Pandemic panic: A network-based approach to predicting social response during a disease outbreak

by

Shannon M. Fast

B.A., Mathematics and Psychology, Saint Olaf College (2012)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Master of Science in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

Signature redacted

Author .....
Sloan School of Management
May 16, 2014

Signature redacted

Certified by .....
Dr. Natasha Markuzon
The Charles Stark Draper Laboratory, Inc.
Technical Supervisor

Signature redacted

Certified by .....
Professor Marta Gonzalez
Assistant Professor of Civil and Environmental Engineering
Thesis Supervisor

Signature redacted

Accepted by .....
Professor Patrick Jaillet
Dugald C. Jackson Professor, Department of Electrical Engineering and Computer Science
Co-director, Operations Research Center

Pandemic panic: A network-based approach to predicting social response during a disease outbreak

by

Shannon M. Fast

Submitted to the Sloan School of Management
on May 16, 2014, in partial fulfillment of the
requirements for the degree of
Master of Science in Operations Research

Abstract

Epidemic trajectories and associated social responses vary widely between populations, with severe reactions sometimes observed. When confronted with fatal or novel pathogens, people exhibit a variety of behaviors from anxiety to hoarding of medical supplies, overwhelming medical infrastructure and rioting. We developed a coupled network approach to understanding and predicting social response to disease spread. We couple the disease spread and panic spread processes and model them through local interactions between agents. The behavioral contagion process depends on the prevalence of the disease, its perceived risk and a global media signal. We verify the model by analyzing the spread of disease and social response during the 2009 H1N1 outbreak in Mexico City, the 2003 SARS and 2009 H1N1 outbreaks in Hong Kong and the 2012-2013 Boston influenza season, accurately predicting population-level behavior. The effect of interventions on the disease spread and social response is explored, and we implement an optimization study to determine the least cost intervention, taking into account the costs of the disease itself, the intervention and the social response. We show that the optimal strategy is dependent upon the relative costs assigned to infection with the disease, intervention and social response, as well as the perceived risk of infection. This kind of empirically validated model is critical to exploring strategies for public health intervention, increasing our ability to anticipate the response to infectious disease outbreaks.

Thesis Supervisor: Dr. Natasha Markuzon
Title: The Charles Stark Draper Laboratory, Inc.
Technical Supervisor

Thesis Supervisor: Professor Marta Gonzalez
Title: Assistant Professor of Civil and Environmental Engineering
Thesis Supervisor

Acknowledgments

I am grateful for all of the assistance and advice I have received during my time at MIT.

I would like to thank Dr. Natasha Markuzon, my supervisor at Draper Laboratory, for her guidance. Her suggestions and her patience were invaluable, as I struggled to formulate the model. She taught me to begin with a picture or a reduced version of the model and to work up from there. Working with Natasha has improved my abilities as a researcher and a writer.

I would like to thank Professor Marta Gonzalez, who has been my MIT adviser. She pointed me to helpful papers that expanded my knowledge and helped me improve the model. She urged me to work on improving the visualizations I created and taught me that being able to show the reader your results in an attractive and straight-forward way is vital to good research.

I am grateful for the extensive biosurveillance database made available to me through AscelBio and for the assistance and expertise of Dr. Jim Wilson. Dr. Wilson has provided operational biosurveillance insight into the problem of social response and helped to ensure that the model's results are valid.

I would like to thank Professor Gerardo Chowell of the School of Human Evolution and Social Change at Arizona State University, who provided access to the case counts I used to approximate the spread of H1N1 in Mexico City.

I would also like to thank Draper Laboratory for supporting my graduate education and to acknowledge partial funding by DTRA contract HDTRA1-12-C-0061.

Finally, I would like to thank my friends, family and loved ones. They have been a sounding board for new ideas and provided support and encouragement when research or classes were difficult.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, we present a network-based model of social response to an infectious disease outbreak. To help in selecting parameters for the model, we analyze when and why social response occurs using data from a global biosurveillance database. We validate the model using real data from disease outbreaks in Hong Kong, Mexico City and Boston. The model replicates the population-level social response for these outbreaks. The effect of interventions, such as contact reduction and vaccination, on disease spread and social response is explored.

1.1 Social response

Advances in modeling and data availability have greatly improved our ability to predict and control epidemics [1]. The next step is to unravel how diseases shape behavioral norms [2, 1]. In most cases, the social norm during a disease outbreak is to be calm, cooperative and adaptable [3]. There are cases, however, when the outbreak of disease can trigger social disturbances. Mild social disturbances include increased anxiety or mild behavioral change such as improvements in hygiene practices. More extreme social disturbances can also occur. These include panic, rioting, hoarding of medical supplies, fight from the area or violence against members of groups believed to have or carry the disease [4, 5, 6, 7, 8, 9]. We call social disturbances observed in response to a disease outbreaks *social response*. More precisely, we define social responses as behavioral or emotional manifestations of concern about disease. To further illustrate what is meant by social response, we will discuss several examples from recent history. Some of these examples are quite dramatic, such as protests and flight; other social response are more mundane, such as economic effects and increased concern about the disease.

Cholera, Haiti: 2010 to present

In 2010, cholera reemerged in Haiti after having been absent from the island for more than a century. Cholera is a diarrheal disease and can cause death if the patient is not properly rehydrated. Cases of cholera were first reported in Haiti on October 19, 2010 in the Artibonite province and then spread throughout the country [10]. As of 2011, the Haitian Ministry of Public Health and Population reported over 93,000 cases and over 2,100 fatalities. At the beginning of the outbreak, protests and violence broke out throughout the country over suspicions (now supported by evidence [10]) that a UN military base was the source of the epidemic. These protests slowed the delivery of treatment and likely cost lives [5].

Ebola, Uganda: 2000 to 2001

The largest Ebola outbreak on record took place in Uganda between September 2000 and February 2001. Ebola is a rare, but highly fatal disease. Early symptoms include fever, headache, diarrhea and nausea. The disease rapidly progresses to include rash, organ failure and sometimes internal or external bleeding. In the Uganda outbreak, 425 people became infected and 53% of those died [6]. The social response was severe [6]. Medical workers fled their posts. Some individuals were reported to have drank and bathed in bleach in an attempt to rid themselves of the virus, and Ebola survivors and their families were stigmatized by the community. In one town, hundreds of protesters, angry over the decision to place a Ebola burial site near the town, converged upon a government office. Government officials and members of the burial team received death threats over their perceived involvement in the spread of disease.

SARS, Hong Kong, special administrative region of China: 2003

Hong Kong was one of the areas most severely affected by the 2002-2003 SARS outbreak. In Hong Kong, 1755 cases were reported during the summer of 2003, with 299 deaths [11]. During the outbreak, residents of Hong Kong reported increased anxiety [12]. There were also behavioral effects. Residents stocked up on food and supplies, while opportunistic retailers took advantage of the situation, selling knock-off Louis Vuitton face masks and snake-oil medicines [7].

SARS had a devastating impact on the Hong Kong economy [13]. Tourists and foreign businesspeople canceled their travel plans, resulting in a 63 percent drop in the total number of visitors during March and April of 2003, compared with the same months in 2002. Retail sales plummeted, as residents stopped non-essential travel outside their homes. Many workers in the most affected sectors of the economy were laid off or had their hours cut as a result of the reduction in demand.

Dengue fever, Singapore: 2013

The 2013 dengue fever season in Singapore was the worst in recorded history. Dengue fever is a mosquito-borne virus that causes severe joint and muscle pain. Prior to 2013, the worst outbreak recorded in Singapore was in 2005, when approximately 14,000 people became infected and 25 people died. In 2013, there were 22,318 reported cases and 7 deaths [14]. There were several signs of increased concern about dengue fever in the population. There was an increase in reporting of potential mosquito breeding sites. Doctors reported paranoid patients with mosquito bites but no symptoms arrived at their offices worried that they would become ill [15].

1.2 Motivation for research

As seen with SARS in Hong Kong, social response can have extreme economic costs. In addition, there are intangible costs associated with social response. Social response can impede the ability of responders to control outbreaks of infectious disease and can result in strain on infrastructure and the economy [9]. Moreover, social responses often herald a significant change in the epidemiological pattern or etiology of an infectious disease event. There is therefore interest in understanding why social response occurs and what can be done to prevent it. In this thesis, we develop a model for the joint transmission of disease and social response. Using this model, we explore the effects of control measures on social response and find least cost strategies for disease control, taking into account the costs of social response.

1.3 Modeling approaches and objectives

The primary model which will be discussed in this thesis is a network-based model describing the temporal spread of disease and social response. The following traits were deemed to be desirable in the model:

1. Be adaptable to different types of disease and different societies.
2. Allow for simulation of high incidence diseases without social response and low incidence diseases with social response.
3. Incorporate the effects of the media on social response.
4. Incorporate insights from social psychology risk and perception theory.
5. Incorporate modern mathematical and computational methods.

We validated the model by simulating disease outbreaks with and without social response. In Hong Kong, we compared the SARS outbreak of 2003 with the H1N1 outbreak of 2009. In Mexico City,

we compared the spring and fall outbreaks of H1N1 in 2009. Finally, in Boston, we simulated the social response resulting from the declaration of a public health emergency during the 2012-2013 influenza season. These simulations showed that the model can accurately replicate the social response observed in real infectious disease outbreaks. We also used this model to study how control measures affect social response. Using an optimization approach, we showed how the least cost disease control strategy is affected by the perceived risk of infection and the relative costs of infection, intervention and social response.

1.4 Thesis organization

In Chapter 2, we consider the causes of social response, looking at both risk perception theories and data. In Chapter 3, we conduct a literature review of models of disease spread and associated social processes and explain how the agent-based model we introduce expands upon these models. In Chapter 4, we introduce our agent-based model of disease spread and social response spread. We describe how disease and social response are spread over time and explain how measures to control the spread of disease are modeled. In Chapter 5, we formulate an optimization problem to find the least cost disease control strategy, accounting for the costs of the intervention, the spread of infection, and the social response. In Chapter 6, we explain how the simulations were implemented and introduce the studies conducted with the model. In Chapter 7, we describe our results. We explain the conditions under which the model produces a social response. We explore the effects of implementing control measures on the disease spread and determine least cost disease control strategies. Finally, we demonstrate the model's ability to replicate social response observed in real outbreaks by simulating the disease spread and social response for SARS and H1N1 in Hong Kong, for the two waves of H1N1 in Mexico City, and for the 2012-2013 influenza season in Boston. In Chapter 8, we summarize our work and discuss directions for future work.

Chapter 2

Causes of Social Response

In order to develop a realistic model of how social response emerges and spreads through populations during a disease outbreak, we needed to understand when social response is most likely to occur. In this chapter, we will discuss theories of risk perception and explain why the social amplification of risk framework [16] is appropriate for modeling reactions to epidemic disease. The social amplification of risk framework predicts that diseases that are novel or unfamiliar to local experts or which have high clinical severity will be more likely than result in social response. Using a global database of disease outbreaks, we will explore the validity of these predictions by comparing the rate of social response for outbreaks with different characteristics.

2.1 Theories of risk perception and health behaviors

There are several social psychology theories that explain how individuals interpret and respond to health risks. The Health Belief Model (HBM) [17] and Protection Motivation Model (PMM) [18] are two popular expectancy-value models. They treat the decision to take protective action against disease as a cost-benefit analysis. The perception of risk is assumed to be proportional to the severity of disease and the likelihood of contracting it. People then take protective action in proportion to their perceived risk. While expectancy-value models have proven useful in studying health risks such as cancer and heart disease and in encouraging use of protective equipment to prevent injury [18], they are not necessarily appropriate for use in modeling epidemic disease [19]. Value-expectancy models assume that individuals are making rational health decisions. These models neglect the emotional component of health decision making and the social context in which individuals make health-related decisions [19]. Moreover, individuals who are aware of and concerned about the disease sometimes take actions that are not protective. For example, rioting can speed the spread of disease rather than slow it.

In light of the shortcomings of other risk perception theories for modeling epidemic diseases, we

decided to use the social amplification of risk framework, originally proposed by Kasperson et al [16]. The framework assumes that perceptions about the risk of disease are formed within a social context, and factors such as the predictability of the disease risk and the society's familiarity with the disease will affect how the disease is perceived. The perceived risk of infection is therefore not necessarily the same as or even proportional to the true likelihood of infection. Biosurveilliance experts associate social responses with a combination of several conditions, including novelty to the society in question, clinical severity, availability of countermeasures and extent of spread [8, 9].The social amplification of risk framework affirms the role of these conditions and suggests a possible mechanism [16]. The framework asserts that risk-related information is often communicated in such a way that the perceived risk of an event is amplified. Amplification of risk is particularly likely when the volume of communication is high, facts are disputed, there is dramatization and the risk is novel or not well understood. In the context of social response to disease outbreaks, Kasperson's framework suggests that commonplace diseases are unlikely to produce risk amplification, whereas novel, severe diseases, such as SARS or MERS, may provoke amplification, since the public feels that the risk is unpredictable and uncontrollable. We incorporate the social amplification of risk framework into our network-based model by creating social response amplification when the characteristics of the disease, including its novelty and clinical severity, led to it being perceived as high risk by the population.

2.2 Analysis of historical biosurvelliance data

2.2.1 Description of Ascel Bio database

In order to gain insight into the relationship between social response and disease, we analyzed historical biosurveillance data. These data were collected by Ascel Bio [20], a biosurveillance company that gathers data on infectious disease outbreaks. The data consist of near-real-time, multi-source reports on 11,926 different disease outbreaks. The reports were assembled by biosurveillance experts and cover a time frame of one year from May 2008 to May 2009. They include over 200 countries and 300 diseases. The reports indicate whether or not a social response was observed and describe other features of the outbreak, such as where the outbreak took place and what steps were taken to contain it. Many of the reports track the progress of a single outbreak over time. We will look at whether social response was reported at any time in the outbreak.

In these data, it is rare to see social response. Only 5% of disease outbreaks had reported social response. Some regions had higher rates of social response than others. Figure 2-1 shows the geographical distribution of reports with and without social response. Africa and the Caribbean had high rates of social response (9.8% and 12.3% respectively), whereas Australasia and North America had relatively low rates (2.9% and 2.6%). It is important to note that a high rate of social

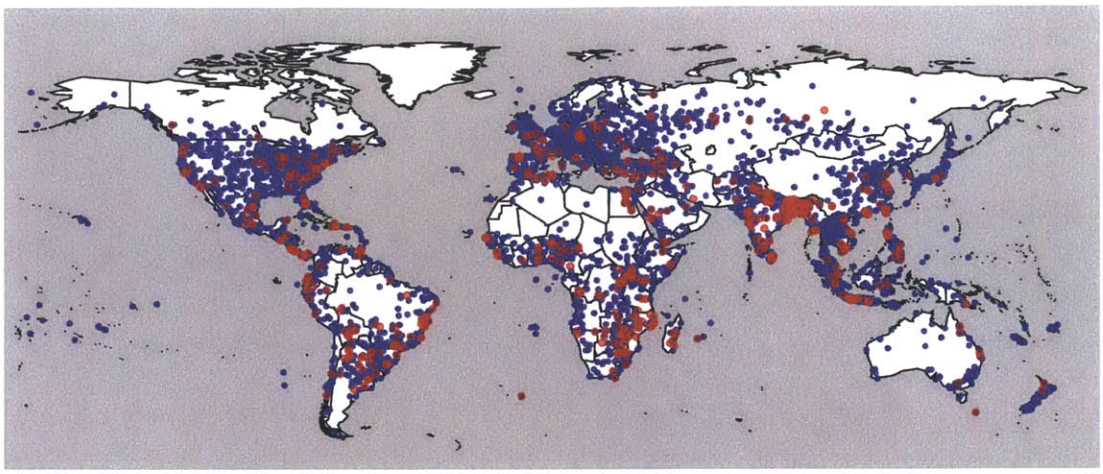


Figure 2-1: Geographical distribution of disease outbreak reports. The 11,926 disease outbreaks recorded by Ascel Bio between May 2008 and 2009 are shown. The majority of events were not associated with social response (blue); there are, however, regional clusters of events with social response (red). Overall, about 5% of outbreaks had reported social response.

response in one region compared to another does not necessarily imply that the populations of the regions would respond differently to identical disease outbreaks. The regions are affected by different types of disease and the size of outbreaks differs between regions. Moreover, our sample is not a random sample of all disease outbreaks. There could be a selection bias in which outbreaks Ascel Bio experts chose to report on. Nevertheless, the Ascel Bio data cover a large number of outbreaks, and, accepting that the data are not comprehensive, we can use them to gain insight why social response might occur.

Figure 2-2 shows the number of outbreak reports broken down by region for ten different diseases. We can see that cholera was most reported on in Africa and the Middle East, influenza received attention in North America, Europe and East Asia, West Nile virus was reported on almost exclusively in North America and plague (yersinia pestis) was not commonly reported in any region. Figure 2-3 shows the proportion of the outbreaks with reported social response. Since the overall rate of social response is 5%, regions with social response rates over 5% can be considered to have above average rates of social response for that disease. Very high rates of social response (above 20%) should be interpreted with caution, since they tend to result from having very few reports for the disease in the region.

When interpreting Figures 2-2 and 2-3, it is important to keep in mind that diseases have different meanings and different importance to public health in different regions of the world. We must interpret the rate of social response in context. For example, almost all regions reported outbreaks of gastroenteritis. However, only Africa and the Caribbean had high rates of social response to gastroenteritis outbreaks. These findings do not imply that residents of Africa or the Caribbean are more prone to social response than other people. Instead, the effects of the disease are fundamentally

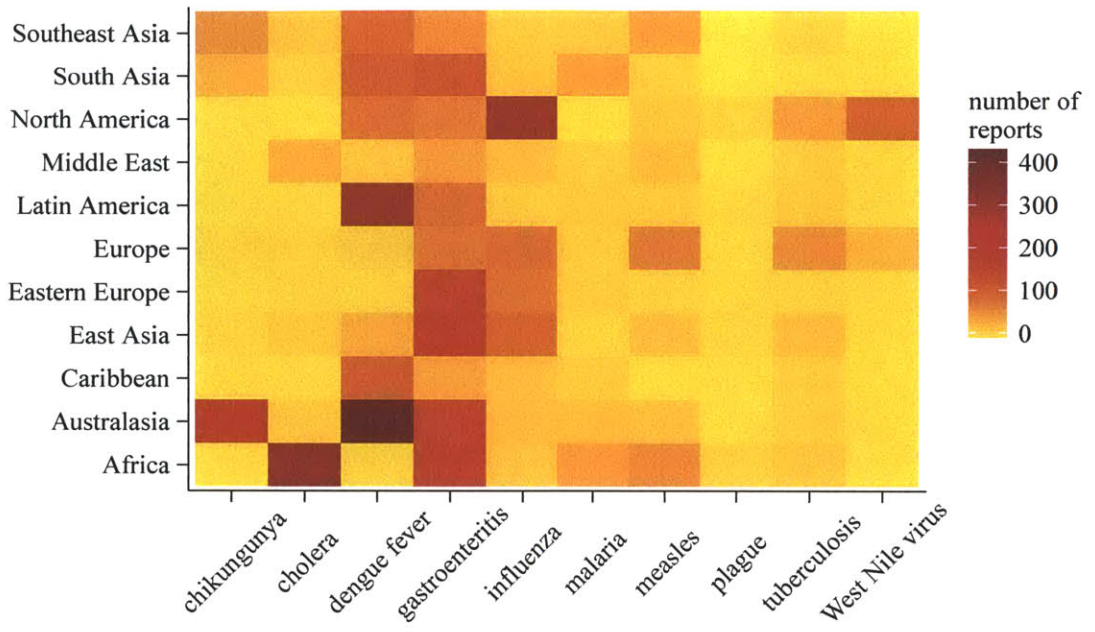


Figure 2-2: Number of outbreak reports by region and disease. The number of Ascel Bio reports in each region is shown for ten different diseases (scale from 0 to 400 reports). The diseases most commonly reported on are different for each region. For example, influenza is commonly reported on in North America, Europe and East Asia, while reports of cholera are more common in Africa and the Middle East. Importantly, greater reporting of a disease in one region does not necessarily imply a greater burden of disease in that area. For example, Europe has a relatively small number of measles cases, but a relatively high number of reports. Since measles has been largely eradicated in Europe, even small pockets of disease receive attention.

different. According to the World Health Organization (WHO), approximately 527,000 children under the age of five die from diarrhea associated with rotavirus (one of several viruses that cause gastroenteritis) every year. Over 85% of these children live in low-income countries in Africa and Asia. Figure 2-4 shows the disability-adjusted years of life lost to diarrheal diseases per 100,000 inhabitants. Africa has a more disability-adjusted years of life lost to diarrheal diseases than any other continent. Since, in Africa, gastroenteritis is the cause of death for many young children, it is reasonable that the rate of social response would be high.

Measles is another example of why the proportion of social response should be interpreted in context. The North America region has a relatively high rate of social response to measles, 7.7%. There are hardly any cases of measles reported in the United States each year. In fact, in 2012, only 55 cases were reported [21]. Measles outbreaks in the United States frequently take place among families and communities that advocate not vaccinating children. Thus, measles outbreaks bring to the forefront an ongoing conflict between parents and health professionals over the safety and efficacy of vaccines.

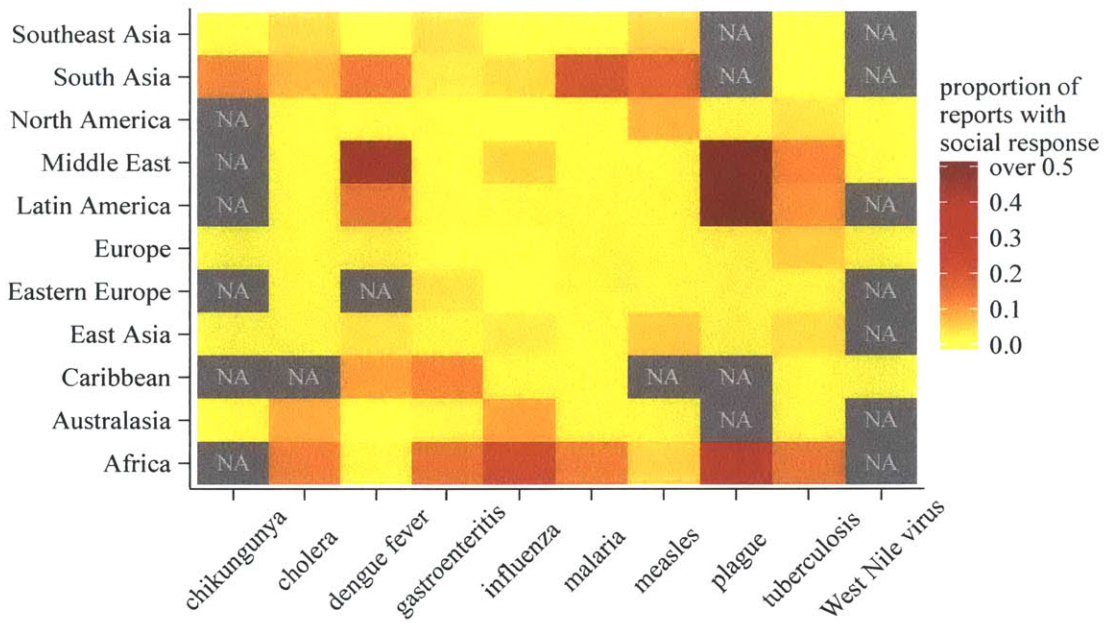


Figure 2-3: Proportion of outbreak reports with social response by region and disease. The proportion of outbreaks with social response is shown (scale from 0.0 to 0.5). If the database contained no reports of the disease in the region, the box is grayed out. The proportion of outbreaks with social response varies by region and disease, because diseases have different meanings in different places. For example, dengue fever outbreaks are more concerning in Latin America and the Middle East, where dengue is an emerging disease than in Southeast Asia, where dengue has been endemic for hundreds of years.

2.2.2 The roles of disease novelty and clinical severity in social response

The social amplification of risk framework predicts that social response is likely when a disease is novel to a region. To determine whether that prediction is consistent with the available data, we looked at the relationship between local experts' familiarity with the disease in question and the social response. Lack of familiarity is usually a mark of novelty. Using text parsing, Ascel Bio coded the reports that the biosurveillance experts generated based on whether the disease was unknown or mysterious to local experts and whether the outbreak was unusual or atypical for the region. As shown in figure 2-5, social response occurred in 16% of outbreaks that involved diseases that were coded as either unknown or mysterious or as unusual or atypical for the region. Social response occurred only 4% of the time for outbreaks of diseases without those properties, a statistically significant difference (Pearson's Chi-square Test: $\chi^2 = 148.1$, df= 1, $p < 0.001$). In the database, the incidence of social response in countries with little prior experience with the disease was consistently higher than in endemic countries, even if the rate of spread was lower. For example, social response occurred in over 10% of dengue fever outbreaks in Latin America where dengue fever is relatively novel, but only in 1% of outbreaks in Southeast Asia where dengue fever is endemic. The observed

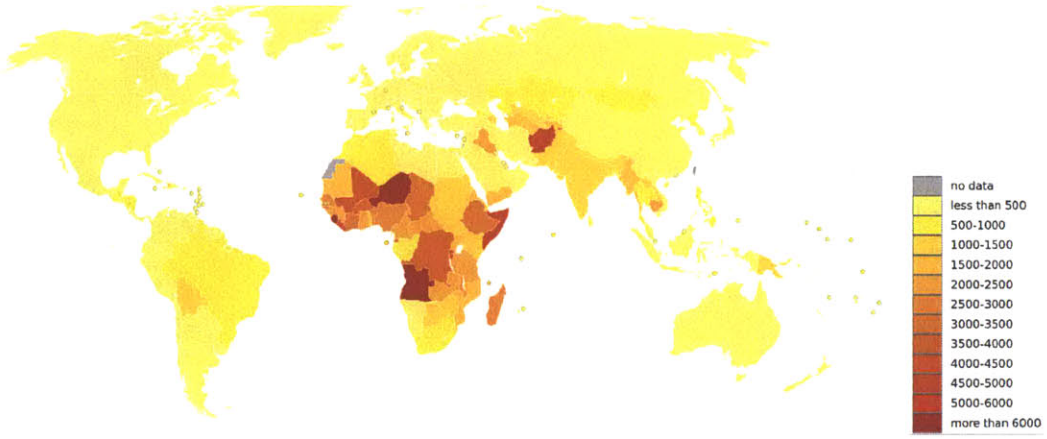


Figure 2-4: Disability-adjusted life years (DALY) lost to diarrheal diseases by country per 100,000 inhabitants. Due to a high number of young children dying from diarrhea, Africa has large number of disability-adjusted life years lost to diarrheal disease, explaining in part the high rate social response to gastroenteritis in Africa. Figure reprinted from Wikimedia Commons. The data were obtained from the WHO.

difference was statistically significant (Pearson's Chi-square Test: $\chi^2 = 11.0$, df= 1, $p < 0.001$). Similar effects were seen with polio. In 1988, the WHO launched the Global Polio Eradication Initiative, which has had spectacular success, reducing the number of cases by over 99%. Today only three countries remain endemic for polio: Afghanistan, Pakistan and Nigeria [22]. Notably, of the 127 polio outbreaks included in the Ascel Bio database, all four outbreaks with reported social response were in India. India shares a border with Pakistan and is at nearly constant risk of having a resurgence of locally-transmitted polio. It is likely that India's unique position on the battlefront against polio helps to explain why its rate of social response is higher than those of other countries.

Kasperson's framework suggests that clinical severity will play a role in social response to disease outbreaks. The Ascel Bio data support this claim. Clinical severity was a significant predictor of social response. Diseases were grouped based on human disease biosafety levels (BSL) [23]. BSL describes the level of danger associated with working with a particular microbe, and ranges from 1 (unlikely to infect human adults) to 4 (severe or fatal disease without available treatment). Table 2.1 shows examples of diseases that fall into each BSL category.

BSL-2	BSL-3	BSL-4
cholera	Creutzfeldt-Jakob disease	Ebola hemorrhagic fever
influenza	SARS coronavirus	Lassa fever
malaria	West Nile virus	Marburg virus
measles	yellow fever virus	Argentine hemorrhagic fever
dengue fever	Chikungunya	Crimean-Congo hemorrhagic fever

Table 2.1: Examples of diseases in each biosafety level (BSL). BSL describes the level of hazard involved in working with a particular disease. BSL ranges from 1 (unlikely to infect human adults) to 4 (severe or fatal disease without available treatment).

Because BSL-1 diseases do not commonly infect humans and almost no outbreaks of BSL-1 diseases were reported in the Ascel Bio database, we have excluded them from the analysis. Diseases with BSL-4 were significantly more likely to have an associated social response than diseases with lower BSL (Pearson's Chi-square Test: $\chi^2 = 26.7$, df= 3, $p < 0.001$; figure 2-5). BSL-4 diseases were associated with social response 17% of the time, whereas social response occurred in only 5% of outbreaks of diseases with lower BSL. In other words, the data indicate that the most severe diseases are more likely to be associated with social response than less severe diseases.

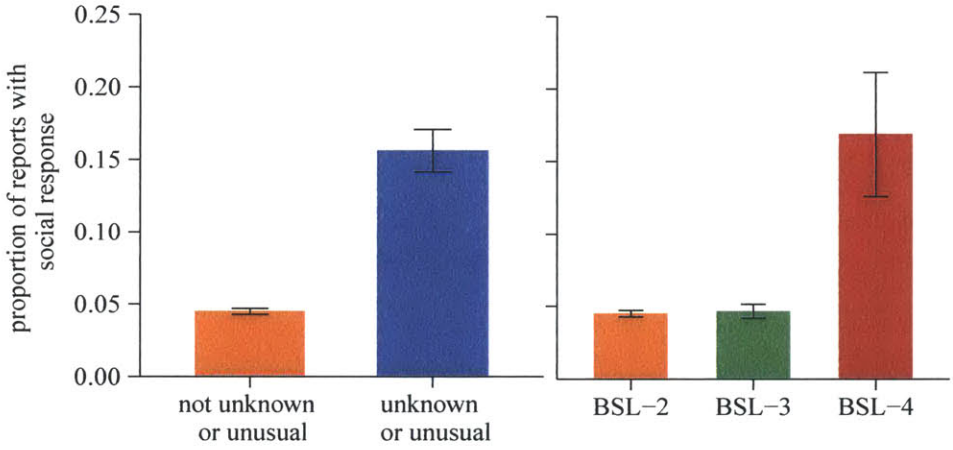


Figure 2-5: Social response most frequently occurs when the disease is novel to the region or clinically severe. If a disease is unusual or atypical in a region or seen as unknown or mysterious to local experts, there is an increased likelihood of social response (left). Outbreaks of diseases with the highest biosafety level (BSL) are more likely to have social responses than diseases with lower BSL (right). BSL-4 diseases are severe and fatal, with no available treatments.

2.3 Dengue fever: Factors contributing to social response

In the previous section, we saw that data support the roles of novelty and clinical severity in social response to disease. In this section, we will discuss a more comprehensive investigation of which factors are associated with social response to dengue fever outbreaks. We were interested in the roles of the incidence of disease and characteristics of the country, such as the economic and political environment, in producing social response.

In this analysis, we focused exclusively on dengue fever outbreaks. By considering only one disease, we did not need to concern ourselves with possible interactions between other predictors and disease type. Therefore, we could more completely isolate the effects of those predictors. For example, we were interested in how the number of cases is associated with social response. While ten cases of SARS might result in more social response than one-hundred cases of dengue fever, it would be unreasonable to conclude that social response is more likely if there are fewer cases. By

comparing the response to 10 cases of dengue fever with the response 100 cases of dengue fever, we were more justified in making inferences about the effect of the number of cases on social response. While ideally it would be possible to control for the effects of different types of diseases by including disease type as an attribute, this suggestion proved infeasible for our data since over 300 diseases were represented.

We selected dengue fever outbreaks to investigate, because dengue is a relatively common disease in the Ascel Bio database and a relatively high frequency of dengue outbreaks resulted in social response. Dengue fever is a mosquito-borne disease, which infects between 50 and 100 million people annually [24]. Symptoms including high-fever, headache, and joint, muscle or bone pain. Occasionally, dengue fever progresses into dengue hemorrhagic fever, which can be dangerous [24]. Dengue is an emerging disease. In the 1950s, cases were primarily limited to Southeast Asia. Now, in addition to Southeast Asia, dengue fever cases are reported in Africa, Latin America, the Caribbean and parts of Australia [24]. Using data from the CDC, WHO, GIDEON, journal articles and news reports, Brady et al. mapped the global footprint of dengue fever as of 2012 [25]. Their map is reprinted in Figure 2-6.

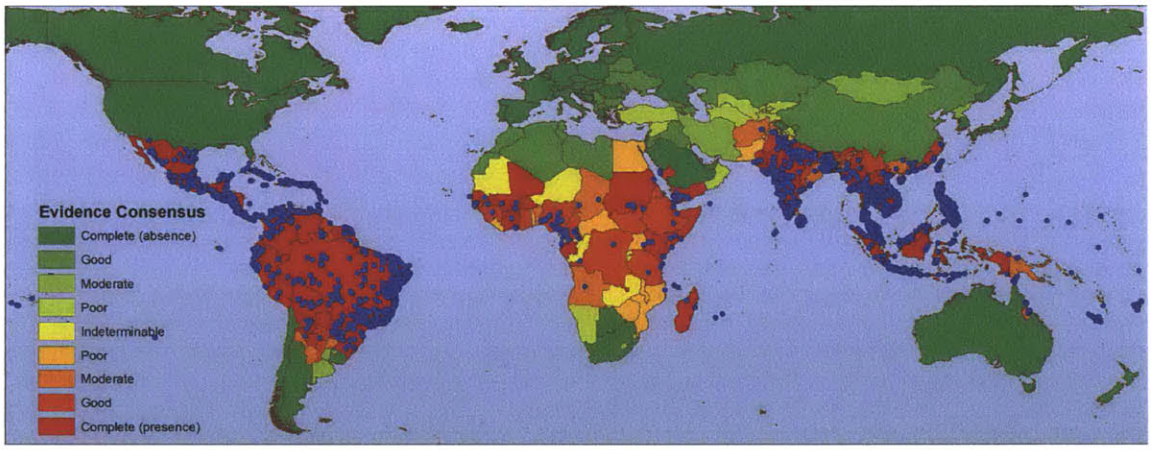


Figure 2-6: The global footprint of dengue fever. While dengue cases were once restricted to Southeast Asia, dengue is now reported in tropical and subtropical locations across the globe. Countries are color-coded based on the evidence of dengue presence (red) or absence (green). Blue dots indicate reported dengue outbreaks. We thank Brady et al. [25] for the use of this figure.

2.3.1 Description of dengue fever outbreak data

The Ascel Bio data contain information suitable to identifying different types of social responses. In this thesis, we will focus on public anxiety. Public anxiety is an indicator for fear, panic or behavioral manifestations of fear or panic being reported in the public. Public anxiety was coded as a binary response. If any type of anxiety was noted, public anxiety was coded as 1 and otherwise public anxiety was coded as 0. In total, 100 predictors were used to differentiate between outbreaks

Event ID	Country	Date	X1	X2	...	X99	X100	Public Anxiety
1	Country 1	6/1/2009	0	0		8000	17	0
1	Country 1	6/2/2009	0	0		8000	17	1
1	Country 1	6/3/2009	1	0		8000	17	0
2	Country 2	8/9/2008	0	1		680	3	0
3	Country 3	1/17/2008	1	1		1300	9	1
3	Country 3	2/3/2008	1	0		1300	9	1

(a) Data prior to collapsing

Event ID	Country	Date	X1	X2	...	X99	X100	Public Anxiety
1	Country 1	6/1/2009	1	0		8000	17	1
2	Country 2	8/9/2008	0	1		680	3	0
3	Country 3	1/17/2008	1	1		1300	9	1

(b) Data following collapsing

Table 2.2: Data collapsing method used for dengue fever prediction. The tables above show how reports related to a single outbreak were collapsed into a single observation. The sample reports were created for the purposes of illustration and are not actually part of the data.

with and without social response. Of these, 40 were binary predictors obtained from Ascel Bio. These predictors were related to the status of the outbreak itself and included information about the number of cases and whether complications or deaths were observed. Since the Ascel Bio data are proprietary, we will refrain from discussing particular predictors. The remaining predictors were country-level economic, geographic and health-related predictors, obtained from the UN, WHO, CIA World Factbook and other public sources. Table A.1 lists the predictors used in the analysis that were obtained from public sources. In general, predictors were selected for inclusion because they were descriptive of the political system, healthcare infrastructure, media infrastructure or culture of the country, or because there could be an association between the predictor and the likelihood of a dengue fever outbreak (annual precipitation is an example).

For many dengue fever outbreaks, the Ascel Bio reports tracked the progress of the disease over time. Since the social response observed later in an outbreak depends upon the disease and social conditions earlier in the outbreak, reports related to the same outbreak were not independent. Therefore, we collapsed together all reports related to a single outbreak into one observation. The 40 binary Ascel Bio predictors changed within an outbreak, as the disease conditions changed. We chose to set these predictors to 1 for the outbreak if they were 1 for any of the reports related to the outbreak. Similarly, public anxiety was coded as 1 for the outbreak if public anxiety was reported at any time during the outbreak. The 60 predictors related to the economic, political and healthcare statuses of the countries did not change during the outbreaks and so were unaffected by collapsing. Table 2.2 illustrates the collapsing process for a set of sample reports created for the purposed of this illustration. Following collapsing, there were 842 outbreaks without reported social response and 64 outbreaks with reported social response.

2.3.2 Modeling techniques

Boosted classification trees

We selected to use boosted classification trees to predict social response. Classification trees follow a greedy algorithm to partition the space into rectangular regions. The goal is to find a partition such that achieves a high level of purity in the regions [26]. Boosting is a ensemble method used to improve the predictive performance of a weak classifier. We used Discrete AdaBoost, as described by Fruend and Shapire [27] and implemented in the R package, ada [28]. Discrete AdaBoost iteratively predicts the response in the training data using a weak classifier, in our case a classification tree. First, the response is coded such that 1 indicates a positive instance and -1 indicates a negative instance. At each boosting iteration, m, a prediction is generated for each observation in the training set. Then the weights of observations that were misclassified are increased for the next iteration, resulting in greater focus on correctly predicting those observations. The m^{th} iteration classifier, $T_m(x)$, is given a weight α_m based on how successful the classifier is at correctly predicting the response. Then, the final classification is given by $T(x) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m T_m(x)\right)$ [27].

The decision to use classification trees was based on the data structure. Many of the predictors of interest contain a a large number of missing values. Thus, removal of observations with missing values would result in a substantial loss of data. We therefore decided to choose a classifier that could handle missing data. Additionally, given the large number of predictors, we desired a classifier that would be insensitive to irrelevant predictors. Classification trees handle missing value and are robust to irrelevant predictors, but have relatively poor predictive performance. Therefore, we used boosting to improve the performance.

We took care in selecting the classifier parameters. Having too many branches in the classification trees or having too many boosting iterations will result in overfitting of the training data [26]. Hastie el al. suggest a maximum tree depth of 6, and we adopted this recommendation. They also suggest that the number of boosting iterations be chose to minimize error on a validation sample [26]. For our data, this suggestion proved infeasible. The responses of interest were very rare, so in order to get meaningful error estimates, we would need a large validation set. We did not have enough observations to permit creating such a validation set. Instead, we selected a relatively low number of boosting iterations, 50, with the hope that this choice would not result in overfitting.

Understanding the importance of predictors

In order to determine whether the incidence of disease or economic and political predictors were helpful in predicting social response, we assembled three different datasets:

1. *Full* - All predictors were included.

2. *Disease Status Excluded* - All predictors were included, except those related to the incidence of disease and whether or not there were deaths or cases of dengue hemorrhagic fever.

3. *Condition of Country Information Excluded* - All predictors except those related to the conditions of the country's health, economy or political environment were included.

We predicted social response for each of the datasets and compared the results to determine what role the types of predictors were playing. If removal of predictors resulted in a reduction in true positives or true negatives, we could conclude that the classifier was making use of these predictors to correctly classify outbreaks as having social response or not. If removal of the predictors did not reduce the number of true positives or true negatives, we could conclude that the predictors were likely not playing a useful role in differentiating between outbreaks with and without social response. We also examined the variable importance of each of the predictors. A variable importance measure for boosted trees is defined in Hastie et al [26]. Intuitively, the variable importance of a predictor is determined by how frequently the predictor is used by the classifiers and how effective splitting on that predictor was at partitioning the data into groups with and without social response.

Error estimation

Error was estimated through five-fold cross-validation. The data were randomly split into five groups, keeping the proportion of observations with social response in each group approximately equal to the proportion in the dataset as a whole. The boosted trees were trained on four groups at a time. Predictions were then generated for the fifth group. This process was repeated five times, so that each of the five groups would serve as the test set once. Due to randomness in the algorithm, boosting can produce different results when run again on the same data. Therefore, we repeated the cross-validation 20 times to obtain average counts of true positives, false positives, true negatives and false negatives as well as standard deviations for each of these counts.

Random oversampling and undersampling

When applied to learning problems such as ours, where there is an imbalance between the number of minority class observations and the number of majority class observations, classifiers tend to perform poorly at predicting the minority rules. This poor performance occurs because the data imbalance leads the classifier to learn rules that describe the majority class, while failing to learn rules that describe the minority class [29]. Sampling methods to improve the balance of the minority and majority classes in the training set have been shown to improve classification performance on a number of problems [29].

While balancing data generally leads to improvements in predictive power, both oversampling and undersampling can introduce problems. Undersampling of the majority class results in the

loss of information, which could be helpful in identifying the majority class. Oversampling of the minority class can lead to overfitting to idiosyncrasies in the training set [29]. We used both random oversampling and random undersampling to balance between these two risks. Assume that before balancing there were N_m minority observations in the training set and N_M majority observations. We included all N_m minority observations in the balanced training set once and then added $2N_m$ additional minority observations by taking a bootstrap sample. We undersampled the majority class by taking a bootstrap sample of size $3N_m$ from the N_M majority class observations. The validation set was left imbalanced, so that the proportion of events with social response in the validation set would be reflective of the observed proportion in the data.

2.3.3 Results

The predictions produced show that the classifier was successful at learning some of the rules that differentiate the minority and majority classes. Tables 2.3a, 2.3b, and 2.3c show the averaged false positives, false negatives, true positives and true negatives for the full dataset, disease status excluded dataset and condition of country excluded dataset respectively.

		Predicted Class	
		No Social Response	Social Response
Actual Class	No Social Response	704.8 (std. dev. 4.5)	137.2 (4.5)
	Social Response	25.6 (1.1)	38.4 (1.1)

(a) Boosted classification trees using full data

		Predicted Class	
		No Social Response	Social Response
Actual Class	No Social Response	708.1 (std. dev. 6.8)	133.9 (6.8)
	Social Response	26.7 (1.0)	37.6 (1.0)

(b) Boosted classification trees using data with disease status predictors excluded

		Predicted Class	
		No Social Response	Social Response
Actual Class	No Social Response	663.0 (std. dev. 7.9)	179.0 (7.9)
	Social Response	29.1 (2.0)	34.9 (2.0)

(c) Boosted classification trees using data with condition of country predictors excluded

Table 2.3: Comparison of performance full data model, disease status excluded model and country condition excluded model. We ran AdaBoost 20 times to obtain average values and standard deviations for the number of true positives, true negatives, false positives and false negatives. The predictive performance of the models run on the data with disease status predictors removed was comparable to the performance of the models run on the full data. Removal of country condition predictors resulted in a reduction in true positives and true negatives compared with baseline.

For imbalanced data, accuracy can be a misleading metric, since high accuracy can be achieved by always predicting the majority class [29]. Therefore, we will also consider alternative metrics to

compare the predictions produced using the three datasets. First, we will define accuracy, precision, recall (sometimes called sensitivity or true positive rate) and specificity (true negative rate). Let TP, TN, FP, and FN refer to true positives, true negatives, false positives and false negatives respectively. Then

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

and

$$\text{Specificity} = \frac{TN}{FP + TN}$$

In context, accuracy is the proportion of the time that we correctly predicted social response or no social response. Precision is the proportion of the time that social response occurred when we predicted that it would. Recall is the proportion of the time that we predicted social response when it occurred. Finally, specificity is the proportion of the time that we predicted that social response did not occur when it did not occur.

Table 2.4 shows the accuracy, precision, recall and specificity for the models built with each of the three datasets. We compared the performance with the expected performance of a baseline naive classifier that randomly predicts the minority class 50% of the time. The models built with all three of the datasets had substantially higher accuracy, precision, recall an specificity than the baseline model. The models built with the full dataset and dataset with disease status excluded had similar performance. The model built on the dataset excluding predictors describing the country condition had lower precision and recall compared with the model built on the full dataset.

	baseline	full dataset	disease status excluded	country condition excluded
accuracy	0.50	0.82	0.82	0.77
precision	0.07	0.21	0.22	0.16
recall	0.50	0.60	0.58	0.55
specificity	0.50	0.84	0.84	0.79

Table 2.4: Comparison of accuracy, precision, recall and specificity with the baseline model. The baseline model randomly predicts the minority class 50% of the time. All of the models outperformed the baseline model on all of the metrics. The models built on the full dataset and dataset with disease status excluded had comparable performance. The model built on the dataset with country condition excluded performed more poorly than the model built on the full dataset for each of the metrics examined.

Table 2.5 lists the publicly available predictors with the highest variable importance measures. Several of these predictors are related to healthcare spending and health outcomes in the country. Others describe the relationship between the people and their government.

ranking	predictor
1	public health expenditure as a percentage of GDP (2007)
2	newspaper subscribers per 1,000 people (most recent year available, 1996 or later)
3	refugee population from the country (2008)
4	percentage of females with secondary education (2007)
5	adult male mortality per 1000 (2007)
6	human development index (2007)
7	corruption perception index (2009)
8	adult female mortality per 1000 (2009)
9	trust in the national government
10	perception of safety

Table 2.5: Important publicly available predictors. Variable importance is a measure of how useful each predictor was at differentiating between the majority and minority classes, given the other predictors in the model. The ten publicly available predictors with the highest variable importance are listed.

2.4 Conclusions from analysis of historical biosurveillance data

For the dengue fever prediction, well over 50% of events with a social response were correctly identified using the full dataset. However, since all reports for an outbreak were collapsed into a single observation, the high predictive accuracy does not mean that future social response could be predicted with as much accuracy. Instead this analysis is most helpful in identifying features of outbreaks that tend to be associated with social response.

The comparison of predictions made with different sets of predictors provides insight into which features are of most interest. Exclusion of country-level health-related and economic predictors resulted in a small decrease in the recall, as well as a substantial decrease in the precision. In other words, the classifier became less effective at identifying social response when it occurred and was much more likely to guess that social response occurred when in fact it did not. This effect is observed because having country-level predictors in the data allows the classifier to subset countries into groups that may experience a social response depending upon the outbreak and groups that are unlikely to respond to dengue fever regardless of the conditions of the outbreak. Examination of predictors identified as important revealed that, in general, countries with stronger public health systems are more likely to have reports of social response to dengue fever than countries struggling with poverty and high mortality and morbidity.

The finding that countries with strong public health systems have heightened social response to dengue spread is interesting and unexpected. We believe that this finding is likely observed because

dengue fever is an emerging disease in these countries. Dengue fever has long been endemic to impoverished, tropical areas of the world. As a result of rising global temperatures and greater amounts of travel to and from dengue-affected areas, it is beginning to spread into wealthier countries that have not experienced dengue outbreaks ever or in many years. In these more developed countries, dengue is a relatively novel disease and is also a serious public health concern. As a result, there is an increase in public awareness, and if the disease is not contained quickly, public anxiety. Health officials must move quickly to contain outbreaks, even small ones.

The Transparency International corruption perception index (CPI), the level of trust in the national government and perception of safety were listed among the most important predictors. This finding suggests that the citizens' relationship with their government has an association with the likelihood of social response. Alternatively, the variable importance of these predictors could be an artifact of multicollinearity among the predictors. Many of the health and economic predictors are highly correlated. For example, countries with high health expenditure tend to have low mortality. The trust and perception of safety predictors are strongly correlated with each other but not with the other predictors in the model. Therefore, they may have appeared as important predictors, because they measure something different than the other predictors included in the model. Additional exploration would be necessary to determine the role of multicollinearity in the importance of these predictors.

Predictions made without including information about the number of infected people or whether deaths or complications were observed were not significantly different than predictions made with this disease severity information, indicating that knowledge of the absolute number of cases and whether there were deaths or complications does not provide additional information beyond that provided by the predictors that remained in the model. That said, other Ascel Bio predictors implicitly include information about the extent of disease, so we did not fully isolate the extent of disease spread by removing the number of infections. The relationship between the number of cases of disease and the social response is interesting and should be explored further.

When formulating the network-based model, we attempted to incorporate our findings from data. In the analysis of the entire AscelBio database, we found that novelty of the disease and its clinical severity are associated with social response. The finding with dengue fever that information about whether there were deaths or complications does not help in predicting social response seems to contradict the role of clinical severity. These findings merit further consideration. However, we believe that the findings about BSL and social response are more reliable and therefore constructed our model assuming that disease severity is related to social response. In the model, we estimate the disease's novelty and perceived severity and include these factors in our model through a key parameter, κ, which reflects the perceived risk of the disease.

With dengue fever, we saw that information about the number of cases was not useful in pre-

dicting social response, and, indeed, there are many examples in the AscelBio database in which a high-incidence disease did not result in social response, while a low-incidence disease did. Therefore, we designed the social response model to be relatively robust against changes in the number of cases. Eventually, we plan to develop a model, similar to the one presented here for dengue fever, that will estimate the risk of social response for every disease and every country. We could then use the output of this predictive model to inform the choice of parameters in the social response model. Such a model would be able to take into account factors such as GDP and health expenditure that we were unable to consider in our manual selection of parameters.

Chapter 3

Literature Review

3.1 Models of disease

Although there are many types of disease models in use today, almost all of them are variations on the susceptible-infected-recovered model introduced by Kermack and McKendrick [30] in 1927. We will first describe this model and then move to discussing metapopulation and agent-based models of disease transmission.

3.1.1 The deterministic SIR model

The Kermack-McKendrick model assumes that there is a population of size N, such that all of the individuals in the population are either susceptible to disease, infected or recovered from disease and immune. This population is fully-mixed, meaning that individuals are in contact with everyone else in the network and interactions take place at random. Let $S(t)$, $I(t)$, and $R(t)$ denote the number of susceptible, infected and recovered people respectively at time t. Let β be the infection rate and let γ be the recovery rate. Then the system is defined by three differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

In summary, transitions take place from the susceptible state to the infected state at a rate proportion to the numbers of infected and susceptible individuals. Similarly, transitions take place from the infected state to the recovered state at a rate proportional to the number of infected individuals.

One topic of interest is whether disease outbreaks will grow to epidemic proportions or whether

they will die out before infecting many people. The basic reproductive number, R_0, is the average number of additional cases generated by one infected person. For Kermack and McKendrick's SIR model [30], R_0 is defined as follows:

$$R_0 = \frac{\beta S}{\gamma}$$

When $R_0 > 1$, the average infected person infects more than one other person, resulting in the disease growing to epidemic proportions. When $R_0 < 1$, the disease dies out and does not develop into an epidemic. The reproductive number continues to be used to compare the epidemic potential of various types of diseases.

While Kermack and McKendrick's model has provided many interesting insights into the spread of infectious disease, it has several unrealistic assumptions. For example, it is obviously not true that an infected person will spread disease to anyone else with equal probability. The infected person is most likely to spread the disease to her family members, friends and colleagues. Additionally, the probability of transmission will depend upon the infected person's stage of illness and the health of the immune system of the exposed person. These observations have lead to the development of new types of models of infectious disease, including metapopulation models and agent-based models.

3.1.2 Metapopulation models

While it is unrealistic to assume random mixing in an entire population, it can be a reasonable assumption for smaller sub-populations. Some types of models assume random mixing within sub-populations and restricted mixing between these sub-populations. The sub-populations are formed based on geographical distance (e.g. modeling each town as a sub-population), behavior (e.g. modeling high-risk groups and low-risk groups as sub-populations) or demographic factors (e.g. modeling children, adults and the elderly as sub-populations) [31].

Metapopulation models are epidemic models in which sub-populations are defined by geographic distance. These models allow researchers to spatially track the spread of disease and have proven useful in studying the dynamics of pandemic diseases. For example, the Global Epidemic and Mobility Model (GLEAM) partitions the world into sub-populations, each with a local airport [32]. Then, by integrating commuting and air travel data into the model, the disease transmission between the sub-populations is simulated. GLEAM has been used to predict the peak timing of 2009 H1N1 on a global scale and to evaluate the effects on interventions, such as travel restrictions, on the spread of H1N1 [32]. Metcalf et al. [33] used a metapopulation approach to explain how measles, mumps rubella and whooping cough, which have nearly been eradicated in developed countries, periodically reemerge. By way of travel and migration, regions can become reseeded with cases of disease, resulting in a resurgence of infection.

3.1.3 Agent-based models

Agent-based models explicitly model every individual in the population. The individuals (agents) are assumed to have a pattern of disease-spreading interaction defined by a social network. Agent-based models allow for incredibly detailed modeling of transmission dynamics and can incorporate data on age, co-location in households and the number and locations of schools and workplaces [34].

Agent-based models are particularly useful when the disease transmission network is complex or dynamic, such as the sexually transmitted infection network [35] or when precision in modeling the disease transmission is necessary for evaluating the effects of interventions or estimating the transmission parameters of the disease. The Framework for Reconstructing Epidemic Dynamics (FRED) [36] is a highly detailed agent-based model that has been used to explore disease dynamics in the United States. FRED constructs a synthetic population using census data. Agents spread disease as they travel to and from work and school and interact within their neighborhoods. FRED has been used to explore the effects of various types of local interventions including school closures [37] and vaccination [38]. There are several similar efforts aimed at developing large-scale, realistic agent-based simulations, including FluTE [39], EpiSims [40] and GSAM [41].

Detailed agent-based simulations are useful, because they are highly realistic models of disease transmission. They are computationally expensive [34], however, and frequently do not allow for simulation of areas other than the United States, since many areas of the world lack the census and transportation network data needed for detailed agent-based simulations. Given these shortcomings, we decided to conduct our simulations using an agent-based model on a random network. We were primarily interested in studying social response, and the choice of a relatively simple disease spread model allowed us to limit the overall complexity of the model. Additionally, several of the cities we were were most interested in modeling, Hong Kong and Mexico City, are located outside of the United States. That said, our decision to use a random social network likely results in a loss in realism. Eubank et al. [42] compared the realistic network generated for EpiSims with various types of randomly generated networks. They found that the Erdős-Rényi random graph and the Baribasi-Albert preferential attachment [43] models were poor approximations of realistic social network degree distributions. The configuration and Chung-Lu [44] models could be parametrized to produce networks similar to the EpiSims network. These models are both methods of generating random networks with prescribed degree distributions.

3.2 Models of social processes

In this section, we will introduce the DeGroot model of opinion diffusion. We use this model, with some modifications, for the diffusion of social response. We will discuss criticisms of the DeGroot model and explain why, despite these criticisms, the DeGroot model is applicable for social response

modeling.

3.2.1 The DeGroot model

The diffusion of ideas, opinions and innovations has long been of interest to researchers. One of the first models of opinion diffusion was proposed by Morris DeGroot in 1974 [45]. We use a variation of this model for the diffusion of social response. DeGroot assumes that each individual, i, in the population has an opinion distribution at time t of F_{it}. Additionally, i assigns a weight (p_{ij}) to herself and each of the other individuals in the population, reflecting the degree of influence each person's opinion has on i's future opinion. If i greatly respects the opinions of j, then p_{ij} would be large, and if i does not place much credence in j's opinions, p_{ij} would be small. The weights are normalized such that, assuming a population of size N, $\sum_{j=1}^{N} p_{ij} = 1$. Then as time progresses from t to $t+1$, i updates her opinion to $F_{i(t+1)} = \sum_{j=1}^{N} p_{ij} F_{jt}$. That is, i's opinion after the update is a weighted sum of her own prior opinion and those of the other people in the population. This process is Markov with transition matrix P, which is constructed such that weight p_{ij} occupies the row i and column j of P for all i and j. It can be shown that the opinions will converge to a consensus assuming that all recurrent states of the Markov chain communicate and are aperiodic.

3.2.2 Criticisms and enhancements of the DeGroot model

The DeGroot model is simple to implement and has received extensive research attention. It has been criticized, however, for being unrealistic [46]. First, in the DeGroot model the individuals do not update the weights they assign to other people's opinions. If there is an individual who never updates his opinion, others will continue to change their opinions to be closer to his. Realistically, individuals who interact repeatedly will disregard an opinion that has already been expressed and does not change. Secondly, the DeGroot model will result in arrival at a consensus under almost all realistic situations, since most societies have strongly connected social networks. In the real-world, the arrival at a consensus is the exception, not the norm.

Acemoglu et al. introduced several models to address these concerns. In one model, they assume that there exist two types of individuals in a society: regular agents, who have average levels of influence over their neighbors and forceful agents, who have disproportionate influence [47]. They show that, if all agents receive at least some information from their neighbors, the agents will eventually arrive at a consensus. However, unlike in the DeGroot model where the consensus opinion is uniquely determined by the weights and original opinions, the consensus opinion in the forceful agent model is a random variable. In a second model, they introduce the idea of stubborn agents, who will not update their beliefs [48]. In this case, opinions do not converge to a consensus and persistent disagreement can take place. The ideas presented by Acemoglu et al. have been used to

study the optimal allocation of non-lethal influence in warzones [49, 50, 51] and, as we will discuss later, to develop an initial model of the joint spread of disease and social response [52].

3.2.3 The use of the DeGroot model for social response communication

While we accept that the DeGroot model has shortcomings, we believe that its use is appropriate in our circumstance. We have modified the DeGroot model, so that the weights assigned by each agent are updated dynamically. This change reduces the likelihood that an agent will continually update its opinion to match that of a neighbor whose opinion never changes. Additionally, the fact that a consensus will almost always occur is not a concern for us. Social response is a transient phenomenon. If a social response occurs, we expect social response levels to rise with the rise of cases of the disease and then decline as the disease stops to spread. Once the disease stops spreading, it is reasonable for the agents in our network to move toward a consensus of no social response.

3.3 Models of disease spread and related social processes

Although there are many there are many models of disease transmission [53] and many models for the spread of information over networks [46], there are relatively few models that address the interaction between disease transmission and social processes [54, 2]. The need to develop complex models that incorporate coupled disease contagion and social contagion processes has been emphasized, since the joint dynamics of such systems often differ from what would be expected from either process operating in isolation [2]. Our model uses the social psychology of risk behavior as well as disease and information spread modeling techniques to explain how disease transmission can give rise to social disruptions.

Other researchers have focused on small-scale actions in response to disease spread, such as self-protective behavior. Epstein et al. [55] explored the effects of fear-induced behavioral changes on the spread of epidemics. They assumed that fear of the disease spreads much like the disease itself. Agents can become "infected" with fear and, assuming that they are fearful, these agents either isolate themselves or flee the area. Epstein et al. found that self-isolation behaviors substantially decreased the scale of the epidemic. If, on the other hand, frightened individuals fled the area, the scale of the epidemic could dramatically increase, since these individuals would seed infections in new susceptible populations. Funk et al. [56] also considered the effects of the spread of concern about disease. They modeled the spread of awareness over a network with individuals taking greater self-protection measures in proportion to their disease awareness. They found that awareness can reduce the size of the outbreak, and, under certain assumptions, stop it altogether.

We conducted a simulation study exploring the effects of using complete quarantine, as in the Epstein model, or partial reduction in transmission, as in the Funk model. Given comparable

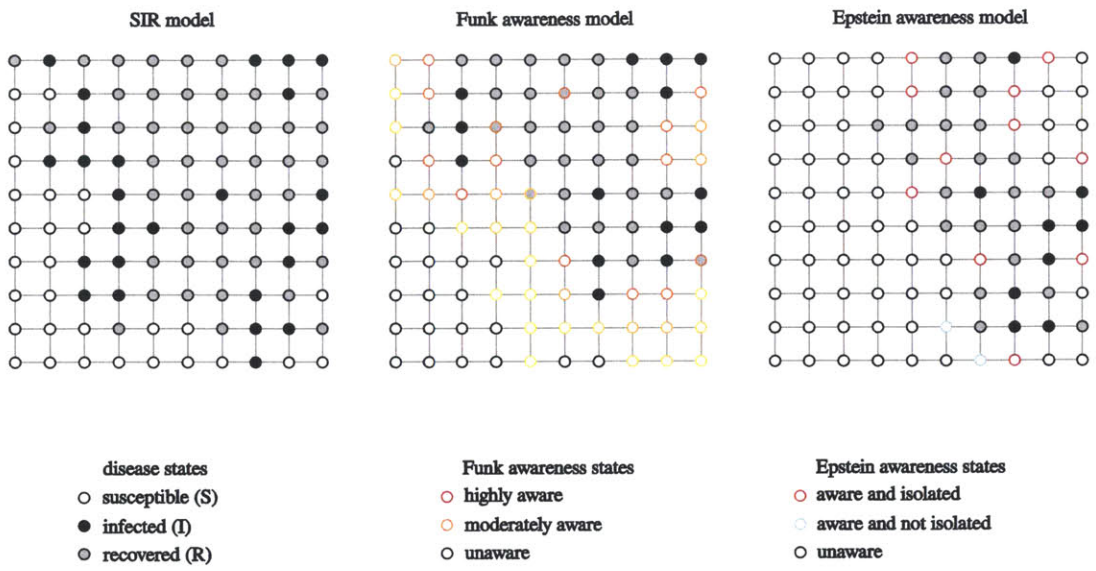


Figure 3-1: Comparison of Funk and Epstein spread of awareness models. We implemented two models of the spread of awareness. Compared with the standard SIR model with no spread of awareness, both models of the spread of awareness resulted in substantial reductions in the number of infections. The Epstein model [55], in which aware nodes isolated themselves, was more effective at halting the spread of disease than the Funk model [56], in which awareness conferred partial protection, even though isolated agents in the Epstein model stopped communicating awareness to their neighbors.

parameters for the spread of awareness, the two models had comparable performance at the early stages of the outbreak. However, overall, the Epstein model isolation formulation resulted in a greater reduction in the spread of disease than the partial protection assumed by the Funk model, even though isolated nodes in the Epstein model stopped communicating awareness to their neighbors. For both models, the individuals in the vicinity of infected persons become aware of the disease, but in the Epstein model this awareness resulted in an absolute stop to the spread of disease, whereas in the Funk model the disease spread was merely slowed. Figure 3-1 shows a standard SIR model on a network with no spread of awareness, an SIR model with the Funk model for awareness and an SIR model with the Epstein model for awareness.

Meloni et al. [57] developed a metapopulation disease model in which the probability of traveling to a different sub-populations and the route taken between sub-populations is determined by the incidence of disease. They assumed that people will be likely to cancel travel to affected areas, and, when they do travel, people will want to route their travel to avoid the disease. In simulations, these measures taken to avoid the spread of disease actually had the opposite effect, since individuals routing their travel to avoid the disease unintentionally spread it to new locations. Meloni et al. concluded that real-time information on a disease spread could reduce the efficacy of disease

containment and mitigation measures.

One recent model explicitly considered the psychology of health behaviors [58]. Working within the framework of the health-belief model, Durham and Casman modeled the likelihood of wearing a face mask during the Hong Kong SARS outbreak as a function of perceived susceptibility to disease, perceived severity of disease, perceived benefits and perceived barriers. They were able to parametrize the model so as to achieve a good fit to the observed prevalence of face mask use during the outbreak. As we noted in Chapter 2, however, the health-belief model model may not be appropriate for epidemic disease modeling, since it disregards the emotional aspects of health-related decision making [19].

3.4 An initial attempt at modeling disease spread and social response

3.4.1 Initial model formulation

The model that we present here builds upon a previous model of disease and social response proposed by Evans [52]. We will briefly describe this model, then explain how the model we introduce modifies and expands upon it. Evans couples an SIR disease model with the information spread models proposed by Acemoglu et al. [48, 47]. Individuals in the population are represented by agents in a network, who spread disease and social response through pairwise interactions. Infected agents act as sources of social response. When an agent becomes infected, its social response is maximized. Then social response is diffused through the network via social interactions. Assume agent i interacts socially with agent j. Then one of four interaction types takes place:

1. Forceful - Agent j adopts agent i's social response.

2. Averaging - Agent j and agent i average their social responses.

3. Decay - Agent j's social response is reduced.

4. Identity - Neither agents social response is changed.

The forceful interaction type is more likely to occur if agent i's social response is elevated. Thus, forceful interactions represent how social response, like disease itself, can be contagious. The averaging interaction type represents the two agents discussing and coming to an agreement about social response. The decay interaction type represents the eventual decline of social response as the disease spread slows. This model was successfully used to model social response to dengue fever outbreaks in Argentina and India [59].

3.4.2 Improvements and expansions

We expand upon Evans' model in several ways. One key change is the introduction of media influence. Some very severe diseases, such as Ebola, require very few cases for a social response to take place. Thus, we wanted to develop a model in which one case of a very severe disease could result in social response in a population of millions of people. Clearly, word-of-mouth communication is insufficient to spread this social response. The signal emanating from a single person is simply too small. Therefore, we needed to introduce media to transmit social response to large numbers of people.

This leads us to the second key change in the model. We needed rules governing which outbreaks would attract attention and which would not. We wanted this decision to be consistent with theories of risk perception and with the available data on when social response takes place. We therefore explicitly introduced the social amplification of risk framework [16] into the model. Amplification takes place when disease are novel, not well understood or have high clinical severity. When amplification occurs, the media sends a large excitatory signal to the population. Additionally, amplification takes place in interpersonal communication in that agents assign greater weight to the opinions of their neighbors with higher social response. This social response update rule is similar in principle to the forceful interaction type in Evans' model. However, it is a more realistic model of opinion formation because the agents are taking into account multiple perspectives, including their own prior views.

Finally, we designed the model to track the progression of disease and social response through time, rather than through interaction number. As a result of the choice to use Acemoglu's opinion spread model, Evans needed to sequentially implement a series of disease-spreading and social response-spreading interactions. For large populations, the tracking of these interactions proves computationally inefficient, and the lack of exact knowledge of the time course of the simulated outbreak complicates the fitting of the model to real-world outbreaks. Through the use of the DeGroot model for social response communication, instead of the Acemoglu model, we were able to construct a model in which all individuals in the population update their disease and social response states simultaneously at each time period. This change greatly improved computational efficiency and allowed us to employ a disease model that more closely resembles other disease-spread models used in the field.

Chapter 4

Methods: Modeling the Spread of Disease and Social Response

We present a model of the joint spread of disease and social response that is flexible and scalable. This model can be used to simulate social response to a number of diseases that spread through contact between humans. The model runs quickly and can be used to simulate populations of hundreds of thousands of people.

4.1 Model overview

The progression of disease and social response through the social network is simulated using two separate network-based models [60], linked by a coupling mechanism. We consider individuals in a population as agents in a network. These individuals are connected to each other through social ties, which are represented by the edges of the network. The agents repeatedly interact according to behavioral rules, which govern the spread of disease and social response. The disease and social response are assumed to spread independently, coupled by an interaction rule; when an agent becomes infected, her social response level is increased to reflect the perceived risk of the disease to the society in question. Specifically, the social response of the agent is set to the disease risk index, $\kappa \in (0.5, 1)$, which represents the society's familiarity with the disease, the degree to which the disease is unexpected, the clinical presentation of the disease and the method and rate of spread. Higher values of κ indicate greater perceived risk, and, consequently, greater amounts of amplification. For example, the common cold in North America, a familiar and mild disease, would have $\kappa = 0.5$, whereas SARS, a novel and highly fatal disease, would have a high value, such as $\kappa = 0.95$. Agents' likelihood of experiencing heightened social response increases when they interact with the sick, with agents who already have heightened social response or when they receive a signal from

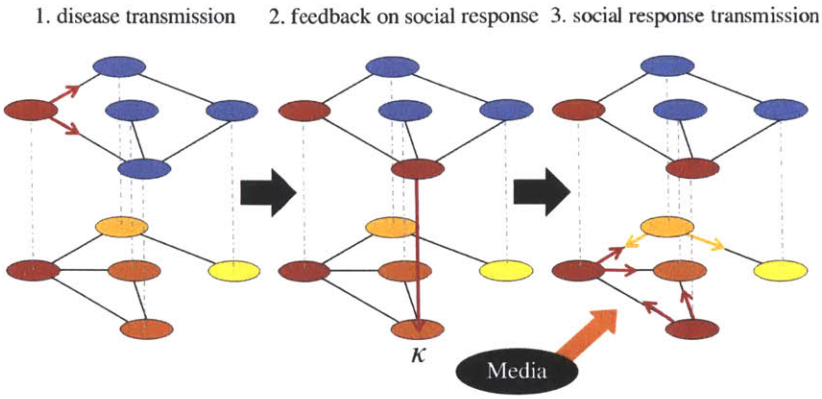


Figure 4-1: Illustration of model dynamics. At every time period: (1) the disease transmission process spreads infection through the network, (2) newly-infected agents increase their social response to κ and (3) social response is communicated between neighbors and via a signal from the media.

the media that the disease is a threat. The dynamics of the model are shown in Figure 4-1 and will be described in more detail in the sections that follow.

4.2 Network structure

The disease spread and social response are modeled on two separate but related networks. Figure 4-2 shows the connection between the disease spread and social response networks. Each agent is represented by one node in each network. The spread of disease fuels the social response through interpersonal communication about the disease and a media signal. Thus, the disease states in the disease spread network affect the social response states in the social response network. The model is formulated such that arbitrary social networks can be used for the disease spread and social response networks and, although each agent must be represented as a node in each network, the networks need not have the same edges. In the studies shown in this thesis, we used random graphs for the disease spread and social response networks. Random graphs are networks generated through a random process [61]. Typically, these networks are designed to have properties similar to those of social networks in the real world.

4.3 Disease spread dynamics

We implement an susceptible-infected-recovered (SIR) model [30], adapted for agent-based modeling [61]. The infection spreads through pair-wise interactions between infected agents and their neighbors on the disease network. Each agent's disease state at time t is represented by $X_i(t) \in \{S, I, R\}$, where S = susceptible, I = infected and R = recovered. The model is initialized with almost all agents in a susceptible state. Agents become infected through contact with infected neighbors on

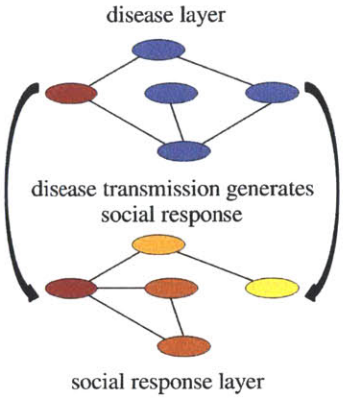


Figure 4-2: Coupling of disease and social processes. Social response and disease are transmitted on two different graphs. The spread of disease fuels increased social response.

the disease network. At time t, an infected agent infects each of her susceptible neighbors, independently, with probability β, that is if $X_i(t) = I$, $X_j(t) = S$ and i and j are neighbors on the disease network, then:

$$X_j(t+1) = \begin{cases} I & \text{with probability } \beta \\ S & \text{with probability } 1-\beta \end{cases} \tag{4.1}$$

Following infection, agents recover after T_R time periods, where the parameter T_R is set to reflect the time that infected persons continue to circulate in the network. Specifically, if $X_i(t-1) = S$ and $X_i(t) = I$, then:

$$X_i(t) = ... = X_i\left(t + T_R - 1\right) = I,\ X_i\left(t + T_R\right) = R \tag{4.2}$$

4.4 Social response spread dynamics

In addition to the disease state, $X_i(t)$, each agent i has a value, $Y_i(t)$, associated with his social response at time t. The social response, $Y_i(t)$, is treated as a continuous random variable in the range 0 to 1. A social response of 0 indicates no anxiety and no behavioral symptoms. A social response of 1 indicates severe anxiety or behavioral manifestations of concern about disease, such as panic buying or participating in a protest. In this model, we do not differentiate between behavioral and non-behavioral responses, but consider both types of social response to be part of the same continuum.

An agent's social response can be changed in one of three ways. First, agents are influenced by their neighbors on the social network and will probabilistically update their social responses to be more similar to their neighbors'. Secondly, if the disease is actively spreading, agents can receive an

excitatory signal from the media, resulting in increased social response. Finally, immediately upon infection, the social responses of infected agents are increased to κ. Thereafter, the social responses of infected agents are allowed to change according to the social response update rules. Heightened social response spreads through the population via two mechanisms: (1) when the disease is novel to the region or is perceived as being particularly threatening, media influence spreads concern through the population and (2) when communicating with their neighbors, agents are biased toward adopting the opinions of their more concerned neighbors, rather than the most calm ones.

4.4.1 Social response update

At time t, agent i's social response updates according to the following two steps, applied sequentially:

Step 1: Social response exchange. We use a modified version of the DeGroot model [45] to describe how social responses are communicated between neighbors. For each neighbor j of node i, with probability q, j communicates with i at time t. If j communicates with i, $I_{ij} = 1$, otherwise $I_{ij} = 0$. For all nodes, i, $I_{ii} = 1$. Let $w_j(t)$ be the weight assigned to the social response of node j at time t. Agents are given a bias toward listening to neighbors that are more concerned about the disease. Therefore,

$$w_j(t) = \begin{cases} 10 & \text{if } Y_j(t) \geq 0.5 \\ 1 & \text{otherwise} \end{cases} \tag{4.3}$$

Then the social response of node i following step 1 is given by:

$$Y_i(t_0) = \tanh\left(M_i(t) + \frac{1}{\sum_{j\in(i,j)} I_{ij} w_j} \sum_{j\in(i,j)} I_{ij} w_j Y_j(t) \right) \tag{4.4}$$

where $M_i(t)$ is the media signal received by agent i at time t, which will be defined in Section 4.3.2. Each agent's social response after Step 1 is a weighted sum of its neighbors' social responses plus the media signal received by that agent. Note that the hyperbolic tangent is used to restrict the social responses to be within the range $[0, 1]$.

Step 2: Social response decay. Decay reflects the eventual cease of panic as the disease spread tapers off. Following Step 2, agent i's social response level is given by:

$$Y_i(t+1) = \alpha \times Y_i(t_0) \tag{4.5}$$

for some constant $\alpha \in [0, 1]$. Following the social response update, the next time period begins and the disease and social response updates are repeated.

4.4.2 Media signal

The media plays an important role in communicating risk messages to the public. Young et al. paired high media exposure and low media exposure diseases on the basis of mode of transmission, symptoms, mortality and prevalence [62]. They found that infectious diseases that receive extensive media attention are considered to be more severe than diseases that do not receive such attention. Thus, we postulate that media attention will increase concern about disease and that diseases that are more novel and severe will have larger media signals.

Let $N_I(t)$ be the number of infected agents at time t. Then, the global media signal at time t, $M(t)$, is defined as follows:

$$M(t) = \begin{cases} 2\kappa - 1 & \text{if } N_I(t) > \frac{1}{2} \times \max\{N_I(0), ..., N_I(t-1)\} \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

Let $M_i(t)$ be the media signal received by agent i. Each agent receives the media signal with probability p. Thus,

$$M_i(t) = \begin{cases} M(t) & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases} \tag{4.7}$$

Media amplifies concern over the disease when many people are infected in the present compared with the past. In general, a non-zero media signal is sent when the disease is actively spreading in the population. At the end of the outbreak, when there are only a few cases remaining, a media signal will not be sent.

4.4.3 The joint diffusion of disease and social response

Agents who are infected generate social response and communicate this response to their neighbors. During the time period in which agent i is infected, i's social response is set to κ. More precisely,

$$Y_i(t) = \kappa \text{ if } X_i(t-1) = S \text{ and } X_i(t) = I \tag{4.8}$$

The model is implemented by simulating the disease state transitions followed by social response state transitions. Thus, at each time t, $X_i(t)$ is updated for each agent i. Then, $Y_i(t)$ is updated, first by setting the social response of newly-infected agents to κ and then by communication among agents and reception of signal from the media.

4.5 Disease spread control measures

Frequently, disease spread leads to government or self-initiated actions intended to slow the spread of disease. We consider two such actions: vaccination of susceptible agents and reduction of disease-spreading contacts.

4.5.1 Vaccination

Vaccination changes the state of an agent from susceptible to recovered. Prior to the beginning of the outbreak, each agent is vaccinated with probability v_p. Thus, v_p is the expected proportion of the population that is vaccinated. For each agent i, such that $X_i(0) = S$,

$$X_i(1) = \begin{cases} R & \text{with probability } v_p \\ S \text{ (or } I \text{ if infected)} & \text{with probability } 1 - v_p \end{cases}$$

Vaccination during the outbreak begins after the cumulative number of cases has surpassed the intervention threshold, τ. Vaccination continues for T_I days following the beginning of the intervention. At each day that vaccination takes place, each susceptible agent is vaccinated with probability v_d. When an agent is vaccinated, its disease state immediately changes to recovered. Thus, if the intervention is taking place at time t and $X_i(t) = S$, then

$$X_i(t+1) = \begin{cases} R & \text{with probability } v_d \\ S \text{ (or } I \text{ if infected)} & \text{with probability } 1 - v_d \end{cases}$$

Vaccination during the outbreak has diminishing returns over time, since on the first day of vaccination each susceptible agent has probability v_d of becoming vaccinated and on subsequent days only susceptible agents who have not yet been vaccinated can become vaccinated.

4.5.2 Contact reduction

Contact reduction begins after the cumulative number of cases has surpassed the intervention threshold, τ, and continues for T_I days following the beginning of the intervention. While the intervention is in place, the agents reduce their disease-spreading contact with one another. Specifically, at each time each edge in the disease graph has probability η of being removed and probability $1 - \eta$ of remaining in the graph for that time, independent of all other edges. Since we are concerned only with edges along which the disease could be spread, that is, edges adjacent to infected nodes, this form of contact reduction can be modeled by reducing the per contact infection rate to $\beta \times (1 - \eta)$ for the duration of the intervention.

4.5.3 Social response surge resulting from the application of control measures

Several researchers have pointed to the phenomenon of "iatrogenic" social response [3, 63, 64]. Iatrogenic social response is social response caused by the actions or communications of officials, rather than by the characteristics of the disease itself. If the officials overreact to the situation or communicate poorly, a social response may occur when it otherwise would not have. We were interested in modeling the situation in which the application of measures to control the spread of disease results in a surge in social response. We assumed that the announcement on an intervention on the disease would lead to a temporary increase in press attention and a temporary increase in the perceived risk of the disease.

Let λ be an indicator for whether iatrogenic social response will be modeled:

$$\lambda = \begin{cases} 1 & \text{if iatrogenic social response is modeled} \\ 0 & \text{otherwise} \end{cases}$$

Assuming that $\lambda = 1$, during the time period when the control measures are initiated (that is, the time period when the cumulative number of cases surpasses τ), we increase the proportion of the population that receives a signal from the media and also increase the size of that signal. Specifically, for this time period we send a media signal equal to $5M(t)$ to each agent with probability $\tanh(10p)$. Following the time period in which the intervention was initiated, the transmission of the media signal returns to normal. The media signal is $M(t)$ and the probability of receiving the signal is p.

Chapter 5

Methods: Finding the Least Cost Disease Control Strategy

There has been great interest in learning how to best control the spread of disease [37, 38, 65]. We were interested in understanding how the optimal strategies for disease control changes if we account for social response. We assumed that there were costs associated with (1) the disease, (2) use of control measures and (3) social response. We then used simulation to find the control strategy that minimizes the expected cost. The control strategies we examined were vaccination prior to the beginning of the outbreak and contact reduction during the outbreak. For contact reduction, we also explored how the least cost strategy changes if we assume that the the initiation of contact reduction leads to a temporary surge in social response.

Let c_D be the cost of one person acquiring the disease. Define the cost of the intervention as c_I. For contact reduction, let c_I be the cost is the cost per person vaccinated prior to the outbreak. For contact reduction, let c_I be the cost of a 100% reduction in contact for T_I days. Let c_{SR} be the cost of the total social response. For simplicity, our simulations always use $c_D = 1$. Thus, we are concerned with relative, rather than absolute costs. With $c_D = 1$, for vaccination, c_I may be interpreted as the cost of being vaccinated relative to the cost of getting the disease. If $c_I = 0.5$, being vaccinated is half as expensive as getting the disease. The interpretations of c_I for contact reduction and c_{SR} for total social response are less straightforward, since both of these are measured in different units than disease.

Now let us define ϕ_D as the total number of people infected by the disease, that is,

$$\phi_D = \sum_{t=0}^{\infty} \sum_{i=1}^{N} I_D \left(X_i(t) = S \text{ and } X_i \left(t+1 \right) = I \right) \tag{5.1}$$

where I_D is an indicator variable for the event that i becomes infected at time $t+1$ and N is the

size of the population. Define ϕ_{SR} as the total social response, that is,

$$\phi_{SR} = \sum_{t=0}^{\infty} \sum_{i=1}^{N} Y_i(t) \tag{5.2}$$

Let the total cost of the disease, intervention and social response be C. We can now express C and formulate optimization problems to minimize $E(C)$. Since the formulations are slightly different for vaccination and contact reduction, we will address the two separately. We will use the notation of linear and nonlinear programming to formulate the problems, though we do not solve the problems using linear or nonlinear programming methods but rather through simulation.

5.1 Vaccination formulation

The cost of disease is given by $c_D\phi_D$, the cost of vaccination is $c_I N_{V_p}$ where N_{V_p} is the number of people vaccinated prior to the outbreak, and the cost of social response is $c_{SR}\phi_{SR}$. The total cost of the disease, vaccination, and social response is the sum of the costs of each of these. Thus,

$$C = c_D\phi_D + c_I N_{v_p} + c_{SR}\phi_{SR} \tag{5.3}$$

Then, the expected cost, given the disease spread and social response parameters is given by:

$$E(C\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) = E\left(c_D\phi_D + c_I N_{v_p} + c_{SR}\phi_{SR}\middle|\right.$$
$$\left.\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda\right) \tag{5.4}$$

This expression simplifies to the following:

$$E(C\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) =$$
$$c_D E(\phi_D\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) +$$
$$c_I v_p N + c_{SR} E(\phi_{SR}\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) \tag{5.5}$$

Observe that the expected cost of disease and the expected cost of social response are dependent upon the model parameters, including v_p. The expected cost of vaccination depends only on c_I, v_p and N. To find the vaccination strategy that minimizes the expected cost, we solve:

$$\arg\min_{v_p \in [0,1]} E(C \,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta, \lambda)$$

subject to:

$$\begin{aligned}
&\beta = \beta_0 \\
&T_R = T_{R_0} \\
&\tau = N \\
&T_I = 0 \\
&\eta = 0 \\
&v_d = 0 \\
&\lambda = 0 \\
&\kappa = \kappa_0 \\
&p = p_0 \\
&q = q_0 \\
&\alpha = \alpha_0
\end{aligned} \tag{5.6}$$

In other words, we find the value of v_p that minimizes the expected cost, while fixing the disease and social response parameters and implementing no vaccination during the outbreak or contact reduction.

5.2 Contact reduction formulation

As before, the cost of disease is $c_D \phi_D$ and the cost of social response is $c_{SR} \phi_{SR}$. The cost of contact reduction is $c_I \eta$. Thus, the expected cost is given by:

$$\begin{aligned}
E(C \,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta, \lambda) = \\
c_D E(\phi_D \,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta, \lambda) + \\
c_I \eta + c_{SR} E(\phi_{SR} \,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta, \lambda)
\end{aligned} \tag{5.7}$$

To find the contact reduction strategy that minimizes the expected cost, we solve:

$$
\begin{aligned}
&\arg\min_{\eta\in[0,1]} \; E\left(C \,|\beta,\; T_R,\; \kappa,\; p,\; q,\; \alpha,\; \tau,\; T_I,\; v_p,\; v_d,\; \eta, \lambda\right) \\
&\\
&\text{subject to:} \\
&\\
&\beta = \beta_0 \\
&T_R = T_{R_0} \\
&\tau = \tau_0 \\
&T_I = T_{I_0} \\
&v_p = 0 \\
&v_d = 0 \\
&\lambda = \lambda_0 \\
&\kappa = \kappa_0 \\
&p = p_0 \\
&q = q_0 \\
&\alpha = \alpha_0
\end{aligned} \tag{5.8}
$$

Here, we find the value of η that minimizes the expected cost, while fixing the disease and social response parameters, fixing the threshold for intervention (τ) and the duration of intervention (T_I) and implementing no vaccination prior to or during the outbreak.

5.3 Summary

We have formulated two optimization problems, one to find the the least cost strategy for vaccination prior to the beginning of the outbreak and the other to find the the least cost strategy for contact reduction. These problems minimize the expected total cost of the disease, given fixed disease spread parameters. We assume that the total cost is the sum of the cost of the disease itself, the cost of the intervention and the cost of total social response. Our formulations differ from previous attempts at finding optimal control measures, because we assume that the spread of disease can have a cost beyond the people infected, namely, the social response. The formulations we have presented are general, but we will show in subsequent chapters how this optimization approach can be used to provide specific insights into how to best control disease spread while accounting for social response.

Chapter 6

Study Design and Implementation

In this chapter, we will discuss the design and implementation of the studies presented in this thesis. In summary, we have three types of studies: (1) numerical experiments, (2) optimization studies of disease control measures and (3) applications of the model to predict social response to actual disease outbreaks. The numerical experiments explore the effects of the model parameters. In particular, we focus on two mechanisms that, through extensive simulation, were found to be important, media and the extent of disease spread. We show how social response spreads through media influence when κ is high or through word-of-mouth when the disease spreads quickly through a large proportion of the population. We also examine the effects of control measures of the spread of disease and social response. The optimization study uncovers how optimal disease control strategies are affected by considering the costs of social response. We find least cost strategies for vaccination and contact reduction for κ equal to 0.5, 0.7 and 0.9. Finally, the applications of the model to real-world outbreaks are designed to show how the model can be used in practice. We demonstrate that we can replicate the social responses observed in the Hong Kong SARS and H1N1 outbreaks and the spring and fall waves of H1N1 in Mexico City. We can also replicate the surge in social response observed in response to the declaration of a public health emergency during the 2012-2013 influenza season in Boston, Massachusetts.

6.1 Implementation

First, we will describe the network structure, the use of the Monte Carlo method and other aspects of simulation implementation that are common to all of the simulations conducted.

6.1.1 Network structure

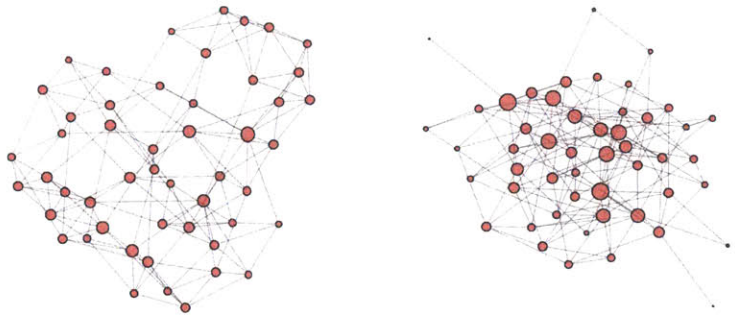
We used random graphs for the disease spread and social response spread networks. The disease spread network and social response spread network differed in structure, with the later having more connections and more non-local connections. The disease spread network was a Watts-Strogatz random graph [66]. The Watts-Strogatz random graph is constructed by arranging the nodes in a circle, then adding edges between nodes and their k closest neighbors and finally randomly rewiring some of these edges to connect to random nodes. The Watts-Strogatz random graph is commonly referred to as a small-world network, because it has strong local clustering but short average path lengths, resulting from the small number of random contacts. These random contacts increase the likelihood that a disease outbreak with grow into an epidemic [53]. For our purposes, the local structure of the graph represents frequent contacts, such as those with classmates, colleagues, and family members. The random contacts represent transmission of disease between strangers.

Like the disease, social response travels through connections between friends and colleagues. However, unlike disease, social response communication is not restricted to interactions in which both parties are physically present. Non-physical interactions through social media, email or telephone are represented in the model by a random scale-free graph. This type of graph has a degree distribution that follows a power law [61]. A few very high degree nodes are connected to much of the population, while most nodes have only a few connections. It is believed that the world wide web and collaboration networks resemble scale-free graphs [61]. Preferential attachment [43] has been proposed as a method to construct random scale-free graphs. We instead used a static random graph with scale-free degree distribution (igraph [67], static.power.law.game), since using this type of graph allowed us to specify the exponent of the degree distribution. Figure 6-1 shows sample small-world and scale-free networks, each with 50 nodes and average degree of 6. These networks differ in their degree distributions and amounts of clustering.

In summary, the social graph was a scale-free graph overlaid on top of the small-world disease graph and included all edges from the disease spread graph as well as additional edges representing remote interactions. All agents in the social graph were given a self-loop, that is for each agent i, (i, i) is an edge in the social graph. The self-loop ensured that agents always included their own opinions in the social response update.

6.1.2 Monte Carlo method

Due to the complexity of the joint dynamics of the disease spread and social response, exact computation of the time course of the disease and social response is intractable. Therefore, we used Monte Carlo simulation to estimate values of interest. In our case, Monte Carlo simulation consisted of repeated modeling the disease spread and social response spread to approximate the distributions of



(a) Small-world network (b) Scale-free network

Figure 6-1: Network structure. A small-world network and a scale-free network are shown. Each network has 50 nodes and average degree of 6. However, the degree distributions are different (nodes are scaled in proportion to degree), and the small-world network has more clustering.

the number of new infections at each time t and the mean social response of the agents at t. As the number of realizations of the model increased, the distributions more closely approximated the true probability distributions. We used the approximate distributions to estimate the expected value and standard deviation of the number of cases and the mean social response. We also obtained estimates of the distributions of total infections (ϕ_D) and total social response (ϕ_{SR}), as well as the peak number of new infections and the peak mean social response. The peak mean social response is the maximum mean social response observed at any time during the outbreak. The mean social response at time t is the average of the social responses of all agents in the network at time t, $\frac{1}{N}\sum_{i=1}^{N} Y_i(t)$. Similarly, the peak number of new infections is the maximum number of new infections simulated during a single time period.

6.1.3 Simulation implementation

Simulations were programmed in python. The igraph package [67] was used for generating the disease spread and social response spread networks. The initial infected agent was selected uniformly at random from among all agents. Then we began the disease transmission and social response spread processes. We allowed simulations to continue running until the disease spread had stopped and the mean social response in the network was less than 0.02. For most simulations that we ran, if the initial infected agent infected no other agents before recovering, we repeated the realization. Since the numerical experiment examining the effect of media influence was intended to have only one infected agent, we did not repeat the realization for simulations in that experiment. Figure 6-2 describes the simulation process.

Table 6.1 shows the number of realizations and number of agents used for each of the simulations conducted. In general, we used more realizations when it was necessary to obtain a more accurate

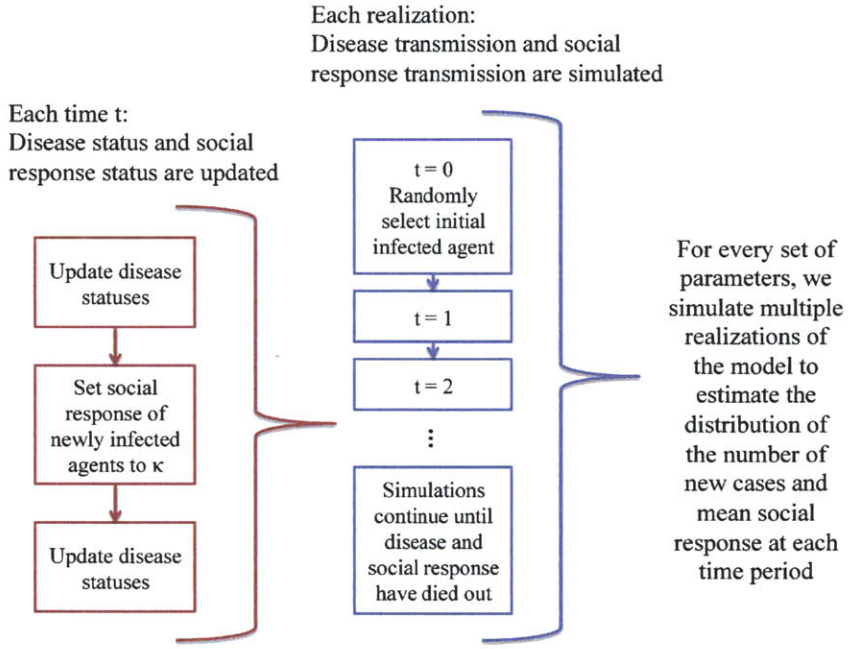


Figure 6-2: Simulation implementation. We used Monte Carlo simulation to estimate the distributions of the number of new cases and the mean social response for each time period. For every set of parameters, we simulated a number of realizations of the model. Each realization consisted of a series of time periods, during which the disease and social response statuses of the agents were updated.

Table 6.1: Number of agents and realizations per set of parameters used in simulations

simulation	number of realizations	number of agents
numerical experiments		
role of media influence	300	15,000
role of disease spread	800	15,000
effects of control measures	500	15,000
finding the least cost disease control strategy		
vaccination prior to outbreak	500	15,000
contact reduction during outbreak	500	15,000
predicting social response to real-world infectious disease outbreaks		
Hong Kong case study	100	400,000
Mexico City case study	100	400,000
Boston case study	100	600,000

estimate of the probability distributions of the number of new cases and the mean social response. We selected the number of agents by weighing the advantages of more realistic population sizes against the computational costs of conducting simulations with large numbers of agents. We chose to use 15,000 agents for the numerical experiments and optimization of control measures. For the Hong Kong and Mexico City cases studies, 400,000 agents were used, and, for the Boston case study, 600,000 agents were used, the approximate population of Boston.

6.2 Numerical experiments

6.2.1 Understanding model behavior

The objective of these simulations was to understand the behavior of the model, absent disease control measures. In particular, we wanted to verify that the model produced a social response when the disease had any of the characteristics predicted by the social amplification risk framework:

- Novelty to the region
- Clinical severity
- Unusual presentation
- Unusual extent of spread

In the model, the novelty, clinical severity and presentation of the disease are represented by the perceived risk index (κ). The extent of spread is controlled by the disease spread parameters, β and T_R. We designed two numerical experiments to parse out the role of these parameters in social response. The first was intended to isolate the effects of the extent of the disease spread. Since the social amplification of risk framework predicts that social response will occur when the disease has

Table 6.2: Disease spread study - disease transmission and social response parameters

disease transmission dynamics		
β	per-contact infection probability	0.030, 0.031, ..., 0.080
T_R	duration of infective period	6 days
τ	intervention threshold	15000 cases
T_I	intervention duration	0 days
η	edge removal probability	0.0
v_p	vaccination prior to outbreak probability	0.0
v_d	vaccination during outbreak probability	0.0
λ	indicator for whether control measures trigger social response surge	0
social response dynamics		
κ	disease risk index	0.5, 0.6, 0.7, 0.8, 0.9, 1.0
p	media penetration	0.0
q	social interaction probability	0.5
α	response decay	0.95

an unusually severe extent of spread, we wanted to confirm that our model could reproduce this result.

The second experiment was intended to determine whether diseases with high perceived risk (κ) could produce social response, even without high incidence. In our model, media sends a social response signal to the population when the number of new cases per day is greater than one half the maximum number of new cases observed previously. Media is therefore not dependent on the absolute number of cases and a media signal can be sent even when the incidence of disease is low. This experiment was intended to explore whether media influence is sufficient to produce a social response when disease incidence is low and the perceived risk of disease is high.

6.2.1.1 The role of disease spread

The goal of this experiment was determine the effects of the extent of disease spread on the social response. In particular, we wished to verify that social response is relatively robust to changes in the incidence of disease, but that social response can occur as a result of disease incidence if the incidence is extremely high. In this experiment, we isolated the effects of the disease spread on spreading social response. So that media influence would have no effect on social response, we set media penetration (p) to 0.0. We then increased the infection probability (β) from 0.030 to 0.080, while holding the duration of the infective period (T_R) constant. We found the peak mean social response for each combination of parameters. The parameters used in these simulations are given in Table 6.2.

Table 6.3: Media influence study - disease transmission and social response parameters

disease transmission dynamics		
β	per-contact infection probability	0.0
T_R	duration of infective period	6 days
τ	intervention threshold	15000 cases
T_I	intervention duration	0 days
η	edge removal probability	0.0
v_p	vaccination prior to outbreak probability	0.0
v_d	vaccination during outbreak probability	0.0
λ	indicator for whether control measures trigger social response surge	0
social response dynamics		
κ	disease risk index	0.5, 0.6, 0.7, 0.8, 0.9, 1.0
p	media penetration	0.00, 0.01, ... 0.19, 0.20, 0.25, ... 0.95, 1.00
q	social interaction probability	0.5
α	response decay	0.95

6.2.1.2 The role of media influence

This experiment was intended to illustrate how a disease that infects very few people can still lead to a social response through media influence. Since we were interested in the effects of media, we wanted to limit variability as a result of the disease spread. Therefore, we set the per-contact infection probability (β) to 0.0. As a result, the agent infected to seed the outbreak was the only agent to become infected. This agent remained infected for 6 days before recovering ($T_R = 6$). Observe that the number of infected people at times 1 to 7 are as follows: $N_I(1) = 1$, $N_I(2) = 1$, ... , $N_I(6) = 1$ and $N_I(7) = 0$. Since a media signal is sent at time t if $N_I(t) > \frac{1}{2} \times \max\{N_I(0), ..., N_I(t-1)\}$, a media signal was sent in this experiment at times 1 through 6. No media signal was sent from time 7 onward. Given the disease spread described above, we were interested finding the peak mean social response for different values of κ and p. The parameters used in these simulations are given in Table 6.3.

6.2.2 Understanding the effects of control measures

The objective of this experiment was to gain insight into how the application of control measures affects the spread of disease and social response. Control measures are used to stop or slow the spread of disease. We considered two control measures: vaccination, applied either before or during the outbreak and contact reduction. Vaccination moves susceptible agents immediately to the recovered state, without becoming infected. Contact reduction temporarily removes edges from the disease transmission network. It can be thought of as a reduction in the per-contact transmission probability of the disease. Contact reduction measures could be as extreme as closing schools and businesses or as mild as promoting improved hygiene practices and encouraging infected persons to remain at

Table 6.4: Control measure study - disease transmission and social response parameters

	control measures			
	baseline, no control measures	vaccination prior to outbreak	vaccination during outbreak	contact reduction
disease transmission dynamics				
β per-contact infection probability	0.045	0.045	0.045	0.045
T_R duration of infective period	6 days	6 days	6 days	6 days
τ intervention threshold	15000 cases	15000 cases	100 cases	100 cases
T_I intervention duration	0 days	0 days	14 days	14 days
η edge removal probability	0.0	0.0	0.0	0.25, 0.5, 0.75
v_p vaccination prior to outbreak probability	0.0	0.1, 0.2, 0.3	0.0	0.0
v_d vaccination during outbreak probability	0.0	0.0	0.01, 0.03, 0.05	0.0
λ indicator for whether control measures trigger social response surge	0	0	0	0
social response dynamics				
κ disease risk index	0.5, 0.7, 0.9	0.5, 0.7, 0.9	0.5, 0.7, 0.9	0.5, 0.7, 0.9
p media penetration	0.1	0.1	0.1	0.1
q social interaction probability	0.5	0.5	0.5	0.5
α response decay	0.95	0.95	0.95	0.95

home.

As a baseline, we set the disease parameters such that the disease would spread through a large proportion of the population. Then we compared the baseline disease spread and social response with the disease spread and social response resulting from various levels of vaccination prior to the outbreak (v_p), vaccination during the outbreak (v_d) and contact reduction (η). Comparisons were made with respect to total new cases and total social response, peak new cases and peak mean social response and the number of days until the peak number of new cases and peak mean social response were obtained. The parameters used in these simulations are presented in Table 6.4.

6.3 Finding the least cost disease control strategy

In this study, we solved the optimization described in Chapter 5. The objective was to learn how the least cost control strategy changes when the cost of social response are considered in addition to the costs of infection and implementing the control measures. The cost of social response, taking into account losses to local business and tourism, can be billions of dollars [13, 68]. By accounting for this cost, public health officials can make better decisions about how to respond infectious disease outbreaks.

Since analytic computation of the joint dynamics of the disease spread and social response spread is complex, we used simulation to estimate the expected cost, given the parameters, for values of v_p

and η ranging from 0.0 to 1.0. Then we determined at which value of v_p or η the minimum expected cost was achieved. In the sections that follow, we will describe in more detail the process for finding the value of v_p or η that minimized the cost.

6.3.1 Simulation approach to finding the least cost strategy

We will begin by explaining how we used simulation to solve the optimization problem formulated in Chapter 5. We set all of the parameters that we were not optimizing over to fixed values. Therefore, all parameters except v_p were fixed for the vaccination optimization and all parameters expect η for the contact reduction optimization. In the optimization formulation in Chapter 5, these fixed parameters are given as β_0, T_{R_0}, etc. Then for values of v_p (alternatively η) ranging from 0 to 1, we found the expected total cost. Recall from Chapter 5, that the expected total cost, given the disease spread and social response parameters is the sum of the cost of the disease spread, the cost of the implementation of the control measures and the cost of the total social response. Thus, the expected cost for vaccination is

$$\begin{aligned} E(C\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) = \\ c_D E(\phi_D\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) \\ + c_I v_p N + c_{SR} E(\phi_{SR}\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) \end{aligned} \quad (6.1)$$

and the expected cost for contact reduction is

$$\begin{aligned} E(C\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) = \\ c_D E(\phi_D\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) \\ + c_I \eta + c_{SR} E(\phi_{SR}\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda) \end{aligned} \quad (6.2)$$

It is trivial to find $c_I v_p N$ and $c_I \eta$, since both terms are products of constants. It is more difficult, however, to find $c_D E(\phi_D\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda)$ and
$c_{SR} E(\phi_{SR}\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda)$, since the total number of infections (ϕ_D) and the total social response (ϕ_{SR}) depend in a complex manner upon all of the disease spread and social response parameters. We used Monte Carlo simulation to estimate
$E(\phi_{SR}\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda)$ and $E(\phi_D\,|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta,\lambda)$ for each value of v_p (η). With these estimates for every value of v_p (η), we could easily calculate the expected total cost for any values of c_D, c_I and c_{SR} and hence find the value of v_p (η) that

minimized the total cost. In summary, the steps to find the value of v_p (η) that minimized the total cost were:

1. Assign fixed values to all the disease spread and social response parameters other than the one being optimized over (v_p or η).

2. For v_p (η) from 0.0 to 1.0, obtain estimates of $E(\phi_{SR}|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta, \lambda)$ and $E(\phi_D|\beta,\ T_R,\ \kappa,\ p,\ q,\ \alpha,\ \tau,\ T_I,\ v_p,\ v_d,\ \eta, \lambda)$.

3. Using formulas 6.1 and 6.2, calculate the expected total cost for each combination of c_D, c_I, c_{SR} and v_p (η).

4. For each combination of c_D, c_I and c_{SR}, find the value of v_p (η) that minimizes the expected total cost.

6.3.2 Vaccination

For the vaccination optimization, we were interested in the effects of v_p on the social response and disease spread. We ran the optimization for each of three values of κ, 0.5, 0.7, and 0.9, and for costs $c_D = 1$, $c_I \in [0,1]$ and $c_{SR} \in [0,0.5]$. The parameters used for the optimization simulations are shown in Table 6.5.

Table 6.5: Least cost vaccination study - disease transmission and social response parameters

disease transmission dynamics		
β	per-contact infection probability	0.045
T_R	duration of infective period	6 days
τ	intervention threshold	15000 cases
T_I	intervention duration	0 days
η	edge removal probability	0.0
v_p	vaccination prior to outbreak probability	0.0, 0.01, ..., 1.00
v_d	vaccination during outbreak probability	0.0
λ	indicator for whether control measures trigger social response surge	0
social response dynamics		
κ	disease risk index	0.5, 0.7, 0.9
p	media penetration	0.1
q	social interaction probability	0.5
α	response decay	0.95

6.3.3 Contact reduction

For the contact reduction optimization, we observed how changes in η affected the expected total cost. The parameters used in these simulations are given in Table 6.6. The contact reduction problem was solved for κ equal to 0.5, 0.7 and 0.9, for costs $c_D = 1$, $c_I \in [0,1]$ and $c_{SR} \in [0,0.5]$,

and for social response surge indicator λ equal to 0 and 1. We ran the optimization for both values of λ, because we were interested in whether assuming that implementation of control measures causes a surge in social response would lead to a change in the least cost level of contact reduction.

Table 6.6: Least cost contact reduction study - disease transmission and social response parameters

disease transmission dynamics		
β	per-contact infection probability	0.045
T_R	duration of infective period	6 days
τ	intervention threshold	100 cases
T_I	intervention duration	14 days
η	edge removal probability	0.0, 0.01, ..., 1.00
v_p	vaccination prior to outbreak probability	0.0
v_d	vaccination during outbreak probability	0.0
λ	indicator for whether control measures trigger social response surge	0, 1
social response dynamics		
κ	disease risk index	0.5, 0.7, 0.9
p	media penetration	0.1
q	social interaction probability	0.5
α	response decay	0.95

6.4 Real-world infectious disease outbreaks: Modeling the spread of disease and social response

The objective of these simulations was to demonstrate the ability of the model to predict social response during actual disease outbreaks. We chose to focus on the H1N1 and SARS outbreaks in Hong Kong, the two waves of H1N1 infection in Mexico City and the 2012-2013 influenza season in Boston, Massachussets. In this section, we will describe the disease outbreaks and corresponding social responses, as well as how we fit the disease spread to the observed cases and estimated the perceived risk, κ.

6.4.1 Approach to parameter selection

We accept that there are many parameters in the model, and if all parameters were allowed to freely vary, it would be very easy to fit the social response for nearly any disease. We therefore decided to set many of the model parameters to fixed values for the simulations of real-world outbreaks. In particular, we decided to fix the social response decay rate (α), media penetration (p) and interaction probability (q). These parameters were fixed, because there is little evidence to suggest that they would change substantially with time. Conceivably, p could be different in different locations, as a result in different levels of attention to media or different amounts of media infrastructure. We decided that, since we do not have data on how media penetration might differ between Mexico

City, Boston and Hong Kong, it was most reasonable to set this parameter to a fixed value. Thus, the only social response parameter that was allowed to vary in these simulations was κ.

The parameter, κ, represents the perceived risk of the disease. We decided that in order for κ to change between simulations, we would need evidence that there were changes in the perceived novelty or severity of the disease to the society in question. News reports and scientific studies informed our choice of κ for each outbreak that we simulated. While ultimately the selection of κ was done manually and was therefore subject to bias, we made every effort to limit that bias.

With the disease spread, we took more liberties in selecting parameters to achieve a good fit to the observed cases. While there have been many studies aimed at accurately estimating important factors in the spread of disease [69, 70], such as the reproductive number, that was not an objective of this study. Our intention was to show that we can accurately model social response to disease outbreaks. Therefore, we selected disease spread and control measure parameters that fit the observed epidemic curve and did not place great concern on the realism of those parameters. We set the time until recovery (T_R) at 6 days, since the infectious period for influenza is approximately 6 days [71]. Then, we selected values of the per-contact infection probability (β) and the control measure parameters, such that the resulting expected new cases per day (or per week in the case of Boston) were a good fit to the observed new cases. Since no vaccine was available for SARS and none was available for H1N1 during the time when it was spreading in Hong Kong or Mexico City, only contact reduction was used for those simulations. For the Boston simulations, vaccination and contact reduction control measures were used.

6.4.2 H1N1 and SARS outbreaks in Hong Kong

6.4.2.1 Background and description

Hong Kong is a special administrative district of China and home to 7 million people. In 2003, Hong Kong became the epicenter of the severe acute respiratory syndrome (SARS) outbreak in South Asia. In total, 1755 cases were reported, with 299 deaths [11]. While SARS was spreading in Hong Kong, Leung et al. conducted a series of surveys of the population [12]. The surveys measured anxiety using the State-Trait Anxiety Inventory (STAI) and asked about self-protective behaviors. It was found that at the height of the outbreak in early April, the residents of Hong Kong were experiencing heightened anxiety in response to the spread of SARS. Over time, as the case count fell, the anxiety scores returned to lower levels. Specifically, at the peak of the outbreak the mean STAI score was 24.8 (on a scale from 10 to 40). By the time the outbreak had ended, the mean score had fallen to 14.5. In addition, anxiety during the SARS outbreak had behavioral manifestations. Many residents wore face masks, stayed home from work or school, and there were reports of panic buying [7]. The economy of Hong Kong was hit hard by the reductions in retail sales and tourism

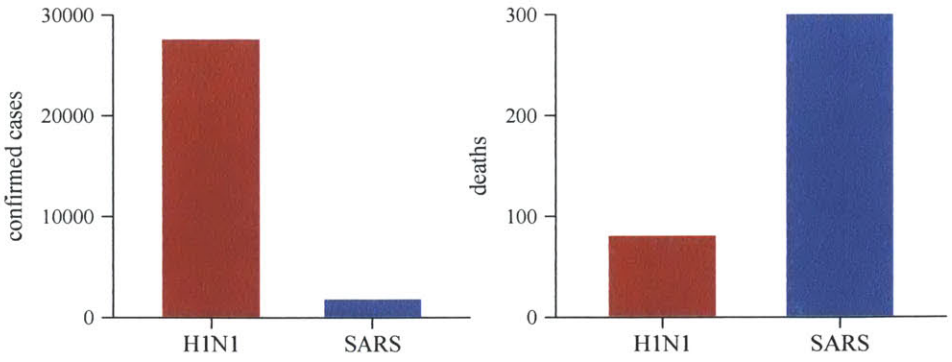


Figure 6-3: Comparison of the severity of the H1N1 and SARS outbreaks in Hong Kong. H1N1 was a prevalent but mild disease, whereas SARS infected few people but had a high fatality rate. During the 2009 H1N1 outbreak in Hong Kong, 27,498 confirmed cases were reported and the total number of cases of H1N1 (confirmed and unconfirmed) is believed to be much higher. However, only 80 deaths were reported. In contrast, 1755 confirmed cases and 299 deaths from SARS were reported in 2003.

[13].

In the summer and fall of 2009, Hong Kong was affected by the global H1N1 pandemic. Compared with SARS, H1N1 affected a very large proportion of the population. Using blood tests taken before and after the outbreak, as well as survey data and reports of confirmed cases, Lee et al. estimated that about 11% of the population of Hong Kong was infected with H1N1 [72]. Another study found that between 11% and 19% of the population was likely infected [73]. Nevertheless, public anxiety was relatively low during the outbreak. One survey found that a majority of residents, 64%, felt that the H1N1 outbreak had had no effect on their daily lives [74]. Another survey found that anxiety was low at all points in the outbreak, with a mean score of about 18 on the STAI [75]. This low level of concern about H1N1 was due to the low perceived severity of H1N1. In late April, when H1N1 was spreading in Mexico but had not yet reached Hong Kong, surveyed residents of Hong Kong estimated H1N1 to be about 60% as severe as SARS. As they learned more, the severity estimates were revised downward. When H1N1 finally hit Hong Kong in June, residents believed H1N1 to be only about 15% as severe as SARS [75]. Figure 6-3 compares the severity of the H1N1 outbreak with that of the SARS outbreak.

The volume of Google searches for "H1N1" in Hong Kong during 2009 supports the findings about perceived disease severity [76]. The most searches were conducted in late April, before H1N1 had begun to spread locally in Hong Kong. Although there was a spike in search interest in mid June when local transmission began, search interest in H1N1 declined in Hong Kong over July and August even as the number of cases rose. The residents of Hong Kong appear to have been interested in the threat of a dangerous disease, but when it actually arrived and was not as severe as feared, interest dissipated. It is also important to note that the arrival of H1N1 in Hong Kong was not a surprise.

Table 6.7: Comparison of SARS and H1N1 outbreaks in Hong Kong

SARS (2003)	H1N1 (2009)
disease characteristics	
1755 cases and 299 deaths; a 17% mortality rate [11].	Approximately 11% of the population infected [72, 73].
Medical personnel had no prior experience with SARS.	Medical personnel were familiar with the treatment of influenza and anti-viral drugs were effective.
The disease's arrival in Hong Kong was unexpected, since there was not widespread knowledge of the SARS outbreak in China [7].	Public health officials had time to prepare a response to H1N1, since it began spreading in Mexico months before it arrived in Hong Kong.
social response characteristics	
High anxiety at the peak of the outbreak, about 25 on the State-Trait Anxiety Inventory [12].	Low anxiety during the outbreak, about 18 on the State-Trait Anxiety Inventory [75].
Behavioural evidence of social response, including panic buying [7].	64% of residents reported that the outbreak had no effect on their daily lives [74].

It was quite clear that H1N1 would have a global footprint, and health officials in Hong Kong had time to prepare their response before H1N1 arrived. In Table 6.7, the SARS and H1N1 outbreaks are compared.

6.4.2.2 Estimating the disease spread and social response parameters

We accounted for the observations about disease severity and incidence in the choice of model parameters. The perceived novelty and severity of H1N1 when it began to spread in Hong Kong were low. Therefore, we set κ equal to 0.6. In contrast, SARS was novel and severe, so we set κ equal to 0.95. For fitting the SARS disease spread, we selected disease parameters and contact reduction parameters (β, T_R, η, τ and T_I) such that the simulated expected number of new infections would approximately match the number of confirmed new cases, rescaled to a population of 400,000. For H1N1, we selected disease parameters and contact reduction parameters such that the simulated expected number of new infections would approximately match the number of confirmed new cases, rescaled assuming a population of 400,000 with 11% of the population eventually infected. We assumed that the proportion of total cases that were confirmed remained constant over time. After setting the disease spread parameters, we modeled the projected social response. The parameters used to simulate the disease transmission and social response for Hong Kong are shown in Table 6.8.

Table 6.8: Hong Kong - disease transmission and social response parameters

		Hong Kong	
		SARS, 2003	H1N1, 2009
disease transmission dynamics			
β	per-contact infection probability	0.05	0.049
T_R	duration of infective period	6 days	6 days
τ	intervention threshold	50 cases	2500 cases
T_I	intervention duration	300 days	300 days
η	edge removal probability	0.50	0.12
v_p	vaccination prior to outbreak probability	0.0	0.0
v_d	vaccination during outbreak probability	0.0	0.0
λ	indicator for whether control measures trigger social response surge	0	0
social response dynamics			
κ	disease risk index	0.95	0.60
p	media penetration	0.1	0.1
q	social interaction probability	0.5	0.5
α	response decay	0.95	0.95

6.4.3 Spring and fall H1N1 outbreaks in Mexico City

6.4.3.1 Background and description

In March of 2009, cases of respiratory illness began to surface in La Gloria, Mexico. By April 10, it was reported that 616 people, or 28.5% of La Gloria's population, had been infected [77]. While it has since been confirmed that multiple respiratory infections were circulating, at the time it appeared that a novel H1N1 influenza virus was responsible for the outbreak. Shortly afterword, influenza cases began to be reported throughout Mexico. The outbreak was particularly intense in Mexico City where, according to data from the Mexican Social Security Institute (Instituto Mexicano del Seguro Social), 203 confirmed cases were reported by June 1 [70]. The Mexican government responded to the outbreak in Mexico City with extreme social distancing measures that have been estimated to have reduced disease transmission by between 29% and 37% [70]. Schools, restaurants and entertainment venues were closed and all public gatherings canceled [77]. Many businesses voluntarily shut down. There were media reports of widespread panic [78, 79]. The economy of Mexico also felt the effects of H1N1. Across Mexico, it is estimated that the spring epidemic cost over 2.3 billion dollars (about 0.3% of Mexico's gross domestic product) [68].

In late August of 2009, there was a second outbreak of H1N1 in Mexico City corresponding with the return of children from the summer school vacation. By that time, fears that the 2009 influenza pandemic could reach the scale of the 1918-1919 pandemic had largely been alleviated. As a consequence, though many more people were infected in the fall wave of infection, the outbreak was met with relative calm. Although the Mexican government continued to encourage hygiene and social distancing of infected individuals, closures were not widespread and the disease was allowed

to run its course [80]. In Mexico City, approximately ten times as many cases were infected in the fall H1N1 outbreak, which lasted from late August to early December, than in the spring outbreak, which lasted from April to May [70]; however, the social response to the fall outbreak was much smaller. The spring and fall waves of infection are compared in Table 6.9.

Table 6.9: Comparison of spring and fall outbreaks of H1N1 in Mexico City in 2009

spring outbreak	fall outbreak
disease characteristics	
203 confirmed cases reported by June 1 [70].	2138 confirmed cases reported between September 1 and December 31 [70].
Marked the emergence of a a strain of influenza that in the past had caused extreme morbidity and mortality.	H1N1 was not as severe as originally feared and there was increased familiarity with the virus.
social response characteristics	
Schools, restaurants and entertainment venues were closed [77].	No widespread closures or activity restrictions [80].
Media reports of widespread panic [78, 79].	Media reports give little indication of panic or anxiety.

Figure 6-4 illustrates how the two outbreaks were treated in the Mexican press. The number of news articles in the Factiva database that were published in Mexico and included references to both Mexico City and influenza was recorded for every day between April and December. During the spring outbreak, there was a spike in interest in H1N1 followed by a slow decline. When H1N1 returned in the fall, it received almost no attention from the press.

6.4.3.2 Estimating the disease spread and social response parameters

We modeled the distinct dynamics of the spring and fall outbreaks by selecting the disease spread parameters (β, T_R, η, τ, and T_I) and the disease risk index (κ). Since no vaccine was available for H1N1 at the time of the outbreaks, contact reduction was the only intervention used. The disease spread parameters and were set such that the expected new infections matched the confirmed new cases per day. The size of disease was scaled, so that the 400,000 agent network could stand in for Mexico City, which has a population of about 9 million. Additionally, since confirmed cases represented approximately 3.5% of total cases in Hong Kong, we assumed that 3.5% of total cases were confirmed in Mexico City. We also assumed that the proportion of total cases that were confirmed remained constant through time.

For the spring wave of H1N1 infection, we set κ equal to 0.75 to reflect the novelty of H1N1, its appearance outside of the normal flu season, its perceived severity in light of the La Gloria outbreak, and the fact that severe cases had been seen in young people, which is indicative of a more virulent form of influenza. For the fall wave of H1N1 infection, the disease risk index was reduced to 0.60, since the population had become more familiar with H1N1 and because it had been shown that

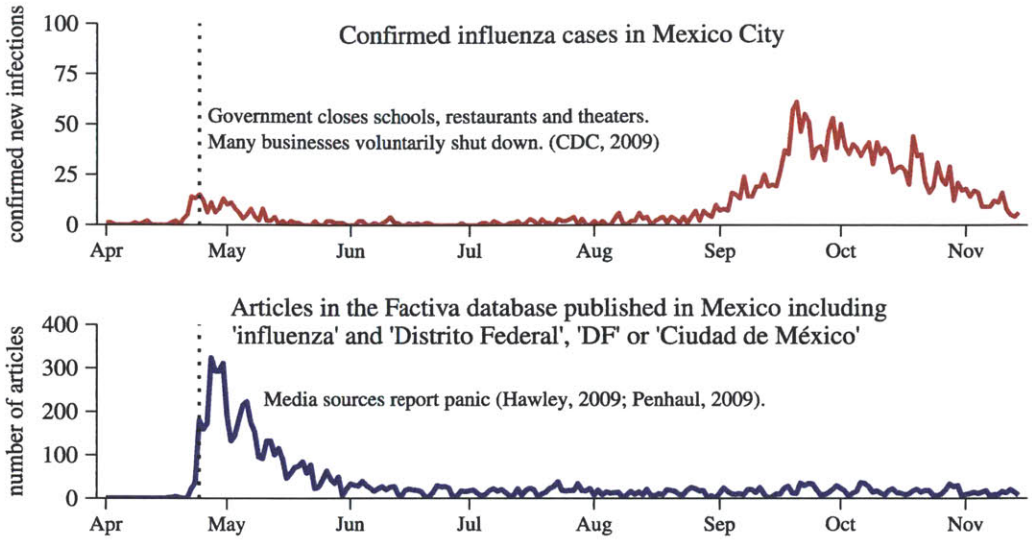


Figure 6-4: Heightened media interest in H1N1 during spring outbreak but not fall outbreak in Mexico City. We posit that media attention to a disease outbreak increases social response. During the spring outbreak of H1N1 in Mexico City, there were over 100 articles published per day on influenza, a large amount of media coverage. During the fall outbreak, although there were more confirmed cases of influenza, there was little press atention.

H1N1 was not as severe as originally feared. The parameters used to simulate the outbreaks in Mexico City are shown in Table 6.10.

6.4.4 2012-2013 influenza season in Boston

6.4.4.1 Background and description

The objective of of this simulation was to model a real-world outbreak in which the application of control measures resulted in a surge in social response. We selected to model the 2012-2013 influenza season in Boston, Massachusetts, because there is a clear association between the declaration of a public health emergency in Boston and attention to the outbreak from the press and from individuals searching the Internet.

The 2012-2013 influenza season arrived in the northeast United States unusually early. The uppermost plot in Figure 6-5 shows the number of new confirmed cases in the US Health and Human Services Region 1 (the northeast) per week for the 2004 through 2012 influenza seasons. The 2012 season is shown in red. The 2012 season in Region 1 had a large number of infections but was not exceptional. The 2004, 2007 and 2010 seasons all had similar numbers of confirmed cases in Region 1 and presumably in the Boston area as well. The 2012 influenza season was unusual, however, in that the peak in cases took place over a month earlier than normal. Boston-area hospitals, Massachusetts General and Beth Israel Deaconess, reported strained capacity due to the increase in influenza cases

Table 6.10: Mexico City - disease transmission and social response parameters

		Mexico City	
		H1N1, spring 2009	H1N1, fall 2009
disease transmission dynamics			
β	per-contact infection probability	0.053	0.047
T_R	duration of infective period	6 days	6 days
τ	intervention threshold	600 cases	2500 cases
T_I	intervention duration	300 days	300 days
η	edge removal probability	0.50	0.25
v_p	vaccination prior to outbreak probability	0.0	0.0
v_d	vaccination during outbreak probability	0.0	0.0
λ	indicator for whether control measures trigger social response surge	0	0
social response dynamics			
κ	disease risk index	0.75	0.60
p	media penetration	0.1	0.1
q	social interaction probability	0.5	0.5
α	response decay	0.95	0.95

[81]. On January, 9 2013, Boston mayor, Tom Menino, declared a public health emergency in Boston. In the week that followed, over 7500 free influenza vaccinations were delivered in 24 clinics around the city [82], and the number of flu cases began to decline.

The declaration of a public health emergency corresponded with a spike in the number of Google searches for "influenza" among residents of the Boston metro area (middle plot, Figure 6-5). The Google search interest was the higher in the week that the emergency was declared than it had been in any other year between 2004 and 2013, with the exception of 2009. A spike in the number of news articles published including the words "Boston" and "influenza" was also observed in the week that the emergency was declared (lower plot, Figure 6-5). It is clear that the emergency declaration was effective at drawing attention to the spread of influenza.

6.4.4.2 Estimating the disease spread and social response parameters

We set the disease spread and control measure parameters such that the simulated weekly cases approximately matched the observed weekly cases, rescaled to account for the difference between total and confirmed cases. We assumed that 3.5% of total cases were confirmed, as in Hong Kong. The confirmed cases per week were obtained from the Boston Public Health Commission [82]. Since the 2012 outbreak was seasonal influenza, we selected a low value of κ, 0.55. We were attempting to simulate the surge in social response resulting from the mayor's declaration of a public health emergency. Therefore, we wanted the intervention to always take place on the same day, January 9th. For the Boston simulation, we ignored the control measure initiation threshold (τ) and instead always initiated the intervention on the 9th if the disease was still spreading at that time. We set

Table 6.11: Boston - disease transmission and social response parameters

		Boston
		influenza, 2012-2013 season
disease transmission dynamics		
β	per-contact infection probability	0.075
T_R	duration of infective period	6 days
τ	intervention threshold	NA, intervention initiated on January, 9 2013 if the disease was still spreading at that time
T_I	intervention duration	14 days
η	edge removal probability	0.2
v_p	vaccination prior to outbreak probability	0.1
v_d	vaccination during outbreak probability	0.03
λ	indicator for whether control measures trigger social response surge	1
social response dynamics		
κ	disease risk index	0.5
p	media penetration	0.1
q	social interaction probability	0.5
α	response decay	0.95

the indicator for whether control measures trigger social response surge (λ) to 1. The parameters used for the Boston simulations are shown in Table 6.11.

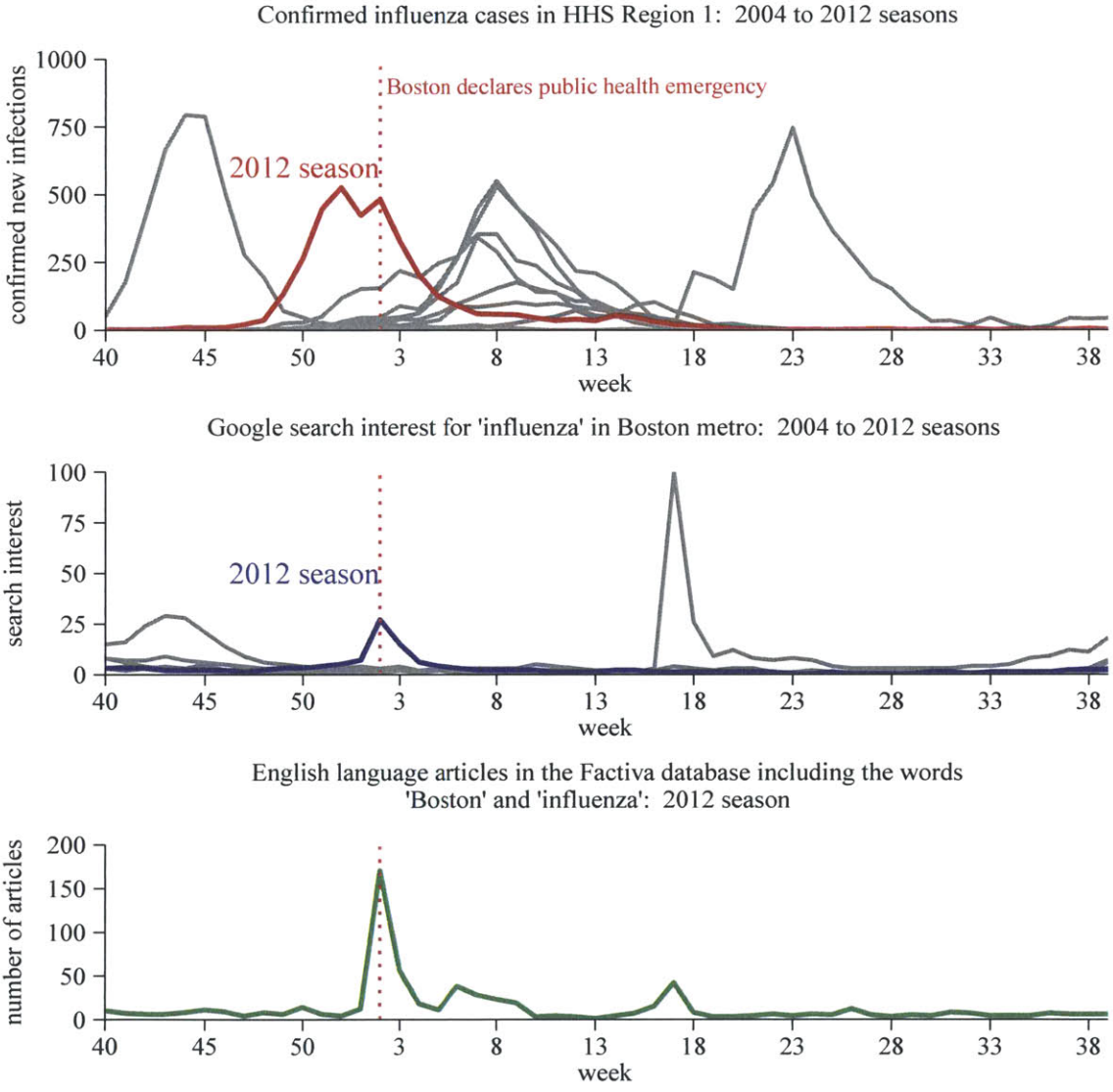


Figure 6-5: Declaration of public health emergency in Boston resulted in spike in Google search interest and media attention. Mayor Menino declared a public health emergency on January 9, 2013 (red dotted lines) following the unusually early onset of the influenza season (solid red line, upper plot). The week that the emergency was declared, there was a spike in Google search interest in "influenza" among residents of the Boston metro area (solid blue line, middle plot) and a spike in the number of English language articles including the words "Boston" and "influenza" (solid green line, lower plot).

Chapter 7

Results

7.1 Numerical Experiments

We analyzed the model behavior under different sets of parameters. Consistent with our observations from data, social response was observed when the disease had high novelty or clinical severity (high κ). Additionally, if the disease spread was exceptionally large, a social response resulted. However, the model was far more sensitive to increases in κ than to increases in β.

7.1.1 Understanding model behavior

7.1.1.1 The role of disease spread

We isolated the effects of disease spread by setting the media penetration parameter p to 0.0. Thus, the observed social response resulted only from word-of-month communication emanating from infected agents, who served as sources of social response since newly infected agents have their social response levels set to κ. Figure 7-1 shows the effect on the expected peak mean social response of increasing the per-contact infection rate (β), while holding the other parameters constant. The expected peak mean social response is the largest mean social response observed at any point in the outbreak, averaged over 300 realizations of the model.

In these simulations, κ had little effect on the expected peak mean social response. For all values of κ, as the infection rate increased, the expected peak mean social response remained level at about 0 for $\beta < 0.04$, gradually increased for $0.04 \leq \beta < 0.07$ and then plateaued as β increased above 0.07. While R_0 cannot be formulated in quite the same way for our model as for the Kermack-McKendrick SIR model, the sudden transition from no social response to increasing social response at $\beta \approx 0.04$ suggests that, given $T_R = 6$, the point at which the disease spread begins to build to epidemic proportions in our model is when β increases past 0.04. For lower values of β, we expect the disease spread to die out before taking hold in the population. Given the formulation of the model, absent

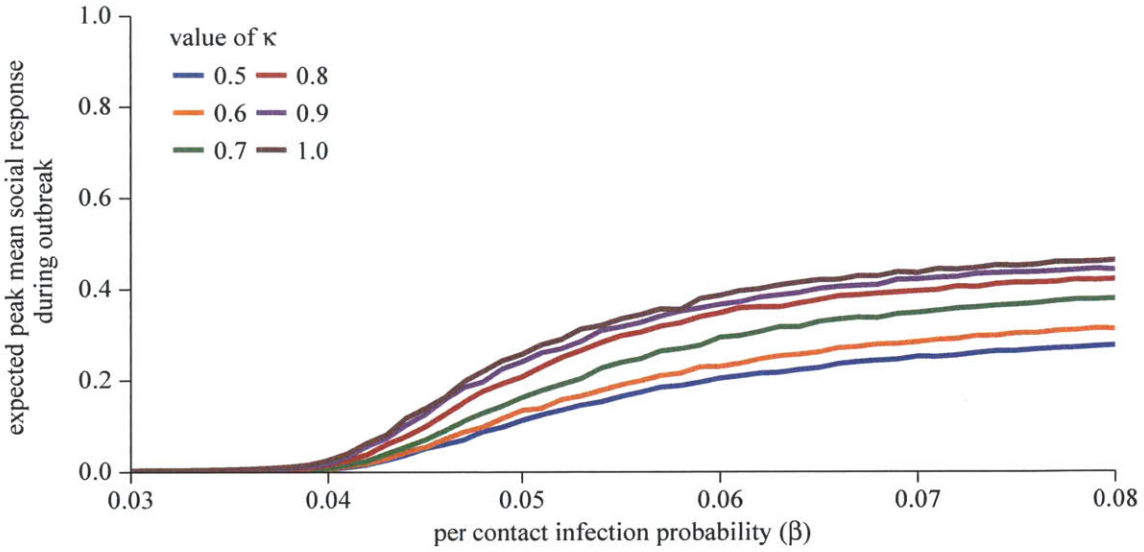


Figure 7-1: Effect of disease spread on social response. The media penetration parameter (p) was set to 0 and the time until recovery (T_R) was fixed at 6 days. Thus, these simulations isolate the effect of changing the per-contact infection probability (β) on the expected peak mean social response. As β increases, the expected peak mean social response gradually rises and then plateaus. Changes in the perceived risk index κ have relatively little effect on the resulting social response, absent media influence.

media influence, the largest social response value that can possibly be achieved is κ. In practice, the expected peak mean social response appears to approach a maximum of approximately $\frac{\kappa}{2}$.

These simulations reflect the case in which the unexpectedly large spread of disease drives the social response. The 2013 dengue fever outbreak in Singapore is an example of such behavior [15]. It is important to note, however, that the model is relatively robust against changes in the infection probability. One of the objectives in developing this model was to limit the effect of disease spread on the resulting social response. It is clear that most diseases, even most diseases with high incidence do not have social response. The model reproduces this behavior. Changes in κ, resulting in changes in the media signal, have a much larger effect on the resulting social response than do changes in the infection rate of the disease. It is difficult to produce a large social response with the model when media has no involvement.

7.1.1.2 The role of media influence

This experiment was intended to isolate the effect of media influence by setting the disease spread parameters such that the spread of disease would result in no social response absent media influence. We then examined the effects of changing κ and p on the expected peak mean social response. The results of this experiment are shown in Figure 7-2.

The effect of increasing media penetration (p) is non-linear. For high novelty or severity of disease

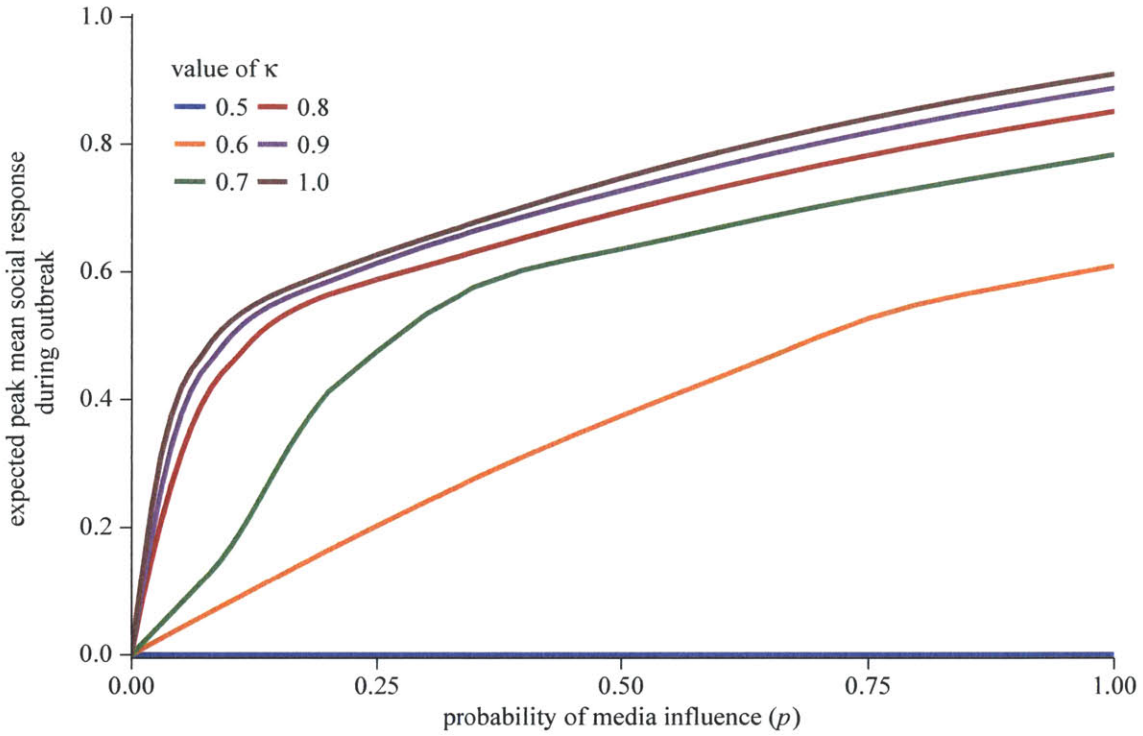


Figure 7-2: Effect of media influence on social response. The disease spread parameters were chosen such that, absent media influence, no social response occurred. When novelty and severity were low (small κ), little social response was observed, regardless of media penetration (p). When the novelty and severity of a disease were high (large κ), media influence resulted in a social response, even at very low levels of media penetration. In other words, if the disease did not not have characteristics that result in a high perceived risk of infection, such as novelty, severity, or lack of knowledge about treatment or transmission, media did not excite a social response.

($\kappa \geq 0.7$), we observed that even at very low levels of media penetration a large social response was observed (Figure 7-2). As p increased past 0.1, the growth in expected peak mean social response slowed. Since the social response of each individual is required to be less than 1, the system became saturated when both p and κ were large. When κ was small ($\kappa < 0.7$), media had little effect on the social response of the agents. For κ equal to 0.6, the expected peak mean social response increased almost linearly as p increased. For κ equal to 0.5, the expected peak mean social response remained approximately 0 as p increased. This result follows from the formulation of the media signal. When $M(t) \neq 0$, $M(t) = 2\kappa - 1$. Thus, when $\kappa = 0.5$, $M(t) = 0$, so no media signal was sent.

Since κ represents the perceived novelty and severity of disease, these simulations demonstrate that when the disease is perceived as highly threatening, a social response can occur, even though incidence may be relatively low. Examples of diseases with high κ include Ebola, avian influenza, SARS and MERS. Outbreaks of these diseases have historically been quite small, but have attracted widespread attention. These simulations also show that, in the model, small outbreaks of diseases

that are commonplace and not severe will not result in social responses. This model behavior is consistent with the research on when social response is most likely to occur.

7.1.2 Understanding the effects of control measures

Since the spread of disease fuels the spread of social response, one of the most effective ways to stop the spread of social response is to stop the spread of disease. We therefore investigated the effect of disease control measures on disease transmission and social response. Two types of control measures were used: vaccination and contact reduction. Vaccination conferred perfect immunity against the disease. The vaccination of a susceptible agent shifted that agent's disease status immediately from susceptible to recovered, without passing through the infected stage. We examined both vaccination prior to the disease and vaccination during the disease after a threshold for intervention had been surpassed. Contact reduction was implemented by randomly removing edges from the disease transmission network. It can be thought of as a reduction in the per-contact infection rate.

Figure 7-3 shows the peak number of new cases per day and peak mean social response. Figure 7-4 shows the total number of cases and the total social response. The control measures we examined had varying levels of effectiveness. Even a low level of vaccination prior to the outbreak ($v_p = 0.1$) resulted in a substantial reduction in the peak new infections and the total new infections. Compared with the baseline simulations the mean number of infections fell by 3869. Vaccination during the outbreak was also highly effective. For 1% vaccination per day ($v_p = 0.01$), the mean number of infections was reduced by 3741 compared with baseline. Contact reduction was much less effective at stopping the spread of disease. A very high level of contact reduction ($\eta = 0.75$) was required to produce reductions in the mean total infections that were comparable to those achieved by 10% vaccination prior to the outbreak.

For contact reduction, the mean number of cases is not the whole story, however. In our simulations, the contact reduction intervention lasted for two weeks after the intervention threshold was reached. If the disease spread was not nearly entirely stopped during that period, a second wave of infection could emerge. In that case, there was little reduction in the total number of infections; the disease spread was merely delayed. Figure 7-5 shows the distributions of the number of days until the peak number of new cases was reached. The baseline distribution, the distribution for 10% vaccination prior to the outbreak and the distribution for contact reduction with η equal to 0.25 are pictured. Vaccination prior to the outbreak consistently reduced the time until the peak in cases was observed. The baseline and contact reduction had a similar number of outbreaks reach their peak in the first one hundred days. Among outbreaks that peaked after one hundred days, those with contact reduction had a longer amount of time pass before the peak compared with the baseline outbreaks. Delaying can sometimes be an effective tactic for disease control, because it allows hospitals to prepare and researchers to learn more about the disease or possibly prepare a

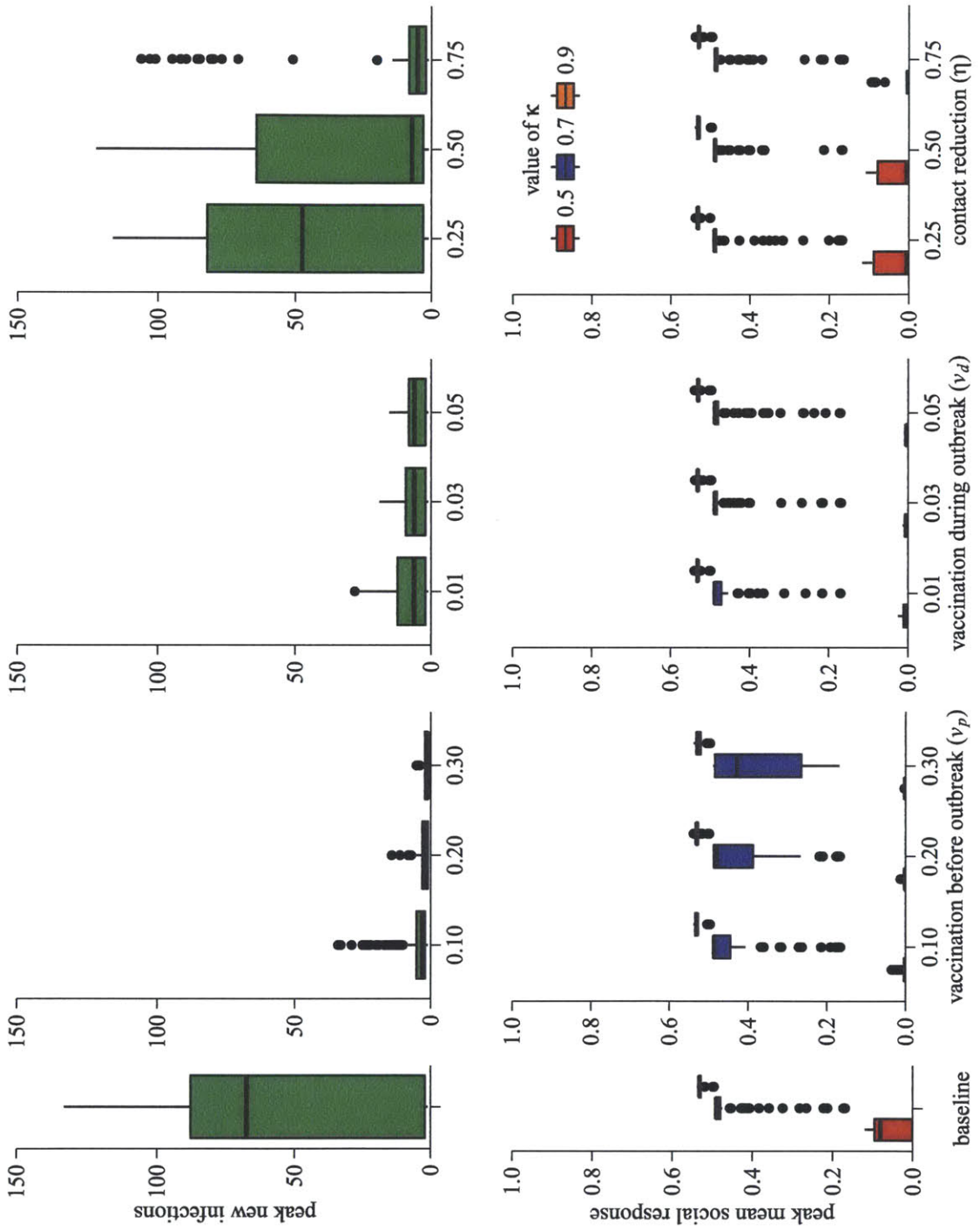


Figure 7-3: Effect of control measures on peak new cases and peak mean social response

Figure 7-4: Effect of control measures on total new cases and total mean social response

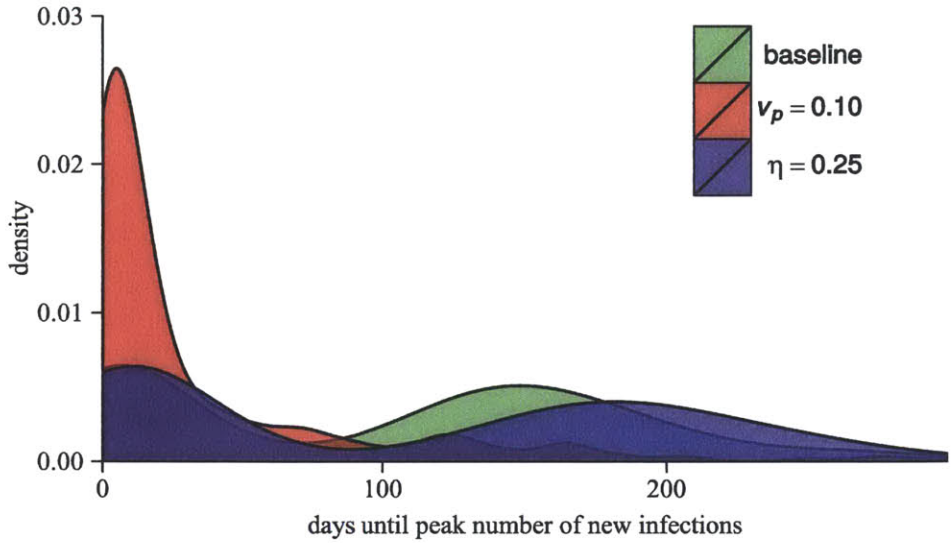


Figure 7-5: Effect of control measures on the number of days until the peak new infections and peak mean social response were reached. Compared with the baseline simulations (green) implementing vaccination prior to the outbreak ($v_p = 0.10$, red) consistently resulted in a reduction in the number of days until the peak number of new infections was reached. In contrast, implementing contact reduction ($\eta = 0.25$, blue) sometimes increased the number of days until the peak number of new infections.

disease. Therefore, although they result in little reduction in the total number of cases, low levels of contact reduction should not be entirely disregarded as tools for disease control.

The H1N1 outbreaks in Mexico City are a real-world example of how contact reduction delayed the peak in cases [70]. The summer school vacation reduced disease spreading contact, leading to a temporary reduction in the number of cases. When students returned to school from summer vacation the outbreak was reseeded and H1N1 resumed spreading. We have found that contact reduction is frequently less effective than vaccination at stopping the spread of disease. However, contact reduction is often the only line of defense against infectious diseases. Vaccines have yet to be developed for many diseases and, as with the H1N1 vaccine, they sometimes do not become available until the disease is already spreading actively. It is therefore important to understand how to best use contact reduction to stop the spread of disease and social response.

We will now consider the effect of control measures on the social response. The effect of control measures on peak social response depended on the value of κ. For κ equal to 0.5, the social response was determined entirely by word-of-mouth communication resulting from the spread of disease. No media signal was sent. Therefore, reducing the spread of disease by implementing control measures resulted in a decrease in peak mean social response. For κ equal to 0.9, reducing the spread of disease had almost no effect on the peak social response. For example, the average baseline peak mean social response was 0.53 when κ was equal to 0.9. Vaccinating 30% of the population before

the outbreak was the most effective control measure that we examined. The average peak mean social response after vaccinating 30% of agents was 0.52, barely different from the baseline level. We found that controlling the spread of disease was somewhat effective at reducing the peak mean social response for κ equal to 0.7.

All control measures that were effective at reducing the spread of disease were effective at reducing the total social response, regardless of the value of κ. This result is one of the reasons why the optimization study focused on the cost of total social response, instead of peak mean social response. The total social response is sensitive to changes in the disease spread for all values of κ, whereas the peak mean social response is robust to changes in the disease spread for high values of κ. The peak mean social response was unaffected by changes in the disease spread for high κ, because, when κ was large, media played a large role in producing social response. Even one or two cases were sufficient to produce a large peak mean social response.

7.2 Finding the least cost disease control strategy

We implemented an optimization study to find the level of disease control that minimized the expected total cost of disease. The expected total cost of disease was assumed to be the sum of the cost of infection with the disease, the cost of the intervention and the cost of the total social response. We found that considering the cost of social response considerably changed the optimal control strategy. In some cases, it became optimal to implement control measures in situations where, absent a cost for social response, the least cost strategy was to do nothing. In all cases, considering the costs of social response resulted in implementing more extreme control measures than when the costs were not considered.

7.2.1 Vaccination

These simulations determined the level of vaccination prior to the outbreak that minimized the expected total cost of the disease. Figures 7-6a, 7-6b and 7-6c show the least cost proportion of the population vaccinated for κ equal to 0.5, 0.7 and 0.9 respectively. For all values of κ, when the cost of vaccinating one individual (c_I) was near 0.0, the least cost strategy was to vaccinate nearly everyone. The least cost amount of vaccination for c_I greater than 0.0 depended on the value of κ. For κ equal to 0.5, the least cost strategy was to vaccinate only a small proportion of the population. The proportion vaccinated increased slighted as the cost of total social response increased and the cost per vaccine decreased. Nevertheless, the least cost proportion vaccinated was always less than 35% of the population when the cost of the vaccine was greater than or equal to 0.02. When the cost per vaccination was exactly equal to the cost of infection, the expected cost was minimized by vaccinating 15% of the population. The simulations exploring the effect of control measures on

disease spread and social response revealed that even such low levels of vaccination prior to the outbreak can be effective at reducing the number of cases of disease. Since the social response was not a substantial factor in the total cost, low levels of vaccination proved sufficient to minimize the cost.

The limited effect of c_{SR} on the least cost amount of vaccination was a result of how diseases with low perceived risk are treated by the media. In our model, the media does not send a social response signal when κ equals 0.5. Therefore, the social response produced by the model was small even when the disease spread was large. As a result, changing the cost of total social response did not result in major changes in the least cost strategy.

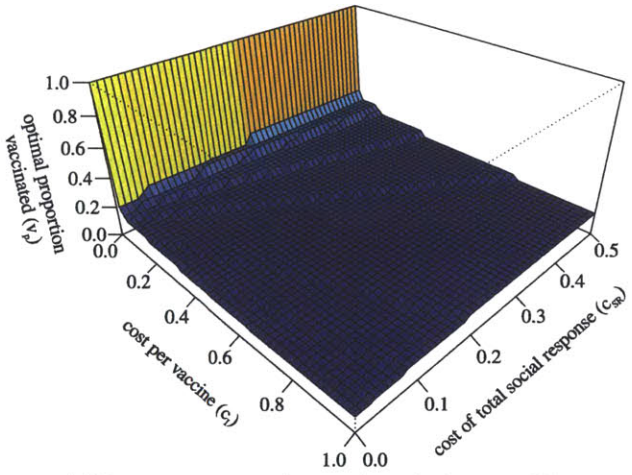
For κ equal to 0.7 or 0.9, the least cost proportion vaccinated was highly dependent upon the cost of total social response. When the cost of total social response was near zero the least cost proportion vaccinated was quite low. The proportion vaccinated rose quickly with increases in the cost of total social response. When the cost of total social response exceeded 0.3, the least cost solution was to vaccinate all agents for all values of the cost of vaccination.

7.2.2 Contact reduction

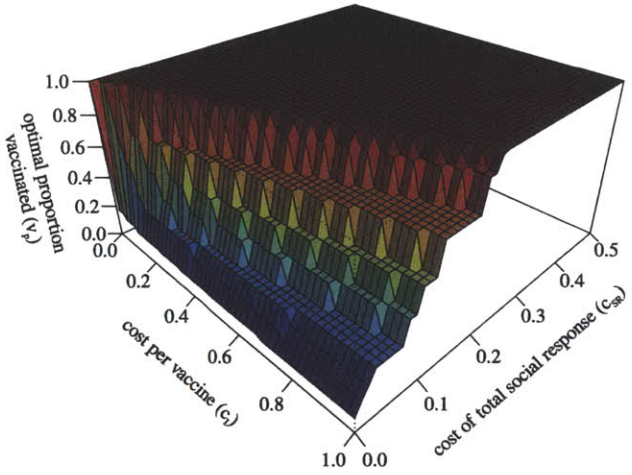
The optimal strategy for contact reduction was sensitive to changes in the cost of social response. Figures 7-7a, 7-7b and 7-7c show the optimal level of contact reduction for κ equal to 0.5, 0.7 and 0.9 respectively. We found that for most combinations of costs, the optimal contact reduction intervention involved a either a very high percentage of contact reduction ($\eta > 0.6$) or no contact reduction at all. In this way, our results mirror those of Maharaj and Kleczkowski [65], who found that the cost of ineffective attempts at social distancing can be higher than the cost of doing nothing. In their words, "if social distancing is being considered, do it well, or not at all."

For κ equal to 0.5, when the cost of contact reduction was 0.0, the least cost contact reduction (η) was over 0.9 for all costs of social response. As the cost of contact reduction increased or the cost of total social response increased, the least cost amount of contact reduction declined. The line $c_{SR} = 0.00001c_I - 0.05$ defined the frontier where the optimal solution shifted from implementing some contact reduction to implementing none at all. For points (c_I, c_{SR}) such that $0.05 < 0.00001c_I - c_{SR}$, the least cost solution was not to implement contact reduction. We discussed previously that implementing low levels of contact reduction frequently does not result in a substantial reduction in the total number of cases but instead merely delays the outbreak. In our optimization formulation, we assigned no value to delaying the onset of an outbreak, so the optimal solutions for contact reduction either involved no contact reduction at all or a high enough level of contact reduction to stop the spread of disease entirely.

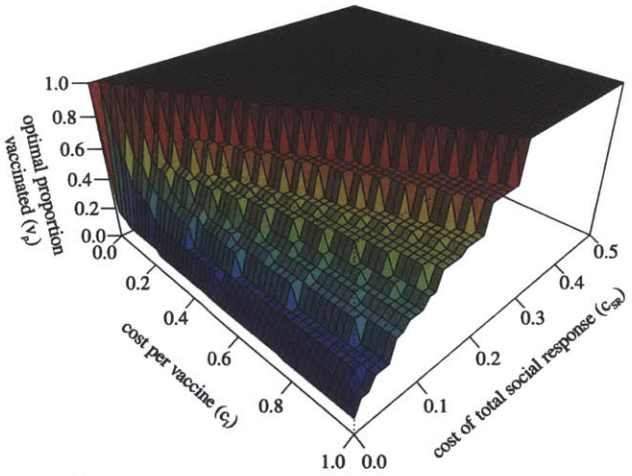
For κ equal to 0.7 or 0.9, the optimal solution for most values of c_I and c_{SR} was to implement a very high level of contact reduction. As when κ was equal to 0.5, there was a line past which the



(a) Least cost proportion vaccinated when $\kappa = 0.5$.



(b) Least cost proportion vaccinated when $\kappa = 0.7$.



(c) Least cost proportion vaccinated when $\kappa = 0.9$.

Figure 7-6: Least cost proportion of the population vaccinated prior to the outbreak. The least cost proportion vaccinated is pictured as a function of the cost of total social response (c_{SR}), the cost per vaccine (c_I) and the perceived risk of infection (κ).

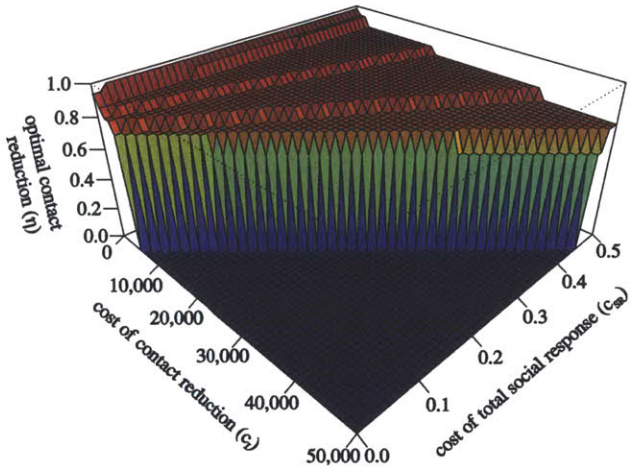
optimal solution was to implement no contact reduction at all. This line was shifted toward lower values of c_I and higher values of c_{SR}, leaving a smaller region in which no contact reduction was the least cost solution.

In this optimization study, we looked at least cost solutions as the cost of contact reduction ranged from 0 to 50,000. This wide range affects the wide range of costs for real socal distancing measures. Low levels of contact reduction could result from people who are infected staying home from work or school. This sort of contact reduction has relatively low cost. Widespread contact reduction involving people who are not infected could be much more costly, since there would be a reduction in retail sales and productive work hours. Despite its high costs, it can sometimes be necessary to initiate widespread contact reduction. Some diseases, including influenza [71], can be transmitted before the infected person becomes symptomatic. Quarantining infected individuals is therefore not sufficient to stop transmission.

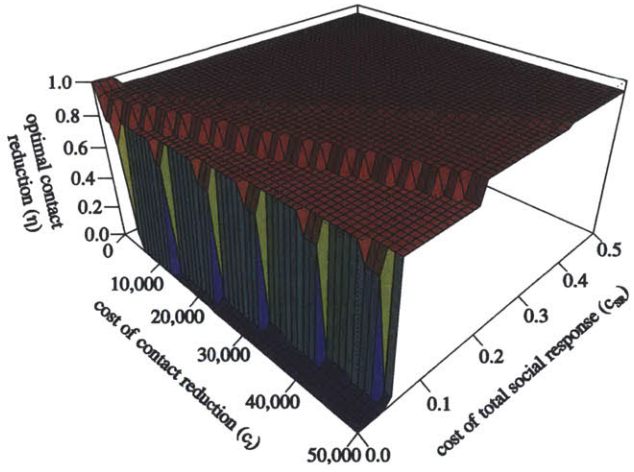
We will now discuss the results of assuming that the initiation of the intervention leads to a temporary surge in social response ($\lambda = 1$). The temporary surge in social response resulting from the intervention introduced an interaction between the cost of intervention (c_I) and the cost of social response (c_{SR}). Effectively, the surge resulted in an increase in c_I, and the size of the increase depended upon c_{SR}. If the cost of social response was low and the intervention led to a surge in social response, there was little increased cost to implementing the intervention. If the cost of social response was high, however, the cost of implementing an intervention that would cause more social response was high.

Figures 7-8a, 7-8b and 7-8c show the least cost level of contact reduction for κ equal to 0.5, 0.7 and 0.9 respectively. When κ was equal to 0.5, there was little change in the least cost amount of contact reduction resulting from the change in λ. As when λ was equal to 0, the optimal solution for λ equal to 1 was to implement no contact reduction for most values of c_I and c_{SR} examined.

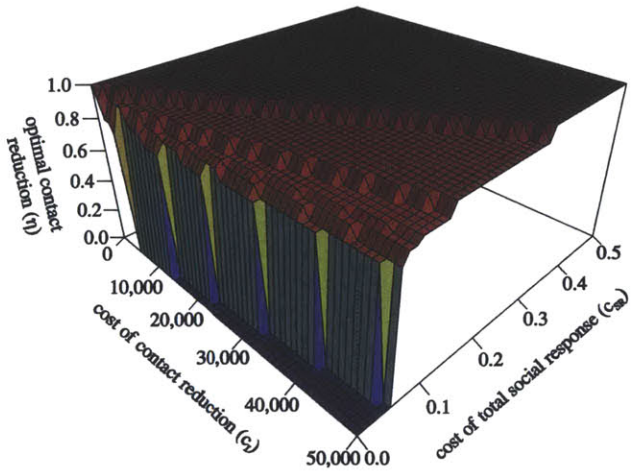
When κ was equal to 0.7 or 0.9, it became even more important to completely stop the spread of disease when contact reduction was attempted. Compared with the optimal strategy when initiating contact reduction had no effect on the social response, the optimal strategy shifted toward more extreme contact reduction measures where measures were implemented at all (compare Figures 7-7b and 7-7c with Figures 7-8b and 7-8c). The region in which the least cost solution was to not implement any contact reduction did not shift as a result of having the implementation of control measures lead to a temporary increase in social response. This result is likely because the region in which the optimal strategy was to not intervene was also the region in which the cost of total social response was low. Since a low cost was assigned to the social response, the increase in social response resulting from the intervention had little effect on the cost of intervention.



(a) Least cost contact reduction when $\kappa = 0.5$

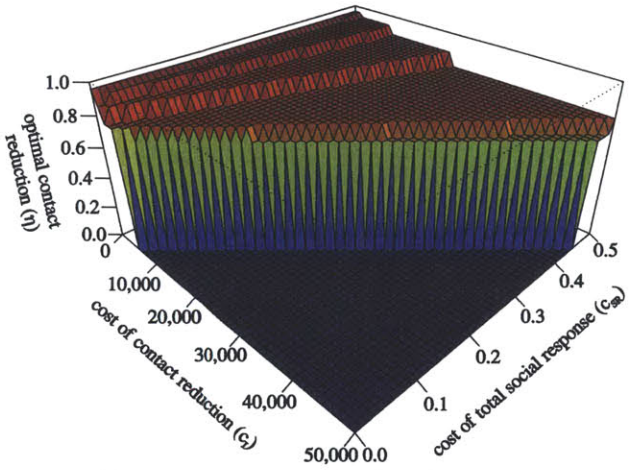


(b) Least cost contact reduction when $\kappa = 0.7$

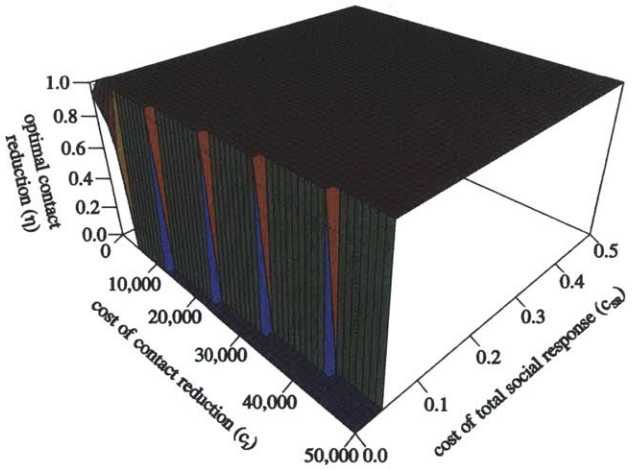


(c) Least cost contact reduction when $\kappa = 0.9$

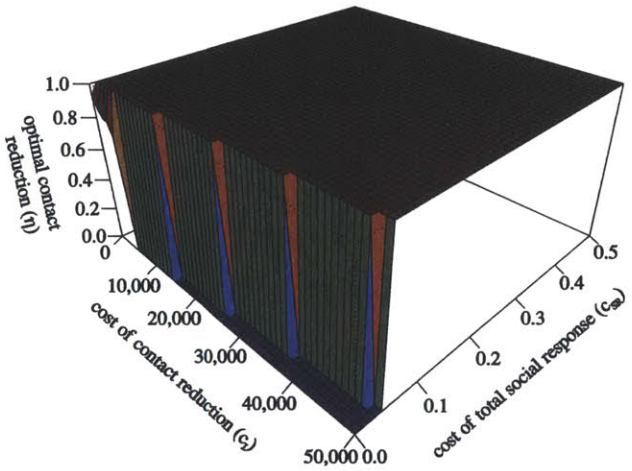
Figure 7-7: Least cost levels of contact reduction when initiation of the intervention had no effect on social response ($\lambda = 0$). The least cost value of contact reduction is pictured as a function of the cost of total social response (c_{SR}), the cost of contact reduction (c_I) and the perceived risk of infection (κ).



(a) Least cost contact reduction when $\kappa = 0.5$



(b) Least cost contact reduction when $\kappa = 0.7$



(c) Least cost contact reduction when $\kappa = 0.9$

Figure 7-8: Least cost levels of contact reduction when initiation of the intervention led to a temporary surge in social response ($\lambda = 1$). The least cost value of contact reduction is pictured as a function of the cost of total social response (c_{SR}), the cost of contact reduction (c_I) and the perceived risk of infection (κ).

7.2.3 Conclusions from least cost control measure study

In all of our simulations, assigning a cost to the total social response resulted in increased disease control measures in the least cost solution. For both vaccination and contact reduction the least cost disease control strategy was highly dependent upon the perceived risk of infection (κ). For κ equal to 0.5, the least cost solution was to implement minimal control measures, especially for high values of c_I and low values of c_{SR}. For κ equal to 0.7 or 0.9, extreme disease control measures were put in place for most values of c_I and c_{SR}. These findings suggest that considering the costs of social response makes in it even more important to quickly stop the spread of diseases with high perceived severity.

When discussing our conclusions, we should note that the costs of disease spread and social response are not independent. We consider the cost of disease spread to include all costs directly associated with the illness. These costs include medicines and hospital beds. The cost of social response includes costs peripheral to the spread of disease, such as people taking time off from work to avoid becoming infected or reduction in tourism revenue from vacationers canceling or delaying their trips. Clearly, both the cost of disease and the cost of social response will be higher for high severity diseases. The cost of intervention may be less dependent on the severity of the disease. The effect of disease severity on the relative costs of disease, intervention and social response should be taken into account when determining the costs for a particular disease outbreak.

7.3 Real-world infectious disease outbreaks: Modeling the spread of disease and social response

Through simulation of the disease spread and social response during the Hong Kong SARS and H1N1 outbreaks, the two waves of H1N1 in Mexico City in 2009 and the Boston 2012-2013 influenza season, we demonstrate that the model's predictions approximate the observed social response for these outbreaks.

7.3.1 H1N1 and SARS outbreaks in Hong Kong

In Hong Kong, there was a consistently low level of social response through the 2009 H1N1 outbreak, while during the the SARS outbreak the social response rose rapidly when cases were first appeared and declined as the number of cases declined. This behavior is replicated by the social response model. The simulated disease spread and social response for the SARS and H1N1 outbreaks are shown in Figure 7-9. The disease spread for SARS was small. Compared with the disease spread for H1N1, the number of new cases per day for SARS was almost imperceptible when viewed on the same scale. Nevertheless, a sizable social response was produced for SARS, but not for H1N1.

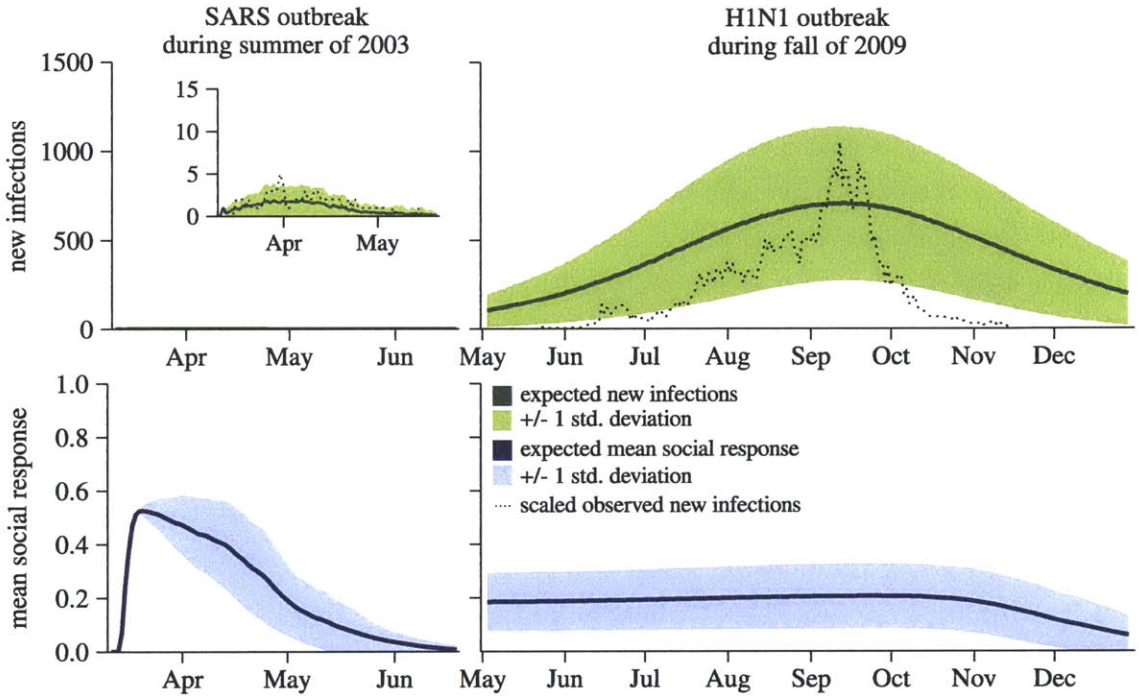


Figure 7-9: The spread of social response to SARS and H1N1 in Hong Kong. Social response and disease spreads are shown for the 2003 SARS outbreak and the 2009 H1N1 outbreak. Disease spread was modeled such that the expected number of new infections per day reflected the observed number of new infections per day. The disease risk index (κ) was set to 0.95 for SARS, reflecting the novelty of the virus, its unexpected appearance in Hong Kong, and the high clinical severity of the disease. Residents of Hong Kong considered H1N1 to be a much less severe disease than SARS. Therefore, we set the disease risk index to be lower, 0.60. Simulations were conducted with 400,000 agents and 100 replications of the model. In these simulations, the widespread media attention directed as SARS helped to fuel a large social response in the population, whereas H1N1 did not produce a social response. In our model, the properties of the disease and how the society perceives the threat determine whether disease will attract media attention and the population will exhibit a social response, and it is possible for diseases that are not highly prevalent to result in large social responses.

During the SARS outbreak in Hong Kong, 1775 people became infected [11]. Since the population of Hong Kong is approximately 7 million, we can assume that the vast majority of residents of Hong Kong had no personal connections to the infected people. In our model, no social response would have been observed for the SARS outbreak if media had not played a role. Word-of-mouth communication is not sufficient in the model to produce a social response when the infected population makes up such a small proportion of the overall population. Since SARS had high perceived risk ($\kappa = 0.95$), a large media signal was sent when cases of SARS began to spread and continued to be sent until the disease spread was close to an end. The media attention to SARS produced a large social response in March and April. Once the last cases were detected in May, the media stopped sending a social response signal to the population, allowing the social response to die out through the decay process.

In a study of anxiety during the H1N1 outbreak, Cowling et al. [75] observed that anxiety remained low throughout the outbreak. Our social response model shows similar effects. The simulated social response is consistently low through the H1N1 outbreak and does not rise with the rise in cases. The case study of the SARS and H1N1 outbreaks in Hong Kong illustrates that our model does not depend strongly on disease incidence. A high incidence disease, such as H1N1, can have a small social response, while a low incidence disease, such as SARS, can have a large social response. The key parameter is the perceived risk of disease. If the perceived risk of disease is high, it is possible for a social response to occur regardless of the incidence of disease.

7.3.2 Spring and fall H1N1 outbreaks in Mexico City

In 2009, the first major urban outbreak of H1N1 took place in Mexico City in April and May. There was widespread concern about the disease and the federal and city governments implemented extreme social distancing measures to prevent the spread of disease. When H1N1 returned to the city in the fall, there was little indication of social response. The spring and fall outbreaks of H1N1 in Mexico City were separated from one another by only three months, but during those three months scientists were able to analyze the data gathered during the spring. It was concluded that H1N1 was less virulent than originally feared. We modeled the disease transmission for the spring and fall outbreaks by fitting the simulated number of new cases per day to the observed number of new cases, appropriately scaled. We then modeled the social response dynamics by setting the perceived risk of disease for both outbreaks.

Figure 7-10 shows that in our simulations, a large social response, driven largely by media influence, was observed for the spring wave of infection. The fall wave of infection was characterized by large disease spread and low social response. In Mexico City, it is clear that media interest in the disease corresponded with the rise in social response during the spring. The number of media articles published on H1N1 during the spring outbreak far exceeded the number published in the fall (see Figure 6-4). In our model, a relatively high perceived risk of infection ($\kappa = 0.75$) led to

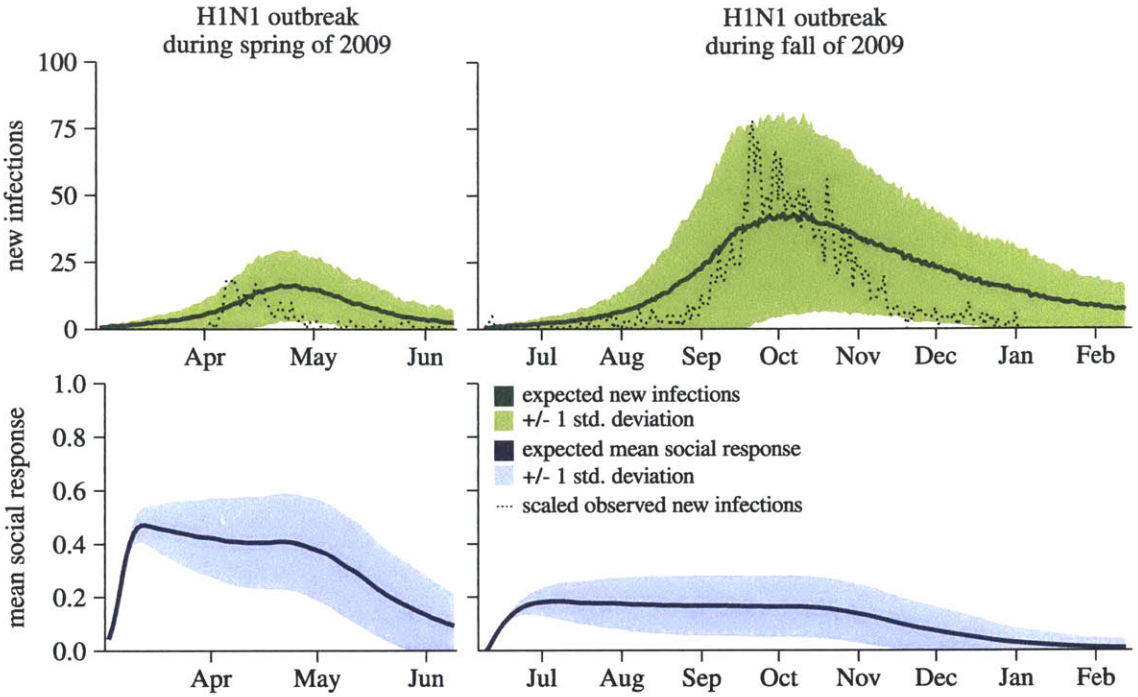


Figure 7-10: The spread of social response to H1N1 during the fall and spring 2009 outbreaks in Mexico City. Social response and disease spreads are shown for the spring H1N1 outbreak and the fall outbreak. Disease spread was modeled such that the expected number of new infections per day reflected the number of new infections per day. For the spring outbreak, the disease risk index (κ) was set to 0.75 to reflect the H1N1's novelty and its perceived severity. For the fall outbreak, we lowed the disease risk index to 0.60, since Mexico City already had experience with H1N1 and H1N1 was shown to not be as severe as originally feared. Simulations were conducted with 400,000 agents and 100 replications of the model. While the number of cases produced in the spring wave of infection was relatively low, a large social response resulted. The fall wave of infection, though larger, was met with relative calm. In our model, the social response in the spring wave was triggered by heightened media attention brought on by the novelty and perceived severity of the disease. In the fall, H1N1 did not attract as much media attention and the resulting social response was much smaller.

increased media attention and consequently social response, in the spring outbreak. By the fall, it had been determined that H1N1 was not as severe as originally feared. Therefore, we set a lower perceived risk of infection ($\kappa = 0.6$), which led to less media attention in the disease and lower social response. These simulations illustrate how changes in the perceived risk of infection, resulting from experience with the disease and increased information, can result in changes in the social response.

As with the Hong Kong case study, the Mexico City case study illustrates that higher incidence of disease does not necessarily imply that the social response will be larger. While the incidence plays a role in the social response, the perceived risk of infection and media influence have a much greater impact. It is important to note that had the Mexican government not implemented extreme contact reduction measures at the beginning of the spring wave of infection, there likely would have been

many more cases in the first wave. Nevertheless, given that these measures were in place, it seems that the social response observed was connected to the *potential* of a large and deadly outbreak, rather than the presence of such an outbreak.

7.3.3 2012-2013 influenza season in Boston

The 2012-2013 influenza season struck Boston unusually early. As the number of cases rose, major area hospitals reported that they were running out of capacity [81]. Mayor Tom Menino declared a public health emergency on January, 9, 2013. Free vaccination clinics were offered around the city and an influenza awareness campaign was launched [82]. The mayor's announcement corresponded with a surge in Google search interest in "influenza" in the Boston metro area and a surge in media reports about influenza in Boston (see Figure 6-5).

We fit the simulated new cases per week to the observed new cases per week, scaled to reflect the fact that only a small proportion of total cases are confirmed. We then modeled the social response. The perceived risk of infection was low ($\kappa = 0.55$), since seasonal influenza is a commonplace and relatively mild disease in Massachusetts. Control measures (vaccination and contact reduction) were initiated on January 9, 2013. On January 9, we modeled a temporary increase in the proportion of the population receiving a signal from the media as well as an increase in the size of the media signal. This increased attention from the media brought about an increase in social response in the week that followed. Then as the number of cases declined, the social response declined.

The Boston public health emergency was selected for simulation because it was an intervention for a mild disease that had received almost no attention from the press prior to the intervention. Therefore, we could tell that the social response was as a result of the intervention and not caused by some other factor. This simulation demonstrates that our model can account for temporary social response resulting from a public official drawing attention to the disease. We do not wish to imply that the declaration of the public health emergency was misguided. Indeed, it is clear that in the weeks following the declaration many people became vaccinated for influenza who may not have otherwise done so. The social response was mild, and there were no indications of panic or detrimental economic consequences.

7.3.4 Conclusions from modeling of real-world disease outbreaks and associated social responses

We have shown through the simulation of disease and social response in Hong Kong, Mexico City and Boston that we can accurately replicate social responses seen in actual disease outbreaks. The model can produce social response when diseases are perceived as high risk, even if they do not have many cases, such as SARS. Adjustments can be made for changes in the perceived severity of

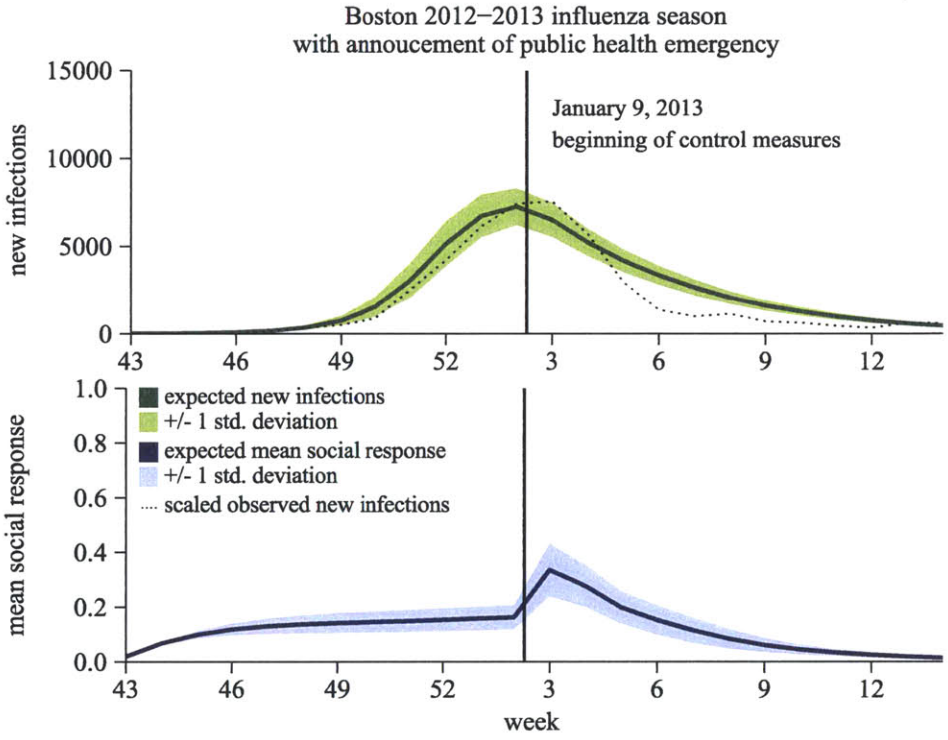


Figure 7-11: The surge of social response resulting from the declaration of a public health emergency during the 2012-2013 influenza season in Boston. Social response and disease spreads are shown for the spring H1N1 outbreak and the fall outbreak. Disease spread was modeled such that the expected number of new infections per week reflected the observed number of new infections per week. The disease risk index (κ) was set to 0.55, since seasonal influenza has a low level of perceived risk. On January 9, when control measures were initiated, we modeled a surge in media attention and an increased media signal. As a result, the social response temporarily rose, before declining with the decline in cases. This simulation was conducted with 600,000 agents and 100 replications of the model. The difference in the variability of the number of new infections in these simulations compared with the Hong Kong and Mexico City simulations is the result of the presentation of cases for Boston on a weekly scale rather than a daily scale.

disease with time, as with H1N1 in Mexico City. Finally, surges in social response resulting from the announcement of disease control measures can be modeled, as with Boston.

Nevertheless, there are some limitations to our results. A primary limitation is the availability of data. We principally relied on news sources to determine whether or not social response had taken place. Scholarly studies into psychological responses to disease outbreaks are rare. The lack of objective data on observed social response has limited our ability to confirm that our model's results are consistent with the observed social response. There is also a lack of data about the spread of disease. We obtained daily or weekly numbers of confirmed cases for each of the diseases examined. However, the total number of cases (both confirmed and unconfirmed) is unknown. In Hong Kong, there was sufficient evidence from serological studies following the H1N1 epidemic to conclude that approximately 3.5% of cases were confirmed. Given that we had no data on the percentage of confirmed cases of influenza in either Mexico City or Boston, we assumed that in these cities 3.5% of cases were confirmed as in Hong Kong. This assumption is likely not accurate. It gives the impression that the H1N1 outbreaks in Mexico City were quite small compared with the Hong Kong outbreak, which is likely not true. Moreover, the scaling of the outbreaks affected our ability to fit the infection curve to the scaled cases, since outbreaks that affect a very large portion of the population are easier to fit than those that infect fewer people. Once the disease has taken hold in the population, it will generally infect a large number of people before dying out. Despite these limitations, our simulations have demonstrated that the model can replicate population-level disease spread and social response with a reasonably high degree of accuracy and have served as an illustration of the variety of disease outbreaks that can be simulated using the model.

Chapter 8

Discussion

8.1 Summary and conclusions

We have formulated a model that can accurately simulate infectious disease outbreaks and their associated social responses. We have demonstrated the flexibility of this model, as well as its ability to provide insight into how disease and social response spread and how they can be prevented. We will now briefly summarize our major findings and discuss how our model contributes to the literature on social processes accompanying the spread of disease.

In Chapter 2, we delved into the theory of health risk perception and analyzed historical biosurveillance data to understand when and why social response occurs. We found that social response is most likely when the disease is novel to the region, not well understood or has high clinical severity. In Chapter 3, we reviewed the literature on models of disease spread and social processes. In Chapter 4, we introduced a model of the joint diffusion of disease and social response. We incorporated our findings from Chapter 2 into the model by introducing a key parameter, κ, which reflects the perceived risk of the disease to the society. The perceived risk determines the degree to which media becomes involved in the spread of social response. In Chapter 5, we formulated an optimization problem intended to minimize the expected total cost of the outbreak, assuming that the cost is the sum of the costs of the disease itself, the intervention to prevent the spread of disease and the social response. In Chapter 6, we described the implementation of the model and introduced a number of studies intended to determine the behavior of the model and to gain insight into how to control the spread of social response. In Chapter 7, we discussed our results. In particular, the numerical experiments to understand model behavior showed that the model will produce a social response either when the disease has characteristics, such as high novelty or severity, that give it high perceived risk or when the disease spread is very large. The numerical experiments in which we compared the social response and disease spread with different types of control measures showed

that reducing the number of cases of disease will reduce the total social response but not the peak mean social response. We also found that, in general, vaccination was more effective at stopping the spread of disease and social response than was contact reduction. The optimization approach to finding the least cost disease control strategy revealed that considering the costs of social response when determining how to intervene in disease outbreaks results in implementing greater control measures, especially for diseases with high novelty or severity. Finally, the simulations of real disease outbreaks demonstrated that the model can accurately model real-world infectious disease and the corresponding social response.

This model is among the first to take into account social processes accompanying the spread of disease. It can help public health officials determine how best to respond to disease outbreaks, by helping them to consider the effects of the spread of disease and government action on the social response. This model contributes to the body of knowledge in several ways. Perhaps most importantly, the present model accounts for discrepancies between perceived and actual risk of infection. One might assume that social response to disease spread is proportional to the overall burden of the disease. Under this hypothesis, we would expect that low probability, high severity diseases and high probability, low severity diseases would elicit little social response, yet this is not the pattern observed. Instead, while high probability, low severity diseases, such as seasonal influenza, do not typically elicit social response, high severity, low probability diseases, such as Ebola and SARS, frequently do. Moreover, the same disease can result in different social responses in different locations or even in the same location at different times, as with H1N1 in Mexico City. Through incorporation of a mechanism for risk amplification, the present model is able to realistically model these scenarios.

The model introduces a mechanism for media influence. Media undoubtedly plays a large role in disseminating information about the threat of disease and, according to the social amplification of risk framework, can be responsible for increasing concern about infection. Bomlitz and Brezis collected data on the number of newspaper articles about different types of diseases [83]. Among the risks that they examined, newspaper coverage of a health risk was negatively correlated with the number of deaths from the risk. This finding is seemingly counter-intuitive, but it is consistent with the research showing that perceived risk and actual risk can differ. Bomlitz and Brezis found that novel, low incidence risks, such as SARS and bioterrorism, garner the bulk of attention, while more commonplace risks are comparatively ignored. Our model captures this behavior. Media actively influences the population when novelty and severity are high, regardless of incidence. We have shown that even when a disease has low incidence and social response does not spread through word-of-mouth, media influence can produce a social response.

8.2 Directions for future research

We have demonstrated that the model can accurately model infectious disease outbreaks with and without social response. We have also shown how the model can be applied to provide insight into how to control the spread of disease and social response. We will now discuss ways in which the model can be improved and expanded upon, as well as suggest future studies that could be completed using the model.

In this thesis, we have explored what happens when disease control measures produce social response. The converse could also occur. That is, social response could compel authorities to take action about the disease. This situation could occur when the government has a vested interest in keeping news of the spread of the disease quiet or when the groups of people most affected by the disease are not a high priority for the government. The key change to the model required to simulate this scenario would be to base the initiation of an intervention off of a threshold for the social response, as opposed to a threshold for the disease spread.

There are several directions for future research related to the structure of the network and the overall formulation of the model. As currently formulated, the model is most applicable to studying diseases that have airborne transmission or are spread via contact. It is less appropriate for vector-borne, water-borne or food-borne illness, since these types of diseases do not spread along human social networks. Changes would need to be made to the disease transmission process and the coupling mechanism between disease spread and social response to make the model more suitable for use with these types of diseases. These changes could consist of changing the structure of the network along which the disease is transmitted. For example, cholera transmission has been modeled as a diffusion process on a hydrological network [84]. Or the changes could be more fundamental. For example, dengue fever, a vector-borne disease, is frequently modeled using compartmental models, which describe the interaction between the mosquito and human populations [85]. An additional direction to explore regarding the model formulation would be to experiment with how the model results are affected by the choice of network. In the literature review, we discussed several agent-based models employing more sophisticated, realistic network structures. While we believe that our decision to use a random network was appropriate since our goal was to formulate a general model, a more realistic network structure might produce more reliable results and could reveal interesting implications about how the structure of social networks affects the transmission of social response.

Another direction for future research is an expansion of the control measures study presented here. We have looked at least cost strategies for vaccination prior to the outbreak and contact reduction during the outbreak. This study could be expanded to include vaccination during the outbreak. Additionally, for the contact reduction study, we fixed the duration of the intervention (T_I) at two weeks and the threshold number of cases (τ) at 100. Optimizing over T_I, in addition to the amount of contact reduction (η), could produce additional insights into how to use disease

control measures to reduce social response.

In the current research, we have considered all types of social response to lie along the same continuum. We know, however, that different types of social response can have different effects. Flight has been shown to further the spread of disease [57, 55]. Reduction in travel might have the opposite effect. One reason why we did not differentiate between different types of social responses is because we do not know why we observe one type of response to one disease and a different type of response to another. Some researchers hypothesize that the public response depends on the way that public officials respond to and communicate about the disease [3]. Additional research is required to differentiate between the different types of social response and determine when each type of response is expected to occur. With that knowledge, it would be possible to consider the feedback effects of the social response on the spread of disease.

Finally, we are interested in developing a more data-driven approach to estimating κ. The current method for estimating κ is to use local doctors' familiarity with the disease and the perception of the likely threat of the disease to arrive at an intuitive estimate of κ. This process could be automated by developing a predictive model that estimates the likely severity of the social response, given the country's characteristics and experience with the disease. The creation of a such a model would reduce the potential for bias in the estimation of κ. Additionally, such a model could better account for the economic and political situation in the country in which the disease outbreak takes place. For example, we found that per capita gross domestic product and health expenditure as a percentage of GDP are useful for predicting social response to dengue fever. These predictors did not play a role in our manual estimation of κ, but could easily be incorporated into a predictive model.

Bibliography

[1] N. Ferguson. Capturing human behaviour. *Nature*, 446(7137):733, 2007.

[2] C. T. Bauch and A. P. Galvani. Social factors in epidemiology. *Science*, 342(6154):47–49, 2013.

[3] T. A. Glass and M. Schoch-Spana. Bioterrorism and the people: How to vaccinate a city against panic. *Clin. Infect. Dis.*, 34(2):217–223, Jan 15 2002.

[4] P. Strong. Epidemic psychology - A model. *Sociol. Health Ill.*, 12(3):249–259, 1990.

[5] J. Zarocostas. Violence hinders efforts to stem Haiti cholera epidemic as deaths continue to rise. *Br. Med. J.*, 341, 2010.

[6] J. Kinsman. "A time of fear": Local, national, and international responses to a large Ebola outbreak in Uganda. *Global. Health*, 8, 2012.

[7] C. Cheng. To be paranoid is the standard? Panic responses to SARS outbreak in the Hong Kong Special Administrative Region. *AP*, 28(1):67–98, 2004.

[8] J. M. Wilson, M. G. Polyak, J. W. Blake, and J. Collmann. A heuristic indication and warning staging model for detection and assessment of biological events. *J. Am. Med. Inform. Assoc.*, 15(2):158–171, 2008.

[9] J. W. McGrath. Biological impact of social disruption resulting from epidemic disease. *Am. J. Phys. Anthropol.*, 84(4):407–419, 1991.

[10] C. Chin et al. The origin of the Haitian cholera outbreak strain. *New Engl J Med*, 364:33–42, 2011.

[11] WHO. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. *GAR*, 2004.

[12] G. M. Leung et al. Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong. *Clin. Infect. Dis.*, 40(12):1713–1720, 2005.

[13] A. Siu and Y. C. R. Wong. Economic impact of SARS: The case of Hong Kong. *Asian Econ.*, 3(1):62–83, 2004.

[14] R. Herriman. Singapore reports 22,318 dengue fever cases in 2013, easily surpassing the 2005 epidemic. Availiable from: `http://www.theglobaldispatch.com`, 2014.

[15] S. Tiong and S. Tan. Singapore residents more worried over dengue. Availiable from: `http://yourhealth.asiaone.com/content/singapore-residents-more-worried-over-dengue`, 2013.

[16] R. E. Kasperson et al. The social amplification of risk - A conceptual framework. *Risk Anal.*, 8(2):177–187, 1988.

[17] I. M. Rosenstock. Why people use health services. *Milbank Q.*, 1966.

[18] D. L. Floyd, S. Prentice-Dunn, and R. W. Rogers. A meta-analysis of research on protection motivation theory. *J. Appl. Soc. Psychol.*, 2006.

[19] A. Leppin and A. R. Aro. Risk perceptions related to SARS and avian influenza: Theoretical foundations of current emperical research. *Int J. Behav. Med.*, 2009.

[20] AscelBio. `http://ascelbio.com`, 2013.

[21] Measles vaccination. Available from: `http://www.cdc.gov/measles/vaccination.html`, 2013.

[22] Poliomyelitis. Available from: `http://www.who.int/mediacentre/factsheets/fs114/en/`, 2013.

[23] L. C. Chosewood and D. E. Wilson, editors. *Biosafety in microbiological and biomedical laboratories*. HHS, 5th edition, 2009.

[24] Dengue. Available from: `http://cdc.gov/dengue/`, 2014.

[25] O. J. Brady et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLOS ONE*, 2012.

[26] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. Springer, 2nd edition, 2008.

[27] Y. Freund and R. E. Schapire. A short introduction in boosting. *Jpn. J. Artif. Intell.*, 14(5):771–780, 1999.

[28] M. Culp, K. Johnson, and G. Michailidis. ada: an R package for stochastic boosting. *J. Stat. Softw.*, 17(2), 2006.

[29] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE*, 21(9), 2009.

[30] W. O. Kermack and A. G. McKendrick. Contribution to the mathematical theory of epidemics. *Proc. R. Soc. A*, 115(772):700–721, 1927.

[31] N. C. Grassly and C. Fraser. Mathematical models of infectious disease transmission. *Nature Rev. Microbiol.*, 6(6):477–487, 2008.

[32] M. Tizzoni et al. Real-time numerical forecast of global epidemic spreading: Case study of 2009 A/H1N1pdm. *BMC Med.*, 10(165), 2012.

[33] J. C. E. Metcalf, K. Hampson, A. J. Tatern, B. T. Grenfell, and O. N. Bjornstad. Persistence in epdemic metapopulations: Quantifying measles, mumps, rubella and whooping cough. *PLOS ONE*, 8(9):e74696, 2013.

[34] M. Ajelli et al. Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infect. Dis.*, 10(1):1–13, 2010.

[35] I. A. Doherty, N. S. Padian, C. Marlow, and S. O. Aral. Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *J. Infect. Dis.*, 191:542–554, 2005.

[36] J. J. Grefenstette et al. FRED (a Framework for Reconstructing Epidemic Dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health*, 13, 2013.

[37] S. T. Brown et al. Would school closure for the 2009 H1N1 influneza epidemic have been worth the cost?: A computational simulation of Pennsylvania. *BMC Public Health*, 11, 2011.

[38] B. Y. Lee et al. A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 pandemic. *Vaccine*, 28(31):4875 4879, 2010.

[39] D. L. Chao, M. E. Halloran, V. J. Obenchain, and I. M. Longini. FluTE, a publicly available stocastic influenza epidemic simulation model. *PLOS ONE*, 2010.

[40] K. R. Bisset, A. M. Aji, M. V. Marathe, and W. Feng. High-performance biocomputing for simulating the spread of contagion over large contact networks. *BMC Genomics*, 13(53), 2012.

[41] J. Parker and J. M. Epstein. A distributed platform for global-scale agent-based models of disease transmission. *ACM T. Model. Comput. S.*, 22(1), 2011.

[42] S. Eubank, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, and N. Wang. Structure of social contact networks and their impact on epidemics. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, 70, 2006.

[43] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286, 1999.

[44] F. Chung and L. Lu. Connected components in a random graph with given degree sequences. *Ann. Comb.*, 6:125–145, 2002.

[45] M. H. DeGroot. Reaching a consensus. *J. Am. Stat. Assoc.*, 69(345):118–121, 1974.

[46] D. Acemoglu and A. Ozdaglar. Opinion dynamics and learning in social networks. *Dyn. Games Appl.*, 1(1):3–49, 2011.

[47] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis)information in social networks. *GEB*, 70:194–227, 2010.

[48] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar. Opinion flucuations and presistent disagreement in social networks. *IEEE*, pages 2347–2352, 2011.

[49] N. J. Howard. Finding optimal strategies for influencing social networks in two player games. Master's thesis, MIT, 2011.

[50] B. W. K. Hung. Optimimization-based selection of influential agents in a rural afghan social network. Master's thesis, MIT, 2010.

[51] M. Bradwick. Belief propagation analysis in two-player games for peer-influence social networks. Master's thesis, MIT, 2012.

[52] J. A. Evans. Modeling social response to the spread of an infectious disease. Master's thesis, MIT, 2012.

[53] M. J. Keeling and K. T. D. Eames. Networks and epidemic models. *J. R. Soc. Interface*, 2(4):295–307, Sep 22 2005.

[54] S. Funk, M. Salathe, and V. A. A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface*, 7(50):1247–1256, Sep 6 2010.

[55] J. M. Epstein, J. Parker, D. Cummings, and R. A. Hammond. Coupled contagion dynamics of fear and disease: Mathematical and computational explorations. *PLOS ONE*, 3(12), Dec 16 2008.

[56] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci. U. S. A.*, 106(16):6872–6877, Apr 21 2009.

[57] S. Meloni, N. Perra, A. Arenas, S. Gomez, Y. Moreno, and A. Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci. Rep.*, 1, 2011.

[58] D. P. Durham and E. A. Casman. Incorporating individual health-protective decisions into disease transmission models: A mathematical framework. *J. R. Soc. Interface*, 9(68):562–570, Mar 7 2012.

[59] J. A. Evans, S. M. Fast, and N. Markuzon. Modeling the social response to a disease outbreak. *LNCS*, 7812:154–163, 2013.

[60] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. U. S. A.*, 99(3):7280–7287, May 14 2002.

[61] M. E. J. Newman. *Networks: An introduction.* Oxford, 2010.

[62] M. E. Young, N. King, S. Harper, and K. R. Humphreys. The influence of popular media on perceptions of personal and population risk in possible disease outbreaks. *Health Risk and Society*, 15(1):103–114, 2013.

[63] L. Bonneux and W. V. Damme. An iatrogenic pandemic of panic. *BMJ*, 332:786–788, 2006.

[64] S. L. Gilman. Moral panic and pandemics. *Lancet*, 375(9729):1866–1867, 2010.

[65] S. Maharaj and A. Kleczkowski. Controlling epidemic spread by social distancing: Do it well or not at all. *BMC Public Health*, 12(679), 2012.

[66] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[67] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, page 1695, 2006.

[68] D. M. Bell, I. B. Weisfuse, M. Hernandez-Avila, C. del Rio, X. Bustamante, and G. Rodier. Pandemic influenza as 21st century urban public health crisis. *Emerg. Infect. Dis.*, 15(12):1963–1969, 2009.

[69] C. Fraser et al. Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science*, 324:1557–1561, 2009.

[70] G. Chowell et al. Characterizing the epidemiology of the 2009 Influenza A/H1N1 pandemic in Mexico. *PLOS Med.*, 8(5), 2011.

[71] How flu spreads. Available from: `http://www.cdc.gov/flu/about/disease/spread.htm`, 2013.

[72] J. T. Wu et al. The infection attack rate and severity of 2009 pandemic H1N1 influenza in Hong Kong. *Clin. Infect. Dis.*, 51(10):1184–1191, 2010.

[73] S. Riley et al. Epidemiological characteristics of 2009 (H1N1) pandemic influenza based on paired sera from a longitudinal community cohort study. *PLOS Med.*, 8(6), Jun 2011.

[74] J. T. F. Lau, S. Griffiths, K. C. Choi, and H. Y. Tsui. Widespread public misconception in the early phase of the H1N1 influenza epidemic. *J. Infect.*, 59(2):122–127, 2009.

[75] B. J. Cowling et al. Community psychological and behavioral responses through the first wave of the 2009 influenza A(H1N1) pandemic in Hong Kong. *J. Infect. Dis.*, 202(6):867–876, Sep 15 2010.

[76] Google trends. Available from: `http://google.com/trends/explore`, 2013.

[77] CDC. Update: Novel influenza A (H1N1) virus infection - Mexico, March-May, 2009. *MMWR*, 58(21):585–589, Jun 5 2009.

[78] C. Hawley. Swine flu spreads panic in Mexico City. Availiable from: `http://usatoday30.usatoday.com/news/world/2009-04-25-mexicocity-flu_N.htm`, Apr 25 2009 Apr 25.

[79] K. Penhaul. Swine flu fears stalk the living and the dead in mexico city. cnn [internet]. Availiable from: `http://www.cnn.com/2009/WORLD/americas/05/01/penhaul.mexico.flu/`, 2009 May 1.

[80] R. Rodriguez and E. Martinez. En escuelas, tercer rebrote de AH1N1. Availiable from: `http://eluniversal.com.mx/primera/33638.html`, 2009.

[81] D. Kotz. Mayor Menino declares public health emergency in Boston because of flu outbreak. Availiable from: `http://www.boston.com/dailydose`, 2013.

[82] Boston Public Health Commission. 2012-2013 influenza season.

[83] L. J. Bomlitz and M. Brezis. Misrepresentation of health risks by mass media. *Journal of Public H*, 30(2):202–204, Feb 2008.

[84] L. Mari, E. Bertuzzo, R. Righetto, M. Gatto, and I. Rodriguez-Iturbe. Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *J. R. Soc. Interface*, 9:376–388, 2012.

[85] M. N. Burattini et al. Modelling the control stategies against dengue in Singapore. *Epidemiol. Infect.*, 136(3):309–319, 2008.

Appendix A

Supplementary materials

Table A.1: Publicly available predictors used in dengue fever prediction

	predictor
1	North America
2	South Asia
3	Southeast Asia
4	East Asia
5	Middle East
6	Australasia
7	Carribbean
8	Latin America
9	Europe
10	Africa
11	latitude
12	longitude
13	human development index (2007)
14	human development index growth from 1990 to 2000
15	percentage of females with secondary education (2007)
16	percentage of males with secondary education (2007)
17	public health expenditure as a percentage of GDP (2007)
18	public spending on education as a percentage of GDP (2007)
19	public spending on the military as a percentage of GDP (2007)
20	health expenditure per capita (2008)
21	gross domestic product per capita (2009)
22	infant mortality per 1000 (2010)
23	under 5 mortality per 1000 (2010)
24	adult female mortality per 1000 (2009)
25	adult male mortality per 1000 (2009)
26	deaths per 1000 due to malaria
27	deaths per 1000 due to cholera
28	deaths per 100,000 due to cardiovascular disease and diabetes
29	number of reported malaria cases (2008)
30	malaria incidence per 100,000 (2008)
31	average annual precipitation from 1960 to 1990
32	DPT immunization rate (2008)
33	endemic for dengue fever (2009, assembled from WHO, CDC, and NIH reports as well as academic sources)

	predictor
34	physicians per 1000
35	hospital beds per 10,000 (2009)
36	percentage satisfied with healthcare
37	percentage that trust in people
38	percentage satisfied with their community
39	percentage that trust in the national government
40	percentage that have a perception of safety
41	homicide rate per 100,000
42	percentage of population with electricity
43	percentage of population with improved sanitation (2008)
44	internet users per 100
45	fixed and mobile phone subscribers per 100
46	newspaper subscribers per 1000 (most recent year available; 1996 to 2013)
47	Bertelsman transformation index (BTI) status index
48	BTI democracy status
49	BTI market economy status
50	BTI management index
51	BTI management performance
52	Transparency International corruption perception index (2009)
53	refugee population from country (2008)
54	refugee population in country (2008)
55	Spanish is the most commonly spoken language
56	English is the most commonly spoken language
57	Arabic is the most commonly spoken language
58	Chinese is the most commonly spoken language
59	Christianity is the most commonly practiced religion
60	Islam is the most commonly practiced religion