazila also stay constant (i.e. people are not permanently moving to other locations when they travel).

6.3.8 Spatial scale (paper III)

In our transmission model for Bangladesh, the spatial scale has to be fine enough to capture the relevant signal, but coarse enough to aggregate out the noise. In addition, the coarser the aggregation, the less privacy concerns. As noted in Wesolowski et al. (2016), on very fine spatial scales, the accuracy of the mobile phone mobility data depends on tower density and calling frequency of individuals. Hence, highly spatially resolved data are noisy and biased. Aggregation helps solving these issues, and aggregated data are less biased by these factors (Wesolowski et al., 2016). However, we want a fine enough spatial scale

to capture the relevant details for the infectious disease transmission. In addition, the smaller the spatial scale, the more plausible is the homogeneous mixing assumption. The computational costs increase with the number of locations – for each additional location, there are five additional compartments and potentially 2(N-1) additional links, where N is the number of locations. Hence, the spatial scale is a trade-off between computational intensity and spatial details, and we wish to find the scale which captures the relevant behaviour of large, complex models, while still being computationally manageable (Keeling and Rohani, 2011). For practical reasons, it is convenient to work with administrative units. Hence, for computational efficiency and since the influenza case data have a limited volume, we have decided to work on the upazila spatial level, as the division and district levels are too coarse, and finer spatial scales are too computationally expensive.

6.3.9 Latent and infectious period (paper III)

As we have many patches in our metapopulation model, we have for computational reasons chosen to work with an exponential distribution for both the infectious and latent periods, even though gamma distributions would probably be more realistic, as the infectious and latent period do not have the memoryless property (that is, it is reasonable to assume that the probability of recovering given that you are still infectious increases with how long you have been infectious). The computational cost is thus a limitation of our study – fitting the model to one influenza season took approximately one week running in parallel on 112 cores. Also for paper II, we chose to use exponential distributions for the infectious and latent periods, as this choice is not likely to affect how the infectious disease spread depends on the underlying population clustering.

When fitting our model parameters to the case data in paper III, we fix the mean latent and infectious periods. Even though there is good knowledge of the latent and infectious periods in the literature, this is a quite strong assumption. However, the scarcity of the data and the strong correlations between the different model parameters prohibit us from estimating the latent and infectious periods. We have little information on the latent period in the data, and the infectious period is strongly correlated with both the transmission parameter and the reporting probability. Also Corbella et al. (2018) fixed their average infectious and latent periods, and claim that detailed individual-level data are needed to estimate these quantities.

6.3.10 Model detail level (paper II and III)

As we are interested in spatial spreading patterns, we obviously include spatial heterogeneity in our model. Occam's razor or the law of parsimony also applies to the case of infectious disease modelling – a model should be suited to its purpose, and be as simple as possible, but not simpler (Keeling and Rohani, 2011). Hence, in our setting, we did not include age structure, even though age is an important determinant for influenza transmission, neither in paper II nor

in paper III. In particular, for paper II, as we are mainly interested in trends when the urbanisation level is varied, age effects are not likely to affect our conclusions. For paper III, we do not have any covariates in our mobility data, and thus we could not have included age without making ad hoc assumptions on the age dependency of travel patterns. Including age would also make the models computationally more intensive. In general, this is true for all the details and structure that would have been included in an agent-based model, which we lose from using a metapopulation model.

6.4 Ethical considerations

6.4.1 Pain tolerance

In this thesis, we have worked with person sensitive data, which could be directly or indirectly identified. Pain tolerance data are sensitive. In the Fit Futures study, the participants provided informed consent, the data collection was approved by the Norwegian Data Protection Authority and the Regional Committee for Medical and Health Research Ethics of Norway, Northern health region, and the procedures were conducted in accordance with the Declaration of Helsinki. In addition, the data have been stored on a secure server.

6.4.2 Mobile phone data

To handle the sensitivity and non-anonymity of the mobile phone data, they are aggregated. Human mobility is highly unique and thus identifiable (de Montjoye et al., 2013). Specifically, de Montjoye et al. (2013) found that if the location of an individual user is specified hourly, four observation points were enough to uniquely identify 95% of the individuals in their study. Even coarse data sets are not fully anonymous (de Montjoye et al., 2013), and aggregation does not guarantee deidentifiability (Eagle, 2009; de Montjoye et al., 2014; Oliver et al., 2015). Another ethical issue with mobile phone data is that it is difficult for individuals to remove their individual data from an aggregated mobile phone data set (Eagle, 2009). The risks have to be balanced versus the public good that can be gained from using the mobile phone data (Eagle, 2009; de Montjoye et al., 2014). This is also the case for other types of more traditional (sensitive) data – there is always some risk of deidentification or malicious use of data. Other sensitive data such as census, household survey, tax and financial data are being used for research purposes for public good (de Montjoye et al., 2018). The difference is that there are long traditions in how to handle these types of data for research purposes, for instance through clear guidelines at data protection agencies and national and regional ethical committees. Hence, there is a need for standardised protocols for usage of mobile phone data, to help guide the trade-off between public good and privacy protection, and to help protect the individual privacy rights (de Montjoye et al., 2014; Eagle, 2009).

6.4.3 Spatially targeted vaccination strategies

In paper II, we considered spatially targeted vaccination strategies. This is of course not ethically unproblematic to use in practice. However, as we considered this in a simulation experiment, our study is not subject to ethical concerns. A spatially targeted vaccination strategy is not fair and equitable, and the CDC ethical guidelines state that allocation of limited resources should be guided by equity (Kinlaw et al., 2009). Policy makers have to balance the trade-off between effectiveness of the vaccination strategies and equity, as the most optimal vaccination strategies are often inconsistent with equity (Keeling and Shattock, 2012; Duijzer et al., 2018a,b; Matrajt et al., 2013). Wu et al. (2007) focussed on the trade-off between equity, simplicity and robustness of spatially targeted allocation policies. In our particular setting, the balance should not be difficult, as we found very little difference between the uniform and fair vaccination strategy, and the (optimal) urban vaccination strategy. However, had the differences been significantly large, the gain in vaccination effectiveness might have outbalanced the wish for equity.

6.5 Sources of bias

6.5.1 Social network definition (paper I)

The definition of the friendship network is a source of bias in the pain tolerance study, as the network is truncated. Moreover, due to the study design, the friendship links are not necessarily the strongest friendships. Friends were defined as the individuals whom they had spent the most time with in the preceding week, thus not necessarily coinciding with the closest friends. If the nominated friends do not correspond to the actual friends, this is a source of bias in the estimated friendship effect.

6.5.2 Selection bias and missing data (paper I)

Even though the response rate was higher than 90% in the Fit Futures study, there is likely some volunteer bias. The individuals who did not participate in the study or in the pain tolerance test may well be a selective group which cannot be regarded as a random sample. In addition, as always with a cohort study, one cannot know whether the results are generalisable to a general population, or whether they are specific to adolescents in that cohort. In our study, the analyses on lifestyle information were conducted restricted to the subset of individuals for whom we had lifestyle information. In doing this, it is assumed that the lifestyle information is missing at random, which might not be the case. There were however only 15 missing cases, and considering their pain tolerance, seven were censored and eight were non-censored, similar to the overall censoring level of the adolescents. In addition, there is bias due to measurement error, especially in the pain tolerance measurements.

6.5.3 Pseudo likelihood approximation (paper I)

Another potential source of bias in our analysis is the pseudo likelihood approximation. In the pseudo likelihood, the observations are assumed to be conditionally independent, which in general does not hold for a network. The reason why we use the pseudo likelihood approximation is due to the censoring. An alternative to using the pseudo likelihood approximation, is to use the dichotomised pain tolerance outcomes, defining the censored individuals as pain tolerant, and the rest as pain sensitive. Dichotomising continuous variables is subject to information loss and bias (Altman and Royston, 2006), and we have thus chosen to instead use the pseudo likelihood approximation, which in general is a good approximation (Kolaczyk and Csárdi, 2014). Lubbers and Snijders (2007) found that the pseudo likelihood approximation did not affect the general conclusions. They did however find that the standard errors estimated by the pseudo likelihood were not reliable. In our approach, we do not use the standard errors directly obtained from the pseudo likelihood, but we use a jackknife approach to estimate the standard errors. In addition, we have a large sample size, so the approximation likely does not matter much, due to consistency properties of the maximum pseudo likelihood estimators (Besag, 1975). In Corander et al. (1998), maximum pseudo likelihood estimators were compared to maximum likelihood estimators, and they found that the performance of the pseudo likelihood approximation depended on the size of the network. They found that the maximum likelihood estimators performed better in small networks of less than 40 nodes, while for larger networks of 40-100 nodes, the bias and mean squared errors of the estimators were nearly equivalent. Our network is much larger, supporting that the pseudo likelihood estimates should be a good approximation in our setting.

6.5.4 Influenza surveillance bias (paper III)

Our case data are likely biased by the fact that we only have surveillance for 12 hospitals, and not all hospitals of Bangladesh. The probability of detecting a case has been found to decrease with distance to the hospital, and rural locations are thus less covered (Nikolay et al., 2017). However, the sentinel hospitals are spread out spatially across the different divisions. In addition, the asymptomatic infections are not likely to be detected by surveillance data. An additional problem identified in Bangladesh was that many sought healthcare in the informal sector, that is, unqualified practitioners such as healers or village doctors (Nikolay et al., 2017).

An advantage of having laboratory-confirmed hospital case data, is that they are less noisy and biased than for instance GP data, as they are less subject to varying healthcare seeking behaviour and noise due to other respiratory infections.

There is in general bias in influenza surveillance data, as not everyone with influenza seeks healthcare, and healthcare seeking behaviour can vary across different factors (Biggerstaff et al., 2014b; Lee et al., 2017; Nikolay et al., 2017).

In Bangladesh, it has been found that the cases captured by the hospital surveillance system differ in age and socioeconomic status from the cases in the general population (Nikolay et al., 2017). Specifically, for fatal respiratory disease, elderlies sought less frequent healthcare, while children under five years and those with high socioeconomic status sought healthcare more frequently. As we are interested in the spatio-temporal pattern, this is particularly a problem if these groups differ in when they are typically infected during an epidemic (i.e. if they typically get infected early or late in the epidemic), since that would bias the temporal pattern of the surveillance data. This can for instance occur due to the different age patterns in the surveillance data from those in the general population, as it has been hypothesised that children drive local transmission, while adults spread the disease to new locations, which can cause shifted epidemic timings across age groups (Lee et al., 2017). This was found for instance during the 2009-2010 influenza pandemic (Apolloni et al., 2013). This effect can also interact with the specific influenza strain, as influenza A(H3) is associated with more frequent cases among the elderly than A(H1N1) seasons, and influenza B is associated with more frequent and earlier cases among children (Lee et al., 2017). Hence, sentinel hospital influenza data are not an unbiased estimate for the general spatio-temporal seasonal influenza pattern of the population, though the bias effects due to differing age patterns are likely to be small, compared to the bias due to the data sparsity.

6.5.5 Bias in mobile phone mobility data (paper III)

Another important potential source of bias comes from the mobile phone data, as the population represented by the mobile phone data might not be representative of the general Bangladeshi population. Firstly, there might be market share bias in the Grameenphone subscriptions. Secondly, there is a bias in mobile phone ownership, and individuals might own multiple SIM cards (Wesolowski et al., 2016). Thirdly, mobile phone data are biased by calling behaviour (Wesolowski et al., 2016), as they depend on when and how often individuals make phone calls – their most frequent location when making a phone call on a day might not coincide with the location in which they spent the most time. Young children are not captured by the mobile phone data (Wesolowski et al., 2016). In addition, women in Bangladesh are less likely to own phones than men (GSMA Intelligence, 2017). In Wesolowski et al. (2012a), the bias in mobile phone usage in Kenya is investigated, and they find that rural, poor women are the least represented group. Wesolowski et al. (2013) compare mobility estimates from mobile phone data to mobility estimates based on census data for Kenya, and find that the mobility estimates from the mobile phone data are "surprisingly robust" to the bias in ownership. Hence, to the extent that results from Kenya can be translated to Bangladesh, the bias in mobile phone ownership does not greatly affect the mobility estimates. A good match between mobility estimates based on mobile phone data and census data is also found in Tizzoni et al. (2014), for Spain, Portugal and France.

Mobile phone mobility data are however likely subject to overestimation

bias, since those who travel most frequently are also those who are most likely to own a phone, as they have fewer economic constraints (Wesolowski et al., 2013). This can overestimate the spatial synchrony of the epidemic. However, overestimation of travel volumes is not likely to affect the relative arrival times in various locations (assuming unbiased mobility estimates, and thus constant overestimation across space and time). Empirical evidence for this claim is found for instance in Tizzoni et al. (2014), who found that mobile phone data overestimated the mobility, but that using the mobile phone mobility data did not greatly affect the relative arrival times in locations. Travel surveys have been found to underestimate both mobility volume and range (Wesolowski et al., 2014). In the case of infectious disease spread, underestimation bias in travel estimates is a more severe problem than overestimation, as underestimation will underestimate the spatial synchrony, which results in too optimistic estimates of the time one has to implement targeted vaccination strategies and to prepare the healthcare system for epidemic introduction in a region. A combination of the two mobility sources would probably result in a better estimate than any of the sources provide alone. In particular, they might complement each other, as travel surveys might be able to estimate mobility for the individuals who do not own mobile phones (Wesolowski et al., 2014).

When we use the mobile phone data to estimate the population sizes and scale up the mobility matrices, we implicitly assume that the proportion subscribing to mobile phone services is uniform in space. In reality, the proportion might depend on different factors, for instance how rural or urban the location is. However, without information on mobile phone subscriptions, we are unable to check the validity of the assumption, or correct for the spatial dependency.

6.6 Limitations

6.6.1 Truncation of social network (paper I)

The fact that we did not find a significant popularity effect in pain tolerance might be due to the study design, both due to the truncation of the social network, and the friendship definition. In the Fit Futures study, the social network consisted of up to five friends, all attending first year of upper secondary school in the Tromsø and Balsfjord municipalities. Hence, the network was truncated in two ways – the maximum number of friends was five, and friends outside first year of upper secondary school in the area were excluded.

6.6.2 Selection or influence (paper I)

Pain tolerance is an unconscious behaviour, so it is difficult to believe that there is a direct friendship selection or influence effect in pain tolerance. However, we tried to control for possible confounding variables like lifestyle factors, socioeconomic status and test sequence, and they could not explain (all of) the peer effect. We argued that a macho effect or a peer pressure masculinity effect are possible explanations for the peer effect that we found. These effects could

also be selection effects, i.e. "macho" boys preferring to be friends with other "macho" boys and vice versa. As we only have a snapshot of the network and the pain tolerances, and not longitudinal data, we cannot separate cause and effect – homophilous selection and social influence, since both processes (and combinations) can lead to the same snapshot and are in general confounded (Shalizi and Thomas, 2011; Steglich et al., 2010). The friendship network and pain tolerance processes evolve simultaneously and are not separate processes, and it is therefore impossible to separate the homophily and social contagion effects without longitudinal data and models (Steglich et al., 2010).

6.6.3 Sensitivity to using Norwegian data (paper II)

In paper II, we aimed at studying the urbanisation phenomenon from a generic viewpoint, and the data are not used to make the country resemble Norway, but rather to include plausible ingredients. We used the Norwegian data, because we had commuting data available on a relatively fine scale, and we believe that it is important to simultaneously inform the commuting pattern and population size distribution based on the same underlying population. However, different ways of defining administrative regions could yield different population size distributions and estimated gravity models, which again could affect our results. In order to address the problem that different ways of dividing the country into administrative regions could affect the results, we also performed the analysis based on the United Kingdom (homogeneous population distribution compared to Norway) and performed sensitivity analysis to the parameters of the gravity model. The latter also handles potential misspecification of the commuting patterns when scaling from municipalities to block units (Section 6.3.3).

6.6.4 Interpretability of κ (paper II)

The clustering algorithm in paper II used a tuning parameter κ to control the clustering. In the clustering algorithm, a non-linear transformation to the random spatial field is used, leaving us unable to invert the transformation and estimate κ. Hence, we are not able to estimate κ for a given country, and the comparisons are thus done purely by visual inspection.

6.6.5 Boundary effects (paper II and III)

In both the urbanisation study in paper II, and in our case study of influenza in Bangladesh, the countries are regarded as isolated systems. The fact that we consider the country in isolation and ignore bordering countries can result in edge effects, since the locations on the border might in reality have a connection to the neighbouring countries.

6.6.6 Limitations of mobile phone mobility data and model validation (paper III)

A limitation related to the mobile phone data is the fact that we cannot track individual trajectories, which is likely to overestimate the spread (Keeling et al., 2010). We let the travelling individuals be random draws from the origin population. However, in reality, it is likely that there is some regularity in who travel on the same connection on different days, as individuals have been found to preferentially visit a few highly frequented locations (González et al., 2008; Song et al., 2010a; Pappalardo et al., 2016). Another limitation is the temporal resolution of the mobility data. As we only have a temporal resolution of 24 hours, we do not capture regular everyday commuting and short duration travel well. We do not know whether the mobility estimated by the mobile phone data is the most relevant mobility for infectious disease transmission (Wesolowski et al., 2016). As the spatial information in the case data is limited, this is impossible to properly validate, though we do find good qualitative agreement with the observed data. The data are not very detailed, and hence we cannot properly assess spatial diffusion performance. However, both in the case data and in our simulations, Dhaka, Chittagong and Barisal were hit before the other divisions in the 2017 seasonal influenza epidemic. The fact that we cannot assess how well our model is able to capture the spatial spread is a limitation, as it is crucial to have a well-tested model when using it for policy decisions. Moreover, we only have mobile phone mobility data for six months, and we are thus not able to capture seasonality or variations throughout the full year.

6.6.7 Climatic drivers (paper III)

We have focussed on human mobility as the main driver of influenza spread. However, there may be other important drivers, like for instance climatic factors. In Brazil, it has even been suggested that environmental factors play a more important role than population factors in driving the influenza timings (Alonso et al., 2007). The driving factors of influenza seasonality are unclear, though it has been suggested that temperature, humidity, solar radiation, seasonal mixing patterns, human mobility and changes in immunity are important mechanisms (Tan et al., 2014; Yu et al., 2013; Tamerius et al., 2013; Weinberger et al., 2012; Tamerius et al., 2010; Lipsitch and Viboud, 2009; Alonso et al., 2007; Lofgren et al., 2007; Altizer et al., 2006; Viboud et al., 2006a,b; Dowell and Ho, 2004; Dushoff et al., 2004; Hope-Simpson, 1981). Also in Bangladesh, influenza activity has been found to be associated with climatic factors like rainfall, temperature, humidity and sunlight (Zaman et al., 2009; Islam et al., 2013). We have not included environmental factors in our spatial transmission model, as such data were not available. It might be that some of the climate effects are indirectly captured by changes in mobility. In addition, it is not obvious how one would include these factors in the model, as their effects are to a large degree debated and unknown. This is an interesting topic for future studies, however outside the scope of this thesis.

6.6.8 Limited influenza case data (paper III)

When we in paper III assess the performance of the mobility approximations, we compare with the deviations from simulations using the daily mobility, assuming that it represents the ground truth. Unfortunately, we cannot assess how well the mobility approximations are able to capture the real, underlying infectious disease spread, due to the low volume and limited spatial extent of the case data. With richer case data, we could have more formally assessed which model performs best in replicating the case data. Nevertheless, even though the mobile phone mobility data are not perfect, there is little reason to suspect that time-averaged mobility would perform worse with higher quality mobility data, by deviating more from the daily mobility.

We did not attempt at predicting the temporal epidemic trajectory, due to the limited case data. This would require estimating the parameters based on only a few weeks of data, and with little data, the credible intervals would be very wide. Moreover, we would have to use other summary statistics, as the peak height, duration and final size would not be known in advance. One could for instance use the early growth rate of the epidemic as a summary statistic in that setting. In addition, the hospital case data are not likely to be available early enough for the predictions to be useful, especially since the predictions would likely only be feasible for short time horizons, as the uncertainty gets large (Woolhouse et al., 2015).

6.7 Implications

6.7.1 Pain treatment (paper I)

This work is a contribution to understanding the social factors of pain, which is essential to prevention and intervention strategies for pain (Craig, 2018). Maniadakis and Gray (2000) recommend that research into interventions to prevent and treat back pain should be a priority. Craig (2018) suggests that social interventions have to be included in pain treatment. As pointed out in the introduction to the thesis, chronic pain is a major economic burden and has substantial quality of life costs (Phillips, 2009; Maniadakis and Gray, 2000; Fillingim, 2017; Nielsen, 2013; Swedish Council on Technology Assessment in Health Care (SBU), 2006; Ashburn and Staats, 1999). Prevention strategies aimed at increasing pain tolerance can reduce the number of people affected by chronic pain.

6.7.2 Inspiration for other studies of urbanisation (paper II)

In paper II, we have proposed a framework which can be used to generate spatial fields with controllable levels of spatial clustering. The framework is general, and we hope it can inspire others in similar applications and problems. We also hope to inspire other, more data-driven and applied studies on urbanisation and the effect on infectious disease transmission for specific populations.

6.7.3 Implications for policy planners (paper II)

Our results have implications for the effects of internal travel restrictions and vaccination strategies. According to our general model, population clustering is an important determinant for the effect of travel restrictions. For the higher clustering levels, travel restrictions were most effective in reducing the final size of the epidemic, and less effective in delaying the epidemic. For the lower clustering levels, travel restrictions were more effective in delaying the epidemic, than reducing final size.

Our results are interesting from a preparedness planning perspective. Many countries already have high population clustering, as found for instance in Australia, where the majority of the population is clustered around the coastal belt (Kapferer, 1990), and in the Turku region of Finland (Vasanen, 2009). In paper II, we assess the population clustering of various European countries by visual inspection, finding that many of them have high population clustering. It is likely that in the future, the population clustering will become even higher. Hence, according to our general model, it is likely that in the future, travel restrictions will be less and less effective in delaying the epidemics, and more and more effective in reducing final size. Whether it is most beneficial to decrease the final size of the epidemic or delay the peak depends on the morbidity and mortality. Delaying the epidemic means more time to plan and implement other interventions and preventive measures. With high morbidity and severe influenza strains, there is more stress on the healthcare system, as the patients require more and/or longer follow-up and care. In these settings, it might be more interesting to delay the epidemic, as to prepare the healthcare system for the peak periods. In addition, a delay in the epidemic can be important to buy time to develop and/or distribute vaccines, as they are more effective when they are distributed early in the epidemic (see for instance Keeling and White (2011); Wood et al. (2009); Tuite et al. (2010); de Blasio et al. (2012)).

We find that vaccination is much more effective than travel restrictions in reducing final size. Hence, travel restrictions are mainly relevant when there are no vaccines available. We found that the rural vaccination strategy was clearly inferior to the urban and uniform vaccination strategies. Hence, it is important to cover the urban locations, and vaccine sentiment campaigns might benefit from targeting urban locations.

6.7.4 Implications for the use of mobile phone mobility data in research (paper III)

Mobile phone mobility data are particularly useful in developing settings where census data are scarce, like Bangladesh. We have successfully modelled infectious disease spread in Bangladesh, and our study is thus an example of how mobile phone data can be used for public good to increase understanding of disease outbreaks. Mobile phone data are not perfect, for instance due to bias in mobile phone ownership. However, there is high adoption of mobile phone ownership in both developed and developing countries (Oliver et al., 2015). In 2011,

the coverage in developing countries was 79%, and in 2014, it had increased to 90% (Oliver et al., 2015). Hence, mobile phone adoption is likely to grow even more in the future, and become even more relevant and representative for the population, for research purposes like ours. Mobile phone data are also likely to be more easily accessible compared to other mobility data sources, as they are automatically generated for other purposes.

Successful use of mobile phone data in public health might make it easier to access mobile phone data both for research purposes and in the case of emerging epidemics. It both puts focus on use of mobile phone data, and shows the possibilities of mobile phone data for public good. Previously, concerns about privacy have stood in the way of use of mobile phone data for real-time pandemic predictions. For instance, privacy concerns hindered the use of mobile phone data during the 2009 pandemic outbreak in Mexico (Eagle, 2009). We find that the time-averaged mobility was a good approximation to the daily mobility. This has important consequences for the privacy concerns, as time-averaged mobility is subject to less privacy concerns than daily mobility, since it contains even fewer temporal details on individual mobility. Hopefully, studies like ours will also encourage efforts to establish standardised protocols for use of mobile phone data in research, as has been requested by the research community (Oliver et al., 2015; de Montjoye et al., 2014; Eagle, 2009; de Montjoye et al., 2018). Standardised protocols are also especially important during emerging outbreaks, in order to speed up the process to provide timely guidelines for decision makers.

We found that the time-averaged mobility performed remarkably well in replicating the simulations using the daily mobility. This is an optimistic result, as it is unlikely that daily mobility could be produced in real-time during an outbreak setting. Hence, our results motivate a more general use of time-averaged mobility, with important implications for future modelling and outbreak control.

6.7.5 Implications for preparedness planning in Bangladesh (paper III)

Our results have implications for policy planners, providing relative arrival times for various seeding scenarios. For example, during the recent Ebola epidemic in West Africa, the CDC used estimates of the probability of introduction into various regions, to help allocate resources for surveillance (Meltzer, 2016). Our model could be used for similar purposes, in a pandemic influenza setting in Bangladesh. The use of mathematical models during the Ebola outbreak was a critical tool for guiding decisions, and helped in combating the epidemic (Meltzer, 2016; Viboud et al., 2018).

Our results have implications for possibilities for targeted interventions in Bangladesh. We find a more coherent spatial spread when we seed in Dhaka or Chittagong division, than for the other spatial seeding scenarios considered. When seeding in Dhaka or Chittagong, the epidemic spreads quickly through the country, and it takes approximately a week between the epidemic sparks in the earliest and the latest division. When seeding in the other divisions the spread is less coherent, implying more time to implement targeted interventions.

6.8 Future perspectives

6.8.1 Longitudinal pain tolerance data

We have found a significant peer effect in pain tolerance, which raises many questions like whether the effect generalises to other social networks, why we find the effect, how the network becomes assortative and what the causal mechanisms behind the effect are. With longitudinal data, this could be explored and tested by using stochastic actor-based models (Snijders et al., 2010). These models can be described as agent-based models, where the dynamic network changes and behaviour of the agents are modelled simultaneously. That is, network links are created and broken, and the covariates of the nodes/actors change through for instance peer influence. The mechanisms behind the network changes can for instance be homophily, reciprocity or transitivity (tendency to form triangles). These models can be used for hypothesis testing, for instance testing whether the similarity in the network is caused mainly by homophily or peer influence, while simultaneously controlling for other confounding variables. Future studies should thus work with longitudinal network data, to address these questions. The problem will be to extend these models to handle a censored variable like pain tolerance.

6.8.2 Exponential random graph models

An alternative way to model the peer effect on pain tolerance, is to use exponential random graph models (Snijders, 2002; Handcock, 2002). In the network autocorrelation model, we model the outcome (pain tolerance) as a function of the network. In exponential random graph models, it is the network that is modelled, as a function of, among other mechanisms, the node covariates (i.e. sex, pain tolerance, age etc.). It would be interesting to study whether the network ties can be explained by similarity in pain tolerance, as this would further support our finding. These models would have to be extended to handle censored node covariates. Preliminary results using the dichotomised pain tolerance in an exponential random graph model indicate that it does indeed seem like similarity in pain tolerance is a significant predictor of friendship ties.

6.8.3 Commuting network versus κ

An interesting alternative approach to the research question in paper II would be to focus more on how the commuting network properties change with κ. It is likely that some of the effects we find can be explained by the changes in the commuting network. By investigating whether and how different network statistics change with κ, we can hope to gain a better understanding of the effect of population clustering.

6.8.4 Richer case data

Traditionally, the lack of data has been a hindrance to the use of mathematical models of infectious diseases to inform public health authorities in developing countries. In paper III, we have used a modern, rich and alternative data source to estimate human mobility, allowing a sophisticated spatial model for infectious disease spread in Bangladesh. However, one of largest limitations of the study is the scarce influenza case data. In recent years, novel data streams have been increasingly used in disease surveillance and prediction (Simonsen et al., 2016; Bansal et al., 2016; Althouse et al., 2015). An interesting topic for further work is thus to combine our framework with novel "big data" sources for disease surveillance, with the scarce, but less biased, traditional surveillance sources, as these novel data streams are particularly useful for countries with little surveillance (Simonsen et al., 2016), both to increase the data volume, and the timeliness of the case data needed for timely predictions. Including multiple data sources is also straightforward in an ABC scheme.

Richer case data allow assessment of the spatial spreading prediction of the daily mobility model, and a more formal model comparison with the mobility approximations. This requires that all the spatial locations are captured in the influenza data. More data also open possibilities to forecast the epidemic trajectory, which is a challenging task (Corbella et al., 2018; Moran et al., 2016; Birrell et al., 2011; Ong et al., 2010). Richer case data also allow even more complex infectious disease models, like testing whether a beta-binomial observation model is more appropriate than the simpler binomial observation model. It would be interesting to know why we are not able to properly estimate the reporting probability. Do the estimates get better with more data, or by using a beta-binomial observation model? Or is it related to the data source itself, that the reporting probability is not identifiable from such data. All of these points would be interesting to address in future work with more case data, either from novel data sources, or in a different setting, for instance using the Norwegian ILI data.

6.8.5 Summary statistics and efficient computation

We have chosen the summary statistics in the ABC-SMC by performing simulations from the model, and exploring which quantities changed when we changed the reporting probability or the transmission parameter. However, it would have been interesting to explore more formally how the posterior distributions depend on the summary statistics, by exploring more summary statistics and combinations. In particular, it would be interesting to try the initial growth rate, as this summary statistic could be valuable for forecasting. It could also be that other summary statistics would better capture the reporting probability. Hence, it would be advantageous to find sufficient summary statistics for the parameters. This is an interesting topic for future work, but would probably require a significant speed-up of the fitting algorithm, for instance through using emulation or uncertainty quantification. Another alternative would be

to use the "Bayesian optimization in likelihood-free inference" (BOLFI), which has been found to accelerate the ABC inference substantially (Gutmann and Corander, 2016).

6.8.6 Generalisability of time-averaged mobility to future seasons

Another important topic for future studies is to assess whether the time-averaged mobile phone mobility is able to capture the essential mobility patterns over time, allowing the same matrix to be used also for future outbreaks. This would require mobile phone mobility data for multiple years. This could also be assessed by detailed, high-quality surveillance data, by assessing the fit of the model with time-averaged mobility for one season to case data for various seasons.

6.8.7 Generalisability of gravity and radiation model results

We have found that the gravity and radiation models were not good at capturing the relevant mobility from the daily mobility for modelling infectious disease spread in Bangladesh. A similar result was found for the gravity model in Finger et al. (2016); Wesolowski et al. (2015b,c), and for the radiation model in Wesolowski et al. (2015b). It would be interesting to further investigate whether this result is generalisable to other developing settings. In high-income settings, the gravity model has been found to capture infectious disease spread well, see for instance Truscott and Ferguson (2012); Xia et al. (2004); Viboud et al. (2006b). Future studies should compare with the same type of data for a high-income setting and a low-income setting, for instance mobile phone mobility data, to investigate whether or not our results are only applicable to developing settings. If the gravity model works well in high-income settings and not in low-income settings, one could further investigate why – what is the difference between the mobility networks?

6.8.8 Higher temporal resolution of mobile phone data

The mobile phone mobility data had a temporal resolution of 24 hours. It would be interesting to compare mobility data with 24 hours resolution to mobility data with a finer temporal resolution, for instance hourly data. That way, it is possible to assess the relative importance of short-distance travel and commuting, which would typically be captured with hourly data, to long-distance travel, which is better captured with daily data. Hence, for future work with more detailed mobile phone mobility data, one could investigate the optimal temporal and spatial scale (or combinations of aggregation scales) for modelling infectious disease dynamics at different levels of resolution. One could also investigate whether the estimation of the epidemiological parameters is consistent for different spatial and/or temporal resolutions.

6.8.9 Effect size of privacy implications

We find that the time-averaged mobility performs comparable to the daily mobility in the infectious disease spread simulations. The time-averaged mobility contains less temporal information on the individual movements than the daily mobility, and is thus subject to less privacy concerns. Imagine for instance that between two small locations, there was only one person travelling, and only on one occasion. Then one would know the date of the movement from the daily mobility, while for the time-averaged mobility, we only know that the event occurred on one of the observation days (Note that this does not occur in our data set, as we are working on the upazila level, and that such rare events would likely be censored from mobile phone mobility data). However, we have not formally tested and compared the identifiability of the two movement sources. That would require individual level data on mobility trajectories, which we have not had access to. A formal comparison of the uniqueness of individuals in daily mobility and time-averaged mobility is a topic for future work.

6.8.10 More sophisticated mobility model

Human mobility is essential to the work in this thesis. We could have used the daily mobile phone mobility data to propose a new, more sophisticated mobility model, or to assess which covariates are important for modelling a mobility network. This can be done by modelling the mobility network using statistical regression methods like the one developed in (Westveld et al., 2011), implemented in the R-package amen (Hoff, 2015). In this way, we could investigate which model best fits our mobile phone mobility data, incorporating both standard terms that are included in the gravity and radiation models like population sizes, distance and population density between the locations, but also investigate new terms. For instance, we could assess the impact of temporal covariates like month, weekday, Ramadan, holiday or weekend, further contributing to the discussion on how to best construct the time-averaged mobility. We could also include more network related terms, like sender and receiver effects (some locations being very active or very attractive), reciprocity, assortativity etc. This is particularly important if mobility in low-income countries is in general not well described by the traditional human mobility models.

Chapter 7

Conclusions

This thesis adds to the growing field of network analysis in public health and epidemic modelling, by developing novel methodology driven by important public health applications. The novel insight gained in this thesis raises new questions that are interesting topics for future studies, and we hope that these questions will stimulate more research. Two of the papers model infectious disease spread on networks, and the other paper is a social network analysis of pain tolerance. Though applied to different phenomena, all three papers are related to contagion processes on networks. The social network analysis of pain focusses on peer influence and social support effects in pain tolerance, rather than the transmission of a pathogen. We have exploited networks to study peer influence, social support, assortativity, infectious disease transmission and interventions. By use of complex large-scale simulations and high-quality data sources, we have demonstrated how network science can contribute to understanding and predicting various phenomena in public health.

Bibliography

Abat, C., Chaudet, H., Rolain, J.-M., Colson, P., and Raoult, D. (2016). Traditional and syndromic surveillance of infectious diseases and pathogens. *International Journal of Infectious Diseases*, 48:22–28.

Ahmed, M., Aleem, M. A., Roguski, K., Abedin, J., Islam, A., Alam, K. F., Gurley, E. S., Rahman, M., Azziz-Baumgartner, E., Homaira, N., Sturm-Ramirez, K., and Iuiaono, A. D. (2018). Estimates of seasonal influenza-associated mortality in Bangladesh, 2010-2012. *Influenza and other respiratory viruses*, 12(1):65–71.

Ahmed, M., Mamun, A., Abedin, J., Homaira, N., Paul, R. C., Iuliano, D., Widdowson, M.-A., Gurley, E. S., Azziz-Baumgartner, E., Luby, S. P., Rahman, M., and Sturm-Ramirez, K. (2013). Incidence of influenza-associated mortality in Bangladesh, 2010–12. *Options for the Control of Influenza*, 8:5–10.

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., Merler, S., and Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC infectious diseases*, 10(1):190.

Ajelli, M., Merler, S., Fumanelli, L., y Piontti, A. P., Dean, N. E., Longini, I. M., Halloran, M. E., and Vespignani, A. (2016). Spatiotemporal dynamics of the Ebola epidemic in Guinea and implications for vaccination and disease elimination: a computational modeling analysis. *BMC medicine*, 14(1):130.

Alirol, E., Getaz, L., Stoll, B., Chappuis, F., and Loutan, L. (2011). Urbanisation and infectious diseases in a globalised world. *The Lancet infectious diseases*, 11(2):131–141.

Alonso, W. J., Viboud, C., Simonsen, L., Hirano, E. W., Daufenbach, L. Z., and Miller, M. A. (2007). Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics. *American journal of epidemiology*, 165(12):1434–1442.

Althouse, B. M., Scarpino, S. V., Meyers, L. A., Ayers, J. W., Bargsten, M., Baumbach, J., Brownstein, J. S., Castro, L., Clapham, H., Cummings, Derek AT Del Valle, S., Eubank, S., Fairchild, G., Finelli, L., Generous, N., George, D., Harper, D. R., Hérbert-Dufresne, L., Johansson, M. A., Konty, K., Lipsitch, M., Milinovich, G., Miller, J. D., Nsoesie, E. O., Olson, D. R., Paul, M., Polgreen, P. M., Priedhorsky, R., Read, J. M., Rodríguez-Barraquer, Smith, D. J., Stefansen, C., Swerdlow, D. L., Thompson, D., Vespignani, A.,

and Wesolowski, A. (2015). Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Science*, 4(1):17.

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., and Rohani, P. (2006). Seasonality and the dynamics of infectious diseases. *Ecology letters*, 9(4):467–484.

Altman, D. G. and Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ*, 332(7549):1080.

Angst, M. S., Phillips, N. G., Drover, D. R., Tingle, M., Ray, A., Swan, G. E., Lazzeroni, L. C., and Clark, J. D. (2012). Pain sensitivity and opioid analgesia: a pharmacogenomic twin study. *Pain*, 153(7):1397–1409.

Apolloni, A., Poletto, C., and Colizza, V. (2013). Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic. *BMC infectious diseases*, 13(1):176.

Araz, O. M., Galvani, A., and Meyers, L. A. (2012). Geographic prioritization of distributing pandemic influenza vaccines. *Health Care Management Science*, 15(3):175–187.

Ashburn, M. A. and Staats, P. S. (1999). Management of chronic pain. *The Lancet*, 353(9167):1865–1869.

Azman, A. S. and Lessler, J. (2015). Reactive vaccination in the presence of disease hotspots. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1798):20141341.

Azziz-Baumgartner, E., Alamgir, A., Rahman, M., Homaira, N., Sohel, B. M., Sharker, M. Y., Zaman, R. U., Dee, J., Gurley, E. S., Al Mamun, A., Mah-E-Muneer, S., Fry, A. M., Widdowson, M.-A., Bresee, J., Lindstrom, S., Azim, T., Brooks, A., Podder, G., Hossain, M. J., Rahman, M., and Luby, S. P. (2011). Incidence of influenza-like illness and severe acute respiratory infection during three influenza seasons in Bangladesh, 2008–2010. *Bulletin of the World Health Organization*, 90(1):12–19.

Bajardi, P., Poletto, C., Ramasco, J. J., Tizzoni, M., Colizza, V., and Vespignani, A. (2011). Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS ONE*, 6(1):e16591.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. (2009a). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of The United States of America*, 106(51):21484–21489.

Balcan, D., Hu, H., Goncalves, B., Bajardi, P., Poletto, C., Ramasco, J. J., Paolotti, D., Perra, N., Tizzoni, M., Van den Broeck, W., et al. (2009b). Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC medicine*, 7(1):45.

Ball, F., Britton, T., House, T., Isham, V., Mollison, D., Pellis, L., and Tomba, G. S. (2015). Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67.

Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., and Viboud, C. (2016). Big data for infectious disease surveillance and modeling. *The Journal of infectious diseases*, 214(suppl_4):S375–S379.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., and Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734:1–74.

Bengtsson, L., Gaudart, J., Lu, X., Moore, S., Wetter, E., Sallah, K., Rebaudet, S., and Piarroux, R. (2015). Using mobile phone data to predict the spatial spread of cholera. *Scientific reports*, 5:8923.

Berke, D. S., Reidy, D. E., Miller, J. D., and Zeichner, A. (2017). Take it like a man: Gender-threatened men's experience of gender role discrepancy, emotion activation, and pain tolerance. *Psychology of Men & Masculinity*, 18(1):62.

Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Histoire de l'Académie Royale des Sciences (Paris) avec les Mémoires de Mathematique & de Physique*, pages 1–45.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179.

Bhuiyan, M. U., Luby, S. P., Alamgir, N. I., Homaira, N., Mamun, A. A., Khan, J. A., Abedin, J., Sturm-Ramirez, K., Gurley, E. S., Zaman, R. U., Alamgir, A., Rahman, M., Widdowson, M.-A., and Azziz-Baumgartner, E. (2014). Economic burden of influenza-associated hospitalizations and outpatient visits in Bangladesh during 2010. *Influenza and other respiratory viruses*, 8(4):406–413.

Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I. C.-H., Hickmann, K. S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., Velardi, P., Vespignani, A., Finelli, L., and for the Influenza Forecasting Contest Working Group (2016). Results from the centers for disease control and prevention's predict the 2013–2014 influenza season challenge. *BMC infectious diseases*, 16(1):357.

Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., and Finelli, L. (2014a). Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC infectious diseases*, 14(1):480.

Biggerstaff, M., Jhung, M. A., Reed, C., Fry, A. M., Balluz, L., and Finelli, L. (2014b). Influenza-like illness, the time to seek healthcare, and influenza

antiviral receipt during the 2010–2011 influenza season—United States. *The Journal of infectious diseases*, 210(4):535–544.

Biggerstaff, M., Johansson, M., Alper, D., Brooks, L. C., Chakraborty, P., Farrow, D. C., Hyun, S., Kandula, S., McGowan, C., Ramakrishnan, N., et al. (2018). Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*, 24:26–33.

Birrell, P. J., de Angelis, D., Wernisch, L., Tom, B. D., Roberts, G. O., and Pebody, R. G. (2016a). Efficient real-time monitoring of an emerging influenza epidemic: how feasible? *arXiv preprint arXiv:1608.05292*.

Birrell, P. J., Ketsetzis, G., Gay, N. J., Cooper, B. S., Presanis, A. M., Harris, R. J., Charlett, A., Zhang, X.-S., White, P. J., Pebody, R. G., and de Angelis, D. (2011). Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proceedings of the National Academy of Sciences of The United States of America*.

Birrell, P. J., Pebody, R. G., Charlett, A., Zhang, X.-S., and de Angelis, D. (2017). Real-time modelling of a pandemic influenza outbreak. *Health technology assessment (Winchester, England)*, 21(58):1.

Birrell, P. J., Zhang, X.-S., Pebody, R. G., Gay, N. J., and de Angelis, D. (2016b). Reconstructing a spatially heterogeneous epidemic: Characterising the geographic spread of 2009 A/H1N1pdm infection in England. *Scientific reports*, 6:29004.

Block, P., Heathcote, L. C., and Heyes, S. B. (2018). Social interaction and pain: An arctic expedition. *Social Science & Medicine*, 196:47–55.

Blum, M. G. and Tran, V. C. (2010). HIV with contact tracing: a case study in approximate Bayesian computation. *Biostatistics*, 11(4):644–660.

Bollen, J., Gonçalves, B., Ruan, G., and Mao, H. (2011). Happiness is assortative in online social networks. *Artificial life*, 17(3):237–251.

Bolton, K., McCaw, J., Moss, R., Morris, R., Wang, S., Burma, A., Darma, B., Narangerel, D., Nymadawa, P., and McVernon, J. (2012). Likely effectiveness of pharmaceutical and non-pharmaceutical interventions for mitigating influenza virus transmission in Mongolia. *Bulletin of the World Health Organization*, 90:264–271.

Britton, T. (2010). Stochastic epidemic models: a survey. *Mathematical biosciences*, 225(1):24–35.

Brockmann, D. and Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342.

Brooks, L. C., Farrow, D. C., Hyun, S., Tibshirani, R. J., and Rosenfeld, R. (2015). Flexible modeling of epidemics with an empirical Bayes framework. *PLoS computational biology*, 11(8):e1004382.

Brooks-Pollock, E., Roberts, G. O., and Keeling, M. J. (2014). A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature*, 511(7508):228.

Brown, G. D., Porter, A. T., Oleson, J. J., and Hinman, J. A. (2018). Approximate Bayesian computation for spatial SEIR (S) epidemic models. *Spatial and Spatio-temporal Epidemiology*, 24:27–37.

Brown, J. L., Sheffield, D., Leary, M. R., and Robinson, M. E. (2003). Social support and experimental pain. *Psychosomatic medicine*, 65(2):276–283.

Brownstein, J. S., Wolfe, C. J., and Mandl, K. D. (2006). Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS medicine*, 3(10):e401.

Cacioppo, J. T., Fowler, J. H., and Christakis, N. A. (2009). Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of personality and social psychology*, 97(6):977–991.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A., and Boelle, P. (2004). A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in medicine*, 23(22):3469–3487.

Cauchemez, S., Donnelly, C. A., Reed, C., Ghani, A. C., Fraser, C., Kent, C. K., Finelli, L., and Ferguson, N. M. (2009). Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *New England Journal of Medicine*, 361(27):2619–2627.

Cauchemez, S., Valleron, A.-J., Boelle, P.-Y., Flahault, A., and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452(7188):750.

Centers for Disease Control and Prevention (2009). Allocation and Distribution Q&A. [Online; accessed 24-November-2018].

Chao, D. L., Halloran, M. E., Obenchain, V. J., and Longini Jr, I. M. (2010). FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS computational biology*, 6(1):e1000656.

Chowell, G., Viboud, C., Wang, X., Bertozzi, S. M., and Miller, M. A. (2009). Adaptive vaccination strategies to mitigate pandemic influenza: Mexico as a case study. *PLoS ONE*, 4(12):e8164.

Christakis, N. A. (2004). Social networks and collateral health effects. *BMJ*, 329(184).

Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379.

Christakis, N. A. and Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258.

Ciancio, B. C. and Kramarz, P. (2014). Surveillance for seasonal and novel influenza viruses. *Concepts and Methods in Infectious Disease Surveillance: John Wiley & Sons, Ltd*, pages 41–57.

Ciofi degli Atti, M. L., Merler, S., Rizzo, C., Ajelli, M., Massari, M., Manfredi, P., Furlanello, C., Scalia Tomba, G., and Iannelli, M. (2008). Mitigation measures for pandemic influenza in Italy: An individual based model considering different scenarios. *PLoS ONE*, 3(3):e1790.

Clancy, D. and O'Neill, P. D. (2007). Exact Bayesian inference and model selection for stochastic models of epidemics among a community of households. *Scandinavian Journal of Statistics*, 34(2):259–274.

Cobbinah, P. B., Erdiaw-Kwasie, M. O., and Amoateng, P. (2015). Africa's urbanisation: Implications for sustainable development. *Cities*, 47:62–72.

Coburn, B. J., Wagner, B. G., and Blower, S. (2009). Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC medicine*, 7(1):30.

Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.-J., and Vespignani, A. (2007a). Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine*, 4(1):e13.

Colizza, V., Barrat, A., Barthélemy, M., and Vespignani, A. (2007b). Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. *BMC medicine*, 5(1):34.

Corander, J., Dahmström, K., and Dahmström, P. (1998). *Maximum likelihood estimation for Markov graphs*. Research Report. Vol. 8. University of Stockholm, Department of Statistics.

Corbella, A., Zhang, X.-S., Birrell, P. J., Boddington, N., Pebody, R. G., Presanis, A. M., and de Angelis, D. (2018). Exploiting routinely collected severe case data to monitor and predict influenza outbreaks. *BMC public health*, 18(1):790.

Cotter, S., Gee, S., and OLLorcain, P. (2010). National summary of influenza pandemic (H1N1) 2009 vaccination in Ireland-provisional data. *Epi insight*.

Craig, K. D. (2009). The social communication model of pain. *Canadian Psychology/Psychologie canadienne*, 50(1):22.

Craig, K. D. (2015). Social communication model of pain. *Pain*, 156(7):1198–1199.

Craig, K. D. (2018). Toward the social communication model of pain. In *Social and Interpersonal Dynamics in Pain*, pages 23–41. Springer.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons.

Danon, L., Ford, A. P., House, T., Jewell, C. P., Keeling, M. J., Roberts, G. O., Ross, J. V., and Vernon, M. C. (2011). Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011.

de Angelis, D., Presanis, A. M., Birrell, P. J., Tomba, G. S., and House, T. (2015). Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*, 10:83–87.

de Blasio, B. F., Iversen, B. G., and Tomba, G. S. (2012). Effect of vaccines and antivirals during the major 2009 A (H1N1) pandemic wave in Norway–and the influence of vaccination timing. *PLoS ONE*, 7(1):e30018.

de Blasio, B. F., Svensson, Å., and Liljeros, F. (2007). Preferential attachment in sexual networks. *Proceedings of the National Academy of Sciences of The United States of America*, 104(26):10762–10767.

de Montjoye, Y.-A., Gambs, S., Blondel, V., Canright, G., de Cordes, N., Deletaille, S., Engø-Monsen, K., Garcia-Herranz, M., Kendall, J., Kerry, C., Krings, G., Letouzé, E., Luengo-Oroz, M., Oliver, N., Rocher, L., Rutherford, A., Smoreda, Z., Steele, J., Wetter, E., Pentland, A. S., and Bengtsson, L. (2018). On the privacy-conscientious use of mobile phone data. *Scientific Data*, 5.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3:1376.

de Montjoye, Y.-A., Kendall, J., and Kerry, C. F. (2014). Enabling humanitarian use of mobile phone data.

Deardon, R., Brooks, S. P., Grenfell, B. T., Keeling, M. J., Tildesley, M. J., Savill, N. J., Shaw, D. J., and Woolhouse, M. E. (2010). Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica*, 20(1):239.

Declich, S. and Carter, A. O. (1994). Public health surveillance: historical origins, methods and evaluation. *Bulletin of the World Health Organization*, 72(2):285.

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*, 2(1):23–41.

Dimitrov, N. B., Goll, S., Hupert, N., Pourbohloul, B., and Meyers, L. A. (2011). Optimizing tactics for use of the US antiviral strategic national stockpile for pandemic influenza. *PLoS ONE*, 6(1):e16094.

Dowell, S. F. and Ho, M. S. (2004). Seasonality of infectious diseases and severe acute respiratory syndrome–what we don't know can hurt us. *The Lancet infectious diseases*, 4(11):704–708.

Duijzer, L. E., van Jaarsveld, W., and Dekker, R. (2018a). Literature review: The vaccine supply chain. *European Journal of Operational Research*, 268(1):174–192.

Duijzer, L. E., van Jaarsveld, W. L., Wallinga, J., and Dekker, R. (2018b). Dose-optimal vaccine allocation over multiple populations. *Production and Operations Management*, 27(1):143–159.

Dushoff, J., Plotkin, J. B., Levin, S. A., and Earn, D. J. (2004). Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(48):16915–16916.

Eagle, N. (2009). Engineering a common good: fair use of aggregated, anonymized behavioral data. In *First international forum on the application and management of personal electronic information*.

Edwards, R., Eccleston, C., and Keogh, E. (2017). Observer influences on pain: an experimental series examining same-sex and opposite-sex friends, strangers, and romantic partners. *Pain*, 158(5):846–855.

Edwards, R. R., Sarlani, E., Wesselmann, U., and Fillingim, R. B. (2005). Quantitative assessment of experimental pain perception: multiple domains of clinical relevance. *Pain*, 114(3):315–319.

Eisenberger, N. I., Jarcho, J. M., Lieberman, M. D., and Naliboff, B. D. (2006). An experimental study of shared sensitivity to physical pain and social rejection. *Pain*, 126(1-3):132–138.

Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., and Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, pages 128–140.

Farah, M., Birrell, P., Conti, S., and de Angelis, D. (2014). Bayesian emulation and calibration of a dynamic epidemic model for A/H1N1 influenza. *Journal of the American Statistical Association*, 109(508):1398–1411.

Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., and Burke, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452.

Fillingim, R. B. (2017). Individual differences in pain: understanding the mosaic that makes pain personal. *Pain*, 158(Suppl 1):S11–S18.

Fillingim, R. B., King, C. D., Ribeiro-Dasilva, M. C., Rahim-Williams, B., and Riley, J. L. (2009). Sex, gender, and pain: a review of recent clinical and experimental findings. *Journal of Pain*, 10(5):447–485.

Finger, F., Genolet, T., Mari, L., de Magny, G. C., Manga, N. M., Rinaldo, A., and Bertuzzo, E. (2016). Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proceedings of the National Academy of Sciences of The United States of America*, 113(23):6421–6426.

Fox, S. J., Miller, J. C., and Meyers, L. A. (2017). Seasonality in risk of pandemic influenza emergence. *PLoS computational biology*, 13(10):e1005749.

Frias-Martinez, E., Williamson, G., and Frias-Martinez, V. (2011). An agent-based model of epidemic spread using human mobility and social network information. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 57–64. IEEE.

Funk, S., Camacho, A., Kucharski, A. J., Lowe, R., Eggo, R. M., and Edmunds, W. J. (2019). Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area Region of Sierra Leone, 2014-15. *PLoS computational biology*, 15(2):e1006785.

Funk, S., Ciglenecki, I., Tiffany, A., Gignoux, E., Camacho, A., Eggo, R. M., Kucharski, A. J., Edmunds, W. J., Bolongei, J., Azuma, P., et al. (2017). The impact of control strategies and behavioural changes on the elimination of Ebola from Lofa County, Liberia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1721):20160302.

Germann, T. C., Kadau, K., Longini, I. M., and Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences of The United States of America*, 103(15):5935–5940.

Gibson, G. J. and Renshaw, E. (1998). Estimating parameters in stochastic compartmental models using Markov chain methods. *Mathematical Medicine and Biology: A Journal of the IMA*, 15(1):19–40.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012.

González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.

Grais, R., Ellis, J. H., Kress, A., and Glass, G. (2004). Modeling the spread of annual influenza epidemics in the US: The potential role of air travel. *Health care management science*, 7(2):127–134.

GSMA Intelligence (2017). Bangladesh: Driving mobile-enabled digital transformation.

Gutmann, M. U. and Corander, J. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *The Journal of Machine Learning Research*, 17(1):4256–4302.

Handcock, M. S. (2002). Statistical models for social networks: Inference and degeneracy. In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 229–240. National Academies Press.

Hay, S. I., Guerra, C. A., Tatem, A. J., Atkinson, P. M., and Snow, R. W. (2005). Opinion – tropical infectious diseases: Urbanization, malaria transmission and disease burden in Africa. *Nature Reviews Microbiology*, 3(1):81.

Heesterbeek, H., Anderson, R. M., Andreasen, V., Bansal, S., de Angelis, D., Dye, C., Eames, K. T., Edmunds, W. J., Frost, S. D., Funk, S., Hollingsworth, T. D., House, T., Isham, V., Klepac, P., Lessler, J., Lloyd-Smith, J. O., Metcalf, J. E., Mollison, D., Pellis, L., Pulliam, J. R. C., Roberts, M. G., Viboud, C., and Isaac Newton Institute IDD Collaboration (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227):aaa4339.

Hill, E. M., House, T., Dhingra, M. S., Kalpravidh, W., Morzaria, S., Osmani, M. G., Yamage, M., Xiao, X., Gilbert, M., and Tildesley, M. J. (2017). Modelling H5N1 in Bangladesh across spatial scales: Model complexity and zoonotic transmission risk. *Epidemics*, 20:37–55.

Ho, L. S. T., Crawford, F. W., Suchard, M. A., et al. (2018). Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease. *The Annals of Applied Statistics*, 12(3):1993–2021.

Hoff, P. D. (2015). Dyadic data analysis with amen. *arXiv preprint arXiv:1506.08237*.

Hope-Simpson, R. (1981). The role of season in the epidemiology of influenza. *Epidemiology & Infection*, 86(1):35–47.

House, J. S., Robbins, C., and Metzner, H. L. (1982). The association of social relationships and activities with mortality: prospective evidence from the Tecumseh Community Health Study. *American journal of epidemiology*, 116(1):123–140.

House, T., Inglis, N., Ross, J. V., Wilson, F., Suleman, S., Edeghere, O., Smith, G., Olowokure, B., and Keeling, M. J. (2012). Estimation of outbreak severity and transmissibility: Influenza A (H1N1) pdm09 in households. *BMC medicine*, 10(1):117.

Hufnagel, L., Brockmann, D., and Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of The United States of America*, 101(42):15124–15129.

Irvine, M. A. and Hollingsworth, T. D. (2018). Kernel-density estimation and approximate Bayesian computation for flexible epidemiological model fitting in Python. *Epidemics*, 25:80–88.

Islam, M. A., Khan, M. K., Ahmed, K. G. S., and Asaduzzaman, M. (2013). Surveillance of influenza in outpatient ILI cases in community based medical college hospital, Bangladesh. *Community Based Medical Journal*, 1(2):38–47.

Jackson, C. H., Jit, M., Sharples, L. D., and de Angelis, D. (2015). Calibration of complex models through Bayesian evidence synthesis: a demonstration and tutorial. *Medical decision making*, 35(2):148–161.

Jit, M., Gay, N., Soldan, K., Hong Choi, Y., and Edmunds, W. J. (2010). Estimating progression rates for human papillomavirus infection from epidemiological data. *Medical decision making*, 30(1):84–98.

Johnson, K. V.-A. and Dunbar, R. I. (2016). Pain tolerance predicts human social network size. *Scientific reports*, 6:25267.

Kapferer, J. L. (1990). Rural myths and urban ideologies. *The Australian and New Zealand Journal of Sociology*, 26(1):87–106.

Kawachi, I., Kennedy, B. P., and Glass, R. (1999). Social capital and self-rated health: a contextual analysis. *American journal of public health*, 89(8):1187–1193.

Kawachi, I., Kennedy, B. P., Lochner, K., and Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American journal of public health*, 87(9):1491–1498.

Keeling, M. J., Danon, L., Vernon, M. C., and House, T. A. (2010). Individual identity and movement networks for disease metapopulations. *Proceedings of the National Academy of Sciences of The United States of America*, 107(19):8866–8870.

Keeling, M. J. and Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.

Keeling, M. J. and Shattock, A. (2012). Optimal but unequitable prophylactic distribution of vaccine. *Epidemics*, 4(2):78–85.

Keeling, M. J. and White, P. J. (2011). Targeting vaccination against novel infections: risk, age and spatial structure for pandemic influenza in Great Britain. *Journal of the Royal Society Interface*, 8(58):661–670.

Keogh, E. (2006). Sex and gender differences in pain: a selective review of biological and psychosocial factors. *Journal of Men's health and Gender*, 3(3):236–243.

Keogh, E. (2018). Sex and gender as social-contextual factors in pain. In *Social and Interpersonal Dynamics in Pain*, pages 433–453. Springer.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721.

Kim, J. J., Kuntz, K. M., Stout, N. K., Mahmud, S., Villa, L. L., Franco, E. L., and Goldie, S. J. (2007). Multiparameter calibration of a natural history model of cervical cancer. *American journal of epidemiology*, 166(2):137–150.

Kinlaw, K., Barrett, D. H., and Levine, R. J. (2009). Ethical guidelines in pandemic influenza: recommendations of the ethics subcommittee of the advisory committee of the director, Centers for Disease Control and Prevention. *Disaster medicine and public health preparedness*, 3(S2):S185–S192.

Kiss, I. Z., Green, D. M., and Kao, R. R. (2008). The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. *Journal of The Royal Society Interface*, 5(24):791–799.

Klovdahl, A. S., Potterat, J. J., Woodhouse, D. E., Muth, J. B., Muth, S. Q., and Darrow, W. W. (1994). Social networks and infectious disease: The Colorado Springs study. *Social science & medicine*, 38(1):79–88.

Kolaczyk, E. D. (2009). Statistical analysis of network data. *Springer Series in Statistics*.

Kolaczyk, E. D. and Csárdi, G. (2014). *Statistical analysis of network data with R*, volume 65. Springer.

Kucharski, A. J., Camacho, A., Flasche, S., Glover, R. E., Edmunds, W. J., and Funk, S. (2015). Measuring the impact of Ebola control measures in Sierra Leone. *Proceedings of the National Academy of Sciences of The United States of America*, 112(46):14366–14371.

Kucharski, A. J., Funk, S., Eggo, R. M., Mallet, H.-P., Edmunds, W. J., and Nilles, E. J. (2016). Transmission dynamics of Zika virus in island populations: a modelling analysis of the 2013–14 French Polynesia outbreak. *PLoS neglected tropical diseases*, 10(5):e0004726.

Kucharski, A. J., Kwok, K. O., Wei, V. W., Cowling, B. J., Read, J. M., Lessler, J., Cummings, D. A., and Riley, S. (2014). The contribution of social behaviour to the transmission of influenza a in a human population. *PLoS pathogens*, 10(6):e1004206.

Kunz, M., Karos, K., and Vervoort, T. (2018). When, how, and why do we express pain? In *Social and Interpersonal Dynamics in Pain*, pages 101–119. Springer.

Kypraios, T., D O'Neill, P., Huang, S. S., Rifas-Shiman, S. L., and Cooper, B. S. (2010). Assessing the role of undetected colonization and isolation precautions in reducing methicillin-resistant Staphylococcus aureus transmission in intensive care units. *BMC infectious diseases*, 10(1):29.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: traps in big data analysis. *Science*, 343(6176):1203–1205.

Lee, E. C., Arab, A., Goldlust, S., Viboud, C., and Bansal, S. (2017). Socio-environmental and measurement factors drive spatial variation in influenza-like illness. *bioRxiv*, 112680.

Lee, J. M., Choi, D., Cho, G., and Kim, Y. (2012). The effect of public health interventions on the spread of influenza among cities. *Journal of theoretical biology*, 293:131–142.

Leenders, R. T. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47.

Liang, X., Zhao, J., Dong, L., and Xu, K. (2013). Unraveling the origin of exponential law in intra-urban human mobility. *Scientific reports*, 3:2983.

Lipsitch, M. and Viboud, C. (2009). Influenza seasonality: lifting the fog. *Proceedings of the National Academy of Sciences of the United States of America*, 106(10):3645–3646.

Lofgren, E., Fefferman, N. H., Naumov, Y. N., Gorski, J., and Naumova, E. N. (2007). Influenza seasonality: underlying causes and modeling theories. *Journal of virology*, 81(11):5429–5436.

Lofgren, E. T., Halloran, M. E., Rivers, C. M., Drake, J. M., Porco, T. C., Lewis, B., Yang, W., Vespignani, A., Shaman, J., Eisenberg, J. N., et al. (2014). Opinion: Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences of The United States of America*, 111(51):18095–18096.

Longini, I. M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D. A., and Halloran, M. E. (2005). Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087.

Longini Jr, I. M., Halloran, M. E., Nizam, A., and Yang, Y. (2004). Containing pandemic influenza with antiviral agents. *American journal of epidemiology*, 159(7):623–633.

Lötsch, J., Ultsch, A., and Kalso, E. (2017). Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis. *BJA: British Journal of Anaesthesia*, 119(4):821–829.

Lu, Z., Mitchell, R. M., Smith, R. L., Karns, J. S., Van Kessel, J. A. S., Wolfgang, D. R., Schukken, Y. H., and Grohn, Y. T. (2013). Invasion and transmission of Salmonella Kentucky in an adult dairy herd using approximate Bayesian computation. *BMC veterinary research*, 9(1):245.

Lubbers, M. J. and Snijders, T. A. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social networks*, 29(4):489–507.

Luke, D. A. and Harris, J. K. (2007). Network analysis in public health: history, methods, and applications. *Annual Review of Public Health*, 28:69–93.

Maniadakis, N. and Gray, A. (2000). The economic burden of back pain in the UK. *Pain*, 84(1):95–103.

Martin, L. J., Hathaway, G., Isbester, K., Mirali, S., Acland, E. L., Niederstrasser, N., Slepian, P. M., Trost, Z., Bartz, J. A., Sapolsky, R. M., et al. (2015). Reducing social stress elicits emotional contagion of pain in mouse and human strangers. *Current Biology*, 25(3):326–332.

Master, S. L., Eisenberger, N. I., Taylor, S. E., Naliboff, B. D., Shirinyan, D., and Lieberman, M. D. (2009). A picture's worth: Partner photographs reduce experimentally induced pain. *Psychological Science*, 20(11):1316–1318.

Masucci, A. P., Serras, J., Johansson, A., and Batty, M. (2013). Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 88(2).

Mateus, A. L., Otete, H. E., Beck, C. R., Dolan, G. P., and Nguyen-Van-Tam, J. S. (2014). Effectiveness of travel restrictions in the rapid containment of human influenza: a systematic review. *Bulletin of the World Health Organization*, 92(12):868–880D.

Matrajt, L., Halloran, M. E., and Longini Jr, I. M. (2013). Optimal vaccine allocation for the early mitigation of pandemic influenza. *PLoS computational biology*, 9(3):e1002964.

McGranahan, G. and Satterthwaite, D. (2014). *Urbanisation: Concepts and Trends*. IIED London.

McKinley, T., Cook, A. R., and Deardon, R. (2009). Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5(1).

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.

Meakin, S., Tildesley, M., Davis, E., and Keeling, M. (2019). A metapopulation model for the 2018 Ebola virus disease outbreak in Equateur province in the Democratic Republic of the Congo. *BioRxiv*, page 465062.

Meloni, S., Perra, N., Arenas, A., Gómez, S., Moreno, Y., and Vespignani, A. (2011). Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports*, 1:62.

Meltzer, M. I. (2016). Modeling in real time during the Ebola response. *MMWR supplements*, 65.

Merler, S. and Ajelli, M. (2010). The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1681):557–565.

Merler, S., Ajelli, M., Pugliese, A., and Ferguson, N. M. (2011). Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: implications for real-time modelling. *PLoS computational biology*, 7(9):e1002205.

Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M., and Brunham, R. C. (2005). Network theory and SARS: predicting outbreak diversity. *Journal of theoretical biology*, 232(1):71–81.

Mogil, J. S. (2018). Friends in pain: pain tolerance in a social network. *Scandinavian Journal of Pain*.

Moran, K. R., Fairchild, G., Generous, N., Hickmann, K., Osthus, D., Priedhorsky, R., Hyman, J., and Del Valle, S. Y. (2016). Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. *The Journal of infectious diseases*, 214(suppl_4):S404–S408.

Moreno, J. L. (1934). Who shall survive? *Washington DC: Nervous and Mental Disease Publishing Company*, pages 1–431.

Moreno, J. L., Whitin, E. S., and Jennings, H. H. (1932). *Application of the group method to classification*. National committee on prisons and prison labor.

National Tuberculosis Controllers Association and Centers for Disease Control and Prevention (2005). Guidelines for the investigation of contacts of persons with infectious tuberculosis. *MMWR*, 54(15):1–47.

Neal, P. J. and Roberts, G. O. (2004). Statistical inference and model selection for the 1861 Hagelloch measles epidemic. *Biostatistics*, 5(2):249–261.

Newman, M. E. (2002). Assortative mixing in networks. *Physical review letters*, 89(20):208701.

Nguyen, C. and Carlson, J. M. (2016). Optimizing real-time vaccine allocation in a stochastic SIR model. *PLoS ONE*, 11(4):e0152950.

Nielsen, C. S. (2013). Chronic pain is strongly associated with work disability. *Scandinavian Journal of Pain*, 4:180–1.

Nielsen, C. S., Stubhaug, A., Price, D. D., Vassend, O., Czajkowski, N., and Harris, J. R. (2008). Individual differences in pain sensitivity: genetic and environmental contributions. *Pain*, 136(1):21–29.

Nikolay, B., Salje, H., Sturm-Ramirez, K., Azziz-Baumgartner, E., Homaira, N., Ahmed, M., Iuliano, A. D., Paul, R. C., Rahman, M., Hossain, M. J., Luby, S. P., Cauchemez, S., and Gurley, E. S. (2017). Evaluating hospital-based surveillance for outbreak detection in Bangladesh: Analysis of healthcare utilization data. *PLoS medicine*, 14(1):1–18.

Noedl, H., Yunus, E., Fally, M., Redlberger-Fritz, M., Starzengruber, P., Swoboda, P., Fuehrer, H., and Khan, W. (2012). Characterization and epidemiology of influenza viruses in patients seeking treatment for influenza-like illnesses in rural Bangladesh. *Journal of Postgraduate Medicine*, 58(4):242.

Norwegian Institute of Public Health (2016). Overvåkingssystemet for influensa. [Online; accessed 3-May-2019].

Nsoesie, E., Mararthe, M., and Brownstein, J. (2013). Forecasting peaks of seasonal influenza epidemics. *PLoS currents*, 5.

Oliver, N., Matic, A., and Frias-Martinez, E. (2015). Mobile network data for public health: opportunities and challenges. *Frontiers in public health*, 3:189.

O'Malley, A. J. and Marsden, P. V. (2008). The analysis of social networks. *Health services and outcomes research methodology*, 8(4):222–269.

O'Neill, P. D. (2010). Introduction and snapshot review: relating infectious disease transmission models to data. *Statistics in medicine*, 29(20):2069–2077.

O'Neill, P. D., Balding, D. J., Becker, N. G., Eerola, M., and Mollison, D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(4):517–542.

Ong, J. B. S., Mark, I., Chen, C., Cook, A. R., Lee, H. C., Lee, V. J., Lin, R. T. P., Tambyah, P. A., and Goh, L. G. (2010). Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore. *PLoS ONE*, 5(4):e10036.

O'Neill, P. D. and Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):121–129.

Pappalardo, L., Rinzivillo, S., and Simini, F. (2016). Human mobility modelling: exploration and preferential return meet the gravity model. *Procedia Computer Science*, 83:934–939.

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925.

Pei, S., Kandula, S., Yang, W., and Shaman, J. (2018). Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences of The United States of America*, 115(11):2752–2757.

Phillips, C. J. (2009). The cost and burden of chronic pain. *Reviews in pain*, 3(1):2–5.

Plan international (2015). Ebola: beyond the health emergency.

Poletto, C., Pelat, C., Levy-Bruhl, D., Yazdanpanah, Y., Boelle, P., and Colizza, V. (2014). Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach. *Eurosurveillance*, 19(23):3.

Poletto, C., Tizzoni, M., and Colizza, V. (2012). Heterogeneous length of stay of hosts' movements and spatial epidemic spread. *Scientific reports*, 2:476.

Poletto, C., Tizzoni, M., and Colizza, V. (2013). Human mobility and time spent at destination: impact on spatial epidemic spreading. *Journal of theoretical biology*, 338:41–58.

Prkachin, K. M., Karos, K., Vervoort, T., and Trost, Z. (2018). Where we've been, where we're at, where do we go from here? In *Social and Interpersonal Dynamics in Pain*, pages 503–520. Springer.

Probert, W. J., Jewell, C. P., Werkman, M., Fonnesbeck, C. J., Goto, Y., Runge, M. C., Sekiguchi, S., Shea, K., Keeling, M. J., Ferrari, M. J., et al. (2018). Real-time decision-making during emergency disease outbreaks. *PLoS computational biology*, 14(7):e1006202.

Punyacharoensin, N., Edmunds, W. J., de Angelis, D., Delpech, V., Hart, G., Elford, J., Brown, A., Gill, N., and White, R. G. (2015). Modelling the HIV epidemic among MSM in the United Kingdom: quantifying the contributions to HIV transmission to better inform prevention initiatives. *Aids*, 29(3):339–349.

Rambhia, K. J., Watson, M., Sell, T. K., Waldhorn, R., and Toner, E. (2010). Mass vaccination for the 2009 H1N1 pandemic: approaches, challenges, and recommendations. *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, 8(4):321–330.

Reis, J. and Shaman, J. (2016). Retrospective parameter estimation and forecast of respiratory syncytial virus in the United States. *PLoS computational biology*, 12(10):e1005133.

Rose, W. J., Bell, J. E., Autry, C. W., and Cherry, C. R. (2017). Urban logistics: Establishing key concepts and building a conceptual framework for future research. *Transportation Journal*, 56(4):357–394.

Rosenquist, J. N., Murabito, J., Fowler, J. H., and Christakis, N. A. (2010). The spread of alcohol consumption behavior in a large social network. *Annals of internal medicine*, 152(7):426–433.

Rvachev, L. A. and Longini Jr, I. M. (1985). A mathematical model for the global spread of influenza. *Mathematical biosciences*, 75(1):3–22.

Salathe, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., Campbell, E. M., Cattuto, C., Khandelwal, S., Mabry, P. L., et al. (2012). Digital epidemiology. *PLoS computational biology*, 8(7):e1002616.

Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., and Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of The United States of America*, 107(51):22020–22025.

Salathé, M., Vu, D. Q., Khandelwal, S., and Hunter, D. R. (2013). The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science*, 2(1):4.

Salomon, J. A., Weinstein, M. C., Hammitt, J. K., and Goldie, S. J. (2002). Empirically calibrated model of hepatitis C virus infection in the United States. *American journal of epidemiology*, 156(8):761–773.

Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., and Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS computational biology*, 11(10):e1004513.

Sattenspiel, L. and Herring, D. A. (1998). Structured epidemic models and the spread of influenza in the central Canadian subarctic. *Human Biology*, pages 91–115.

Sattenspiel, L. and Herring, D. A. (2003). Simulating the effect of quarantine on the spread of the 1918–19 flu in central Canada. *Bulletin of mathematical biology*, 65(1):1–26.

Schwartz, E. J., Choi, B., and Rempala, G. A. (2015). Estimating epidemic parameters: Application to H1N1 pandemic data. *Mathematical biosciences*, 270:198–203.

Shalizi, C. R. and Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*, 40(2):211–239.

Shrestha, S., Chatterjee, S., Rao, K. D., and Dowdy, D. W. (2016). Potential impact of spatially targeted adult tuberculosis vaccine in Gujarat, India. *Journal of The Royal Society Interface*, 13(116):20151016.

Signorini, A., Segre, A. M., and Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS ONE*, 6(5):e19467.

Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484(7392):96.

Simonsen, L., Clarke, M. J., Williamson, G. D., Stroup, D. F., Arden, N. H., and Schonberger, L. B. (1997). The impact of influenza epidemics on mortality: introducing a severity index. *American journal of public health*, 87(12):1944–1950.

Simonsen, L., Gog, J. R., Olson, D., and Viboud, C. (2016). Infectious disease surveillance in the big data era: towards faster and locally relevant systems. *The Journal of infectious diseases*, 214(suppl_4):S380–S385.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of The United States of America*, 104(6):1760–1765.

Small, M. and Tse, C. K. (2005). Small world and scale free model of transmission of SARS. *International Journal of Bifurcation and Chaos*, 15(05):1745–1755.

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.

Snijders, T. A. and Borgatti, S. P. (1999). Non-parametric standard errors and tests for network statistics. *Connections*, 22(2):161–170.

Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social networks*, 32(1):44–60.

Song, C., Koren, T., Wang, P., and Barabási, A.-L. (2010a). Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818.

Song, C., Qu, Z., Blumm, N., and Barabási, A.-L. (2010b). Limits of predictability in human mobility. *Science*, 327(5968):1018–1021.

Steglich, C., Snijders, T. A. B., and Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1):329–393.

Sun, L., Lee, C., and Hoeting, J. A. (2015). Parameter inference and model selection in deterministic and stochastic dynamical models via approximate Bayesian computation: modeling a wildlife epidemic. *Environmetrics*, 26(7):451–462.

Swedish Council on Technology Assessment in Health Care (SBU) (2006). Methods of treating chronic pain. SBU report no 177/1+2. *Stockholm, SBU*.

Tamerius, J., Nelson, M. I., Zhou, S. Z., Viboud, C., Miller, M. A., and Alonso, W. J. (2010). Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environmental health perspectives*, 119(4):439–445.

Tamerius, J. D., Shaman, J., Alonso, W. J., Bloom-Feshbach, K., Uejio, C. K., Comrie, A., and Viboud, C. (2013). Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS pathogens*, 9(3):e1003194.

Tan, K.-X., Jacob, S. A., Chan, K.-G., and Lee, L.-H. (2015). An overview of the characteristics of the novel avian influenza A H7N9 virus in humans. *Frontiers in microbiology*, 6:140.

Tan, Y., Lam, T. T.-Y., Wu, C., Lee, S.-s., Viboud, C., Zhang, R., and Weinberger, D. M. (2014). Increasing similarity in the dynamics of influenza in two adjacent subtropical Chinese cities following the relaxation of border restrictions. *The Journal of general virology*, 95(Pt 3):531.

Tatem, A. J., Qiu, Y., Smith, D. L., Sabot, O., Ali, A. S., and Moonen, B. (2009). The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. *Malaria Journal*, 8(1):287.

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145(2):505–518.

Teytelman, A. and Larson, R. C. (2013). Multiregional dynamic vaccine allocation during an influenza epidemic. *Service Science*, 5(3):197–215.

The World Bank (2018). Population, total. [Online; accessed 5-June-2018].

Tizzoni, M., Bajardi, P., Decuyper, A., King, G. K. K., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., and Colizza, V. (2014). On the use of human mobility proxies for modeling epidemics. *PLoS computational biology*, 10(7):e1003716.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202.

Trost, Z., Strachan, E., Sullivan, M., Vervoort, T., Avery, A. R., and Afari, N. (2015). Heritability of pain catastrophizing and associations with experimental pain outcomes: a twin study. *Pain*, 156(3):514.

Truscott, J. and Ferguson, N. M. (2012). Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS computational biology*, 8(10):e1002699.

Tuite, A. R., Fisman, D. N., Kwong, J. C., and Greer, A. L. (2010). Optimal pandemic influenza vaccine allocation strategies for the Canadian population. *PLoS ONE*, 5(5):e10520.

United Nations, Department of Economic and Social Affairs, Population Division (2018). 2018 revision of world urbanization prospects. Accessed: 2018.07.26.

US Department of Health and Human Services (2017). HHS Pandemic influenza plan 2017 UPDATE. https://www.cdc.gov/flu/pandemic-resources/pdf/panflu-report-2017v2.pdf.

Valente, T. W. (2010). *Social networks and health: Models, methods, and applications*, volume 1. Oxford University Press New York.

Valente, T. W., Fujimoto, K., Chou, C.-P., and Spruijt-Metz, D. (2009). Adolescent affiliations and adiposity: a social network analysis of friendships and obesity. *Journal of Adolescent Health*, 45(2):202–204.

Vasanen, A. (2009). Deconcentration versus spatial clustering: changing population distribution in the Turku urban region, 1980–2005. *Fennia-International Journal of Geography*, 187(2):115–127.

Vervoort, T., Karos, K., Trost, Z., and Prkachin, K. M. (2018). Social and interpersonal dynamics in pain.

Viboud, C., Alonso, W. J., and Simonsen, L. (2006a). Influenza in tropical regions. *PLoS medicine*, 3(4):e89.

Viboud, C., Bjørnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., and Grenfell, B. T. (2006b). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312(5772):447–451.

Viboud, C., Miller, M. A., Grenfell, B. T., Bjørnstad, O. N., and Simonsen, L. (2006c). Air travel and the spread of influenza: important caveats. *PLoS medicine*, 3(11):e503.

Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., Vespignani, A., and the RAPIDD Ebola Forecasting Challenge group (2018). The rapidd ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21.

Watts, D. J., Muhamad, R., Medina, D. C., and Dodds, P. S. (2005). Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences of The United States of America*, 102(32):11157–11162.

Weinberger, D. M., Krause, T. G., Mølbak, K., Cliff, A., Briem, H., Viboud, C., and Gottfredsson, M. (2012). Influenza epidemics in Iceland over 9 decades: changes in timing and synchrony with the United States and Europe. *American journal of epidemiology*, 176(7):649–655.

Wesolowski, A., Buckee, C. O., Engø-Monsen, K., and Metcalf, C. (2016). Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *The Journal of infectious diseases*, 214(suppl_4):S414–S420.

Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., and Buckee, C. O. (2012a). Heterogeneous mobile phone ownership and usage patterns in Kenya. *PLoS ONE*, 7(4):e35319.

Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., and Buckee, C. O. (2013). The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface*, 10(81):20120986.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., and Buckee, C. O. (2012b). Quantifying the impact of human mobility on malaria. *Science*, 338(6104):267–270.

Wesolowski, A., Metcalf, C., Eagle, N., Kombich, J., Grenfell, B. T., Bjørnstad, O. N., Lessler, J., Tatem, A. J., and Buckee, C. O. (2015a). Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proceedings of the National Academy of Sciences of The United States of America*, 112(35):11114–11119.

Wesolowski, A., O'Meara, W. P., Eagle, N., Tatem, A. J., and Buckee, C. O. (2015b). Evaluating spatial interaction models for regional mobility in sub-Saharan Africa. *PLoS computational biology*, 11(7):e1004267.

Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rasheed, S. B., Engø-Monsen, K., and Buckee, C. O. (2015c). Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38):11887–11892.

Wesolowski, A., Stresman, G., Eagle, N., Stevenson, J., Owaga, C., Marube, E., Bousema, T., Drakeley, C., Cox, J., and Buckee, C. O. (2014). Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. *Scientific reports*, 4:5678.

Westveld, A. H., Hoff, P. D., et al. (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics*, 5(2A):843–872.

Wood, J., McCaw, J., Becker, N., Nolan, T., and MacIntyre, C. R. (2009). Optimal dosing and dynamic distribution of vaccines in an influenza pandemic. *American journal of epidemiology*, 169(12):1517–1524.

Woolhouse, M. E., Rambaut, A., and Kellam, P. (2015). Lessons from Ebola: Improving infectious disease surveillance to inform outbreak management. *Science translational medicine*, 7(307):307rv5–307rv5.

World Health Organization (2016). Influenza (seasonal). http://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal). Accessed: 2016-08-15.

World Health Organization (2019). Zoonotic influenza. [Online; accessed 06-June-2019].

Wu, J. T., Riley, S., and Leung, G. M. (2007). Spatial considerations for the allocation of pre-pandemic influenza vaccination in the United States. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1627):2811–2817.

Xia, Y., Bjørnstad, O. N., and Grenfell, B. T. (2004). Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281.

Yang, W., Lipsitch, M., and Shaman, J. (2015a). Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 112(9):2723–2728.

Yang, Y., Sugimoto, J. D., Halloran, M. E., Basta, N. E., Chao, D. L., Matrajt, L., Potter, G., Kenah, E., and Longini, I. M. (2009). The transmissibility and control of pandemic influenza A (H1N1) virus. *Science*, 326(5953):729–733.

Yang, Y., Zhang, Y., Fang, L., Halloran, M., Ma, M., Liang, S., Kenah, E., Britton, T., Chen, E., Hu, J., et al. (2015b). Household transmissibility of avian influenza A (H7N9) virus, China, February to May 2013 and October 2013 to March 2014. *Eurosurveillance*, 20(10).

Yu, H., Alonso, W. J., Feng, L., Tan, Y., Shu, Y., Yang, W., and Viboud, C. (2013). Characterization of regional influenza seasonality patterns in china and implications for vaccination strategies: spatio-temporal modeling of surveillance data. *PLoS medicine*, 10(11):e1001552.

Zaman, R. U., Alamgir, A. S. M., Rahman, M., Azziz-Baumgartner, E., Gurley, E. S., Sharker, M. A. Y., Brooks, W. A., Azim, T., Fry, A. M., Lindstrom, S., Gubareva, L. V., Xu, X., Garten, R. J., Hossain, M. J., Khan, S. U., Faruque, L. I., Ameer, S. S., Klimov, A. I., Rahman, M., and Luby, S. P. (2009). Influenza in outpatient ILI case-patients in national hospital-based surveillance, Bangladesh, 2007–2008. *PLoS ONE*, 4(12):e8452.

Zeman, J. and Garber, J. (1996). Display rules for anger, sadness, and pain: It depends on who is watching. *Child development*, 67(3):957–973.

Zhang, P. and Atkinson, P. M. (2008). Modelling the effect of urbanization on the transmission of an infectious disease. *Mathematical biosciences*, 211(1):166–185.

Ziersch, A. M. (2005). Health implications of access to social capital: findings from an Australian study. *Social Science & Medicine*, 61(10):2119–2131.

Zimmer, C., Yaesoubi, R., and Cohen, T. (2017). A likelihood approach for real-time calibration of stochastic compartmental epidemic models. *PLoS computational biology*, 13(1):e1005257.

Zipf, G. K. (1946). The P_1P_2/D hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677.

Papers

Paper I

The peer effect on pain tolerance

I

Observational study

Solveig Engebretsen*, Arnoldo Frigessi, Kenth Engø-Monsen, Anne-Sofie Furberg, Audun Stubhaug, Birgitte Freiesleben de Blasio and Christopher Sivert Nielsen

The peer effect on pain tolerance

https://doi.org/10.1515/sjpain-2018-0060
Received March 23, 2018; revised April 4, 2018; accepted April 4, 2018; previously published online May 19, 2018

Abstract

Background and aims: Twin studies have found that approximately half of the variance in pain tolerance can be explained by genetic factors, while shared family environment has a negligible effect. Hence, a large proportion of the variance in pain tolerance is explained by the (non-shared) unique environment. The social environment beyond the family is a potential candidate for explaining some of the variance in pain tolerance. Numerous individual traits have previously shown to be associated with friendship ties. In this study, we investigate whether pain tolerance is associated with friendship ties.

Methods: We study the friendship effect on pain tolerance by considering data from the Tromsø Study: Fit Futures I, which contains pain tolerance measurements and social network information for adolescents attending first year of upper secondary school in the Tromsø area in Northern Norway. Pain tolerance was measured with the cold-pressor test (primary outcome), contact heat and pressure algometry. We analyse the data by using statistical methods from social network analysis. Specifically, we compute pairwise correlations in pain tolerance among friends. We also fit network autocorrelation models to the data, where the pain tolerance of an individual is explained by (among other factors) the average pain tolerance of the individual's friends.

Results: We find a significant and positive relationship between the pain tolerance of an individual and the pain tolerance of their friends. The estimated effect is that for every 1 s increase in friends' average cold-pressor tolerance time, the expected cold-pressor pain tolerance of the individual increases by 0.21 s (p-value: 0.0049, sample size $n=997$). This estimated effect is controlled for sex. The friendship effect remains significant when controlling for potential confounders such as lifestyle factors and test sequence among the students. Further investigating the role of sex on this friendship effect, we only find a significant peer effect of male friends on males, while there is no significant effect of friends' average pain tolerance on females in stratified analyses. Similar, but somewhat lower estimates were obtained for the other pain modalities.

Conclusions: We find a positive and significant peer effect in pain tolerance. Hence, there is a significant tendency for students to be friends with others with similar pain tolerance. Sex-stratified analyses show that the only significant effect is the effect of male friends on males.

Implications: Two different processes can explain the friendship effect in pain tolerance, selection and social transmission. Individuals might select friends directly due to similarity in pain tolerance, or indirectly through similarity in other confounding variables that affect pain tolerance. Alternatively, there is an influence effect among friends either directly in pain tolerance, or indirectly through other variables that affect pain tolerance. If there is indeed a social influence effect in pain tolerance, then the social environment can account for some of the unique environmental variance in pain tolerance. If so, it is possible to therapeutically affect pain tolerance through alteration of the social environment.

***Corresponding author: Solveig Engebretsen,** Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Post box 1122 Blindern, 0316 Oslo, Norway, Phone: +47 470 83 876, E-mail: solveig.engebretsen@medisin.uio.no; and Department of Infectious Disease Epidemiology and Modelling, Norwegian Institute of Public Health, Oslo, Norway
Arnoldo Frigessi: Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway; and Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway
Kenth Engø-Monsen: Telenor Research, Telenor Group, Fornebu, Norway
Anne-Sofie Furberg: Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway; and Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway
Audun Stubhaug: Department of Pain Management and Research, Oslo University Hospital, Oslo, Norway; and Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
Birgitte Freiesleben de Blasio: Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway; and Department of Infectious Disease Epidemiology and Modelling, Norwegian Institute of Public Health, Oslo, Norway
Christopher Sivert Nielsen: Department of Pain Management and Research, Oslo University Hospital, Oslo, Norway; and Department of Ageing, Norwegian Institute of Public Health, Oslo, Norway

Keywords: pain tolerance; social network; assortativity; cold-pressor test; social influence.

1 Introduction

One hundred and fifty years of twin research has established that most human phenotypes are heritable. Pain is no exception, and the genetic contributions to both clinical and experimental pain are considerable [1]. However, genetic influences do not explain all the variability, and environmental causes are typically of equal or greater importance. The often overlooked and frequently disbelieved finding from twin studies is that shared (family) environmental influences have negligible effect on most phenotypes, leaving non-shared (unique) environmental influences as the major source of variance beyond genetics [2]. This is also the case for pain sensitivity. Nielsen et al. [3] found that 54% of the variance in cold-pressor pain was due to genetic factors, with the remaining variance explained by unique environment. Trost et al. [4] found that genetic factors accounted for 55% of the variance in cold-pressor tolerance, and again no evidence of shared environment effects. Likewise, Angst et al. [5] reported 49% heritability and no shared environmental effect for cold-pressor tolerance, though a possible minor effect of shared environment was observed for pain threshold. Finally, a twin study of nine experimental pain assays found shared environmental effects for only one of these (flare area after burn injury) [6].

Though such findings are ubiquitous in the twin literature, an explanation of what exactly the unique environmental factors influencing human health and behaviour are, is lacking. One potential candidate for explaining remaining variance is the social environment beyond the family of origin. Plomin and Daniels [2] hypothesises that the unique environmental variance might increase with age, due to expansion of the social network beyond the family. In a large meta-analysis, Nan et al. [7] demonstrate that the unique environmental influences on body mass index increase with age. They also find that while the family environment has some influence on body mass index in children, the effect is negligible after puberty (18–22 years). This might be due to a shift where adolescents are increasingly affected by their peers rather than their parents.

When ties in a network tend to be between individuals with similar traits, such that connected nodes are more similar, the network is said to be assortative [8]. Newman [8] found that in social networks, individuals with many friends tend to be friends with individuals who have many friends. Sex is known to be assortative in social networks [9]. Other examples of traits that are assortatively mixed in social networks are obesity, smoking, loneliness and alcohol consumption [10–14]. It has also been shown that some genotypes are correlated with social links [15]. Two possible processes can explain these observations. Either individuals are attracted to, and therefore form friendships with, others with similar traits (homophily), or there is a social contagion effect where friends influencing each other cause the similarity. The two processes might also operate simultaneously.

To our knowledge, social network analysis has not been applied in the study of pain phenotypes. In this paper, we investigate whether pain tolerance is assortative in a friendship network of young adolescents.

2 Methods

2.1 Study population

The study sample comprises participants in the Tromsø study: Fit Futures I (TFF1), conducted in the Tromsø and Balsfjord municipalities in Northern Norway in 2010–2011. TFF1 was executed as a single site study, at the Clinical Research Unit, University Hospital of Northern Norway in Tromsø. The study included physical examinations, questionnaire screening and interviews. All first year students at the eight upper secondary schools in the area were invited to participate ($n = 1{,}117$). Of these, 1,038 participated, for a response rate of 93%. From this sample, students with cognitive disabilities ($n = 20$) were removed, leaving a final sample of $n = 1{,}018$. However, not all the participants completed the different pain tests, resulting in a sample size of $n = 997$ individuals for the main outcome measure (the cold-pressor test).

2.2 Social network survey

A study nurse interviewed all the students. As part of the interview, they were asked to name up to five friends from their own school or the seven other schools in the area. The friends were defined to be the five individuals whom they had spent the most time with in the preceding week. Follow-up questions were asked where necessary, to ensure unique identification of the named individuals. Friends were then recoded to personal identification

number, which uniquely identifies all Norwegian residents. If a participant named friends who did not participate, these were retained as anonymous nodes in the social network.

2.3 Demographic variables and lifestyle factors

Among the interview and questionnaire data collected on the adolescents, we will in this analysis use information about sex, age, school programme (vocational versus general studies programme), smoking and physical activity. For some of the $n = 997$ individuals, lifestyle information (smoking and physical activity) is missing, resulting in a sample size of $n = 982$ adolescents for the analysis where lifestyle is considered.

In the statistical analysis in this study, the following coding of the variables is used: sex is 0 if the adolescent is a girl, and 1 if the adolescent is a boy. School programme is 0 if the pupil is attending a vocational programme, 1 if the pupil is attending a general studies programme. Physical activity is an ordered categorical (frequency) variable. The adolescents were asked "If you are actively doing sports or physical activity outside school, how many days a week are you active?", with response options "never", "less than once a week", "1 day a week", "2–3 days a week", "4–6 times a week" and "almost every day". We use a coding from 0 ("never") to 5 ("almost every day"). Smoking is coded as 0 if the adolescents never smoke, 1 if they smoke sometimes and 2 if they smoke daily.

2.4 Experimental pain assessment

The experimental pain procedure encompassed 1) heat pain threshold and tolerance, 2) pressure pain threshold and tolerance on the fingernail and trapezius muscle and 3) the cold-pressor test, in that order.

2.4.1 Heat pain

Heat pain was induced on the ulnar side of the forearm with a MEDOC ATS somatosensory stimulator with a 30×30 mm thermode (Medoc, Ltd, Israel). Stimuli started at a baseline of 32 °C and increased by 1 °C/s until the student pressed a button, at which the temperature was recorded and the thermode returned to baseline by 8 °C/s. Three threshold tests were conducted, where the students were told to press the button as soon as the sensation changed from warm to pain. This was followed by two tolerance tests, where the instruction was to press the button when the pain became unbearable. An upper safety limit was set at 50 °C for all stimuli.

2.4.2 Pressure pain

Pressure pain thresholds and tolerances were assessed with an AlgoMed computerised pressure algometer with a 1 cm^2 probe (Medoc, Ltd, Israel) on the cuticle of the ring finger nail of the non-dominant hand and on the non-dominant trapezius muscle, midway between the neck and shoulder joint. The baseline pressure was 0 kPa, and was increased by a rate of 30 kPa/s until the student pressed a button, at which point the pressure was recorded and the probe was removed. As for heat, thresholds were assessed three times and tolerances were assessed twice. The threshold instruction was to press the button when the sensation changed from pressure to pain. The tolerance instruction was to press the button when the pain became unbearable. An upper safety limit of 1,000 kPa was set for all stimuli.

2.4.3 Cold-pressor test

The student was asked to place her/his non-dominant hand and wrist in 3 °C water in a 13-L plexi-glass container connected with a circulating water bath (Julabo PF40-HE; Julabo Labortechnik GmbH, Germany) and maintain it there as long as she/he was able or until the maximum time of 105 s was reached. Pain tolerance was defined as the time (in seconds) the participant kept the hand in the water.

2.4.4 All stimuli

For heat pain and pressure pain thresholds, the two last measurements were averaged. In a few cases, one of the threshold measures is censored while the other is not. In these cases, we set the threshold to be the minimum value. For heat and pressure pain tolerance, we used the last measurement. For illustrative purposes, the pain tolerance measures were also dichotomised, with subjects reaching the maximum stimulus limit (105 s, 50 °C, 1,000 kPa for cold-pressor, heat and pressure stimuli, respectively) defined as pain tolerant and the remaining participants defined as pain sensitive.

2.5 Statistical analysis

In order to assess the relationship between friendship ties and individual traits, social network analysis is used. Two different approaches are taken. We calculate pairwise correlations of pain tolerance of two friends, taking into account one friendship at the time. In addition, we fit network autocorrelation models to the data, which relate the pain tolerance of an individual to the average pain tolerance of its friends, considering all the friendships of the individual simultaneously. Though these two measures are clearly not independent, they are not equivalent. It is possible to have a significant effect for one but not for the other. However, if both are significant, then there is stronger evidence of a peer effect. We will use a 0.05 significance level in the statistical analysis.

2.5.1 Definition of the social network

In the friendship network, nodes represent the individuals. There is a directed link going from node *i* to node *j*, if individual *i* nominated individual *j* as a friend. The node *i* is the start node of the link, and the node *j* is the end node of the link. This gives us an adjacency matrix $\boldsymbol{W}$, which contains the friendship information, where the element in the *i*th row and the *j*th column is 1 if individual *i* nominated individual *j* as a friend, and 0 otherwise.

2.5.2 Pairwise correlations

Since there is an upper bound on the pain tolerance measurements, they are right-censored. We thus measure correlations with Kendall's rank correlation coefficient, Kendall's τ, which is straightforward to compute when there is only one fixed censoring point, as in our setting [16]. We compute the correlation between the pain tolerance of the start node and the end node of all the links.

In order to assess the significance of the correlation, we perform permutation tests. This is done by randomly shuffling the pain tolerance observations for all the individuals, while keeping the network structure fixed. The purpose of the permutation is to remove the correlation between the outcome and the network structure, and hence generate a random reference model. Our observed value of τ can then be compared with the correlations obtained for the shuffled data sets. We use 50,000 permutations. The correlations from the permuted data sets are observations obtained under the null hypothesis of no correlation between friends and pain tolerance. We can thus obtain an estimated confidence interval of τ under the null hypothesis, or test if the correlation is significant.

2.5.3 Network autocorrelation model

In order to estimate the effect of friends' pain tolerance on the individual's pain tolerance, we fit a network autocorrelation model [17] to the data. The model is given by

$$\boldsymbol{Y} = \rho \boldsymbol{W}_{\mathrm{N}} \boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon,$$

where $\boldsymbol{Y}$ is the vector of pain tolerance for all the individuals, ρ is the autocorrelation between an individual's pain tolerance and the friends' pain tolerance, $\boldsymbol{W}_{\mathrm{N}}$ is the adjacency matrix, normalised by the number of friends for each individual (so that the rows sum to one), $\boldsymbol{X}$ is a matrix with other explanatory variables, $\boldsymbol{\beta}$ is the corresponding vector of coefficients and ϵ is a vector of random normal noise with mean 0 and variance σ^2. The autocorrelation coefficient ρ is a measure of the degree to which an individual has a similar pain tolerance to the overall pain tolerance of her/his friends, and can be interpreted as the expected increase of the individual's pain tolerance with an increase of the average pain tolerance of the individual's friends by one unit (1 s for cold-pressor pain, 1 °C for heat, 1 kPa for pressure).

When fitting this model, we have to take into account both the fact that the observations are dependent (due to the network structure) and censored. In the following, we describe the details of the fitting procedure. We focus on the cold-pressor pain tolerance, and let Y_i be the cold-pressor pain tolerance of individual *i*. We estimate the parameters of the network autocorrelation model as maximisers of a likelihood function, which is a function of the parameters given the observations. Since the full likelihood is intractable, we approximate the likelihood by the pseudo likelihood [18]. In the pseudo likelihood approximation, the observations are assumed to be conditionally independent. In addition, we have to take into account that we do not observe the response variable itself, but we observe $Z_i = \min(Y_i, 105)$. Let δ_i be a censor indicator, so δ_i is one if observation *i* is censored, zero otherwise. The pseudo likelihood is given by

$$L(\boldsymbol{z}, \rho, \boldsymbol{\beta}, \sigma, \boldsymbol{y}) = \prod_{i=1}^{n} \left(\frac{1}{\sigma} \varphi\left(\frac{z_i - x_i^T \boldsymbol{\beta} - \rho y_{-i}}{\sigma} \right) \right)^{1-\delta_i} \times \left(1 - \Phi\left(\frac{105 - x_i^T \boldsymbol{\beta} - \rho y_{-i}}{\sigma} \right) \right)^{\delta_i},$$

where y_{-i} is the average pain tolerance of the friends of individual *i*, $\boldsymbol{x}_i$ is the vector of explanatory variables for

individual i, n is the number of individuals, φ is the standard normal density and Φ is the standard normal cumulative probability function. The maximum pseudo likelihood estimators are in general good approximations [19]. The y_{-i} cannot be evaluated directly, due to the censored observations. Hence, we use an expectation-maximisation (EM) algorithm to maximise the pseudo likelihood. In an EM algorithm, the estimation is performed iteratively, until we have convergence in the estimates. Let $\boldsymbol{\theta}=(\rho, \boldsymbol{\beta}, \sigma)$ be the vector of model parameters. We start with an initial guess for the censored observations as random draws from a truncated normal on [105, ∞). We denote the initial guess by $\hat{\boldsymbol{y}}_0$. Note that the observed y_i that are not censored are kept at their observed values. A first estimate for the model parameters, $\hat{\boldsymbol{\theta}}_1$, is then found by

$$\hat{\boldsymbol{\theta}}_1 = \text{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{z}, \boldsymbol{\theta}, \hat{\boldsymbol{y}}_0).$$

This estimate for $\boldsymbol{\theta}$ is then used to find a first estimate for $\boldsymbol{y}$, $\hat{\boldsymbol{y}}_1$, by

$$\hat{y}_i^1 = \begin{cases} y_i, & \text{if } \delta_i = 0, \\ \text{mean of a truncated N}(\hat{\rho}_1 \boldsymbol{W}_{\text{N}i} \hat{\boldsymbol{y}}_0 + \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_1,\ \hat{\sigma}_1) \text{on } [105, \infty) & \text{otherwise,} \end{cases}$$

where $\hat{y}_i^1$ is the ith element of $\hat{\boldsymbol{y}}_1$ and $\boldsymbol{W}_{\text{N}i}$ is the ith row of the matrix $\boldsymbol{W}_{\text{N}}$. We continue updating using $\hat{\boldsymbol{y}}_k$ to find $\hat{\boldsymbol{\theta}}_{k+1}$ by

$$\hat{\boldsymbol{\theta}}_{k+1} = \text{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{z}, \boldsymbol{\theta}, \hat{\boldsymbol{y}}_k).$$

Then $\hat{\boldsymbol{\theta}}_{k+1}$ is used to find $\hat{\boldsymbol{y}}_{k+1}$ by

$$\hat{y}_i^{k+1} = \begin{cases} y_i, & \text{if } \delta_i = 0, \\ \text{mean of a truncated N}(\hat{\rho}_{k+1} \boldsymbol{W}_{\text{N}i} \hat{\boldsymbol{y}}_k + \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_{k+1},\ \hat{\sigma}_{k+1}) \text{on } [105, \infty) & \text{otherwise.} \end{cases}$$

The updating steps are repeated until we have reached convergence. In order to obtain standard errors for the estimates, we use jackknife procedures which are specific for networks, as explained in [20]. The idea behind the jackknife approach is to leave out one individual at a time, and compute the estimates based on the network with this one individual removed. The variance between these estimates is used to estimate the variance of the estimators.

3 Results

Descriptive statistics for the experimental pain measures are given in Table 1. All pain tolerance and threshold

Table 1: Descriptive statistics for the various pain measures.

	n	Girls	Boys	Cens.	Med.	Min.
Cold-pressor tol.	997	481	516	502	105	9.30
Heat pain tol.	983	478	505	278	48.40	37.90
Press. tol. f.	902	440	462	316	798	182
Press. tol. t.	886	428	458	228	570	84
Heat pain th.	983	479	504	13	44.90	35.45
Press. th. f.	905	443	462	21	400	99
Press. th. t.	882	428	454	3	253.75	55
	Med. girls	**Med. boys**	**Min. girls**	**Min. boys**		
Cold-pressor tol.	78.10	105	11	9.30		
Heat pain tol.	47.50	49.40	37.90	39.40		
Press. tol. f.	667	994.50	182	252		
Press. tol. t.	478	684	84	136		
Heat pain th.	44.75	45.15	36.30	35.45		
Press. th. f.	364	459	99	116		
Press. th. t.	227	278	55	92		

The Tromsø Study: Fit Futures I: Number of participants, number of girls, number of boys, number of censored individuals (cens.), median (med.), minimum value (min.), median for the girls, median for the boys, minimum value for the girls and minimum value for the boys, for the different pain measurements. Cold-pressor tol. is the cold-pressor pain tolerance, heat pain tol. is the heat pain tolerance, press. tol. f is the pressure pain tolerance at the fingernail, press. tol. t is the pressure pain tolerance at the trapezius, and th. is the threshold.

measures were right censored, though this was most pronounced for the cold-pressor test. A histogram of the measured cold-pressor pain tolerance values is given in Fig. 1.

3.1 Cold-pressor pain tolerance

3.1.1 Main results

A plot of the friendship network, coded by cold-pressor pain tolerance, is shown in Fig. 2. At first glance, there seems to be some clustering of individuals with similar pain tolerance. In Fig. 3, we plot the proportion of cold-pressor pain tolerant individuals as a function of the proportion of friends who are pain tolerant. The probability of being pain tolerant seems to be an increasing function of the proportion of friends who are pain tolerant.

We compute Kendall's τ for the continuous cold-pressor pain tolerances, and obtain a value of $\tau = 0.13$, indicating a positive correlation between an individual's pain tolerance and the individual's friends' pain tolerance. There is thus a tendency for friendships among individuals with similar pain tolerance. In order to assess whether this tendency is significant, we perform a permutation test, resulting in τ estimates in the range (−0.072, 0.067). Thus, our observed correlation is much higher than would be expected merely by chance, with a p-value less than 0.00004.

Because sex is assortative in social networks, sex is a potential confounding variable for the association between friendship ties and pain tolerance, since it is also correlated with pain tolerance [21]. The estimated partial correlation coefficient of cold-pressor pain tolerance and friendship ties, adjusted for sex, is 0.12. The permutation test for the partial correlation results in values in the range (−0.061, 0.074), and thus there is still significant correlation in friends' pain tolerance after controlling for sex.

We also fit a network autocorrelation model for cold-pressor pain tolerance. The estimated coefficients in a model that only includes the autocorrelation term and an intercept are given in Table 2. The estimated

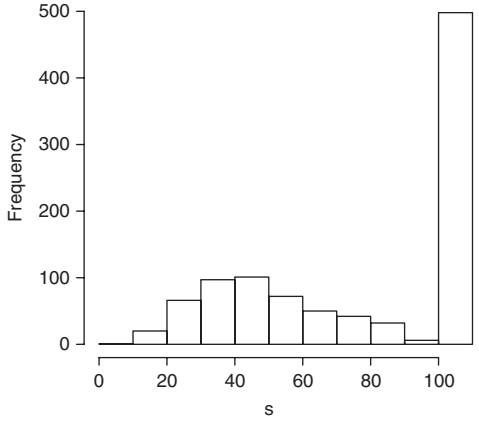

Fig. 1: Histogram of the cold-pressor pain tolerance. Observed pain tolerance in seconds in the Tromsø Study: Fit Futures I.

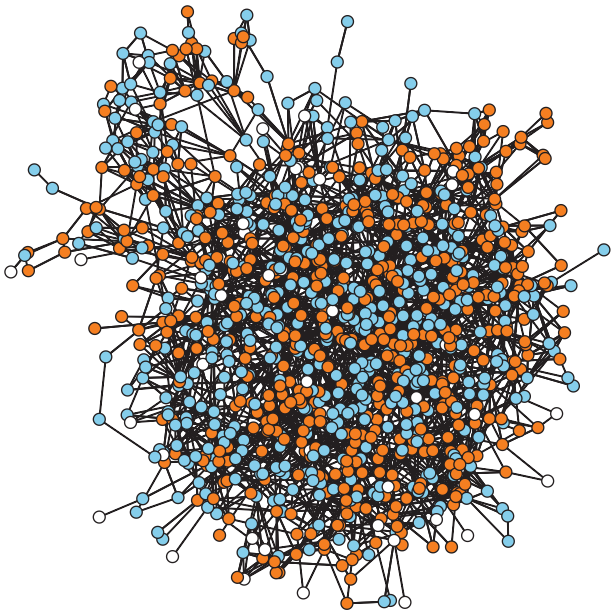

Fig. 2: Friendship network. Social network of the adolescents in the first year of upper secondary school in Tromsø in the Tromsø Study: Fit Futures I, where the nodes are coloured by cold-pressor pain tolerance. Orange: pain tolerant. Blue: pain sensitive. White: not measured (non-participants or not tested). The isolates (i.e. individuals with no friendship ties) have been removed to reduce visual clutter. Out of the isolates, there were three pain sensitive individuals and two pain tolerant individuals.

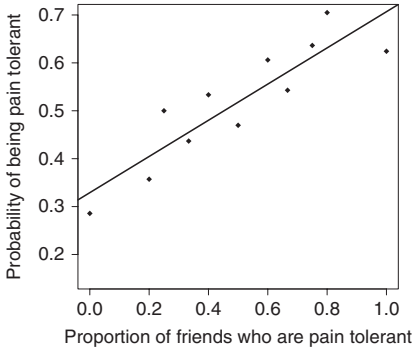

Fig. 3: Cold-pressor pain tolerance versus friends. Proportion of cold-pressor pain tolerant individuals as a function of the proportion of their friends who are pain tolerant in the Tromsø Study: Fit Futures I. The best linear fit to the points is also plotted.

Table 2: Fitted network autocorrelation model.

	ρ	Intercept
Estimate	0.44	55.45
Std. dev	0.062	6.19
p-Value	$1.40 \cdot 10^{-12}$	$<2.0 \cdot 10^{-16}$

Estimated coefficients in the network autocorrelation model using data from the Tromsø Study: Fit Futures I, standard deviation (Std. dev) and p-values.

autocorrelation coefficient ρ is highly significant. The estimated size of the effect is that an increase in the average pain tolerance of friends by 1 s increases the expected pain tolerance of the individual by 0.44 s.

In addition, we fit a network autocorrelation model where we control for sex, age and school programme (vocational versus general studies). The resulting coefficients are given in Table 3. The adjusted autocorrelation coefficient ρ is significant and positive. Sex and school programme are highly significant, while age is not, which is not so surprising since most of the adolescents are of same age (15–17 years old). The estimated effect is that by increasing the average pain tolerance of an individual's friends by 1 s, the expected pain tolerance of the individual increases by 0.21 s. For example, consider two individuals with the same sex, age and school programme, but the friends of the first individual have an average pain tolerance of 30 s, whereas the friends of the second individual have an average pain tolerance of 90 s. The expected difference in pain tolerance for the two individuals is then $0.21 \times 60\ \text{s} = 12.6\ \text{s}$.

3.1.2 Popularity

In order to examine whether the popular individuals have higher cold-pressor pain tolerance, we fitted network autocorrelation models with different network centrality measures of the nodes as covariates. The out-degree of node i is the number of links starting at node i, thus the number of friends nominated by individual i (truncated at five, since the individuals could only nominate up to five friends). The in-degree of node i is the number of individuals who nominated individual i as a friend, hence this is the number of individuals who consider individual i as one of their (likely) top five friends. We also examine whether closeness (how close the individual is to the other individuals in the network), betweenness (a measure of to which extent an individual lies between other individuals) and eigenvector centrality (a measure of the influence of a node in a network), see for instance [22] for details about these measures, have an effect on pain tolerance. The in-degree distribution and out-degree distribution of the network are given in Fig. 4. We fitted one network autocorrelation model for each centrality measure and sex, age and school programme were included in all the models. We found no statistically significant relationship between any of the centrality measures and pain tolerance. However, there was a statistical trend for the effect of out-degree on pain tolerance (estimate 1.79, standard deviation 1.04, p-value 0.084), suggesting that individuals who report having more friends may be somewhat more pain tolerant.

Table 3: Fitted network autocorrelation model, controlling for sex, age and school programme.

	ρ	Sex	Age	School p	Intercept
Estimate	0.21	18.44	−1.19	32.58	72.22
Std. dev	0.073	2.54	0.95	3.29	18.44
p-Value	0.0049	$3.61 \cdot 10^{-13}$	0.21	$<2.0 \cdot 10^{-16}$	$9.08 \cdot 10^{-5}$

Estimated coefficients in the network autocorrelation model using data from the Tromsø Study: Fit Futures I, standard deviation (Std. dev) and p-values. School p. denotes school programme.

3.1.3 Lifestyle factors

We know that lifestyle factors are assortative in social networks. Since lifestyle factors can have an effect on pain tolerance, they constitute possible confounders for the assortativity of pain tolerance in social networks. We therefore control for lifestyle factors in the analysis, by controlling for physical activity and smoking.

The analysis is performed on the subset of the data for whom we have complete information on lifestyle factors.

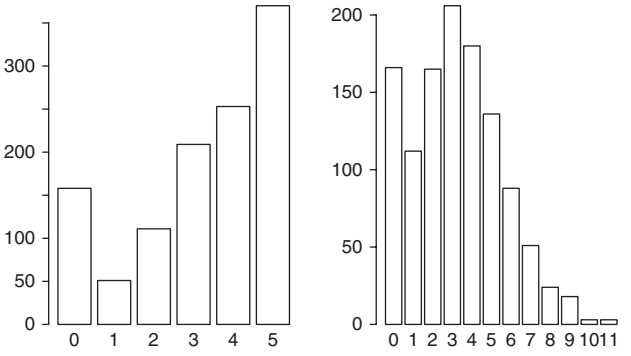

Fig. 4: Degree distributions. Degree distributions for the adolescent network in the Tromsø Study: Fit Futures I. Left: out-degree. Right: in-degree.

We first refit the network autocorrelation model without lifestyle factors on this subnetwork, in order to compare the fitted model with lifestyle factors on the same subjects. The estimated coefficients are given in Table 4. The estimated coefficients for the network autocorrelation model where we control for lifestyle factors are given in Table 5. We see that smoking has a significant negative effect on pain tolerance, while physical activity has a significant and positive effect. However, controlling for these factors has a minimal effect on the autocorrelation coefficient, and it is still significant. Hence, lifestyle similarities between friends do not explain the association between friendship ties and pain tolerance.

3.1.4 Sex differences in the autocorrelation

We investigate whether the autocorrelation coefficient, ρ, is sex dependent, so that the effect of friends on girls' pain tolerance (ρ_g) can differ from the effect of friends on boys' pain tolerance (ρ_b). The estimated coefficients for this model are given in Table 6. We see that ρ_g (girls) is not significant. The estimated effect for boys is larger than the estimated effect for girls, and it is significant. We thus find a significant correlation between boys and their friends in pain tolerance, but not for girls.

We also investigate whether same-sex friendships have a different effect than different-sex friendships. In the friendship network, there are 3,111 same-sex friendships and 582 different-sex friendships. We thus have one autocorrelation coefficient for the friendship effect of female friends on females ($\rho_{g,ss}$), one for male friends on males ($\rho_{b,ss}$), one effect for male friends on females ($\rho_{g,ds}$) and one effect for female friends on males ($\rho_{b,ds}$).

The estimated coefficients for the model are given in Table 7. The only significant effect is the effect of same-sex friendships for boys ($\rho_{b,ss}$), that is, the effect of boys on boys. Neither the effect of girls on girls nor the effect of boys on girls were significant.

Table 4: Fitted network autocorrelation model, controlling for sex, age and school programme, on the individuals for whom we have lifestyle information.

	ρ	Sex	Age	School p.	Intercept
Estimate	0.24	17.31	−1.36	31.40	73.32
Std. dev	0.072	2.50	0.87	3.24	17.39
p-Value	0.0010	$4.62 \cdot 10^{-12}$	0.12	$<2.0 \cdot 10^{-16}$	$2.47 \cdot 10^{-5}$

Estimated coefficients in the network autocorrelation model using data from the Tromsø Study: Fit Futures I, standard deviation (Std. dev) and *p*-values. School p. denotes school programme.

Table 6: Fitted network autocorrelation model with sex-dependent ρ, controlling for sex, age, school programme, smoking and physical activity.

	ρ_g	ρ_b	Sex	Age
Estimate	0.15	0.27	5.30	−1.36
Std. dev	0.10	0.089	11.80	1.70
p-Value	0.16	0.0022	0.65	0.42
	School p	**Smoke**	**Phys.act**	**Intercept**
Estimate	27.80	−5.62	1.92	79.09
Std. dev	3.40	2.28	0.64	33.47
p-Value	$2.22 \cdot 10^{-16}$	0.14	0.0026	0.018

Estimated coefficients in the network autocorrelation model using data from the Tromsø Study: Fit Futures I, standard deviation (std. dev) and *p*-values. School p. denotes school programme and phys. act denotes physical activity. ρ_g is the effect of friends on girls' pain tolerance and ρ_b is the effect of friends on boys' pain tolerance.

3.1.5 Test sequence

One possible explanation for the effect of friends' pain tolerance on the individual's pain tolerance is the competition between peers. The adolescents are tested consecutively, so they are able to brag to their peers about being able to endure the maximum testing time. This might induce peer pressure, where others also want to perform as well. Thus, we hypothesise that if test sequence has an effect on pain tolerance, then the effect will be positive, so that those who are tested later have a higher pain

Table 5: Fitted network autocorrelation model, controlling for sex, age, school programme, physical activity and smoking.

	ρ	Sex	Age	School p.	Phys. act	Smoke	Intercept
Estimate	0.22	17.00	−1.27	27.75	1.87	−5.29	71.53
Std. dev	0.074	2.53	1.44	3.32	0.62	2.28	26.75
p-Value	0.0035	$1.94 \cdot 10^{-11}$	0.38	$<2.0 \cdot 10^{-16}$	0.0028	0.020	0.0075

Estimated coefficients in the network autocorrelation model using data from the Tromsø Study: Fit Futures I, standard deviation (Std. dev) and *p*-values. School p. denotes school programme and phys. act denotes physical activity.

Table 7: Fitted network autocorrelation model with sex-dependent ρ, and separating the effect of same-sex and the effect of different-sex, controlling for sex, age, school programme, smoking and physical activity.

	$\rho_{g,ss}$	$\rho_{b,ss}$	$\rho_{g,ds}$	$\rho_{b,ds}$	Sex
Estimate	0.12	0.25	0.065	−0.030	7.70
Std. dev	0.10	0.077	0.048	0.055	11.39
p-Value	0.27	0.0012	0.17	0.59	0.50
	Age	**School p.**	**Smoke**	**Phys. act**	**Intercept**
Estimate	−1.38	29.12	−5.71	1.84	79.73
Std. dev	1.97	3.45	2.36	0.64	37.37
p-Value	0.48	$<2.0 \cdot 10^{-16}$	0.016	0.0041	0.032

Estimated coefficients in the network autocorrelation model using data from the Tromsø Study: Fit Futures I, standard deviation (Std. dev) and *p*-values. School p. denotes school programme and phys. act denotes physical activity. $\rho_{g,ss}$ is the effect that female friends have on a female, $\rho_{b,ss}$ is the effect that male friends have on a male, $\rho_{g,ds}$ is the effect of male friends on a female and $\rho_{b,ds}$ is the effect that female friends have on a male.

tolerance. We therefore control for test sequence in the network autocorrelation model. The results are given in the Supplementary Materials. We find that test sequence has a positive and significant effect. When controlling for test sequence, the estimated correlation coefficient ρ decreases, but it is still significant.

It is also possible that the test sequence effect is larger among friends. We therefore also fit a network autocorrelation model, where we control for the test order among friends. We find that the test order among friends does not have a significant effect on the pain tolerance of the individual (see Supplementary Materials). In addition, including this covariate does not seem to affect the network autocorrelation coefficient. Hence, there seems to be an effect of the number of adolescents who have been tested before the individual, and it is not confined to friends. However, even though controlling for this induced peer pressure effect reduces the estimated effect of friends' average pain tolerance on the individual's pain tolerance, there is still a significant autocorrelation effect.

3.2 Other pain modalities

In order to study whether pain tolerance is also assortative for other pain tolerance measures, we repeated the analysis for heat pain tolerance, pressure pain tolerance at the fingernail and pressure pain tolerance at the trapezius muscle. The full results are given in the Supplementary Materials. We find a positive and significant pairwise correlation between friendship ties and pain tolerance for all pain modalities, also after controlling for sex. Considering network autocorrelation effects, we find that for pressure pain tolerance at the fingernail, there is a significant effect of friends' pressure pain tolerance on the individual's pressure pain tolerance. For pressure pain tolerance at the trapezius muscle, the effect is not significant, but borderline significant on the 0.05 level, indicating that there is possibly an effect. As an exception, we find no indication of an effect of the friends' average pain tolerance on the individual's heat pain tolerance.

3.2.1 Pain threshold

We also study the friendship effect on pain threshold. When considering the pairwise correlations of friendship ties and pain threshold, there is a significant effect for heat pain threshold and pressure pain threshold at the fingernail, but no effect for pressure pain threshold at the trapezius muscle. For the network autocorrelation effects, there is a significant effect of the average pain threshold of friends on the individual's pain threshold for pressure pain threshold at the fingernail, but not for heat pain threshold and pressure pain threshold at the trapezius muscle.

3.2.2 Popularity

We examined whether popularity or centrality had a positive effect on pain tolerance or pain threshold, for heat pain, pressure pain at the fingernail and pressure pain at the trapezius muscle. For heat pain tolerance and heat pain threshold, the in-degree has a significant effect. For pressure pain tolerance at the trapezius muscle, none of the centrality measures are significant, but out-degree is borderline significant with a *p*-value slightly above 0.05 (0.051).

4 Discussion

We studied the effect of friendship ties on cold-pressor pain tolerance in a group of upper secondary school students, and found that their social network is assortative in terms of pain tolerance. Hence, there is a significant tendency for pupils with high pain tolerance to be friends with other pupils with high pain tolerance, and vice versa.

Investigating further how this effect depends on the sex of the individual, and whether the friendships are

same-sex friendships or different-sex friendships, we find that the only significant effect is the effect that male friends have on males. We find no significant effect of friends for girls on pain tolerance in stratified analyses.

Considering other pain modalities, we found that even though the relationship was most significant for cold-pressor pain tolerance, it is clear that there is a positive association between friends and pain tolerance. For pain threshold, we found that the relationship between friendship ties and pain threshold was not as clear as the relationship between friendship ties and pain tolerance.

There are two possible reasons why we find a significant and positive correlation between individuals and their friends in pain tolerance. Either individuals select friends based on similarity in pain tolerance (homophily), or it could be due to friends influencing each other so that one becomes more similar to one's friends (social transmission). If the latter is true, this would provide an explanation for some of the unique environmental contribution to pain, which is not explained by genes and family environment. Compelling evidence of both homophily and social transmission has been reported for several other phenotypes, including obesity, alcohol consumption and smoking [11–14]. To the extent that these phenotypes influence pain tolerance, it is quite plausible that social transmission of pain tolerance could occur as a downstream effect.

Pain tolerance is not a directly observable trait and, in our experience, research subjects are frequently surprised by their own response to the cold-pressor test. The direct selection of friends based on their pain tolerance therefore seems unlikely. However, there is good reason to believe that such selection might occur indirectly, for instance by individuals selecting friends with similar lifestyle, or other characteristics that are also associated with pain tolerance. In other words, the lifestyle arguments that hold for social transmission also hold for homophily. Therefore, it is reasonable to assume that at least part of the observed assortativity is due to homophily, and not explained by social transmission alone.

We have partially controlled for lifestyle by controlling for smoking and physical activity, as a substitute for lifestyle. Though we do have information on multiple lifestyle factors (smoking, snuff use, alcohol consumption, obesity and physical activity), we do not include all of these lifestyle factors in the analysis, because they are highly correlated with each other. To avoid multicollinearity issues, we only include smoking and physical activity. Since the lifestyle covariates are highly correlated, the effect is likely to be similar for other lifestyle factors. Though such adjustments reduce the effect of friends' pain tolerance on the individual slightly, it is still significant. However, there may be other factors which we have not controlled for, which are both associated with pain tolerance and friendship ties. One potential candidate is socioeconomic status. We did try to control for education of parents as a substitute for socioeconomic status. We found no significant effect of the education of parents on the pain tolerance of the individual, and the friendship effect did not decrease when controlling for the education of the parents (results not shown). The fact that the effect that boys have on boys is the only significant effect in stratified analysis, indicates that there is possibly a "macho effect" present, where "tough" boys tend to be friends with other "tough" boys and vice versa, or that boys influence each other's pain tolerance, perhaps as a peer pressure effect. Here it is worth noting that participants may report to each other whether or not they endured the full time for the cold-pressor test, thus inducing peer pressure on friends tested later. We controlled for test sequence, and found that it had a positive, significant effect. Hence, adolescents who are tested later have, on average, higher pain tolerance than adolescents tested earlier. Controlling for test sequence does decrease the effect of friends' pain tolerance on the individual's pain tolerance, but it does not explain all of the friendship effect. The friendship effect is still significant after controlling for test sequence.

For cold-pressor pain tolerance, there was no significant effect of any of the centrality measures, but the estimated effect of out-degree (the number of nominated friends by the individual) showed an indication of a possible effect. Out-degree was also borderline significant for pressure pain tolerance at the trapezius muscle, and in-degree (the number who nominated the individual as their friend) had a significant effect on heat pain tolerance and heat pain threshold. Thus, there is an indication of more central individuals being more pain tolerant, but we did not find significant evidence for this. This supports (but does not confirm) previously published results indicating that pain tolerance is higher for individuals with more friends [23].

One limitation of the study is the fact that we use the pseudo likelihood approximation and not the full likelihood, to estimate the parameters in the network autocorrelation models. The reason why we use the pseudo likelihood approximation is due to the censoring. An alternative to using the pseudo likelihood approximation is to use the dichotomised pain tolerance outcomes instead of the continuous measurements. We have chosen not to do this, due to the information loss related to dichotomising continuous variables [24]. Another limitation of the study is that we do not have the full social network

for the individuals, since they could only nominate up to five friends, all within the upper secondary school in the Tromsø area. In addition, the adolescents were not asked to name their five closest friends, but they were asked to name the five individuals with whom they had spent most time the preceding week. This does not necessarily coincide with the top five friends, causing uncertainty in the strength of these friendship ties. The friendships are also not weighted, so we cannot distinguish strong friendships from weak ones. Another source of error in the friendship network is the fact that the network is self-reported. Last, and most importantly, without longitudinal information, we are unable to determine to what the degree pain tolerance is assortative due to homophily (similar people becoming friends) or due to social transmission.

Authors' statements

Research funding: None declared.

Conflict of interest: We have no conflicts of interest to declare.

Informed consent: All participants provided written informed consent before inclusion. For participants younger than 16 years, written or oral consent from one parent was also obtained. In cases of oral consent, two study nurses provided confirmation that oral consent from a parent was obtained, in accordance with procedures laid out in the ethics approval.

Ethical approval: Data collection for TFF1 was approved by the Norwegian Data Protection Authority, and the Regional Committee for Medical and Health Research Ethics of Norway, Northern health region. The study procedures were conducted in accordance with the Declaration of Helsinki.

References

[1] Nielsen CS, Knudsen GP, Steingrimsdottir OA. Twin studies of pain. Clin Genet 2012;82:331–40.
[2] Plomin R, Daniels D. Why are children in the same family so different from one another? Int J Epidemiol 2011;40:563–82.
[3] Nielsen CS, Stubhaug A, Price DD, Vassend O, Czajkowski N, Harris JR. Individual differences in pain sensitivity: genetic and environmental contributions. Pain 2008;136:21–9.
[4] Trost Z, Strachan E, Sullivan M, Vervoort T, Avery AR, Afari N. Heritability of pain catastrophizing and associations with experimental pain outcomes: a twin study. Pain 2015;156:514–20.
[5] Angst MS, Phillips NG, Drover DR, Tingle M, Ray A, Swan GE, Lazzeroni LC, Clark JD. Pain sensitivity and opioid analgesia: a pharmacogenomic twin study. Pain 2012;153:1397–409.
[6] Norbury TA, MacGregor AJ, Urwin J, Spector TD, McMahon SB. Heritability of responses to painful stimuli in women: a classical twin study. Brain 2007;130:3041–9.
[7] Nan C, Guo B, Warner C, Fowler T, Barrett T, Boomsma D, Nelson T, Whitfield K, Beunen G, Thomis M, Maes HH, Derom C, Ordoñana J, Deeks J, Zeegers M. Heritability of body mass index in pre-adolescence, young adulthood and late adulthood. Eur J Epidemiol 2012;27:247–53.
[8] Newman ME. Assortative mixing in networks. Phys Rev Lett 2002;89:208701.
[9] McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Sociol 2001;27:415–44.
[10] Cacioppo JT, Fowler JH, Christakis NA. Alone in the crowd: the structure and spread of loneliness in a large social network. J Pers Soc Psychol 2009;97:977–11.
[11] Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008;358:2249–58.
[12] Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. Ann Intern Med 2010;152:426–33.
[13] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007;357:370–9.
[14] Steglich C, Snijders TAB, Pearson M. Dynamic networks and behavior: separating selection from influence. Sociol Methodol 2010;40:329–93.
[15] Fowler JH, Settle JE, Christakis NA. Correlated genotypes in friendship networks. Proc Natl Acad Sci 2011;108:1993–7.
[16] Helsel DR. Statistics for censored environmental data using Minitab and R, 2nd ed. 77. New Jersey: John Wiley & Sons, 2011.
[17] O'Malley J, Marsden PV. The analysis of social networks. Health Serv Outcomes Res Methodol 2008;8:222–69.
[18] Besag J. Statistical analysis of non-lattice data. Statistician 1975:179–95.
[19] Kolaczyk ED, Csárdi G. Statistical analysis of network data with R. 65. New York: Springer-Verlag, 2014:145.
[20] Snijders T, Borgatti SP. Non-parametric standard errors and tests for network statistics. Connect 1999;22:161–70.
[21] Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL. Sex, gender and pain: a review of recent clinical and experimental findings. J Pain 2009;10:447–85.
[22] Kolaczyk ED. Statistical analysis of network data: methods and models. New York: Springer-Verlag, 2009.
[23] Johnson KV-A, Dunbar RI. Pain tolerance predicts human social network size. Sci Rep 2016;6. Doi: 10.1038/srep25267.
[24] Altman DG, Royston P. The cost of dicotomising continuous variables. BMJ 2006;332:1080.

Supplemental Material: The online version of this article offers supplementary material (https://doi.org/10.1515/sjpain-2018-0060).

Paper II

A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions

II

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions

Solveig Engebretsen[1,2]*, **Kenth Engø-Monsen**[3], **Arnoldo Frigessi**[1,4], **Birgitte Freiesleben de Blasio**[1,2]

1 Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway, **2** Department of Infectious Disease Epidemiology and Modelling, Division for Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway, **3** Telenor Research, Telenor Group, Fornebu, Norway, **4** Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

* solveig.engebretsen@medisin.uio.no

Citation: Engebretsen S, Engø-Monsen K, Frigessi A, Freiesleben de Blasio B (2019) A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions. PLoS Comput Biol 15(3): e1006879. https://doi.org/10.1371/journal.pcbi.1006879

Editor: Adam J Kucharski, London School of Hygiene & Tropical Medicine, UNITED KINGDOM

Received: September 6, 2018

Accepted: February 18, 2019

Published: March 7, 2019

Data Availability Statement: The population size data for Norway are available at Statistics Norway (http://www.ssb.no/en/statbank/table/11342/). The commuting data for Norway are available at Statistics Norway (https://www.ssb.no/statbank/table/03321). The population sizes for Iceland are available at Statistics Iceland (https://px.hagstofa.is/pxis/pxweb/is/Ibuar/Ibuar__mannfjoldi__2_byggdir__Byggdakjarnarhverfi/MAN03200.px/). Population data for the UK are available at the Office for National Statistics GB (https://www.ons.gov.uk/peoplepopulationandcommunity/population

Abstract

The world is continuously urbanising, resulting in clusters of densely populated urban areas and more sparsely populated rural areas. We propose a method for generating spatial fields with controllable levels of clustering of the population. We build a synthetic country, and use this method to generate versions of the country with different clustering levels. Combined with a metapopulation model for infectious disease spread, this allows us to in silico explore how urbanisation affects infectious disease spread. In a baseline scenario with no interventions, the underlying population clustering seems to have little effect on the final size and timing of the epidemic. Under within-country restrictions on non-commuting travel, the final size decreases with increased population clustering. The effect of travel restrictions on reducing the final size is larger with higher clustering. The reduction is larger in the more rural areas. Within-country travel restrictions delay the epidemic, and the delay is largest for lower clustering levels. We implemented three different vaccination strategies—uniform vaccination (in space), preferentially vaccinating urban locations and preferentially vaccinating rural locations. The urban and uniform vaccination strategies were most effective in reducing the final size, while the rural vaccination strategy was clearly inferior. Visual inspection of some European countries shows that many countries already have high population clustering. In the future, they will likely become even more clustered. Hence, according to our model, within-country travel restrictions are likely to be less and less effective in delaying epidemics, while they will be more effective in decreasing final sizes. In addition, to minimise final sizes, it is important not to neglect urban locations when distributing vaccines. To our knowledge, this is the first study to systematically investigate the effect of urbanisation on infectious disease spread and in particular, to examine effectiveness of prevention measures as a function of urbanisation.

andmigration/populationestimates/datasets/populationestimatesforukengland andwalesscotlandandnorthernireland).

Funding: SE and AF acknowledge partial funding from the Norwegian Research Council centre BigInsight project 237718. The funders had no role in study resign, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

We study the interplay between urbanisation and infectious disease spread. As part of the worldwide urbanisation process, people are continuously moving to urban areas, and the cities are growing in size. This causes clusters of areas with high population density and clusters of areas with low population density, which is what we call population clustering. By simulating infectious disease spread in a synthetic country where we vary this population clustering, we explore the consequences of urbanisation on infectious disease spread. Our qualitative results have direct implications for infectious disease control guidelines and policies. We find that implementing internal travel restrictions have greater impact on the final number ill in the most urbanised countries than in the less urbanised countries. The effect is largest in the more rural parts of the country. According to our model, travel restrictions are more effective in delaying the epidemic in the less urbanised countries than in the more urbanised countries. We investigate vaccination strategies, where locations are targeted depending on how urban or rural they are. We find that it is important to vaccinate the urban locations—if the most urban locations are not covered by the vaccine, the final number ill will be a lot larger.

Introduction

We are living in a world which is continuously urbanising. From a United Nations report [1], we know that in 2014, 54% of the population was living in urban areas. By 2050, they estimate that 66% of the population will be living in urban areas. The number of "megacities" is also increasing [1]. Urbanisation involves clustering of people within a geographical area [2]. Migration from rural to urban areas is one of the key drivers of urbanisation, and results in spatial expansion of urban centers [3]. Though there are also other characteristics of urbanisation, like urban sprawl [2, 3] and population growth, we here focus on this population clustering. By population clustering, we mean that large cities are often surrounded by other cities or suburbs with large population sizes. Rural areas also tend to appear in clusters (i.e. positioned closely together). This has for instance been found to be the case for the Turku region in Finland, where regions of both high and low population densities are clustered [4]. Another example is Australia, where the majority of the population is clustered around the coastal belt [5]. From here on, the term urbanisation will be used to refer to the population clustering.

We aim at studying the urbanisation phenomenon and its effect on infectious disease spread in a general setting. Our purpose is not to describe single outbreaks in specific populations, or finding the "best model" or strategy for a specific country, but rather study the phenomenon from a more generic and principled point of view.

In this paper, we explore the effect of internal travel restrictions and vaccination on infectious disease spread, when the clustering of population is continuously varied. By internal travel restrictions, we will refer to restrictions on non-commuting travel within the country. If the disease dynamics is different for different levels of population clustering, this can have important implications for the effectiveness and design of interventions. In order to study this, we need a continuous series of countries where everything is fixed, apart from the urbanisation, which changes between the countries in a controllable and continuous manner. This is difficult in practice, and we therefore construct a fictional country for this purpose, however trying to maintain some realism. Our aim is not to develop a precise model for urbanisation in a country. We aim for a simple model, where urbanisation is controlled by one single tuning parameter, which captures key features of urbanisation. We use this model to generate a

fictional country with a plausible population distributed in urban and rural areas. More specifically, we use a plausible population size distribution, plausible commuting patterns modelled by a gravity law and a plausible infectious disease model. We use a metapopulation infectious disease model, where an SEIR (susceptible, exposed, infectious, recovered) process [6] governs the disease dynamics in each location, and the different locations are coupled through individuals travelling between them. This framework allows us to investigate in silico how urbanisation affects various aspects of infectious disease spread. As a motivating example, we will study an influenza-like illness spreading in a single country, where we assume that the pathogen has already been imported to the country. We investigate how the infectious disease dynamics depends on the underlying population clustering in the country and focus on the effect of internal travel restrictions and three different vaccination strategies.

The plausibility of the synthetic country is obtained by conserving the population size distribution of Norway and a gravity law fitted to data on commuting between Norwegian municipalities. We use Norwegian data to ensure reasonable population sizes, and plausible commuting for those population sizes, and because we have commuting data available on a relatively fine spatial scale. We are studying the urbanisation phenomenon generically, and we do not claim, nor is it our aim, that these results are directly applicable to Norway. This framework is a simulator, and it is not our purpose to model and provide results for a specific country. Our results allow a theoretical description of how urbanisation affects interventions to control epidemics.

To our knowledge, this is the first modelling and simulation study which systematically investigates the effect of population clustering on the spread of infectious diseases. In particular, it is the first study to consider the effect of internal travel restrictions and vaccination in relation to urbanisation. This is a study of the urbanisation phenomenon represented in a very simplified and theoretical way, yet with important elements of realism. We focus on these two control strategies, because they are clearly affected by urbanisation. We do not consider international air travel restrictions, school closure or other sanitary measures, which were used for instance during the 2009 pandemic [7, 8]. Travel restrictions have a long history, and date back at least to the 14th century, where people were prevented from leaving or entering specific communities during the plague epidemics [9]. Travel bans were also used in many cities and countries during the 1918-1919 influenza epidemic [9]. More recently, internal travel restrictions were used as a mitigation measure against influenza virus transmission during the 2009 pandemic in Mongolia, through interruption of provincial rail and road travel [10], and during the recent Ebola outbreaks [11]. There has been some work on infectious disease spread and urbanisation focussing on the improved health conditions in urban areas compared to rural areas. For the developed countries, health has overall improved with increased urbanisation [12]. In low-income settings, health conditions are on average better in urban areas than in rural areas, but there are also significant challenges relating to inequities and heterogeneity in health among the urban population (favelas, slums, etc.). High population density increases exposure to infectious diseases [12]. In Africa, the urban population has better nutritional status, fewer morbid events, increased vaccine coverage and better access to healthcare services compared to the rural population, and have reduced levels of malaria transmission and other severe diseases [13]. There is also one study, [14], which develops a model for the effect of urbanisation on the transmission of infectious diseases, focussing on population growth and land use development. However, the infectious disease spread model is very simple. In addition, they do not consider the effect of interventions.

The effectiveness of both vaccination and internal travel restrictions on mitigating an infectious disease have been studied in various settings, e.g. [9, 10, 15–20]. Germann et al. [15] study the spread of a hypothetical pandemic influenza, with a basic reproductive number in

the range 1.6-2.4, in the United States, and find that (domestic) travel restrictions do not have an effect on the final size of the epidemic, but might be able to slightly delay the time course. They also find that vaccination drastically reduces the final number of cases and delays the spread. Ferguson et al. [16] find that reducing long distance travel within the United States (domestic air travel) only slightly delays the influenza epidemic, for a hypothetical pandemic influenza strain with varying transmissibility. They consider vaccination in both the United Kingdom and the United States, and find that vaccination significantly reduces the final size of the epidemic. In accordance with these studies, the US Department of Health and Human Services also claims that vaccination is the most effective way of preventing the public health impact of (pandemic) influenza [21]. In a review study, Mateus et al. [17] find that domestic travel restrictions can delay the influenza epidemics by one week, and that extensive travel restrictions can reduce the final size of the epidemic. Both seasonal and pandemic influenza strains are considered. They also find that travel restrictions have minimal impact in urban centers with dense population and high mobility. Brownstein et al. [18] consider the influenza epidemic following the travel ban after 9/11. They claim that the decrease in air traffic in the United States caused a delayed and prolonged influenza season (however, note also the rebuttal in [22]). For a hypothetical influenza strain in Korea, it was found that 50% internal travel restrictions delayed the peak timing and had a slight reduction effect on the peak [19]. Some work has also been done on mathematical explanations and expressions for delay in epidemic spreading due to travel restrictions, for border control on international travel [23], and for more general mobility networks [7, 24].

During the 2009 influenza pandemic, uniform vaccination guidelines were given (i.e. pro rata), for instance in Norway [25], Ireland [26] and the United States [27, 28], and the default vaccination strategies are usually uniform [29]. However, the effectiveness of the different vaccination strategies and which strategy is best in terms of minimising attack rate, possibly depend on the population clustering of the underlying country. We thus simulate the epidemic with three different vaccination strategies—uniform vaccination, prioritising urban locations and prioritising rural locations. This allows us to compare the three vaccination strategies as a function of population clustering and provides us with a better understanding and possibilities for refined vaccination strategies. There are numerous examples of spatially targeted vaccination and antiviral strategies in the literature [29–39]. Some of these are based on dynamic optimisation strategies [29, 31, 33, 39], while others are based on prioritising locations with high prevalence (i.e. geographically targeting hotspots) [30, 34]. In [38], pro rata vaccination strategies are compared to vaccination strategies prioritising locations sequentially by population size (among other vaccination strategies). This is similar to the vaccination strategy we investigate, targeting urban and/or rural locations.

We investigate how domestic travel restrictions and vaccination affects the epidemic timing, spread and final size for the various levels of population clustering. For policy planners, the timing of peak dates and initial dates are important because they indicate how much time there is to implement interventions and preventive measures. If peak dates for spatially proximate regions are close in time, the efficiency of the health care organisation is challenged. This is extra problematic if there is in addition clustering of high peak incidences, since that would imply that spatially proximate locations have high disease activity at the same time. We further disentangle the effect of travel restrictions on final size of the epidemic by examining how the effect depends on how urban or rural the location is.

We first describe the simulation set-up with details on the properties of the fictional country, the clustering algorithm, the disease dynamics model, the travel restrictions and vaccination strategies. We then use this tool to simulate the infectious disease spread for the various levels of clustering, investigate the effect of various amounts of internal travel restrictions,

simulate and investigate the effect of the three vaccination strategies and finally a combination of travel restrictions and vaccination. We end with discussion and concluding remarks.

Models

In order to investigate how population clustering affects infectious disease dynamics, we build a series of countries with varying population clustering, where everything else is fixed. Imagine something like taking one country and reshuffling its population, so to increase monotonically its level of urbanisation. We construct a fictional country, where the spatial clustering of population sizes is controlled by a design parameter. However, we try to conserve some realism in the different ingredients of the framework. The realism is obtained by using a population size distribution and commuting law fitted to data from Norwegian municipalities.

In this section, the different parts of the simulation framework will be described. We first describe how to generate the geographical areas, with a plausible population distribution. We then introduce the clustering algorithm that is used to generate different versions of the geographical area, with different levels of clustering. Then we describe the models for the two (coupled) dynamical processes for the geographical areas—the infectious disease process and the mobility process. Finally, we describe the interventions—the internal travel restrictions and the vaccination strategies.

Generating the geographical areas

We construct a country, where the spatial clustering of population sizes is controlled by a design parameter. The country is a square, consisting of many small block units.

The construction process consists of two steps. First, we generate the population sizes in the block units of the country by drawing a random sample from a reasonable population distribution. In the second step, we apply a clustering algorithm, generating different versions of the country with various levels of clustering. The clustering is done by rearranging the block units according to a mapping rule.

Population sizes. The block units are assigned populations which are drawn from a gamma distribution. We use 6561 block units with unit area 81 km^2, corresponding to a total area of approximately 530 000 km^2. We rescale our resulting population sizes to sum up to 6 000 000 (slightly larger than Norway). The data are from Statistics Norway for 2016, and we use the 428 municipalities, excluding Svalbard. We use the municipality scale, because this is the finest spatial scale for which we have both population size data and commuting data. The block units can be interpreted as a discretisation of administrative units, so that multiple block units can make up one administrative unit. We do not use the population sizes in the administrative units directly, because a finer scale is necessary for the clustering algorithm to generate smooth distributions. The number of block units is chosen as a trade off between a high resolution, and computational feasibility.

We fit a gamma distribution to the logarithm of the population sizes, since the distribution is clearly very skewed. We draw random realisations from this distribution, providing reasonable population sizes in the block units. The histogram of the population size for the municipalities of Norway and the fitted distribution are given in S1 Fig.

We fit distributions to normalised population data from various other countries, to assess similarity across countries. The fitted population distributions are given in Fig 1. For France, we use the 334 arrondissements [40], for Italy the 107 provinces [41], for Netherlands the 388 municipalities [42], for Denmark the 98 municipalities [43], for Iceland the 129 post codes (data for 2017 from Statistics Iceland) and for UK the population sizes in the 424 local authorities (Office for National Statistics GB, estimates for 2016). The population sizes in the

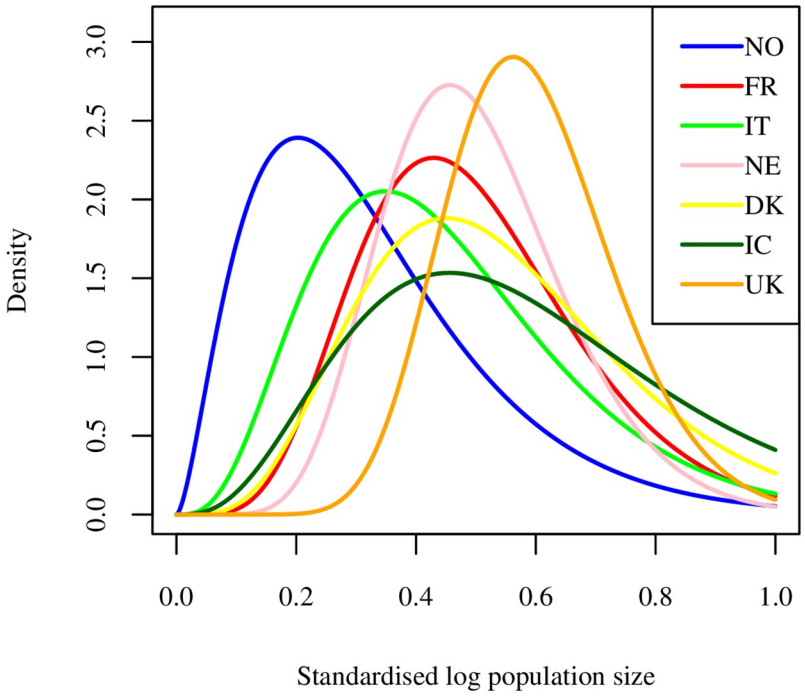

Fig 1. Population sizes. Population size distribution for Norway, France, Italy, Netherlands, Denmark, Iceland and United Kingdom, in different administrative units: 428 municipalities for Norway, 334 arrondissements for France, 107 provinces for Italy, 388 municipalities for the Netherlands, 98 municipalities for Denmark, 129 post codes for Iceland and 424 local authorities for the UK.

https://doi.org/10.1371/journal.pcbi.1006879.g001

Norwegian municipalities seem to be more heterogeneous than for the other countries, with a higher occurrence of smaller population sizes. The UK distribution seems most homogeneous. Noticeably, all the population distributions are skewed. Note that these administrative levels are not standardised across the different countries. Therefore comparison between countries must be treated with caution.

In some of the results, the locations are grouped according to population size. We let Q1 denote areas which have population size smaller than the 25% quantile, Q2 denote areas with population size between the 25% quantile and the median, Q3 denote areas with population size between the median and 75% quantile, and Q4 denote areas with population size larger than the 75% quantile. So Q1 are the most rural areas, while Q4 are the most urban areas in our simulations.

Clustering algorithm. The population clustering is done by rearranging the populated locations in space. The population sizes are fixed for all the levels of clustering.

In order to induce positive correlations between the neighbouring locations, we use a geostatistical model to generate a random spatial field where we control the correlations. We use a model with a Matérn covariance function with range parameter 5.0 and process variance 0.1, so that the covariance, C, between population sizes in two locations with distance d apart is [44], p. 126, section 4.1]

$$C(d,\kappa) = 0.1(2^{\kappa-1}\Gamma(\kappa))^{-1}(d/5.0)^{\kappa}K_{\kappa}(d/5.0),$$

where K_{κ} is a modified Bessel function of the second kind of order κ and the parameter κ is varied to control the smoothing of population sizes. We use the euclidean distance in block units. The larger κ, the stronger the tendency for block units of similar population size to

cluster together. This generates a smoother population size distribution in space across the country. We map the sampled population sizes to the random spatial field by respecting the corresponding order of the locations, so that the largest population size is mapped to the location of the largest number in the random spatial field, the second largest population size is mapped to the location of the second largest number in the random spatial field and so on. The resulting versions of the country for various values of κ are given in Fig 2. The version in the upper left of Fig 2 is the country without any clustering. We clearly see that population clustering increases with κ.

In order to assess what a reasonable range for κ is, we perform a visual comparison to some real countries (S2 Fig, S1 Text). Iceland seems to have the highest population clustering level (similar to $\kappa = 3.0$). Norway and France also seem to have a high population clustering,

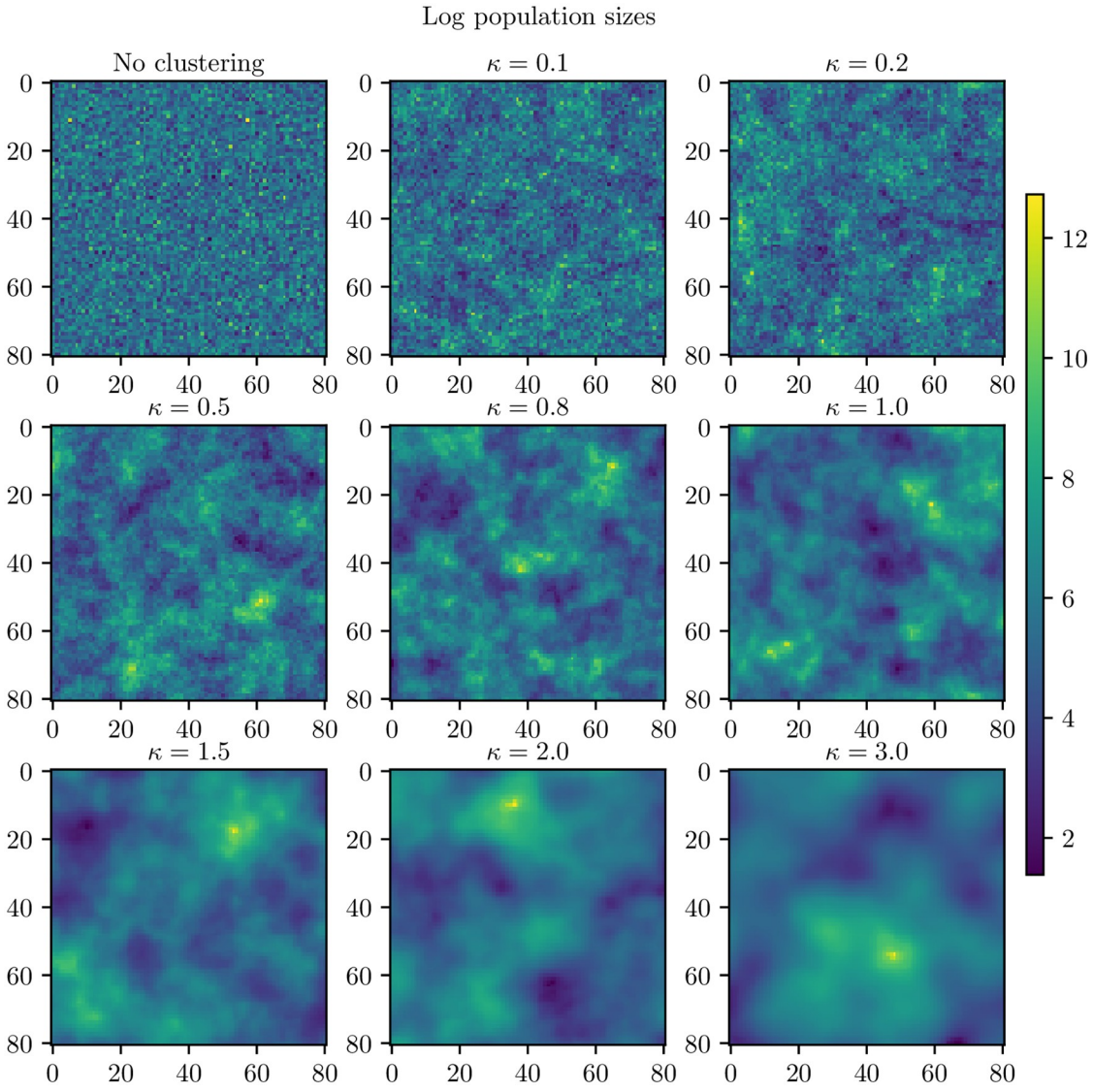

Fig 2. Clustering levels. Generated versions of the country for various κ. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$.

https://doi.org/10.1371/journal.pcbi.1006879.g002

similar to $\kappa = 1.5$ or $\kappa = 2.0$. The United Kingdom seems to be somewhere between $\kappa = 1.5$ and $\kappa = 1.0$. Germany seems to be somewhere between $\kappa = 0.5$ and $\kappa = 0.8$, while the Netherlands seem similar to something between $\kappa = 0.8$ and $\kappa = 1.0$. Similarities are found by visual inspection. A more formal estimation or assessment of κ is beyond the scope of this paper. The transformation that we use from the random spatial field to the resulting versions of the countries with different clustering levels is not linear, and it is thus not possible to formally estimate κ. However, we are not aiming at providing quantitative results for specific countries, nor is our framework the best choice for that purpose. We are interested in general trends when κ is varied, both qualitative and quantitative, in order to be able to make comparisons between different cases. However, the absolute measures (the peak dates, final sizes etc.) are not translatable to specific real settings. When a quantity like the non-infected area during an epidemic changes monotonically when the (synthetic) level of urbanisation increases, then we will hypothesise that urbanisation affects the quantity. When we do not discover a monotone effect, we will conclude that a dependency is less likely.

The disease dynamics model

The disease dynamics model is a metapopulation model which can be described as a network where every node represents a location, and every edge between locations represents people who travel between the locations and thus can spread the disease further. In every location there is a separate set of stochastic difference equations governing the local disease dynamics, but the processes are coupled through travellers. Similar disease dynamics models are used for instance in [45] and [46] for modelling the global spread of influenza-like illnesses.

Local infection dynamics. In every location unit, a stochastic SEIR model [6] is used to describe the infection dynamics. Let $S^i(t)$, $E^i(t)$, $I^i(t)$ and $I^i_a(t)$ denote the number of susceptible individuals, exposed individuals, symptomatic infectious individuals and asymptomatic infectious individuals at time t in location i, respectively. As in [45], we let the probability of being asymptomatic (given infectious) be 0.33 and the transmission probability of the asymptomatic infectious be reduced by 50%. The stochastic SEIR equations are given by:

$$S^i(t+\Delta t) = S^i(t) - \text{Binom}(S^i(t), \beta\Delta t I^i(t)/N_i + 0.5\beta\Delta t I^i_a(t)/N_i),$$

$$E^i(t+\Delta t) = E^i(t) + \text{Binom}(S^i(t), \beta\Delta t I^i(t)/N_i + 0.5\beta\Delta t I^i_a(t)/N_i) - \text{Multinom}(E^i(t), 0.33\lambda\Delta t, 0.67\lambda\Delta t),$$

$$I^i(t+\Delta t) = I^i(t) + \text{Binom}(E^i(t), 0.67\lambda\Delta t) - \text{Binom}(I^i(t), \gamma\Delta t),$$

$$I^i_a(t+\Delta t) = I^i_a(t) + \text{Binom}(E^i(t), 0.33\lambda\Delta t) - \text{Binom}(I^i_a(t), \gamma\Delta t),$$

where β is the transmission probability per unit time, N_i is the population in location i, $1/\lambda$ is the average latent period, $1/\gamma$ is the average infectious period, $\text{Binom}(n, p)$ is the binomial distribution with n trials and success probability p and $\text{Multinom}(n, p_1, p_2)$ is the multinomial distribution with n trials and success probabilities p_1 and p_2. The equation for $R^i(t)$ (the number of recovered/removed individuals in location i at time t) is redundant, since we assume that the total population size remains constant during the epidemic. Using this model, the estimated basic reproductive number (the average number of new cases caused by an infectious individual in a fully susceptible population) is given by $R_0 = \frac{\beta}{\gamma}(0.5 \cdot 0.33 + 0.67)$ [46]. We choose a time step Δt of 12 hours, in order to distinguish day time from night time.

Mobility and global infection dynamics. In the model, the disease dynamics in the block units are coupled through travelling individuals. Every individual has a defined home and work location. During day time, the individuals mix at their work location, while at night, they mix at their home location. We stress the importance of keeping track of the commuting individuals, that is, making sure the same individuals are commuting every day. Keeling et al. [47] find that by keeping track of the individuals who regularly commute, the spread rate of the epidemics is substantially reduced.

Commuting is implemented by a gravity law [48]. The parameters of the gravity law are fitted to data from Statistics Norway [49] on commuting between Norwegian municipalities.

The fitted gravity model is

$$w_{ij} \propto \frac{N_i^{0.73} N_j^{0.51}}{d_{ij}^{1.22}},$$

where w_{ij} is the number of commuters from location i to location j, N_i is the population size in location i, N_j is the population size in location j and d_{ij} is the distance (in meters) between the locations i and j. The resulting number of commuters are scaled in order to approximately match the proportion of commuters in Norway, which is 0.177. The scaling is approximate, since the number of commuters between any two pairs of locations has to be an integer number, so we round down to the nearest integer.

In addition to commuting, we consider non-commuting travel. Non-commuting travel is implemented in the model by allowing all the non-commuting individuals to travel to a random location, with some fixed probability. If the individual travels, the destination location is random, with probabilities proportional to the population size in the locations. The non-commuting individuals who travel, mix in the destination population for one day and one night (24 hours), before returning to their home location. The number of people in location i during day time, $N_i^{\text{day},t}$, and night time, $N_i^{\text{night},t}$, are given by

$$N_i^{\text{day},t} = N_i^{\text{home}} + \sum_j w_{j,i} + \sum_j u_{j,i}^t - \sum_j u_{i,j}^t,$$

$$N_i^{\text{night},t} = N_i^{\text{home}} + \sum_j w_{i,j} + \sum_j u_{j,i}^t - \sum_j u_{i,j}^t,$$

where N_i^{home} is the number of people living in location i who do not commute, $u_{i,j}^t$ is the number of non-commuters travelling from location i to location j on day t, and the t indicates that the number of people varies from day to day.

The travel probability is set to control the ratio of non-commuting to commuting. We denote this ratio by τ. A ratio of non-commuting travel to commuting of 1/10 is likely to be the most plausible travel ratio, in the baseline scenario with no travel restrictions. Hence, $\tau = 1/10$ is the baseline scenario. By using the number of yearly domestic flights in Norway as a proxy for non-commuting travel, we find a ratio of non-commuting travel to commuting of approximately 1/12. This is in agreement with [45], where they state that commuting flows are one order of magnitude larger than airline flows.

Simulation set-up and seeding. The process is stochastic, and we perform 100 disease simulations for each clustering level and report the average. The epidemic is dynamically seeded by placing one infectious individual in ten different locations with large population size (above the 80% quantile), and two infectious individuals in the location with largest population size, every day. This corresponds to 12 seeding events daily, which is 0.04% of the number of arriving international air travel passengers in Norway in 2016, as in [15]. The seeding locations

have the same population sizes for all the different clustering levels. The disease parameters are set to mimic an influenza with high transmissibility, since implementations of interventions are more relevant for influenza strains with large transmission potential. The average latent period is set to 1.9 days, the average infectious period is set to 3.0 days (as in [45]) and the transmission parameter is set to 0.60, corresponding to a basic reproductive number of 1.50.

Interventions

The following sections describe the two intervention measures examined in this work. They are applied both in separate simulations to isolate their effects, and in a combination strategy.

Travel restrictions. Travel restrictions are varied through the amount of non-commuting travel in the simulations, for the various levels of population clustering. The scenarios we consider are $\tau = 0$ (only commuting), $\tau = 1/1000$, $\tau = 1/100$ and $\tau = 1/10$ (baseline scenario).

Vaccination. We simulate the epidemic with three different vaccination strategies—vaccinating uniformly in space, preferentially vaccinating rural locations and preferentially vaccinating urban locations.

We assume that we have enough vaccines to vaccinate 40% of the population, similar to the vaccine coverage in Norway during the 2009 pandemic [25]. We let the individuals eligible for the vaccine be the susceptible individuals. We assume that the vaccine efficacy is 70%, so that 70% of the vaccinated individuals become immune to the virus, while the remaining 30% remain susceptible, in line with estimates of the efficacy of the 2009 pandemic vaccine [50]. Susceptible individuals who are vaccinated but still susceptible, are likely to experience fewer symptoms and be more resistant to the virus if they become infected, so we assume that they are 20% less infectious (in S1 Text, we also analyse the more optimistic setting where they are 80% less infectious). We assume that the vaccine is introduced 75 days after the first influenza case introduction, and that the vaccines are distributed uniformly (in time) each day for six weeks.

In the setting with uniform vaccination, we vaccinate (approximately) 40% of the population in each location. The approximation is due to the integer approximations, so we vaccinate to the nearest integer. Since we do not vaccinate immediately, there might be locations where the number of susceptibles is too small to vaccinate 40%. The remaining vaccination doses are then uniformly distributed in the other locations. In the urban vaccination strategy, we vaccinate preferentially in urban locations, by only vaccinating the 50% largest locations, using the same total number of vaccines as in the uniform vaccination setting. In the rural vaccination strategy, we vaccinate preferentially in rural locations, by leaving out the 2% locations with highest population size, using the same total number of vaccines as in the uniform vaccination setting.

Model outcomes

The epidemics are compared by examination of final size, peak date, peak prevalence and the proportion of area that is not infected during the epidemic. The final size is defined as the total number who were infected during the epidemic. The peak date is the date with the highest number infected, and the peak prevalence is the proportion infected on the peak date. The area not infected is the proportion of block units where the prevalence was never larger than a threshold for seven consecutive days. The threshold is 1.0%, except from in the locations where the population size is less than 100, then the threshold is one case.

We compare quantitative properties to compare the effect sizes between different clustering levels in order to assess whether or not the clustering plays an important role. We are thus

interested in whether there is a monotone relationship between such a quantity and κ, while the specific values are less interesting.

Results

Baseline scenario

In order to compare the dynamics for the different levels of population clustering, the disease spread was simulated for the various clustering levels. It is intuitive that with a large enough amount of non-commuting travel, there will be sufficient mixing between all the block units in the country for the population structure not to play a role in the disease dynamics. The mean global prevalence curves for the different clustering levels are given in Fig 3 for the baseline scenario with no vaccination and no travel restrictions. The curves were visually very similar, but the higher clustering levels tended to have a higher and earlier peak. Note that the confidence bands were overlapping for most clustering levels. This indicates that with the baseline amount of non-commuting, the mixing was so high that the underlying population clustering seemed to have little effect on the disease dynamics. Slightly less area was infected for the higher clustering levels, and the final sizes were slightly smaller the more clustering (Table 1). The peak dates (the day with the largest number of infected symptomatic individuals) in the different block units are given in Fig 4. The epidemic was able to spread throughout the whole

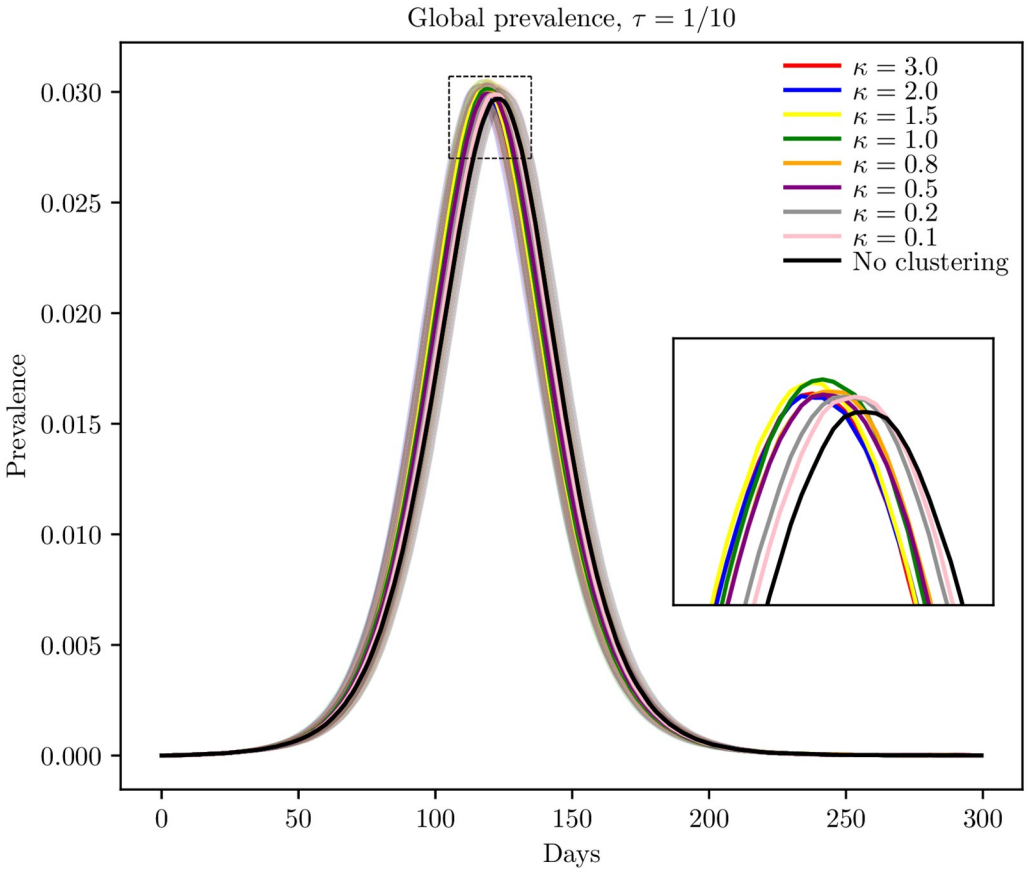

Fig 3. Global prevalence in baseline scenario. Estimated global prevalence for the various clustering levels in the baseline scenario with no interventions, with 95% confidence bands around the mean.

https://doi.org/10.1371/journal.pcbi.1006879.g003

Table 1. Baseline scenario.

κ	Area not infected	Final size
No clustering	0.187 (0.00392)	0.575 (0.000791)
0.1	0.193 (0.00324)	0.574 (0.000837)
0.2	0.200 (0.00323)	0.574 (0.000811)
0.5	0.209 (0.00401)	0.573 (0.000895)
0.8	0.209 (0.00399)	0.573 (0.000815)
1.0	0.211 (0.00435)	0.572 (0.000880)
1.5	0.213 (0.000379)	0.572 (0.000932)
2.0	0.214 (0.00388)	0.572 (0.000903)
3.0	0.214 (0.00400)	0.572 (0.000845)

Percentage of non-infected area and final sizes, for different levels of clustering. Standard deviations are given in parenthesis.

https://doi.org/10.1371/journal.pcbi.1006879.t001

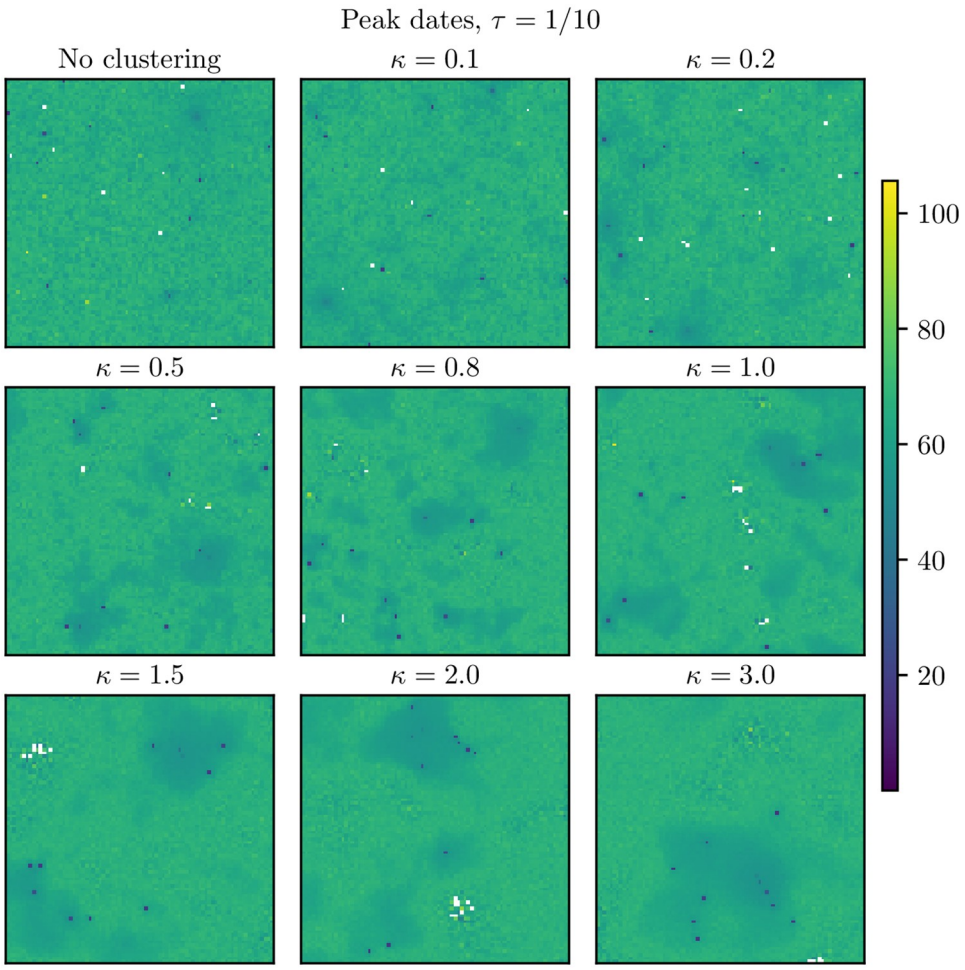

Fig 4. Peak dates for baseline scenario. Peak dates for the various clustering levels. These are averages over the simulations where an epidemic occurred in the respective block units. The white locations never experienced the epidemic. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$. The seeding locations are seen as blue dots.

https://doi.org/10.1371/journal.pcbi.1006879.g004

Table 2. $\tau = 0$.

κ	Area not infected	Final size
No clustering	0.511 (0.00418)	0.525 (0.00138)
0.1	0.542 (0.00455)	0.519 (0.00159)
0.2	0.606 (0.00428)	0.501 (0.00228)
0.5	0.694 (0.00427)	0.480 (0.00267)
0.8	0.700 (0.00473)	0.479 (0.00291)
1.0	0.696 (0.00387)	0.489 (0.00195)
1.5	0.777 (0.00361)	0.458 (0.00310)
2.0	0.802 (0.00391)	0.440 (0.00335)
3.0	0.794 (0.00304)	0.454 (0.00185)

Percentage of area not infected and final sizes. Standard deviations are given in parenthesis.

https://doi.org/10.1371/journal.pcbi.1006879.t002

country. We found spatial clustering in peak dates (by visual inspection), and the spatial clustering was larger the more clustered the country in terms of population size.

Travel restrictions

Three scenarios were considered: $\tau = 0$ (only commuting), $\tau = 1/100$ (90% travel restrictions) and $\tau = 1/1000$ (99% travel restrictions).

Table 2 shows the mean percentage of area not infected and the mean final size for $\tau = 0$. The final sizes of the epidemic decreased with increased clustering, and the area which did not experience the infection increased with increased clustering. The final size for the country with no clustering was 14% higher than for the most clustered ($\kappa = 3.0$).

The global prevalence curves for the different clustering levels for $\tau = 0$ are given in Fig 5a. The higher clustering levels experienced an earlier peak and a higher peak prevalence. Considering the confidence bands, these curves are significantly different—not between every clustering level, but the prevalence curves for the higher clustering levels are significantly different from the prevalence curves for the lower clustering levels. Interestingly, the higher peak prevalence did not imply a larger final size, as could be expected. Instead, the higher clustering levels had both higher peak prevalences and lower final sizes.

In Fig 6, we plotted the peak dates for $\tau = 0$. In addition, the initial date, peak prevalence and the probability of experiencing the epidemic in each location are given in S3, S4 and S5 Figs. There was spatial clustering in both the initial dates, peak dates, peak incidences and probabilities of experiencing the epidemic, and the spatial clustering increased with increased (population) clustering levels. The more clustered the country, the fewer locations were infected on average (Table 2). In the peak dates plot in Fig 6, we see that we had some locations which were not infected in any of the 100 simulations (coloured white). For the higher κ, these non-infected locations were clustered, and the cluster sizes increased with κ. The epidemic did not spread throughout the whole country for the highest clustering levels, but seemed to be restricted to the highly populated area.

The mean global prevalence curves for the settings with $\tau = 1/1000$ and $\tau = 1/100$ are given in Fig 5b and 5c, respectively. In the prevalence curves for $\tau = 1/1000$, we got similar results as for $\tau = 0$, with an earlier and sharper peak for the higher clustering levels. The same was found for $\tau = 1/100$, but the differences between the curves were smaller, as expected.

For $\tau = 1/1000$, fewer locations were infected and the final size decreased with increased clustering, just as in the setting with $\tau = 0$ (cf. Table 3). The same was found for $\tau = 1/100$ in

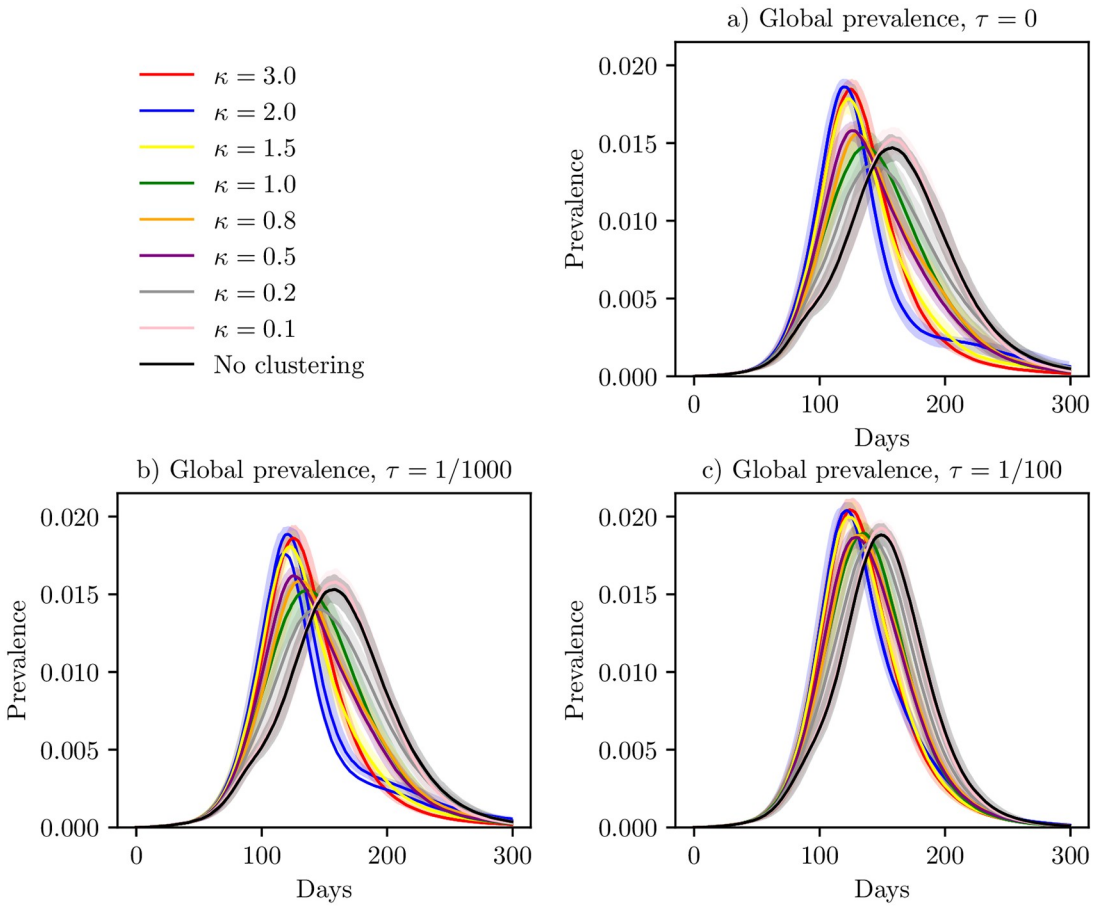

Fig 5. Global prevalence under travel restrictions. Global prevalence curves in the situation with travel restrictions included in the model, with $\tau = 0$ (a), $\tau = 1/1000$ (b), $\tau = 1/100$ (c), with 95% confidence bands around the mean.

https://doi.org/10.1371/journal.pcbi.1006879.g005

Table 4, but the differences were smaller. The effect of travel restrictions on reducing the final size was larger with higher clustering. The $\tau = 0$, $\kappa = 3.0$ scenario had 21% reduced final size, while the $\tau = 0$, no clustering-scenario had a 9% reduction. For the 99% travel restrictions, the reductions were 19% for $\kappa = 3.0$ and 8% for the scenario without clustering. For the 90% travel restrictions, the corresponding reductions were 12% and 6%.

The peak dates for $\tau = 1/1000$ and $\tau = 1/100$ are given in Figs 7 and 8, respectively. Fig 7 shows that for $\tau = 1/1000$, there were some clusters of protected locations which did not experience the infection in any of the simulations. Comparing with the situation with $\tau = 0$ in Fig 6, we found that when adding the non-commuting travel, the infection was no longer trapped in the larger hubs for the higher levels of clustering, but was able to infect a larger area of the country. There were also some protected clusters of locations with $\tau = 1/100$ (as opposed to the baseline scenario).

We have plotted the peak date for the mean global prevalence curve, peak prevalence, mean area not infected and mean final size for the various levels of clustering, for the baseline scenario and the travel restriction scenarios (Fig 9). The peak dates occurred earlier for increased levels of clustering, for all the travel ratios. In addition, the curves were very similar for the three travel restriction scenarios, while the peak dates occurred earlier for the baseline scenario. Hence, implementing travel restrictions delayed the epidemic peak. For the peak

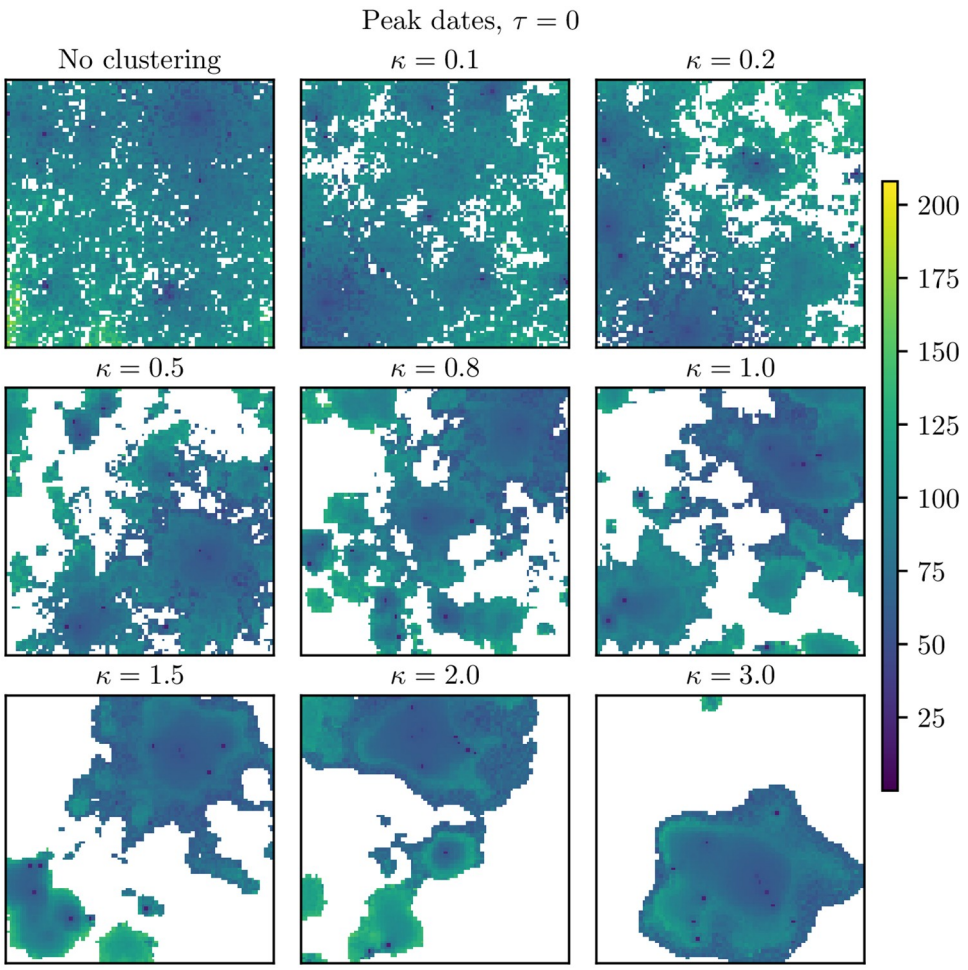

Fig 6. Peak dates for $\tau = 0$. Peak dates for infection when $\tau = 0$. These are averages over the simulations where an epidemic occurred in the respective block units. The white locations never experienced the epidemic. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$.

https://doi.org/10.1371/journal.pcbi.1006879.g006

Table 3. $\tau = 1/1000$.

κ	Area not infected	Final size
No clustering	0.501 (0.00441)	0.527 (0.00130)
0.1	0.532 (0.00397)	0.522 (0.00155)
0.2	0.591 (0.00416)	0.507 (0.00182)
0.5	0.674 (0.00389)	0.489 (0.00212)
0.8	0.679 (0.00411)	0.488 (0.00213)
1.0	0.679 (0.00324)	0.494 (0.00159)
1.5	0.753 (0.00372)	0.467 (0.00193)
2.0	0.772 (0.00389)	0.455 (0.00248)
3.0	0.772 (0.00307)	0.461 (0.00173)

Percentage of non-infected area and final sizes, for different levels of clustering. Standard deviations are given in parenthesis.

https://doi.org/10.1371/journal.pcbi.1006879.t003

Table 4. $\tau = 1/100$.

κ	Area not infected	Final size
No clustering	0.430 (0.00406)	0.543 (0.00107)
0.1	0.454 (0.00425)	0.540 (0.00118)
0.2	0.497 (0.00459)	0.532 (0.00140)
0.5	0.556 (0.00451)	0.522 (0.00131)
0.8	0.559 (0.00443)	0.521 (0.00156)
1.0	0.566 (0.00486)	0.522 (0.00123)
1.5	0.610 (0.00438)	0.508 (0.00161)
2.0	0.618 (0.00468)	0.506 (0.00153)
3.0	0.625 (0.00414)	0.504 (0.00150)

Percentage of non-infected area and final sizes, for different levels of clustering. Standard deviations are given in parenthesis.

https://doi.org/10.1371/journal.pcbi.1006879.t004

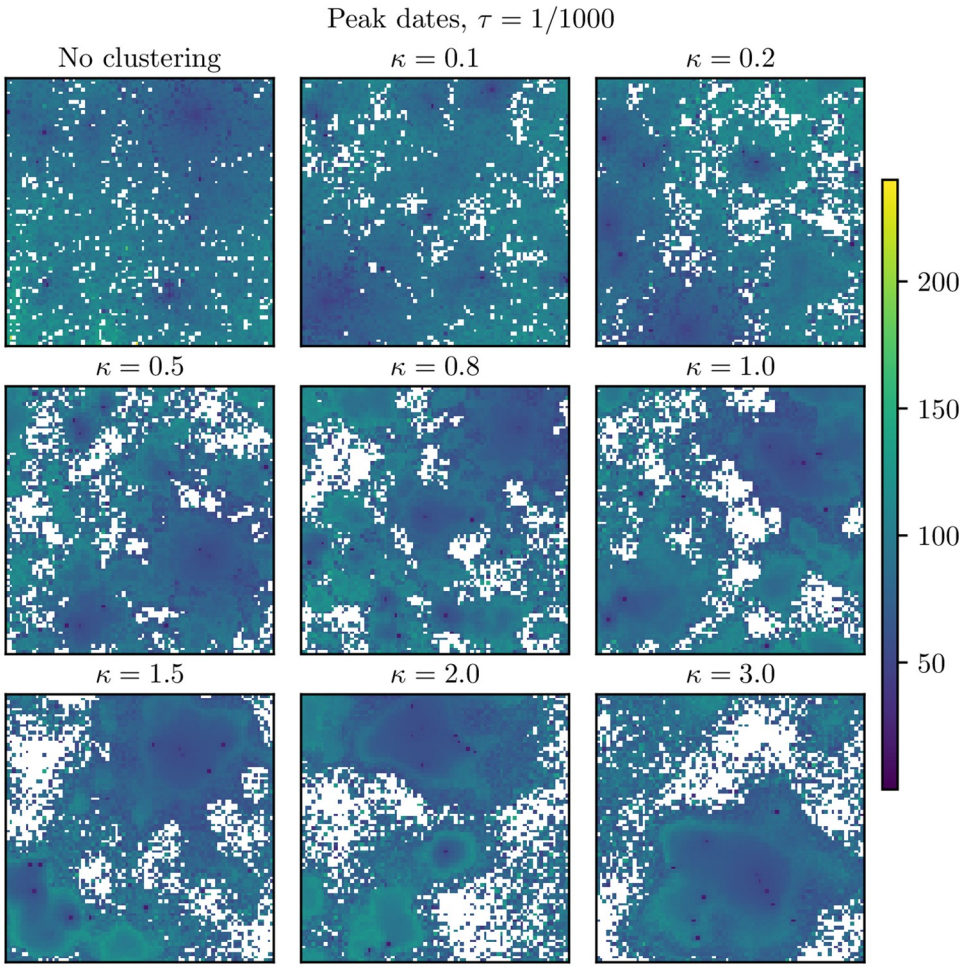

Fig 7. $\tau = 1/1000$: Peak dates. Peak dates in the setting with $\tau = 1/1000$ for the various clustering levels. These are averages over the simulations where an epidemic occurred in the respective block units. The white locations never experienced the epidemic. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$.

https://doi.org/10.1371/journal.pcbi.1006879.g007

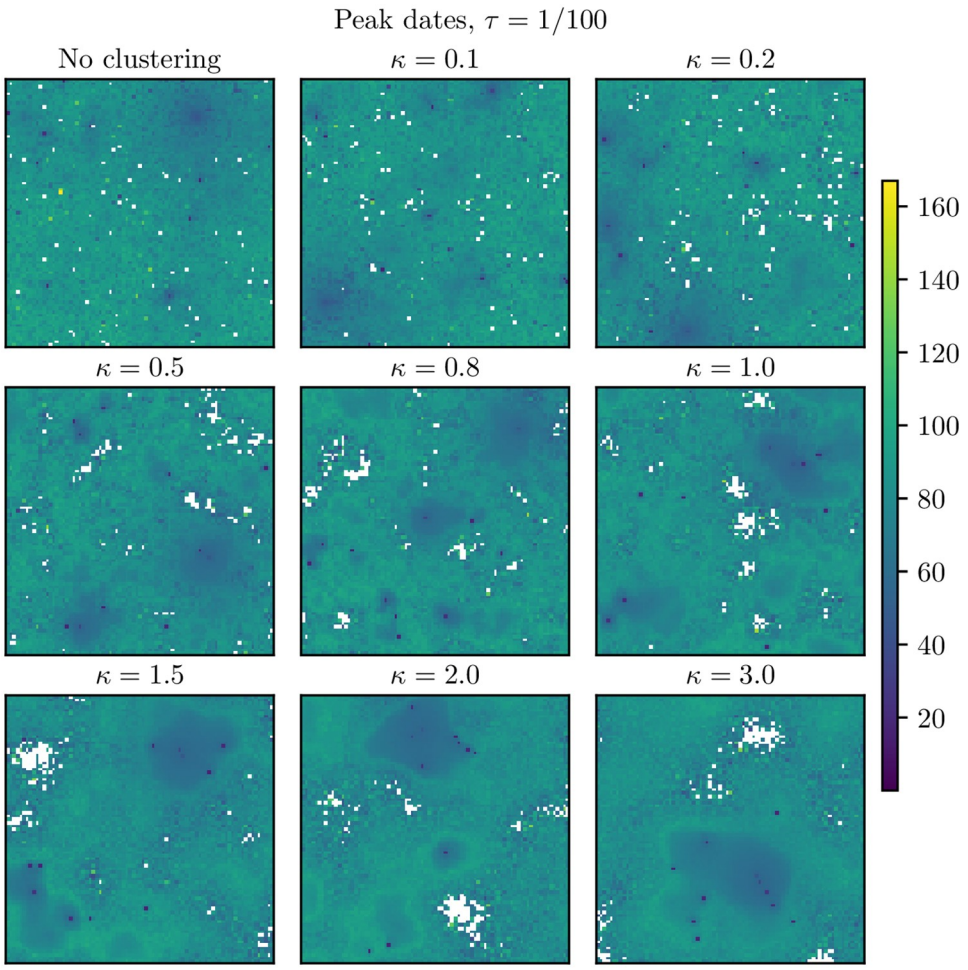

Fig 8. $\tau = 1/100$: Peak dates. Peak dates in the setting with $\tau = 1/100$ for the various clustering levels. These are averages over the simulations where an epidemic occurred in the respective block units. The locations which were never infected are coloured in white. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$.

https://doi.org/10.1371/journal.pcbi.1006879.g008

prevalence, we note that the higher levels of travel restrictions, the lower the peak. The decrease in peak prevalence was larger for the lower clustering levels. The decrease in peak prevalence for 99% travel restrictions compared to the baseline scenario was 38% for the highest clustering level and 48% for the lowest clustering level. There was almost no difference between the 99% travel ban scenario and the 100% travel ban scenario. The peak prevalence increased with increased clustering. The difference in peak prevalence with clustering was more prominent the more extensive the travel restrictions. The peak prevalence for $\kappa = 3.0$ was 20% higher than the peak prevalence for the "no clustering"-level for the complete travel ban scenario. For the mean area not infected, there was little difference between the 99% travel restrictions and the full travel ban setting. The more travel restrictions, the more area was protected. In addition, the amount of area which was protected increased with increased clustering, and the effect of clustering was stronger the more travel restrictions. For the final sizes, we found that the more travel restrictions, the lower the final size. In addition, as we have seen, for the travel restriction scenarios, the final size was lower for higher clustering levels. The more extensive the travel restrictions, the larger the difference between the various clustering levels.

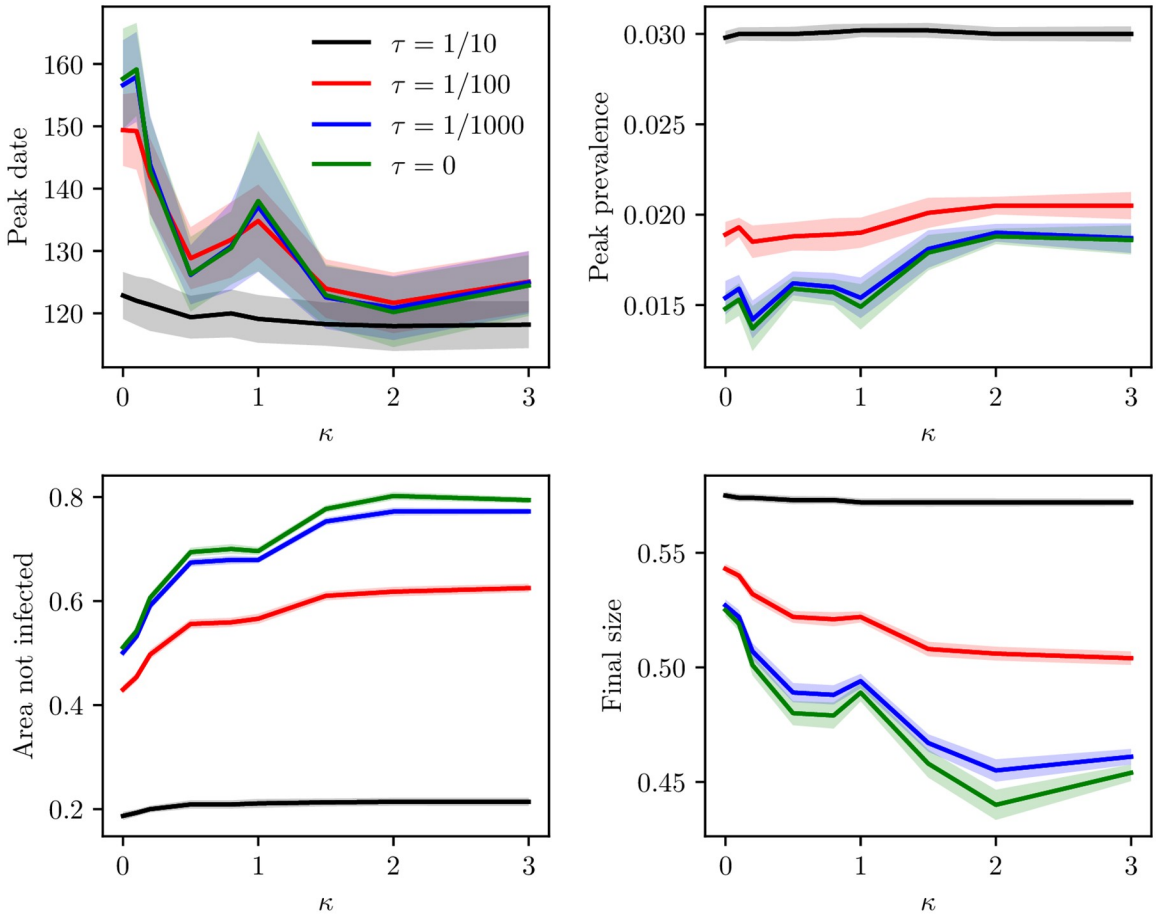

Fig 9. Peak date, peak prevalence, area not infected and final size. Peak dates for the global prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with corresponding 95% confidence bands. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.

https://doi.org/10.1371/journal.pcbi.1006879.g009

Final size versus travel ratio. To further investigate the relationship between final size and travel ratio for the various clustering levels, we also considered $\tau = 1/400$, $\tau = 1/200$, $\tau = 1/80$, $\tau = 1/40$ and $\tau = 1/20$, in addition to $\tau = 0$ (only commuting), $\tau = 1/1000$, $\tau = 1/100$ and $\tau = 1/10$. The final size of the epidemic versus travel ratio for the different levels of clustering are given in Fig 10a. For all the clustering levels, the final size increased with increased travel ratio, as expected, and the confidence bands are insignificant compared to the variation over the τ range. We found a much larger increase in final size when increasing the amount of non-commuting travel for the higher levels of clustering, than for the lower clustering levels. Hence there was a rapid increase in final size for the highest clustering levels with increased amounts of non-commuting travel. Further inspecting the final size in the various locations, we found that the decrease in final size with travel restrictions was mainly due to fewer locations being infected, but there was also some decrease in final size beyond this effect. In particular, comparing the final size in the locations which are hit in every τ-scenario, the difference in final size is largest for $\kappa = 2.0$, where the average final size decreases from 0.584 ($\tau = 1/10$) to 0.575 ($\tau = 0$). The smallest difference is for $\kappa = 0$, where the average final size is slightly larger for $\tau = 0$ than for $\tau = 1/10$ (0.583 versus 0.582). For other κ values, there was a decrease,

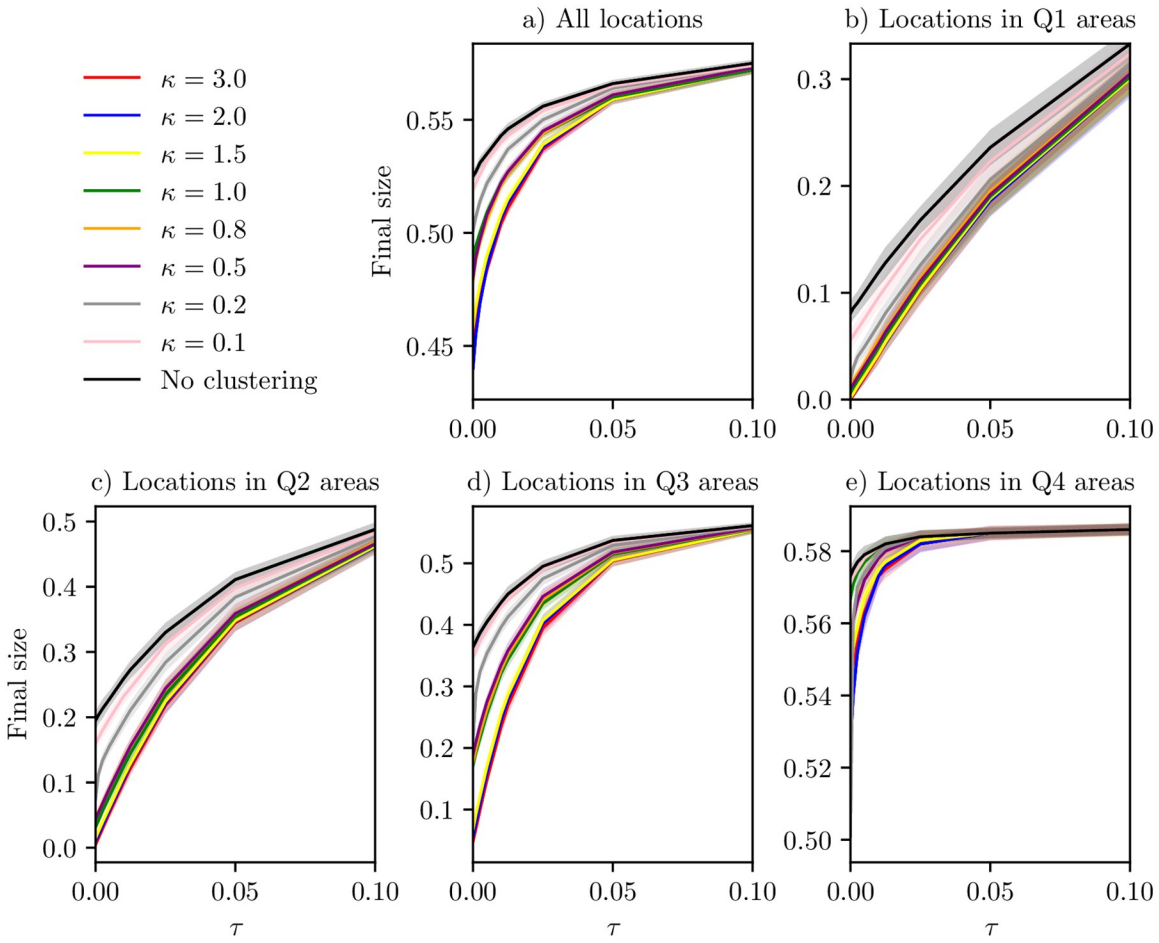

Fig 10. Final size versus travel ratio. Final size versus ratio of non-commuting travel to commuting for various clustering levels, κ, with corresponding 95% confidence bands. a) All locations. b) Locations with population size smaller than the 25% quantile. c) Locations with population size between the 25% and 50% quantile. d) Locations with population size between the 50% quantile and the 75% quantile. e) Locations with population size larger than the 75% quantile.

https://doi.org/10.1371/journal.pcbi.1006879.g010

however, the decrease was much smaller than the total decrease in final size, due to fewer locations being hit.

From the plots of the peak dates for the various travel ratios, there seemed to be some protected clusters in the settings with a low τ, where the epidemic seemed to be restricted to the hubs. We therefore expect the epidemic risk to be different for locations with different population sizes. We plotted the mean final size versus travel ratio, for various levels of κ, in areas of different population size. The final sizes for the Q1, Q2, Q3 and Q4 areas are given in Fig 10b–10e. For the Q1, Q2 and Q3 areas, we found that the higher levels of clustering were more affected by the travel restrictions. The final size in these areas increased more with increasing τ for large κ. This coincides with high population clustering having protected clusters for lower values of τ. For the Q4 areas, we also found a more rapid increase in final size for the higher clustering levels, but the curve quickly levels off to a point where a further increase in τ does not affect the final size much. In S1 Text, we fit functions for final size versus τ for all the clustering levels, and find a monotone and positive relationship between growth rate and κ for the Q1, Q2 and Q3 locations.

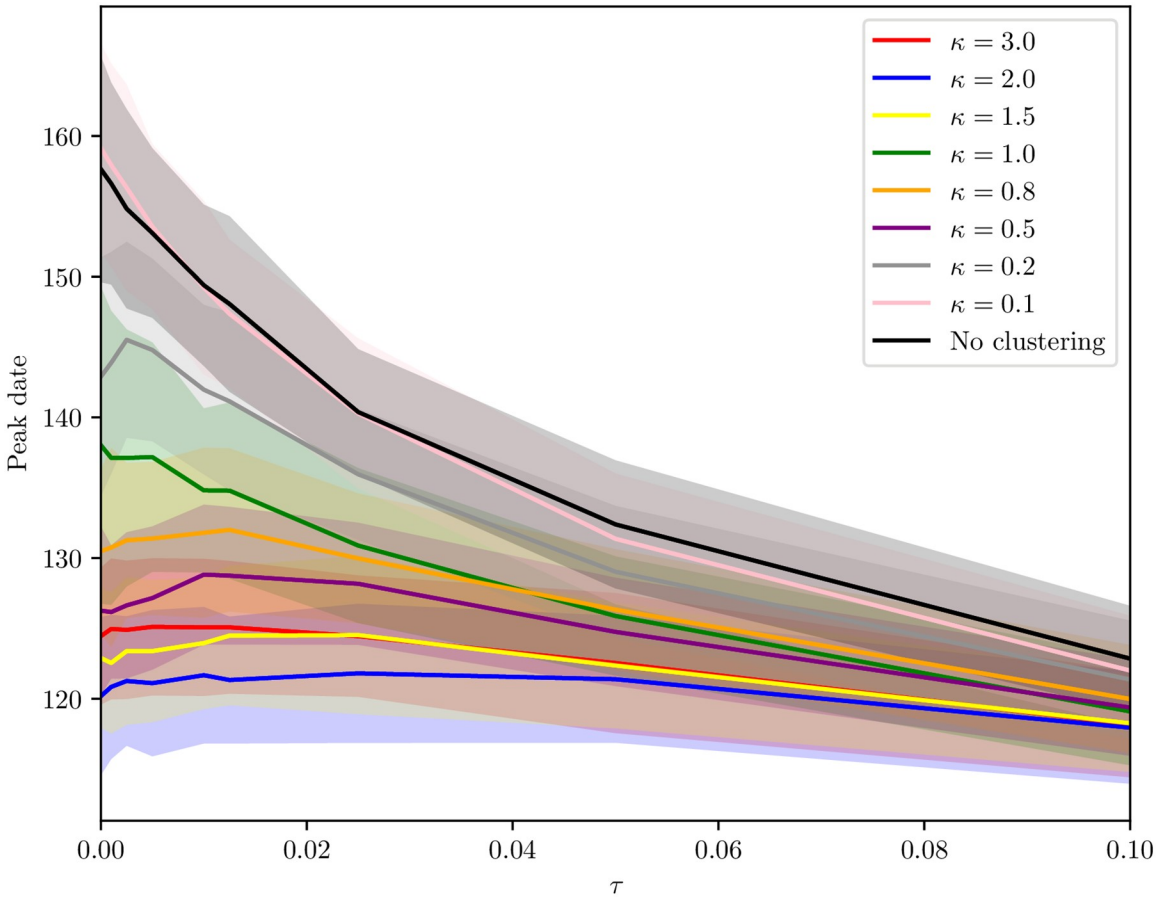

Fig 11. Peak date versus travel ratio. Peak date versus ratio of non-commuting travel to commuting for various clustering levels, κ, with 95% confidence bands.

https://doi.org/10.1371/journal.pcbi.1006879.g011

Peak date versus travel ratio. Overall, the epidemic peaked earlier with higher levels of clustering. We plotted the peak date versus travel ratio, for the different κ. The plot is given in Fig 11. The difference in peak date between the clustering levels decreased with increased non-commuting travel, as expected, and there was a rapid decrease for the low clustering levels. Travel restrictions delayed the epidemic with up to one week for the highest clustering levels and up to five weeks for the lowest clustering levels.

Sensitivity analysis. We performed sensitivity analysis with respect to delay in implementation of travel restrictions, the length of stay for the non-commuter travellers, the disease parameters and the country the population data is based on (population size distribution and gravity law). We performed sensitivity analysis with respect to the parameters in the gravity law and the shape of the gravity law. We performed the analysis using the radiation law [51] to model commuting, instead of the gravity law, and an analysis with an increased range parameter of the Matérn covariance function. In addition, we performed simulations in the setting where only the symptomatic infectious individuals were restricted from travelling. In general, the qualitative results for the various sensitivity analyses were similar to the results in the main analysis, and the details are provided in S1 Text. The effect sizes differ between the various settings. We will also comment on the robustness of the different qualitative results to these settings in the Discussion section.

Vaccination

In the setting with only regular commuting, there were some clusters of protected locations for the highest clustering levels, and the epidemic seemed to be restricted to the hubs. It might therefore be a more effective use of vaccines to allocate all the resources to the most urban locations, since the more rural locations are more protected from the epidemic. However, a different strategy would be to preferentially allocate resources to exactly these rural locations, since the vaccination is more likely to successfully eliminate the risk in these locations.

The global prevalence curves in the uniform (pro rata) vaccination setting are given in Fig 12a. The peak timing of the epidemic was similar for the different clustering levels, but there was a higher peak for the higher clustering levels (however note that the confidence bands are overlapping). The mean area not infected and final sizes are given in Table 5. Uniform vaccination reduced the final size substantially for all the clustering levels, compared to the baseline scenario and the travel restriction settings (cf. Tables 1–4). The reduction was slightly larger for the lower levels of clustering (i.e. a 65-66% reduction in final size for the no clustering, $\kappa = 0.1$ and $\kappa = 0.2$ versions, compared to a 63-64% reduction for the $\kappa \geq 1$ versions).

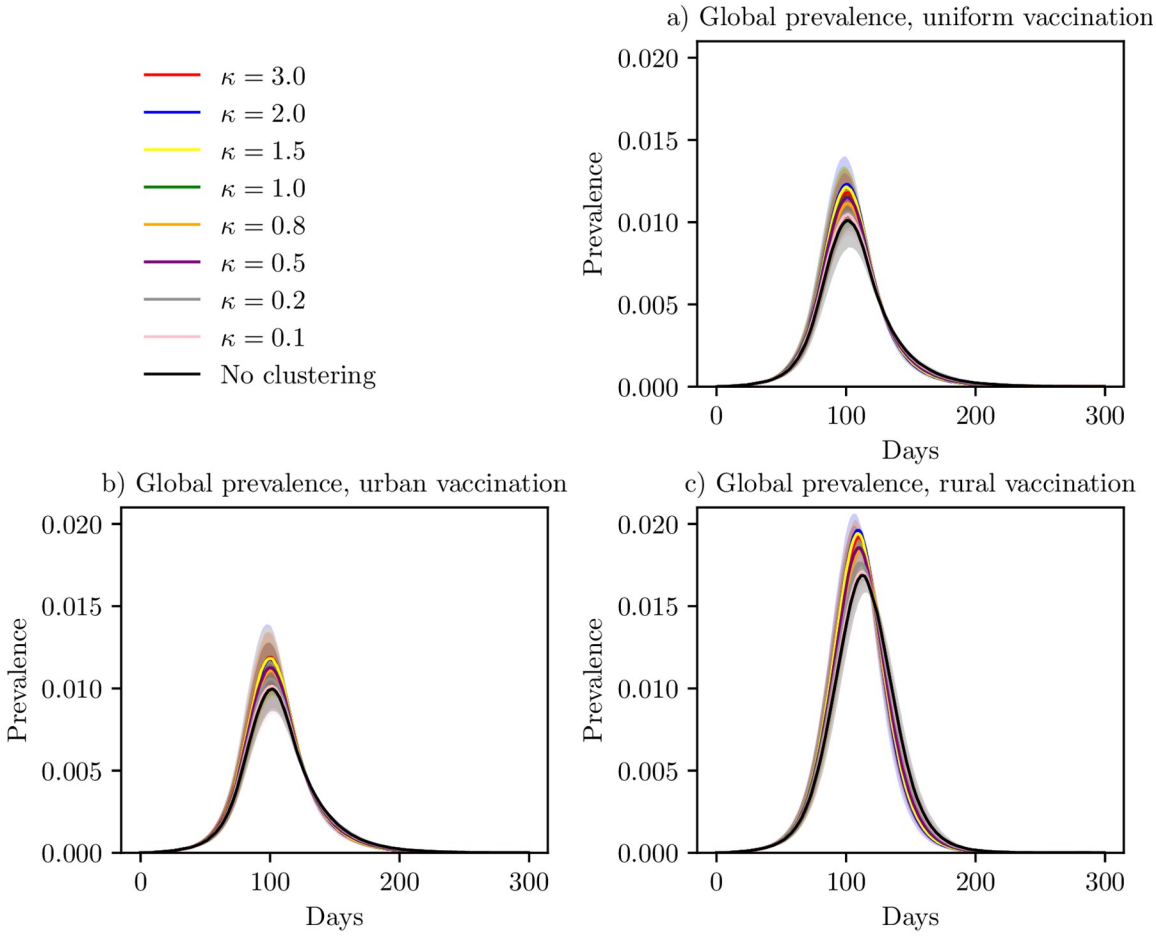

Fig 12. Global prevalence under vaccination. Global prevalence curves for the various clustering levels, under three different vaccination schemes, with 95% confidence bands around the mean. (a) Uniform vaccination, (b) preferential vaccination in urban locations and (c) preferential vaccination in rural locations.

https://doi.org/10.1371/journal.pcbi.1006879.g012

Table 5. Uniform vaccination.

κ	Area not infected	Final size
No clustering	0.762 (0.0234)	0.195 (0.0112)
0.1	0.759 (0.0208)	0.198 (0.00978)
0.2	0.764 (0.0183)	0.200 (0.00855)
0.5	0.761 (0.0222)	0.207 (0.00968)
0.8	0.767 (0.0225)	0.205 (0.0115)
1.0	0.760 (0.0253)	0.208 (0.0102)
1.5	0.764 (0.0200)	0.212 (0.00906)
2.0	0.768 (0.0236)	0.212 (0.0103)
3.0	0.777 (0.0217)	0.207 (0.00988)

Percentage of area not infected and final sizes in the situation with a uniform vaccination scheme. Standard deviations are given in parenthesis.

https://doi.org/10.1371/journal.pcbi.1006879.t005

The global prevalence curves for the urban vaccination strategy are given in Fig 12b. The prevalence curves were very similar to the uniform vaccination setting. The mean area not infected and mean final size are given in Table 6. More area was infected in this setting compared to the uniform vaccination setting. The final sizes were slightly smaller in the urban vaccination setting compared to the uniform vaccination setting, for all clustering levels except $\kappa = 3.0$. The effectiveness of this vaccination strategy compared to the uniform vaccination strategy did not seem to depend on the underlying population clustering of the country. The reduction in final size for the country without clustering was 67% while the reduction under the $\kappa = 3.0$ scenario was 63%.

The global prevalence curves for the rural vaccination strategy are given in Fig 12c. Again, the peak timing seemed to be quite similar for all the clustering levels, with a higher peak for the higher clustering levels. The mean area not infected and mean final size are given in Table 7. With the rural vaccination strategy, less area was infected, but the final size was a lot larger than for the urban and uniform vaccination strategies, due to both a higher peak prevalence and a longer epidemic. There was only a 42% reduction in final size for the $\kappa = 3.0$ version, and 44% for the least clustered version. Comparing the final sizes for the rural

Table 6. Urban vaccination.

κ	Area not infected	Final size
No clustering	0.695 (0.0207)	0.192 (0.0105)
0.1	0.699 (0.0238)	0.194 (0.0113)
0.2	0.701 (0.0233)	0.198 (0.0114)
0.5	0.703 (0.0242)	0.204 (0.0110)
0.8	0.706 (0.0236)	0.202 (0.0109)
1.0	0.703 (0.0227)	0.204 (0.0103)
1.5	0.706 (0.0284)	0.208 (0.0139)
2.0	0.710 (0.0252)	0.208 (0.0121)
3.0	0.706 (0.0236)	0.209 (0.0107)

Percentage of area not infected and final sizes in the situation with a vaccination scheme which preferentially vaccinates urban locations. Standard deviations are given in parenthesis.

https://doi.org/10.1371/journal.pcbi.1006879.t006

Table 7. Rural vaccination.

κ	Area not infected	Final size
No clustering	0.865 (0.0126)	0.323 (0.00399)
0.1	0.860 (0.0119)	0.325 (0.00398)
0.2	0.857 (0.0114)	0.327 (0.00380)
0.5	0.857 (0.0139)	0.328 (0.00463)
0.8	0.856 (0.0135)	0.329 (0.00489)
1.0	0.851 (0.0126)	0.330 (0.00458)
1.5	0.854 (0.0123)	0.332 (0.00450)
2.0	0.859 (0.0126)	0.331 (0.00447)
3.0	0.860 (0.0101)	0.330 (0.00367)

Percentage of area not infected and final sizes in the situation with a vaccination scheme which preferentially vaccinates rural locations. Standard deviations are given in parenthesis.

https://doi.org/10.1371/journal.pcbi.1006879.t007

vaccination strategy and the urban vaccination strategy yields that the difference was larger for the lowest clustering levels.

We have plotted the peak date for the mean global prevalence curve, peak prevalence, mean area not infected and mean final size for the various levels of clustering, for the baseline scenario and the different vaccination strategies. The plot is given in Fig 13. The peak dates were slightly later for the lower clustering levels than for the higher clustering levels, for the baseline and the rural vaccination strategy. The peak date was similar for the urban and uniform vaccination strategy, while it occurred later for the rural vaccination strategy. All vaccination strategies reduced the peak. The reduction was larger for the uniform and urban vaccination strategies (which were very similar), than for the rural vaccination strategy. The peak prevalence increased slightly with increased clustering under all vaccination schemes. The area infected seemed robust to the population clustering. For the final sizes, we clearly see a reduction with all the vaccination strategies, and the uniform and urban vaccination strategies were the most effective (and very similar), while the rural was a lot less effective. In addition, the final size was quite robust to the underlying clustering, but there was a slightly larger final size for the higher clustering levels.

In S1 Text, we repeated the vaccination simulations, where we assumed that the vaccinated non-immune individuals were 80% less infectious (and not 20% as in the results presented here). The qualitative results were the same, and the final sizes were only slightly smaller for this more optimistic scenario. The urban vaccination strategy was the most effective in reducing final size for all clustering levels, and the uniform vaccination strategy performed similarly.

Combination strategy

The impact of the vaccination strategies were quite robust to the underlying population clustering. Since the travel restriction intervention resulted in more protected rural area the higher the clustering, we investigated whether the performance of the vaccination strategies in combination with travel restriction differed from the performance without any travel restrictions. The peak dates, peak prevalences, areas not infected and final sizes for the different strategies and clustering levels are given in Fig 14. The urban and the uniform vaccination strategies were very similar, also when travel restrictions were included. The peak dates were slightly

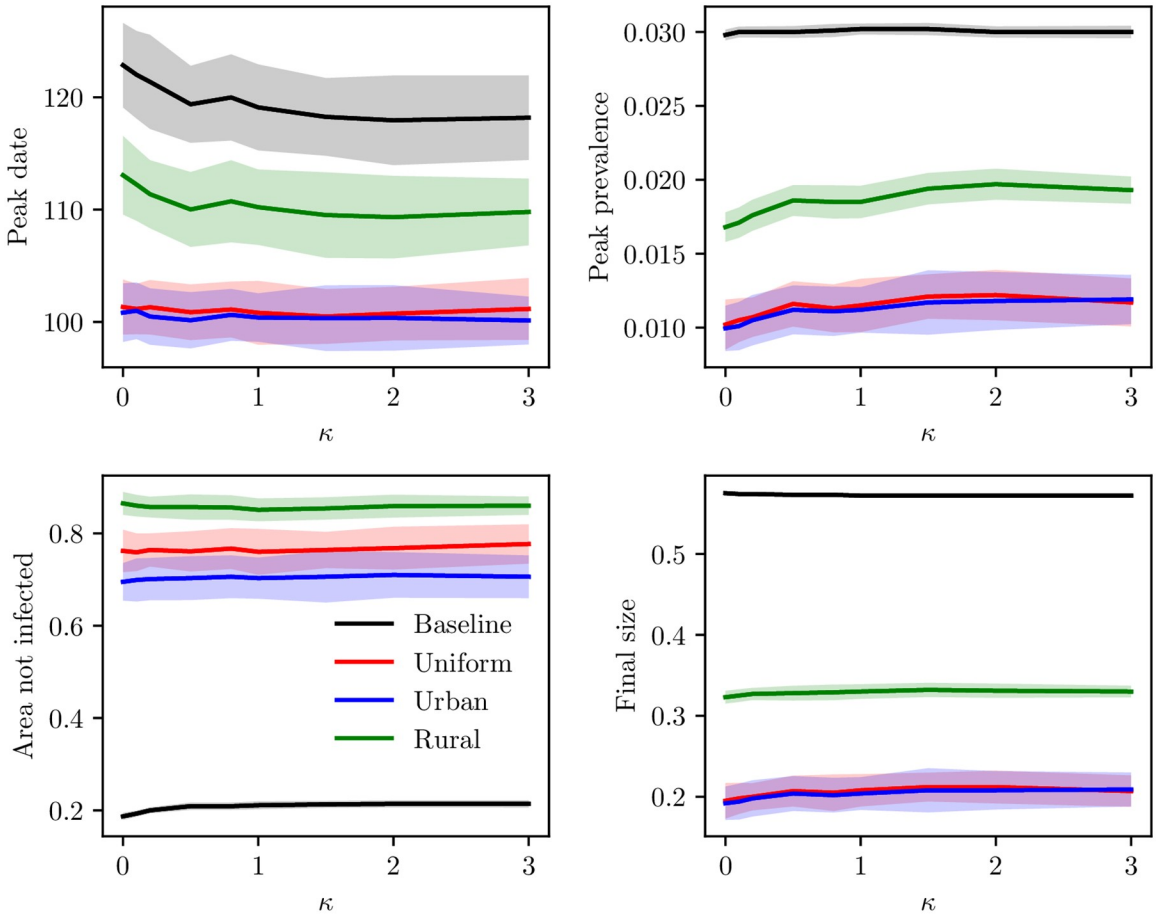

Fig 13. Peak date, peak prevalence, area not infected and final size under vaccination. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands. The lines correspond to the baseline scenario, uniform vaccination, urban vaccination and rural vaccination. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.

https://doi.org/10.1371/journal.pcbi.1006879.g013

delayed (only with a couple of days) with increased clustering for the urban and the uniform vaccination strategies. Under the rural vaccination strategy, the peak dates occurred earlier for the higher clustering levels than for the lower clustering levels. For the area not infected, the different vaccination strategies were much more similar when travel restrictions were included, and the protected area increased compared to vaccination only or travel restrictions only. The combination strategy further reduced the final size and the peak prevalence for all vaccination strategies, but the qualitative relationship with population clustering was similar to the vaccination only setting. The urban vaccination strategy combined with travel restrictions reduced the final size with 87% for the lowest clustering level, and with 78% for the highest clustering level. For the uniform vaccination strategy, the corresponding reductions were 85% and 76%. Though they were very similar, the difference between the final size for the urban vaccination strategy and the uniform vaccination strategy was larger for the higher clustering levels. Hence for all clustering levels, there was a (small and non-significant) benefit in using the urban vaccination strategy instead of the uniform vaccination strategy, but the benefit was (slightly) larger for higher clustering levels.

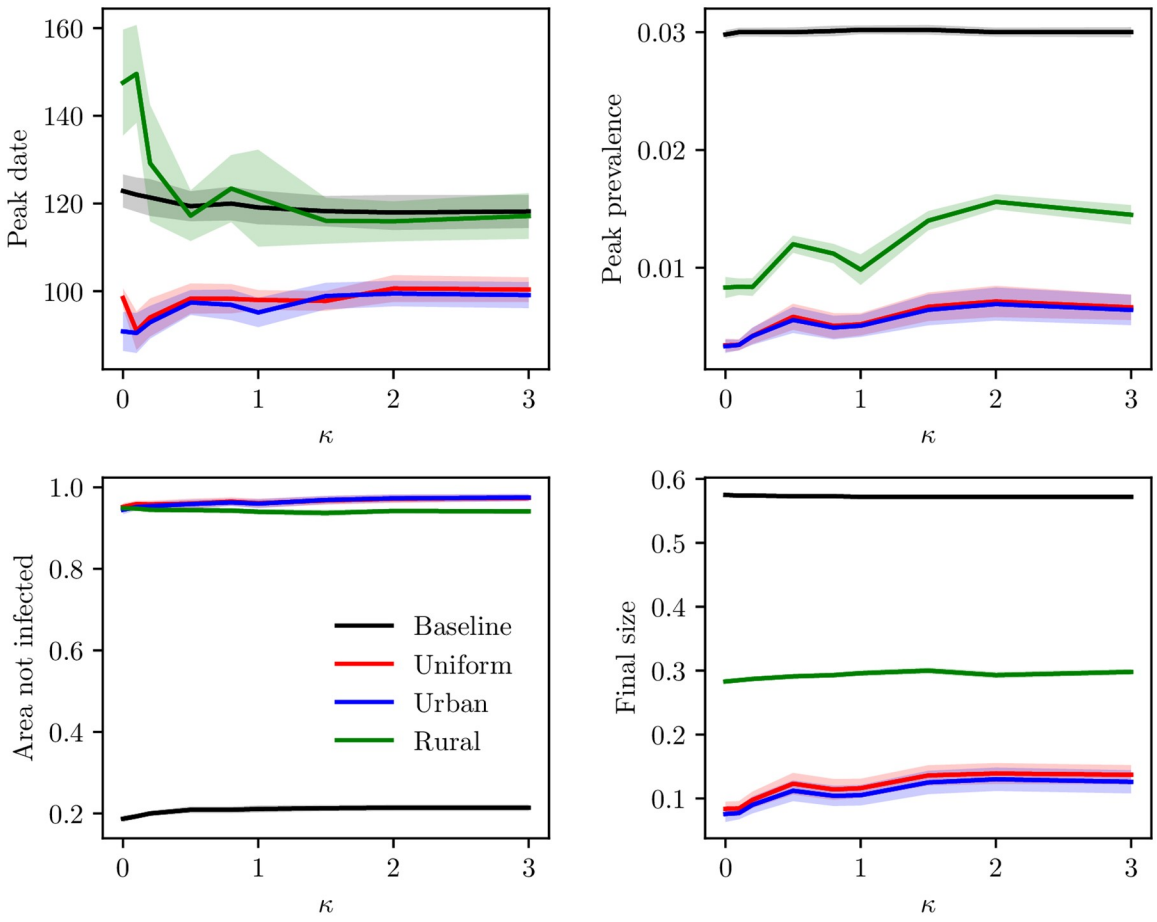

Fig 14. Peak date, peak prevalence, area not infected and final size under combined interventions. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands around the mean. The lines correspond to the baseline scenario, uniform vaccination, urban vaccination and rural vaccination. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.

https://doi.org/10.1371/journal.pcbi.1006879.g014

Discussion

We have proposed a method for generating spatial fields with controllable levels of clustering of the population. The clustering is controlled by a design parameter κ. Combined with an SEIR model for infectious diseases, we have used this tool to investigate the interplay between infectious disease spread, the effectiveness of interventions and population clustering. This framework is more general and could also easily be applied to the problem of simultaneously investigating the tendency for people to move to cities and the population growth. We have studied the urbanisation phenomenon from a theoretical viewpoint, and hope that this will inspire other, more data-driven and applied studies on urbanisation in specific settings for specific populations.

Choice of commuting model

In our main analysis, we have chosen to model commuting with a gravity law. The gravity law is a popular choice for modelling influenza/infectious disease spread [52], and has been found to capture well spatio-temporal influenza [53] and measles [54] dynamics. An alternative to

the gravity law is the radiation law [51], which is parameter free, and therefore claimed to be more universal. Some analytical inconsistencies of the gravity law are also pointed out in [51]. The radiation law does not directly depend on distance, but rather on the population density between locations. The fact that the radiation law is parameter free is attractive in this setting, since we are modelling commuting in a fictional country and thus obviously do not have commuting data available for that country. However, the gravity law has been shown to have a better fit [55], especially for finer scales, which is the case in our setting. For further information on the gravity law, the radiation law, extensions of these and other mobility models, we refer to the thorough review in [56]. We have chosen to use the gravity law in our main analysis since it had a better fit to our commuting data ($R^2 = 0.82$ versus $R^2 = 0.67$ for the radiation law), and because it is commonly used in spatial models of infectious disease transmission. Regarding the shape of the gravity law, we have chosen to use a power function of distance, as in for instance [57], [58], [59], [53] and [54]. Some commuting features that are not well captured by the gravity model are pointed out in [52], and they suggest some extensions to the gravity law for solving these issues. However, for reasons of simplicity, we have chosen to work with the more standard model. In S1 Text, we perform sensitivity analysis on the parameters of the gravity law, the shape of the distance function, and we redo the analysis using a radiation law. The qualitative patterns are robust to the choice of commuting model. However, the different effect sizes varied with the shape of the commuting law. Comments on discrepancies in the effect sizes under the different scenarios are provided below.

Findings—Discussion, robustness, implications and relation to previous literature

Baseline scenario. In a baseline scenario (no travel restrictions, no prior immunity and no vaccination), we found that the population clustering did not play an important role in outcome measures. This finding was robust across a range of different parameter values and choices—for a different set of disease parameters, for a fictional country based on a UK gravity law and population distribution, when varying the length of stay for the long distance travellers, when varying the shape and parameters of the gravity law and when modelling the commuting with a radiation law.

Travel restrictions. ***Final size***. In contrast with the baseline scenario, population clustering had an effect on the outcome measures under travel restrictions. Most importantly, the final size decreased with increased clustering. The effect of travel restrictions on decreasing the final size was stronger for higher clustering levels. This finding was quite robust to the various parameters and choices, but we note some discrepancies. When halving the distance parameter of the gravity law, travel restrictions effectively reduced the final size for both high and low clustering levels, and hence clustering mattered less in this setting. The decrease in final size under travel restrictions was larger in this setting than for the main analysis. When the distance parameter is decreased, the destination and origin population increase in importance relative to the distance, and hence the commuting patterns become more similar for the different clustering levels. In the limit of $\lim_{x \to 0} d^x$, the commuting networks are equal, and so we expect clustering to be less important for a lower distance parameter. For the UK based country, the reduction in final size with travel restrictions was larger than for the main analysis for higher clustering levels, and smaller for the lower clustering levels. The effect of travel restrictions on reducing final size was smaller for an exponential distance function in the gravity law, and there was no effect for the lowest clustering levels. Doubling the distance parameter in the gravity law, there was also no effect of travel restrictions on reducing final size for the lowest clustering levels. In these two commuting models, the distance is punished more, hence, there

is more commuting to proximate locations compared to the gravity law used in the main analysis. Therefore, the country is more well-connected through the commuting network, especially for the lower clustering levels, and travel restrictions on non-commuting travel are most likely therefore less effective.

When commuting was modelled by a radiation law, there was little effect of travel restrictions on reducing final size, and no effect for the lowest clustering levels. This is likely due to the construction of the radiation law, which does not directly depend on distance, but rather on the population density between the locations, suppressing some of the clustering effects that are found with the gravity law. Less area is protected for the radiation law, especially for the lowest clustering levels, indicating that the countries are more well-connected through the commuting network, decreasing the effect of travel restrictions.

We investigated further how the reduction of final size varied by the type of location, that is, how rural or urban the location is. For all but the most urban locations, the final size clearly increased more with increased amount of travel for the higher clustering levels. The effect of travel restrictions on epidemic final size was thus larger for the higher clustering levels, but the effect was most prominent in the less urban areas. This was also the case when varying the length of stay for the long distance travellers, for the different set of disease parameters and for the country based on UK commuting and population size distributions. This is in accordance with [17], where they find no effect of travel restrictions in urban areas. The fact that the effect of travel restrictions was most prominent in the rural areas, might make them less attractive to implement. On the other hand, health care services are more scarce in rural areas than in urban areas, so it might also be attractive that the travel restrictions are most effective in the rural areas. This should be taken into account by policy planners and decision makers.

We note that a 99% travel restriction is a quite strong restriction, and the effect size on the final size is surprisingly low. In addition, there is a problem with compliance, economic costs and how this would work in practice, and such extreme travel restrictions are not very realistic. The vaccination intervention was much more effective in reducing the final size than internal travel restrictions, in agreement with [15, 16]. In [15, 16], they found no effect of internal travel restrictions on final size. We found that internal travel restrictions had some effect on final size, as in [17].

Peak prevalence and timing. Internal travel restrictions delayed and reduced the peak. The reduction in peak prevalence was robust to all the settings considered. Under travel restrictions, we found that the higher the clustering, the higher the peak. This finding held for a range of parameter choices and assumptions, but the peak prevalence did not increase with clustering when commuting was modelled by a radiation law. For the exponential distance function in the gravity law, this was only the case for the full travel ban setting.

The peak occurred earlier for higher clustering levels for all the settings, except when the distance parameter was doubled. Travel restrictions were found to delay the peak for most settings. However, for the UK based country, with doubled destination population parameter and the radiation law settings, this was only the case for the lower clustering levels, while there was no delay effect for the higher clustering levels. Overall, the delay effects were smaller for the UK based country, the radiation law setting and when doubling the destination population parameter. Travel restrictions were more effective in delaying the peak for the lower clustering levels. This was true for all the settings considered, but the effect sizes depended on the various parameter choices and settings.

Our results are in agreement with [10, 15–17], where travel restrictions were found to delay the epidemic. In [10], it was found that a 50% travel restriction delayed the epidemic by 1.5 weeks, and we find a delay of 3-10 days, depending on the clustering level. With 90% travel restrictions, [15] found a delay in the epidemic by a few days and [16] found a one week delay

in the spread. In our simulations, we found a delay of 4-27 days with 90% travel restrictions, depending on the clustering level. Our results are thus similar to the results in [15] and [16] for the higher clustering levels (more specifically, $\kappa = 3.0$, $\kappa = 2.0$ and $\kappa = 1.5$), while we find larger delay effects for the lower clustering levels.

Vaccination. Preferentially vaccinating urban locations was the single most effective strategy in reducing final size, though only slightly better than vaccinating uniformly. This is in accordance with the result in [38], indicating little difference in the final size with a pro rata (uniform) vaccination strategy, and a sequential vaccination by population size-strategy. Preferentially vaccinating the more rural locations was clearly the most inferior strategy. The vaccination was slightly more effective for the lower clustering levels, but the epidemic progression was robust to the population clustering under all three vaccination strategies (in line with the baseline scenario), as opposed to under the travel restriction settings. This may be due to the travel restrictions breaking down specific infection routes, as opposed to the vaccination.

Combination strategy. We implemented a combination strategy, combining 99% travel restrictions with the three different vaccination strategies. We found that the most effective strategy for reducing the final size of the epidemic was a combination of travel restrictions with urban vaccination, for all the clustering levels. The difference between the final size for the urban vaccination strategy and the uniform vaccination strategy was (slightly) larger for the higher clustering levels. This is likely because when extensive travel restrictions were imposed, the most rural areas were protected, as we have seen in the analyses. The combination strategy was very efficient in reducing the final size. Travel restrictions further reduced the final size with almost 20 percentage points for the lowest clustering level and 15 percentage points for the highest clustering level. For lower κ, the travel restrictions were much more efficient in reducing final size when combined with vaccination, than in isolation.

Ethical issues. There are of course ethical issues with spatially targeted vaccination strategies. The CDC (Centers for Disease Control and Prevention) ethical guidelines state that allocation of limited resources should be guided by equity [60]. There are often inconsistencies between the optimal vaccination strategy and the most equitable vaccination strategy [29, 35, 37, 61]. In [32], focus is given to the trade off between equity, simplicity and robustness for efficient spatial allocation policies. Hence, the policy makers have to balance the effectiveness of the vaccine strategy and equity. In our setting, the urban vaccination strategy performed only slightly better than the uniform (fair) vaccination strategy, and hence the uniform vaccination strategy would be recommendable, taking equity into account.

Limitations

Our study is subject to limitations. The scaling from Norwegian municipalities to block units should be handled with care. The data that we used to fit the population size distribution and gravity law were on a different scale than the block units used in the fictional country, and we can not assess how well these models generalise to the finer block unit scale. In addition, different ways of dividing the population into administrative units could yield different population distributions, which could affect the results. We believe that it is of key importance to model the commuting and the population sizes for the same scale and population, and since commuting data are not available on a finer scale (due to for instance privacy regulations), we have chosen to use the finest scale available for the commuting data. The block units can then be interpreted as a discretisation of administrative units. Census data are often collected for administrative units, but mobile phone data could be used as an information source for commuting on a square gridded fine scale resolution, and would have been an interesting alternative. The fine scale resolution of the fictional country is necessary for the clustering algorithm

to provide smooth transitions between the different clustering levels. The fact that we consider the country in isolation and ignore bordering countries can result in edge effects, since the locations on the border might in reality have a connection to the neighbouring countries.

We address some of these limitations in our sensitivity analysis in S1 Text. To handle the fact that different ways of defining administrative regions could affect the population distribution and in turn the results, we also perform the analysis on a country based on population data from the United Kingdom. The population size distribution for Norway was quite heterogeneous compared to other European countries examined, while United Kingdom had a more homogeneous population size distribution. Here, a gravity law fitted to commuting data for the United Kingdom is also used. We also perform sensitivity analysis on the parameters and shape of the gravity law. Though the qualitative patterns were not so sensitive to the choice of commuting model, the effect sizes were not robust to the shape of the commuting law.

Practical implications

According to the general model developed here, population clustering is an important determinant for the effect of travel restrictions. For high levels of clustering, internal travel restrictions decrease the final size of the epidemic, and this is most prominent in the rural areas. There is no large effect on the peak date for the high clustering levels. For the lower clustering levels, there is less of a benefit in terms of final size reduction. However, for lower clustering levels, the peak date is delayed when implementing internal travel restrictions. That means more time to plan and implement interventions and preventive measures. Internal travel restrictions reduce the peak prevalence for all the clustering levels, reducing the stress on the health care systems. In addition, we found that the higher clustering levels have larger spatial clustering in peak dates. The more spatial clustering in peak dates, the more stress on the health care systems. Whether it is more attractive to delay an epidemic or decrease the final size, depends on the specific influenza strain. If there is high morbidity, such that the patients require more and/or longer health care, it might be of importance to delay the epidemic, to prepare the health care system for the peak period. In addition, if there exists vaccines or other interventions, a delay in the epidemic can be of key importance for the effectiveness of the vaccination programme, since there is more time to distribute (and develop/improve) the vaccine or implement other interventions. The impact of vaccines depends on how early they are introduced (see for instance [25, 33, 62, 63]). For the European countries we considered (by visual inspection), the minimum clustering level was around $\kappa = 0.5$, and the highest around $\kappa = 3.0$, and they will likely become even more clustered in the future. This means that according to our model, internal travel restrictions are likely to be less and less effective in delaying epidemics, while they will be more effective in decreasing final sizes (especially in rural areas). In addition, proximate regions will, to an even higher extent than today, experience their peak simultaneously. Hence, it will be even more important to be able to predict the peak timing, in order to prepare the health care system for the peak period. In addition, in order to minimise the final sizes of the epidemic, it is important not to neglect the urban locations for vaccination, and thus specific vaccination sentiment campaigns might target urban locations.

Supporting information

S1 Text. Additional remarks and sensitivity analysis.
(PDF)

S1 Fig. Population size distribution for the municipalities of Norway.
(TIF)

S2 Fig. Population densities. Population density in administrative units in Norway, Iceland, Germany, France, Netherlands and United Kingdom.
(TIF)

S3 Fig. $\tau = 0$: Initial dates. Initial dates for infection when $\tau = 0$. These are averages over the simulations where an epidemic occurred in the respective block units. The white locations never experienced the epidemic. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$.
(TIF)

S4 Fig. $\tau = 0$: Peak prevalence. Peak prevalence for $\tau = 0$. These are averages over the simulations where an epidemic occurred in the respective block units. The white locations never experienced the epidemic. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$.
(TIF)

S5 Fig. $\tau = 0$: Probability of epidemic. Probability of infection for $\tau = 0$. Upper left: No clustering. Upper center: $\kappa = 0.1$. Upper right: $\kappa = 0.2$. Middle left: $\kappa = 0.5$. Middle center: $\kappa = 0.8$. Middle right: $\kappa = 1.0$. Bottom left: $\kappa = 1.5$. Bottom center: $\kappa = 2.0$. Bottom right: $\kappa = 3.0$.
(TIF)

S6 Fig. Global prevalence under delayed travel restrictions. Estimated global prevalence for the various smoothing levels with corresponding 95% confidence bands, in the setting with delay in implementation of travel restrictions.
(TIF)

S7 Fig. Final size versus travel ratio with varying travel duration. Final size versus τ for various clustering levels, κ, when the length of stay for non-commuting travellers varies, with corresponding 95% confidence bands. a) All locations. b) Locations with population size smaller than the 25% quantile. c) Locations with population size between the 25% and 50% quantile. d) Locations with population size between the 50% quantile and the 75% quantile. e) Locations with population size larger than the 75% quantile.
(TIF)

S8 Fig. Peak dates versus travel ratio with varying travel duration. Peak date versus τ for various clustering levels, κ, when the length of stay for non-commuting travellers varies, with corresponding 95% confidence bands.
(TIF)

S9 Fig. Final size and peak date with travel ban targeting infectious symptomatic. Final size and peak date versus τ for various clustering levels, κ, with only travel restrictions for the infectious symptomatic, with corresponding 95% confidence bands.
(TIF)

S10 Fig. Final size versus travel ratio with alternative disease parameters. Final size versus τ for various clustering levels, κ, for the alternative disease parameters, with corresponding 95% confidence bands. a) All locations. b) Locations with population size smaller than the 25% quantile. c) Locations with population size between the 25% and 50% quantile. d) Locations with population size between the 50% quantile and the 75% quantile. e) Locations with population size larger than the 75% quantile.
(TIF)

S11 Fig. Peak date versus travel ratio for alternative disease parameters. Peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands. The results are for the alternative disease parameters.
(TIF)

S12 Fig. Final size versus travel ratio for the UK based country. Final size versus τ for various clustering levels, κ, with corresponding 95% confidence bands. The results are in the countries based on data from the United Kingdom. a) All locations. b) Locations with population size smaller than the 25% quantile. c) Locations with population size between the 25% and 50% quantile. d) Locations with population size between the 50% quantile and the 75% quantile. e) Locations with population size larger than the 75% quantile.
(TIF)

S13 Fig. Peak date versus travel ratio for the UK based country. Peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands. The results are in the country based on data from the United Kingdom.
(TIF)

S14 Fig. Sensitivity analysis (halving distance parameter): Peak dates, peak prevalence, area not infected and final size. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, when the distance parameter of the gravity law was halved. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S15 Fig. Sensitivity analysis (halving distance parameter): Final size and peak date. Final size and peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands, when the distance parameter of the gravity law was halved.
(TIF)

S16 Fig. Sensitivity analysis (doubling distance parameter): Peak dates, peak prevalence, area not infected and final size. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, when the distance parameter of the gravity law was doubled. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S17 Fig. Sensitivity analysis (doubling distance parameter): Final size and peak date. Final size and peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands, when the distance parameter of the gravity law was doubled.
(TIF)

S18 Fig. Sensitivity analysis (halving destination population parameter): Peak dates, peak prevalence, area not infected and final size. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, when the destination population parameter of the gravity law was halved. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel

restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S19 Fig. Sensitivity analysis (halving destination population parameter): Final size and peak date. Final size and peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands, when the destination population parameter of the gravity law was halved.
(TIF)

S20 Fig. Sensitivity analysis (doubling destination population parameter): Peak dates, peak prevalence, area not infected and final size. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, when the destination population parameter of the gravity law was doubled. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S21 Fig. Sensitivity analysis (doubling destination population parameter): Final size and peak date. Final size and peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands, when the destination population parameter of the gravity law was doubled.
(TIF)

S22 Fig. Peak dates, peak prevalence, area not infected and final size, exponential distance function. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, with an exponential function of distance in the gravity law. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S23 Fig. Final size and peak date with exponential distance function. Final size and peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands, with an exponential function of distance in the gravity law.
(TIF)

S24 Fig. Peak dates, peak prevalence, area not infected and final size, range parameter 10.0 in the covariance function. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, when the range parameter of the Matérn covariance function was increased from 5.0 to 10.0. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S25 Fig. Final size and peak date with range parameter 10.0 in the covariance function. Final size and peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands, when the range parameter of the Matérn covariance function was increased from 5.0 to 10.0.
(TIF)

S26 Fig. Peak dates, peak prevalence, area not infected and final size, when the radiation law was used to model commuting. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, when the commuting was implemented by the radiation law. The lines correspond to the baseline scenario, 90% travel restrictions, 99% travel restrictions and 100% travel restrictions. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S27 Fig. Final size and peak date when the radiation law was used to model commuting. Final size and peak date versus τ for various clustering levels, κ, with corresponding 95% confidence bands, for the results where commuting was implemented by the radiation law.
(TIF)

S28 Fig. Peak dates, peak prevalence, area not infected and final size, 20% reduced infectiousness. Peak dates for the global mean prevalence curve, peak prevalence, mean area not infected and mean final size as a function of clustering, with 95% confidence bands, when the assumed infectiousness of non-immune vaccinated is reduced by 20%. The lines correspond to the baseline scenario, uniform vaccination, urban vaccination and rural vaccination. Top left: peak date. Top right: peak prevalence. Bottom left: area not infected. Bottom right: final size.
(TIF)

S1 Table. Estimated *a* for urban and rural locations. Estimated power a for final size = $\tau^a + b$, for different levels of clustering, for the Q1 (most rural), Q2, Q3 and Q4 (most urban) locations.
(PDF)

S2 Table. Estimated *a* for all locations. Estimated power a for final size = $\tau^a + b$, for different levels of clustering.
(PDF)

S3 Table. Delayed travel restrictions. Global peak day, global peak prevalence, percentage of area not infected and final sizes in the situation with a delay in the implementation of the travel restrictions. Standard deviations are given in parenthesis.
(PDF)

Acknowledgments

The authors thank Gianpaolo Scalia Tomba for interesting discussions.

Author Contributions

Conceptualization: Solveig Engebretsen, Kenth Engø-Monsen, Arnoldo Frigessi, Birgitte Freiesleben de Blasio.

Formal analysis: Solveig Engebretsen.

Methodology: Solveig Engebretsen, Kenth Engø-Monsen, Arnoldo Frigessi, Birgitte Freiesleben de Blasio.

Supervision: Kenth Engø-Monsen, Arnoldo Frigessi, Birgitte Freiesleben de Blasio.

Visualization: Solveig Engebretsen.

Writing – original draft: Solveig Engebretsen.

Writing – review & editing: Solveig Engebretsen, Kenth Engø-Monsen, Arnoldo Frigessi, Birgitte Freiesleben de Blasio.

References

1. United Nations, Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2014 revision, Highlights (ST/ESA/SER.A/352). 2014.
2. Rose WJ, Bell JE, Autry CW, Cherry CR. Urban Logistics: Establishing Key Concepts and Building a Conceptual Framework for Future Research. Transp J. 2017; 56(4):357–394. https://doi.org/10.5325/transportationj.56.4.0357
3. McGranahan G, Satterthwaite D. Urbanisation: Concepts and Trends. IIED London; 2014.
4. Vasanen A. Deconcentration versus spatial clustering: changing population distribution in the Turku urban region, 1980–2005. Fennia-International Journal of Geography. 2009; 187(2):115–127.
5. Kapferer JL. Rural myths and urban ideologies. Aust N Z J Sociol. 1990; 26(1):87–106. https://doi.org/10.1177/144078339002600105
6. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford university press; 1992.
7. Bajardi P, Poletto C, Ramasco JJ, Tizzoni M, Colizza V, Vespignani A. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. PloS One. 2011; 6(1):e16591. https://doi.org/10.1371/journal.pone.0016591 PMID: 21304943
8. Cruz-Pacheco G, Duran L, Esteva L, Minzoni A, Lopez-Cervantes M, Panayotaros P, et al. Modelling of the influenza A (H1N1) v outbreak in Mexico City, April-May 2009, with control sanitary measures. Euro Surveill. 2009; 14(26):19254. PMID: 19573510
9. Sattenspiel L, Herring DA. Simulating the effect of quarantine on the spread of the 1918–19 flu in central Canada. Bull Math Biol. 2003; 65(1):1–26. https://doi.org/10.1006/bulm.2002.0317 PMID: 12597114
10. Bolton K, McCaw J, Moss R, Morris R, Wang S, Burma A, et al. Likely effectiveness of pharmaceutical and non-pharmaceutical interventions for mitigating influenza virus transmission in Mongolia. Bull World Health Organ. 2012; 90:264–271. https://doi.org/10.2471/BLT.11.093419 PMID: 22511822
11. Plan international. Ebola: beyond the health emergency. 2015.
12. Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. Urbanisation and infectious diseases in a globalised world. Lancet Infect Dis. 2011; 11(2):131–141. https://doi.org/10.1016/S1473-3099(10)70223-1 PMID: 21272793
13. Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW. Opinion—tropical infectious diseases: Urbanization, malaria transmission and disease burden in Africa. Nat Rev Microbiol. 2005; 3(1):81–90.
14. Zhang P, Atkinson PM. Modelling the effect of urbanization on the transmission of an infectious disease. Math Biosci. 2008; 211(1):166–185. https://doi.org/10.1016/j.mbs.2007.10.007 PMID: 18068198
15. Germann TC, Kadau K, Longini IM, Macken CA. Proc Natl Acad Sci U S A. Proceedings of the National Academy of Sciences. 2006; 103(15):5935–5940.
16. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006; 442(7101):448–452. https://doi.org/10.1038/nature04795 PMID: 16642006
17. Mateus AL, Otete HE, Beck CR, Dolan GP, Nguyen-Van-Tam JS. Effectiveness of travel restrictions in the rapid containment of human influenza: a systematic review. Bull World Health Organ. 2014; 92(12): 868–880D. https://doi.org/10.2471/BLT.14.135590 PMID: 25552771
18. Brownstein JS, Wolfe CJ, Mandl KD. Empirical Evidence for the Effect of Airline Travel on Inter-Regional Influenza Spread in the United States. PLoS Med. 2006; 3(10):e401. https://doi.org/10.1371/journal.pmed.0030401 PMID: 16968115
19. Lee JM, Choi D, Cho G, Kim Y. The effect of public health interventions on the spread of influenza among cities. J Theor Biol. 2012; 293:131–142. https://doi.org/10.1016/j.jtbi.2011.10.008 PMID: 22033506
20. Wood JG, Zamani N, MacIntyre CR, Becker NG. Effects of internal border control on spread of pandemic influenza. Emerg Infect Dis. 2007; 13(7):1038–1045. https://doi.org/10.3201/eid1307.060740 PMID: 18214176
21. US Department of Health and Human Services and others. HHS Pandemic influenza plan 2017 UPDATE. US Department of Health and Human Services Web site; 2017.

22. Viboud C, Miller MA, Grenfell BT, Bjørnstad ON, Simonsen L. Air travel and the spread of influenza: important caveats. PLoS Med. 2006; 3(11):e503. https://doi.org/10.1371/journal.pmed.0030503 PMID: 17132057
23. Tomba GS, Wallinga J. A simple explanation for the low impact of border control as a countermeasure to the spread of an infectious disease. Math Biosci. 2008; 214(1-2):70–72. https://doi.org/10.1016/j.mbs.2008.02.009
24. Gautreau A, Barrat A, Barthelemy M. Global disease spread: statistics and estimation of arrival times. J Theor Biol. 2008; 251(3):509–522. https://doi.org/10.1016/j.jtbi.2007.12.001 PMID: 18222486
25. de Blasio BF, Iversen BG, Tomba GS. Effect of vaccines and antivirals during the major 2009 A (H1N1) pandemic wave in Norway–and the influence of vaccination timing. PLoS One. 2012; 7(1):e30018. https://doi.org/10.1371/journal.pone.0030018
26. Cotter S, Gee S, OLLorcain P. National Summary of Influenza Pandemic (H1N1) 2009 vaccination in Ireland-Provisional data. Epi insight. 2010.
27. Centers for Disease Control and Prevention. Allocation and Distribution Q&A; 2009. Available from: https://www.cdc.gov/h1n1flu/vaccination/statelocal/centralized_distribution_qa.htm. Cited 24 November 2018
28. Rambhia KJ, Watson M, Sell TK, Waldhorn R, Toner E. Mass vaccination for the 2009 H1N1 pandemic: approaches, challenges, and recommendations. Biosecur Bioterror. 2010; 8(4):321–330. https://doi.org/10.1089/bsp.2010.0043 PMID: 21043791
29. Matrajt L, Halloran ME, Longini IM Jr. Optimal vaccine allocation for the early mitigation of pandemic influenza. PLoS Comput Biol. 2013; 9(3):e1002964. https://doi.org/10.1371/journal.pcbi.1002964 PMID: 23555207
30. Shrestha S, Chatterjee S, Rao KD, Dowdy DW. Potential impact of spatially targeted adult tuberculosis vaccine in Gujarat, India. J R Soc Interface. 2016; 13(116):20151016. https://doi.org/10.1098/rsif.2015.1016 PMID: 27009179
31. Dimitrov NB, Goll S, Hupert N, Pourbohloul B, Meyers LA. Optimizing tactics for use of the US antiviral strategic national stockpile for pandemic influenza. PloS One. 2011; 6(1):e16094. https://doi.org/10.1371/journal.pone.0016094 PMID: 21283514
32. Wu JT, Riley S, Leung GM. Spatial considerations for the allocation of pre-pandemic influenza vaccination in the United States. Proc R Soc Lond B Biol Sci. 2007; 274(1627):2811–2817. https://doi.org/10.1098/rspb.2007.0893
33. Keeling MJ, White PJ. Targeting vaccination against novel infections: risk, age and spatial structure for pandemic influenza in Great Britain. J R Soc Interface. 2011; 8(58):661–670. https://doi.org/10.1098/rsif.2010.0474 PMID: 20943682
34. Azman AS, Lessler J. Reactive vaccination in the presence of disease hotspots. Proc R Soc B. 2015; 282(1798):20141341. https://doi.org/10.1098/rspb.2014.1341 PMID: 25392464
35. Duijzer LE, van Jaarsveld WL, Wallinga J, Dekker R. Dose-Optimal Vaccine Allocation over Multiple Populations. Prod Oper Manag. 2018; 27(1):143–159. https://doi.org/10.1111/poms.12788
36. Nguyen C, Carlson JM. Optimizing real-time vaccine allocation in a stochastic SIR model. PloS One. 2016; 11(4):e0152950. https://doi.org/10.1371/journal.pone.0152950 PMID: 27043931
37. Keeling MJ, Shattock A. Optimal but unequitable prophylactic distribution of vaccine. Epidemics. 2012; 4(2):78–85. https://doi.org/10.1016/j.epidem.2012.03.001 PMID: 22664066
38. Araz OM, Galvani A, Meyers LA. Geographic prioritization of distributing pandemic influenza vaccines. Health Care Manag Sci. 2012; 15(3):175–187. https://doi.org/10.1007/s10729-012-9199-6 PMID: 22618029
39. Teytelman A, Larson RC. Multiregional dynamic vaccine allocation during an influenza epidemic. Serv Sci. 2013; 5(3):197–215. https://doi.org/10.1287/serv.2013.0046
40. Wikipedia contributors. List of arrondissements of France—Wikipedia, The Free Encyclopedia; 2018. Available from: https://en.wikipedia.org/wiki/List_of_arrondissements_of_France. Cited 24 April 2018
41. Wikipedia contributors. Provinces of Italy—Wikipedia, The Free Encyclopedia; 2018. Available from: https://en.wikipedia.org/wiki/Provinces_of_Italy. Cited 24 April 2018
42. Wikipedia contributors. List of municipalities of the Netherlands—Wikipedia, The Free Encyclopedia; 2018. Available from: https://en.wikipedia.org/wiki/List_of_municipalities_of_the_Netherlands. Cited 24 April 2018
43. Wikipedia contributors. List of municipalities of Denmark—Wikipedia, The Free Encyclopedia; 2018. Available from: https://en.wikipedia.org/wiki/List_of_municipalities_of_Denmark. Cited 24 April 2018
44. Cressie N, Wikle CK. Statistics for Spatio-Temporal Data. John Wiley & Sons; 2011.

45. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci U S A. 2009; 106(51):21484–21489. https://doi.org/10.1073/pnas.0906910106 PMID: 20018697

46. Colizza V, Barrat A, Barthelemy M, Valleron AJ, Vespignani A. Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. PLoS Med. 2007; 4(1):e13. https://doi.org/10.1371/journal.pmed.0040013 PMID: 17253899

47. Keeling MJ, Danon L, Vernon MC, House TA. Individual identity and movement networks for disease metapopulations. Proc Natl Acad Sci U S A. 2010; 107(19):8866–8870. https://doi.org/10.1073/pnas.1000416107 PMID: 20421468

48. Zipf GK. The P_1P_2/D Hypothesis: On the Intercity Movement of Persons. Am Sociol Rev. 1946; 11(6): 677–686. https://doi.org/10.2307/2087063

49. Statistics Norway. 03321: Employed persons (aged 15-74), by municipality of work and municipality of residence. 4th quarter (M) 2000-2017 Available from: https://www.ssb.no/en/statbank/table/03321.

50. Hardelid P, Fleming D, McMenamin J, Andrews N, Robertson C, SebastianPillai P, et al. Effectiveness of pandemic and seasonal influenza vaccine in preventing pandemic influenza A (H1N1) 2009 infection in England and Scotland 2009-2010. Euro Surveill. 2011; 16(2):19763. PMID: 21251487

51. Simini F, González MC, Maritan A, Barabási AL. A universal model for mobility and migration patterns. Nature. 2012; 484(7392):96–100. https://doi.org/10.1038/nature10856 PMID: 22367540

52. Truscott J, Ferguson NM. Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. PLoS Comput Biol. 2012; 8(10):e1002699. https://doi.org/10.1371/journal.pcbi.1002699 PMID: 23093917

53. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. Science. 2006; 312(5772):447–451. https://doi.org/10.1126/science.1125237 PMID: 16574822

54. Xia Y, Bjørnstad ON, Grenfell BT. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. Am Nat. 2004; 164(2):267–281. https://doi.org/10.1086/422341 PMID: 15278849

55. Masucci AP, Serras J, Johansson A, Batty M. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Phys Rev E Stat Nonlin Soft Matter Phys. 2013; 88(2): 022812. https://doi.org/10.1103/PhysRevE.88.022812 PMID: 24032888

56. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M. Human mobility: Models and applications. Phys Rep. 2018; 734:1–74. https://doi.org/10.1016/j.physrep.2018.01.001

57. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proc Natl Acad Sci U S A. 2015; 112(38): 11887–11892. https://doi.org/10.1073/pnas.1504964112 PMID: 26351662

58. Merler S, Ajelli M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. Proc R Soc Lond B Biol Sci. 2010; 277(1681):557–565. https://doi.org/10.1098/rspb.2009.1605

59. Ciofi degli Atti ML, Merler S, Rizzo C, Ajelli M, Massari M, Manfredi P, et al. Mitigation Measures for Pandemic Influenza in Italy: An Individual Based Model Considering Different Scenarios. PLoS One. 2008; 3(3):e1790. https://doi.org/10.1371/journal.pone.0001790 PMID: 18335060

60. Kinlaw K, Barrett DH, Levine RJ. Ethical guidelines in pandemic influenza: recommendations of the Ethics Subcommittee of the Advisory Committee of the Director, Centers for Disease Control and Prevention. Disaster Med Public Health Prep. 2009; 3(S2):S185–S192. https://doi.org/10.1097/DMP.0b013e3181ac194f PMID: 19675459

61. Duijzer LE, van Jaarsveld W, Dekker R. Literature review: The vaccine supply chain. Eur J Oper Res. 2018; 268(1):174–192. https://doi.org/10.1016/j.ejor.2018.01.015

62. Wood J, McCaw J, Becker N, Nolan T, MacIntyre CR. Optimal dosing and dynamic distribution of vaccines in an influenza pandemic. Am J Epidemiol. 2009; 169(12):1517–1524. https://doi.org/10.1093/aje/kwp072 PMID: 19395691

63. Tuite AR, Fisman DN, Kwong JC, Greer AL. Optimal pandemic influenza vaccine allocation strategies for the Canadian population. PloS One. 2010; 5(5):e10520. https://doi.org/10.1371/journal.pone.0010520 PMID: 20463898

Paper III

Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh

III