ABSTRACT

Title of dissertation: INTERNATIONAL EXTERNALITIES IN PANDEMIC INFLUENZA MITIGATION

Stephen R. Hutton, Doctor of Philosophy, 2011

Dissertation directed by: Professor Maureen Cropper
Department of Economics

A serious influenza pandemic could be devastating for the world. Ideally, such a pandemic could be contained, but this may be infeasible. One promising method for pandemic mitigation is to treat infectious individuals with antiviral pharmaceuticals. While most of the benefits from treatment accrue to the country in which treatment occurs, there are some positive spillovers: when one country treats more of its population this both reduces the attack rate in the other country and increases the marginal benefit from additional treatment in the other country. These externalities and complementarities may mean that self-interested rich countries should optimally pay for some AV treatment in poor countries.

This dissertation demonstrates the presence of antiviral treatment externalities in simple epidemiological SIR models, and then in a descriptively realistic Global Epidemiological Model (GEM). This GEM simulates pandemic spread between cities through the international airline network, and between cities and rural areas through ground transport.

Under the base case assumptions of moderate transmissibility of the flu, the distribution of antiviral stockpiles from rich countries to poor and lower middle income countries may indeed pay for itself: providing a stockpile equal to 1% of the population of poor countries will reduce cases in rich countries after 1 year by about 6.13 million cases at a cost of 4.62 doses per rich-country case avoided. Concentrating doses on the outbreak country is, however, even more cost-effective: in the base case it reduces the number of influenza cases by 4.76 million cases, at the cost of roughly 1.92 doses per case avoided. These results depend on the transmissibility of the flu strain, the efficacy of antivirals in reducing infection and on the proportion of infectious who can realistically be identified and treated.

INTERNATIONAL EXTERNALITIES IN PANDEMIC INFLUENZA MITIGATION

by

Stephen R. Hutton

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory committee:

Professor Maureen Cropper (Chair)
Professor Peter Cramton
Professor John Rust
Professor Ginger Jin
Professor Donald Milton

Dedication:

To Erica, for standing with me all these years.

Acknowledgements

This research is based on the Global Epidemic Model (GEM) developed by Michael Goedecke and Georgiy Bobashev (RTI International). I am also grateful for assistance and support from Maureen Cropper (University of Maryland and Resources for the Future), Joshua Epstein (Johns Hopkins University), and Mead Over (Center for Global Development), who have contributed to this research. This work was supported in part by the Pilot Studies of Modeling of Infectious Disease Agents Study (MIDAS) cooperative agreement from NIGMS (1 U01 GM0700698). I thank the World Bank Research Board and the KCP Trust Fund for research support, and seminar participants at the Center for Global Development and World Bank for useful comments.

Table of contents

List of figures

Chapter 1 Introduction

In the 20th century the world experienced three influenza pandemics (1918, 1957 and 1968), which had significant economic costs and caused millions of illnesses and deaths. The 1918 pandemic killed an estimated 3% of the world's population, despite being slow to spread between countries in an age before air travel. With today's global transportation network, an outbreak of influenza could quickly reach pandemic proportions. There have been concerns since the 1990s about the potential for an H5N1 "avian flu" strain mutating into a form that can be transmitted between humans. The 2009 H1N1 "swine flu" pandemic and the SARS outbreak of 2003 remind us of the continued risks to the world should a pandemic occur. They also remind us that there is relatively little cooperation and coordination between countries, and that wealthy nations prioritize stockpiling doses of vaccines or antivirals for their own citizens before considering treatment in other countries.

Whether such an inwardly focused policy is optimal depends on the nature and magnitude of externalities in treating pandemic flu. How much do policies to slow the spread of the flu in one county reduce attack rates in other countries? Does treatment of infected persons in one country increase the marginal benefits of treatment policies in other countries (i.e., are there treatment complementarities?). The overarching goal of this dissertation is to answer both questions in a realistic model of the spread of influenza through the global air transport network.

Simulating the impact of control strategies in a global epidemiological model allows examination of two questions regarding international cooperation to mitigate a

pandemic: (1) Is it cost-effective for wealthy nations to pay for the purchase and distribution of antivirals in poor countries to slow the spread of the pandemic? (2) What global allocation rules are most effective in reducing attack rates? In the case where treatment with antiviral drugs alone can contain a potential pandemic, there is an obvious case for wealthy nations to pay for pandemic containment. But what about the case where a pandemic cannot be detected early enough or treatment is not effective enough for containment to be possible? Pandemic epidemiology involves two types of externalities that suggest that it might be in the self-interest of wealthy countries to fund such a scheme. The treatment policy in one country will affect the rate at which the pandemic spreads to other countries, so treatment provides a positive externality. At the same time, the increasing marginal effectiveness of treating more people can lead to complementarities across countries. The question is: how large are these effects?

To investigate these issues this dissertation uses a detailed Global Epidemiological Model to simulate influenza pandemics under a range of conditions and antivirus treatment policies. The model divides the world into 106 regions and models travel among 288 cities and 101 rural areas in these regions. The flu spreads from one city to another via air travel, and from cities to a rural area in each region via land travel. Within each city or rural area the flu spreads via a model in which people transition from susceptible to exposed to infectious to recovered or dead. The model distinguishes among age groups and uses age-specific contact rates to model the spread of the flu. The probability of infection given contact in a particular region varies with latitude (it decreases as one moves away from the equator) and season (it is higher during the winter than during the summer).

With or without control policies, the number of people infected in each region depends on when the flu begins (influenza that peaks in Northern Hemisphere winter infects more people) and on the infectiousness of the flu (i.e., the reproductive rate R_0, which measures the number of people an infectious person would infect in an otherwise totally susceptible population).

Ideally, a serious influenza outbreak could be contained and a pandemic prevented, but the difficulty in detecting a potential pandemic-strain before it spreads widely means that containment is generally infeasible. However, control policies can reduce number of people infected by a pandemic; this dissertation considers the effects of administering antiviral (AV) pharmaceuticals to symptomatic infectious individuals. It is not feasible to calculate optimal AV policies, but it is possible to consider a set of plausible AV scenarios.

In the absence of international cooperation it is assumed that antiviral stockpiles, as a percent of population, vary with per capita income. Poor countries are assumed to have no stockpiles. Compare two rules for rich countries distributing stockpiles in poor countries: one under which each country receives a fixed number of doses (in proportion to population) and another under which antivirals are allocated to the country in which the flu begins, which is assumed to be a poor country. The success of a control strategy is measured in terms of its impact on the attack rate (percent of the population infected) at the end of a year. Does it (collectively) pay rich countries to make a donation in terms of the impact it has on their own attack rate?

The benefits from collective action in the form of influenza treatment depend on the size of treatment externalities. Treatment externalities are large if a pandemic can be contained in the source country (Ferguson, et al., 2005; Longini, et al., 2005). In the more likely case in which the pandemic will spread through air travel the externalities associated with anti-viral treatment are smaller: treating infectious people in one's own country reduces the domestic attack rate, but has a proportionately much smaller impact on other countries. The question is whether the cost of purchasing and distributing antivirals to other countries pays for itself in terms of reducing a country's own attack rate.

It is always in the interest of wealthy countries to purchase and distribute antivirus doses in the outbreak country when doing so can contain a pandemic. In other cases, the marginal private benefits from using limited antivirus supplies to treat domestic patients exceed the marginal benefits from donating those doses abroad. But, if wealthy countries retain a stockpile of antiviral drugs sufficient to treat their own cases, then they can increase their welfare by paying for purchase and distribution of additional doses to the outbreak source country.

The benefits to rich countries of paying for antivirals in poor countries under "midrange" assumptions of influenza transmissibility cover the costs: donation of antivirus to the outbreak source country reduces the number of influenza cases in rich countries after 1 year by 4.76 million cases, at the cost of roughly 1.92 doses per case avoided. This donation policy is welfare-enhancing for wealthy countries even at a zero percent case fatality rate; at any positive fatality rate the policy is even more valuable.

The dissertation is organized as follows. Chapter 2 briefly reviews the literature on policies to control pandemic flu. Chapter 3 discusses the dynamics of pandemic influenza in simple one-city and two-city S-I-R models. These models are used to illustrate the nature of externalities and complementarities present in treating influenza epidemics through antivirals, and to contrast the Nash equilibrium in treatment strategies with the socially optimum treatment strategy. Chapter 4 presents the Global Epidemiological Model, and describes the behavior of the model in a baseline scenario with no policy interventions. Chapter 5 presents the results of simulating pandemic flu under various antiviral stockpile assumptions. Chapter 6 concludes.

Chapter 2 Literature on pandemic influenza

To put this research into context, it is useful to consider three avenues of research in the existing literature: 1) the health literature on optimal vaccination policy for epidemic diseases, 2) a set of spatially detailed epidemiological models that model influenza spread within a city or country, and 3) a set of less detailed but global models that track the spread of the flu through the international airline network.

The literature on optimal vaccination policy considers the externality effects of vaccination policies, either domestically or internationally. Vaccination provides significant benefits for an individual who is vaccinated, but also large external benefits to others in the same region; sufficiently high levels of vaccination lead to a "herd immunity" effect, where the reproductive rate of the disease is reduced to the point where epidemics can be prevented.

Francis (2004) considers optimal use of vaccines to control an influenza epidemic using a single region SIR model. The model assumes that a perfectly effective vaccine exists at the beginning of the epidemic, and demonstrates that there can be a significant positive externality from vaccination. The model assumes a decentralized setting, where agents optimally choose to vaccinate or not in each period as a function of the current model state variables. There is thus there is an insufficient level of vaccination in a private market relative to the socially optimal vaccination level. Francis develops a system of optimal vaccine subsidies (or taxes for avoiding vaccination) that internalize

this externality, and so induce agents to purchase an efficient level of vaccine; the optimal subsidy pays less than 100% of the price of vaccination.

Boulier et al. (2007) also consider vaccine use in a single region SIR model, focusing on the magnitude of the externalities in vaccine use. They consider a setting where a vaccine exists at the start of an epidemic, but is not 100% effective. They demonstrate that the marginal value to society from vaccination is positive and increases with more people vaccinated as long as the effective reproductive rate of the epidemic is greater than 1, and declines thereafter. They estimate the average value of vaccination for influenza, and find that the size of the externality from vaccination can be greater than 1 case reduced per person vaccinated.

Barrett (2003) considers international vaccination externalities for disease eradiation of endemic diseases (as opposed to epidemic diseases). Such a disease can be eradicated if a sufficiently large proportion of people are vaccinated, but this vaccination must occur in every country, and thus eradication is a "weakest link" game. Barrett shows that eradication will often not be the Nash equilibrium of this game, because poor countries may derive a lower local benefit from vaccination relative to vaccination costs, and because the herd immunity effect means that the benefits from vaccination to the last countries to vaccinate will be low, since vaccination elsewhere will have dramatically reduced the global disease prevalence. A cooperative international regime with enforcement mechanisms may be needed to achieve the eradication outcome.

Spatially detailed models of avian flu transmission at the country level have been used to compare a wide set of policy interventions. These models are able to use a much

greater level of detail and structure in metapopulations, and so are effective in analyzing policy interventions aimed at containing an influenza outbreak, such as ring quarantines or prophylaxis, or contact tracing strategies that use vaccination or antiviral strategies to target those likely to be contacted by infectious individuals. However, these models cannot capture the international dimensions of pandemic control policies.

Ferguson et al. 2005 use a spatially explicit model of the 85 million people living in Thailand and a 100km radius of neighboring contiguous countries. The model tracks influenza spread through households, workplaces, schools and random community contact. They consider targeted prophylactic use of antivirus to contain a pandemic. They find that purely social targeting (blanket prophylaxis of all household, workplaces and schools that have any new infectious case) has a 90% probability of containing a pandemic with reproductive rate 1.25, so long as it can be implemented after only 20 cases have been observed. They find that ring prophylaxis (targeting everyone within a given radius of a new infectious case) with a 5 kilometer radius can contain a pandemic with reproductive rate of up to 1.5 at an average cost of 2 million AV courses. They also consider physical distancing measures, where schools and workplaces are closed; these measures combined with prophylaxis can contain a pandemic of reproductive rate up to 1.7 with 90% probability. However, all policies require early detection of a pandemic strain and extremely rapid response.

Longini et al. 2005 use a stochastic model to track influenza spread through a structured geographically distributed population of 500,000 people in rural SE Asia to examine the effectiveness of antivirus, quarantine and pre-vaccination with a low efficacy non-strain-specific vaccine. They use data from Thailand to model the number of close

and casual daily contacts between individuals of various age-groups to build up a contact rate matrix. The model includes a structured social network that maps households, household clusters, workplaces, communal mixing and a regional hospital; the network structure affects the mixing probabilities, and allows consideration of targeted policy interventions. They find that targeted use of antivirals have a high chance of containing a pandemic for reproductive rates of up to 1.4, if implemented within 21 days of the first case. Antivirals can be effective for reproductive rates up to 1.7 if accompanied by pre-vaccination.

Germann et al. 2006 use a stochastic agent-based simulation model to track spread of an influenza pandemic through the United States. They use US Census and Department of Transport data to construct a detailed spatial model of the United States, with a metapopulation of 281 million people distributed among 65,334 census tracts, each organized into 2,000 communities. The model maps households, household clusters, preschools, playgroups, schools, workplaces neighborhoods and communities. They consider a range of policy interventions including targeted prophylaxis, mass vaccination, school closure, and quarantine. To achieve reductions in attack rate down to 10% through targeted antivirus use alone, 10 million doses are needed for an influenza strain with a reproductive rate of 1.7, and 51 million doses for a reproductive rate of 1.8. But in all cases, this targeted treatment requires rapid and effective targeting, with a knowledge of population clusters that may not be available to public health agencies without significant upfront investment.

A series of global models follow Rvachev and Longini (1985) in using and refining an SIR model core to track the progress of an influenza pandemic. Rvachev and

Longini develop a deterministic difference equation model to simulate travel within a network of 52 cities, and use this to reproduce the 1968-9 pandemic that started in Hong Kong. They find that the predicted local outbreak peak timing roughly matches that of the actual pandemic.

Grais et al. (2003) use a global influenza pandemic model with the same 52 cities from Rvachev and Longini. Their model adds seasonal variation in infectiousness; they use monthly seasonal scaling factors, one for the northern hemisphere and one for the southern hemisphere. They consider the dynamics of a pandemic that starts in Hong Kong, and show that the pandemic outbreaks occur concurrently in northern hemisphere and southern hemisphere cities despite the seasonal scaling factors. They show that the pandemic reaches northern hemisphere cities an average of 111 days earlier using year 2000 travel volumes rather than 1968 travel volumes. Policy interventions are not considered.

Hufnagel et al. (2004) use a global pandemic model to track a generic pandemic disease, and demonstrate the example to model the 2003 SARS outbreak. They use a probabilistic rather than deterministic model, where disease transmission and travel are stochastic. A stochastic model is potentially valuable in considering the early stage of a pandemic, where the variation in behavior by a few infected individuals can alter the course of the pandemic. However, they find that probabilistic variation has little impact at the global level, and that the pandemic outcomes remain predictable in spite of variation in the behavior of those individuals infected early on. They consider quarantine-like restrictions on individuals and travel restrictions on cities. Isolating only 2% of individuals can be highly effective in reducing attack rates (from 78.45% to

37.50%) but travel restrictions are less effective; it would require shutting down the 27.5% of the most highly trafficked air connections to achieve a similar effect.

Colizza et al. (2007) consider a global influenza pandemic model covering 3,100 urban areas (but no rural areas). They consider allocations of antiviral stockpiles across countries, but focus on containment rather than mitigation. Containment is possible in their model if outbreak cities have an antivirus stockpile, but this relies on optimistic assumptions about antivirus treatment; they assume that 50-70% of new infectious cases can be treated, and that treatment occurs immediately, and so antivirals are very effective in reducing influenza transmission. Their conception of global cooperation is one where AV doses are shared freely between countries, rather than one where rich countries retain most doses for domestic use. They conclude that pandemic strains with reproductive rates up to 1.9 can be controlled with global stockpiles sufficient to treat 2%-6% of the world population, and that there is a strong case for cooperative use of AV where rich countries share their stockpile with poorer countries, but this is contingent on their model's ability to contain the pandemic.

Epstein at al. (2007) use an earlier version of the Global Epidemiological Model (with 155 cities) used in this dissertation to examine the effect of travel restrictions. They find that travel restrictions alone provide only a small delay in the time it takes for a pandemic to arrive in the United States (2-3 weeks for an outbreak starting in Hong Kong; no delay for an outbreak starting in London), but note that delays can be longer when travel restrictions are combined with other policy measures. They also note the critical effect of seasonality; travel restrictions that delay the pandemic can increase the

severity of local outbreaks if the outbreaks are pushed into the high infectiousness season (winter).

This dissertation draws on the previous literature, but advances it in several important ways. Though the positive externalities from vaccination are high and are well understood, the externalities from pandemic mitigation through antiviral use are not well understood, particularly in the realistic case where antiviral use alone is incapable of containing a pandemic. This dissertation uses an epidemiological model similar to global models in the previous literature, but covers the entire global population of 6.41 billion rather than merely those in selected cities. It adopts an age structure and age-specific contact rates similar to those used in the country-level models, unlike previous global models.

Chapter 3 Basic Influenza Models

The main class of epidemiological models used to model the spread of an epidemic through a large population are known as "compartmental" models, where at any given time the population is divided amongst a set of mutually exclusive categories. This chapter describes the common core to these models, and uses simple one-city and two-city simulation examples to demonstrate key properties of model dynamics, and the potential for positive externalities in this framework.

i) SIR models

The simplest compartmental model is known as the SIR model, where each member of the population is Susceptible (S), Infectious (I) or has Recovered (R)[1]. The model is ideal for tracking spread of a disease such as a specific pandemic strain of influenza, where re-infection is not possible because people who recover develop antibodies that protect against the particular strain.

The SIR model tracks the number of people in each category in each time period, such that $S(t) + I(t) + R(t) = N$ (the population size).[2] Given initial values, the number of people in each category changes over time as Susceptible cases are infected by Infectious cases, and as Infectious cases Recover. Assume a very small but positive value of I(0), and assume $R(0) = 0$, so $S(0) \approx N$. The model operates in continuous time as described by:

[1] Sometimes this category is referred to as "Removed", since some proportion of infectious cases may die rather than recover. For most of this analysis we assume a zero fatality rate, but adding a small positive fatality rate has little impact on model dynamics.

[2] The model can also be written with S, I and R as proportions of the population, such that they sum to 1.

$$\frac{dS(t)}{dt} = -\beta I(t)S(t) \quad (1)$$

$$\frac{dI(t)}{dt} = \beta I(t)S(t) - \delta I(t) \quad (2)$$

$$\frac{dR(t)}{dt} = \delta I(t) \quad (3)$$

β represents the probability of transmission conditional on exposure to an infectious person, or the average rate of infection per susceptible.[3] βS(t) represents the rate at which an infectious individual infects susceptible persons. δ is the rate at which an infected individual recovers, and so 1/ δ is the average duration of infection. Disease prevalence increases over time (dI(t)/dt > 0) if and only if [β/δ] S(t) > 1. Simply put, prevalence increases only if the number of infections caused by an infectious person during the time he is infectious exceeds 1 (i.e., only if he can replace himself). Thus, a necessary condition for an epidemic to begin is for β/δ > 1. β/δ is also the average number of persons infected by an infectious person in an otherwise totally susceptible population (i.e., when S(0) = N) and is termed the basic reproductive rate, R_0.[4]

The system of equations does not have a closed form solution, but is known to have certain basic properties (Kermack and McKendrick 1927). It can be shown that:

$$S(t) = S(0)e^{-R_0(R(t)-R(0))} \quad (4)$$

[3] This implicitly assumes that the rate at which people contact each other remains constant throughout the epidemic; there is no scope for prevalence-elastic behavior, where individuals might take preventative actions to reduce exposure if much of the population is infected.

[4] We denote the reproductive rate at time t as R_t. R with a subscript will always refer to the reproductive rate, not the number of Recovered cases in period t, which we denote R(t) for both continuous and discrete time models. This notation is unfortunate, but mirrors that used in the literature where R is used both for the stock of Recovered and for the Reproductive rate.

And that in the limit as t→∞ the number of Recovered is described by a transcendental equation:

$$R(\infty) = N - S(0)e^{-R_0\left(R(\infty)-R(0)\right)} \qquad (5)$$

where R(∞) is the limiting value of R(t) as t →∞. With R(0) = 0, S(0) ≈ N and N = 1, (5) can be simplified as:

$$\ln\big(S(\infty)\big) = R_0(S(\infty) - 1) \qquad (6)$$

R(∞) < N, so the pandemic ends endogenously and the entire population is never infected. The reproductive rate R_t declines over time because there are fewer Susceptible cases left in the population, and dI(t)/dt is increasing in S(t). In practical terms, more of the people that a given Infectious person contacts have already Recovered and are immune to re-infection, so each successive Infectious case causes fewer new cases. As long as the reproductive rate R_t is greater than one, then the number of infectious cases I(t) is increasing. When R_t drops below one, then I(t) is decreasing, and the epidemic will end in the limit as I(t) and dI(t)/dt approach zero.

A slight variation on the core SIR model is the SEIR model, which adds an Exposed category. Susceptible cases that are contacted by an Infectious case move to an Exposed incubation state, where they are infected but not yet Infectious, before proceeding to Infectious after a fixed time period, and then on to Recovered. The model performs in much the same way, but is more descriptively realistic and allows us to fit observed real-world parameter values more easily. It is this SEIR model that underlies the GEM, but in this chapter I focus on simpler SIR models, which share the same

general properties. The SIR model's behavior can be difficult to demonstrate analytically, so instead I examine a series of simulation models to illustrate key properties of this class of models. This informal and qualitative illustration will inform the discussion of the full GEM results.

ii) One city model baseline

Consider a discrete time version of the SIR model in a single city of population N. The model is described by:

$$S(t+1) = S(t) - \beta I(t)S(t) \quad (7)$$

$$I(t+1) = I(t) + \beta I(t)S(t) - \delta I(t) \quad (8)$$

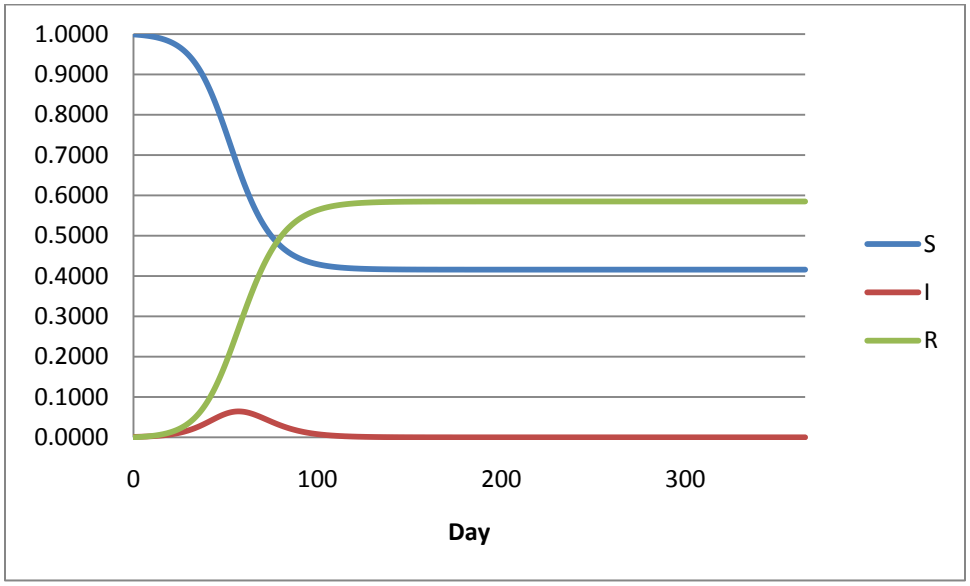
$$R(t+1) = R(t) + \delta I(t) \quad (9)$$

Assume R(0) = 0, assume I(0) is very small relative to N, and so S(0) = N – I(0) ≈ N. Specifying parameter values for β, δ and N and assuming a seed value for I_0 the model behavior to be observed, either analytically or by running the model using a simulation program. For example, suppose:

- β = 0.3 (each infectious person infects 0.3 people day when the entire population is susceptible)
- δ = 0.2 (an expected infectious duration of 5 days)
- Together these imply $R_0 = 1.5$
- N = 1 (which lets us interpret S, I and R as proportions)
- I(0) = 0.001 (very small relative to the population size)

Figure 3.1 shows how S(t), I(t) and R(t) behave over the course of an epidemic.

Figure 3.1: One city model, baseline



S, I and R are well-behaved[5]. The epidemic is effectively over in less than 4 months, with only 0.01N new infectious cases occurring after day 104. Much of the population remains uninfected at the end of the epidemic: the "attack rate"[6] (R(t)/N) after 1 year is 0.584; that is, 58.4% of the population have been infected, while 41.6% remain Susceptible. The peak of the epidemic occurs on day 57, where 6.38% of the population is Infectious. This implies that the reproductive rate on day 57 $R_{75} \approx 1$, and that $R_t < 1$ for $t > 57$.

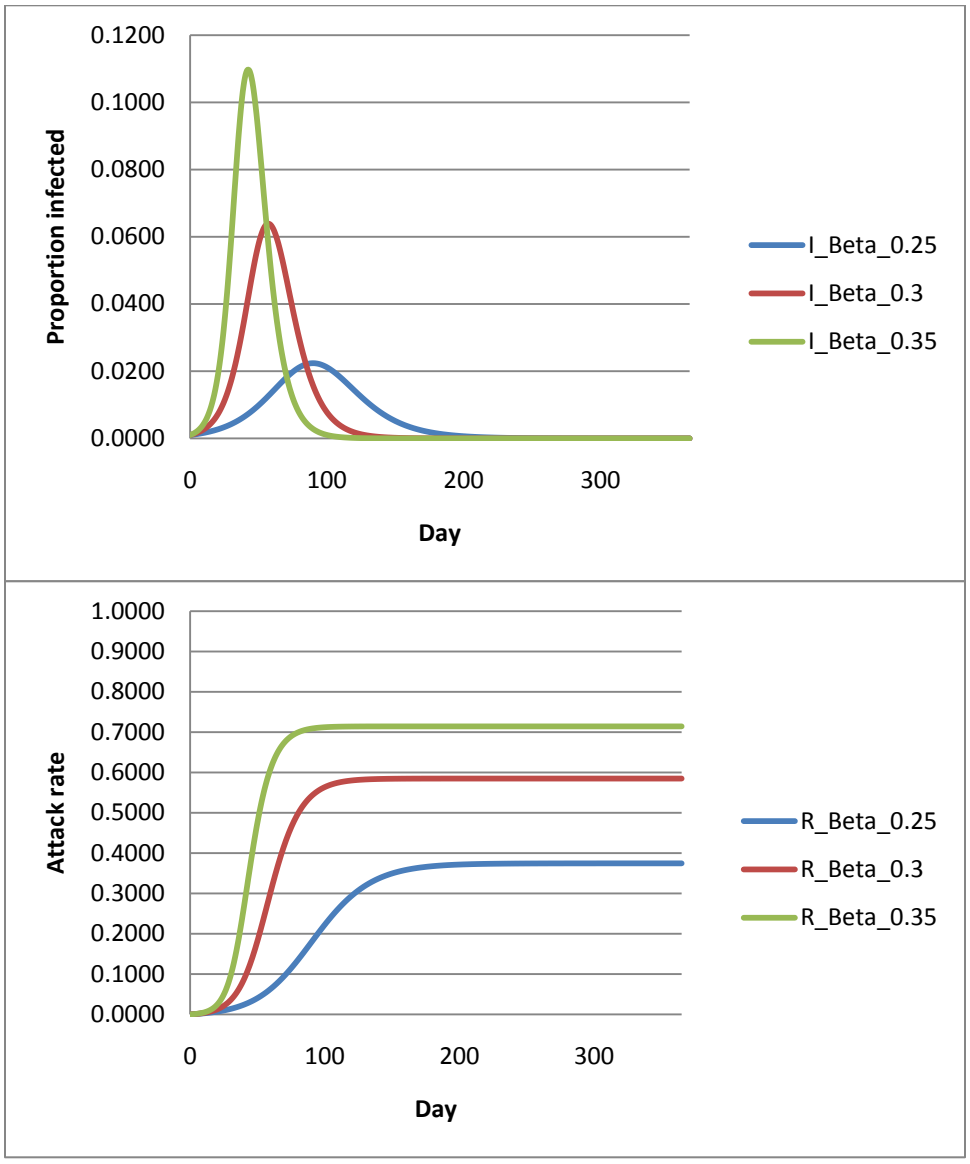
Changing the parameter assumptions changes the observed outcome in an intuitive fashion. For example, suppose $\beta = 0.25$ or $\beta = 0.35$, giving R_0 values of 1.25 or 1.75, respectively. Figure 3.2 shows the behavior of I(t) and R(t) under these

[5] S, I and R are not continuous functions of time – in a discrete time model these values are only defined at discrete values of t – so we cannot formally talk about behavior of some continuous function S(t). Nonetheless, we can observe what behavior is occurring, and what the limiting function as the length of each interval →0 would look like, and what properties this limiting function would have. This limiting function would be quasiconcave for S, I and R; increasing for R and decreasing for S.

[6] Technically the attack rate is(R(∞) – R(0))/N, but since R(0) = 0 is assumed in all cases it is simpler to refer to R(t)/N as the attack rate throughout.

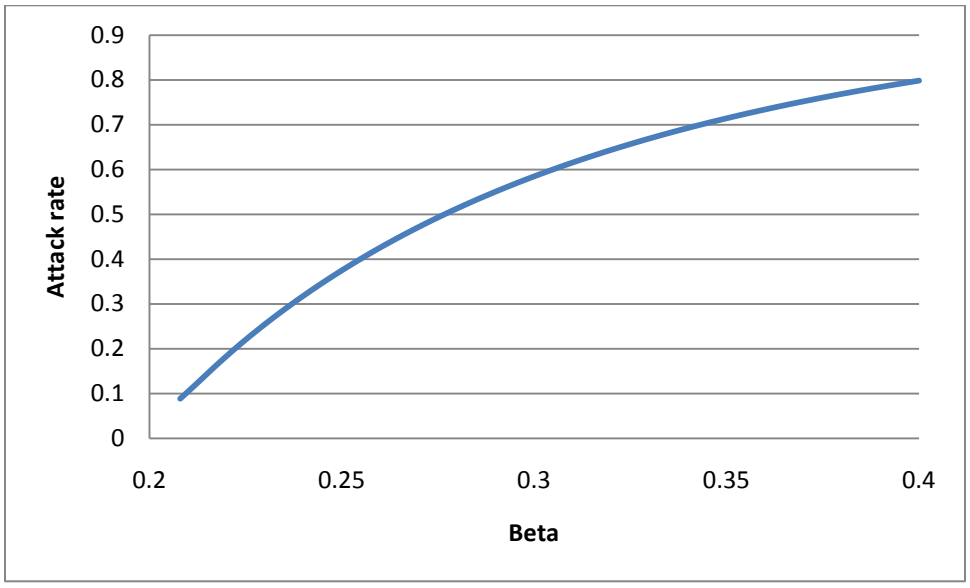
assumptions. Model outcomes are highly sensitive to β; higher values of β mean a more severe epidemic, with faster onset, higher peak, and higher attack rate.

Figure 3.2: One city model, Sensitivity to Beta



A key property of the model is how the attack rate changes as R_0 declines (see Figure 3.3). Recall that in the example with δ=0.2, R_0 approaches 1 as β approaches 0.2.

Figure 3.3: One city model, Attack rate sensitivity to Beta



When $\beta > \delta$ the attack rate is an increasing function of β and is strictly concave (above a threshold)[7]. This drives a key property of the SIR model: policy measures that reduce R_0 have increasing returns to scale as they drive R_0 towards 1 from above. This means that, in general, a policy that reduces R_0 from 1.7 to 1.6 will have a smaller impact on the gross attack rate than a policy that reduces R_0 from 1.4 to 1.3. This means that policies that reduce R_0 by a given amount are more effective in countries with lower disease transmission rates, due to density, social activity or climate. Many policies could potentially reduce R_0: treatment of Infectious cases with antivirus will reduce β and could increase δ, vaccinating Susceptible individuals (or providing them with prophylactic

[7] The function is not quite strictly concave when β is close to δ (ie when R_0) is close to 1, there is an inflection point in the function. See Appendix 1 for more details.

antivirus) would reduce β, using quarantine or other physical distancing measures to reduce contact rates would reduce β.

iii) One city model with anti-viral treatment

The SIR model can be used to analyze the impact of policy interventions, such as a policy to treat Infectious individuals with antiviral pharmaceuticals[8] (AV). This chapter considers a model where the policy choice variable is the size of an exhaustible stockpile of AV doses, mirroring the AV policy interventions used in the GEM (see Chapter 5). This model cannot be solved analytically, but its properties can be demonstrated through use of a simple simulation model. Appendix 1 presents an alternative model (where the choice variable is the proportion of population that is treated) where model outcomes can be computed directly.

Assume that there exists a fixed AV stockpile with P* doses, $0 \leq P^* \leq N$, and let P(t) be the number of doses consumed by day t (with P(0) = 0). Also assume that a fixed proportion *p* of new infectious cases are immediately treated with AV, for as long as P(t) < P*. Each case treated reduces the stockpile by one dose, so:

$$P(t+1) = P(t) + p\beta(t)I(t)S(t) \text{ if } P(t) < P^* - p\beta(t)I(t)S(t) \qquad (10)$$

$$P(t+1) = P^* \text{ otherwise}$$

[8] The most common antivirals are Zanamivir and Oseltamivir, sold and marketed as Tamiflu. Treatment involves taking a course of one of these drugs, typically one dose a day for 5-10 days. In this research a "dose" is used to refer to sufficient antivirals for a full course of treatment, not just a single capsule.

Treatment has the effect of reducing the effective infection rate by modifying the infectiveness parameter β. There is an underlying β^o determined by infectiousness and contact rates , and then an effective β(t) in any time period. Assume:[9]

$$\beta(t) = (1 - ep)\beta^o \text{ if } P(t) < P^* - p\beta(t)I(t)S(t) \qquad (11)$$

$$\beta(t) = \beta^o \text{ otherwise}$$

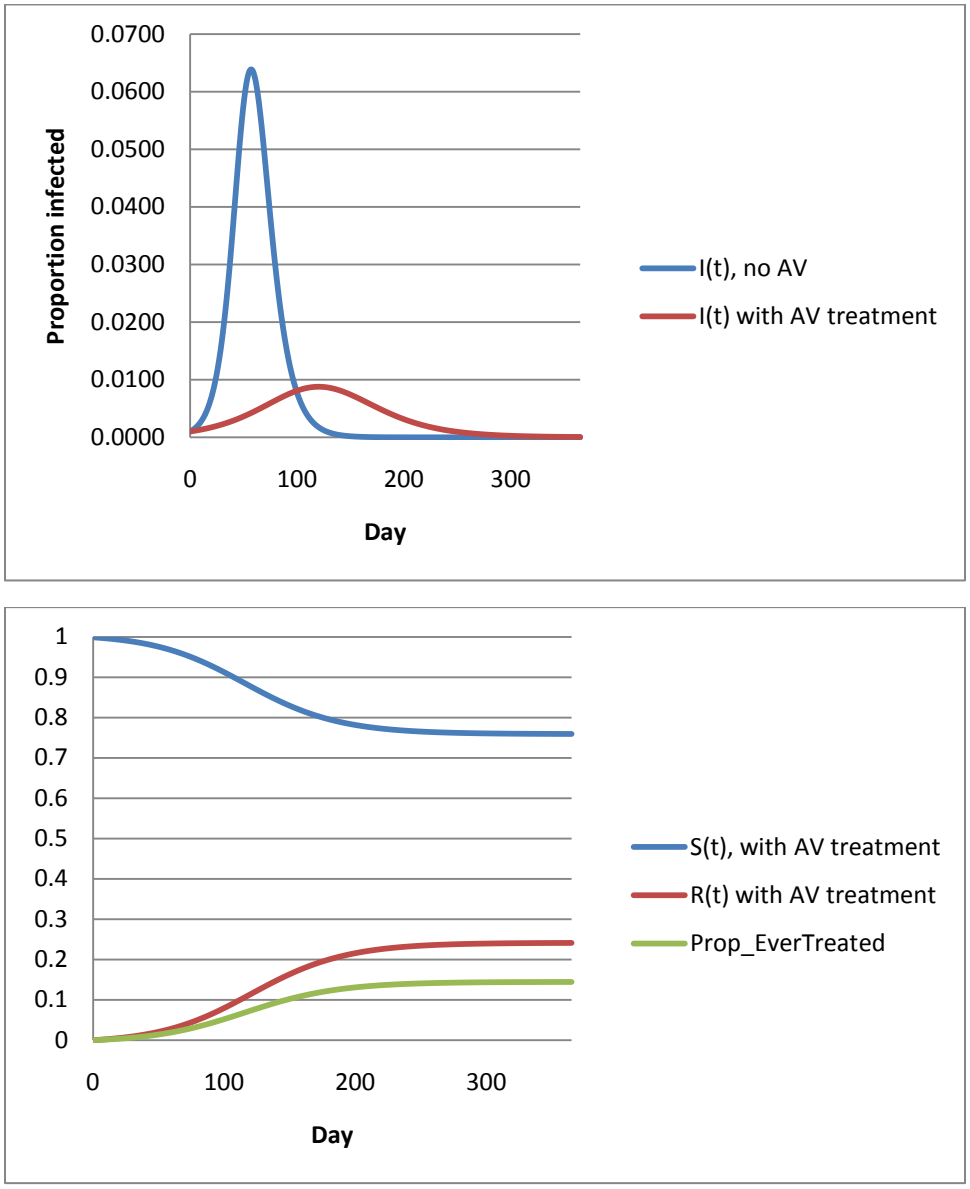
where e is the efficacy of the antivirus treatment. Suppose for example that:

- $\beta^o = 0.3$, $\delta = 0.2$, $N = 1$, $I_0 = 0.001$, as before.
- $P^* = 0.2$, i.e. the stockpile is sufficient to treat 20% of the entire population
- $p = 0.6$, i.e. 60% of newly infectious cases can be treated
- $e = 0.4$, i.e. AV reduces the infectiousness of patients by 40%
- These imply that as long as the AV stockpile has not been exhausted, the effective beta $\beta = 0.228$

Figure 3.4 shows the model dynamics in this case, where the stockpile is not exhausted.

[9] Note that this simple formulation ignores slight discreteness issues that occur in the period in which the stockpile is exhausted; total doses distributed will be slightly greater than the stockpile size, but this difference will be very small as long as the number of periods is large and the number of new infectious cases per period is small relative to P.

Figure 3.4: One city model with unexhausted AV stockpile

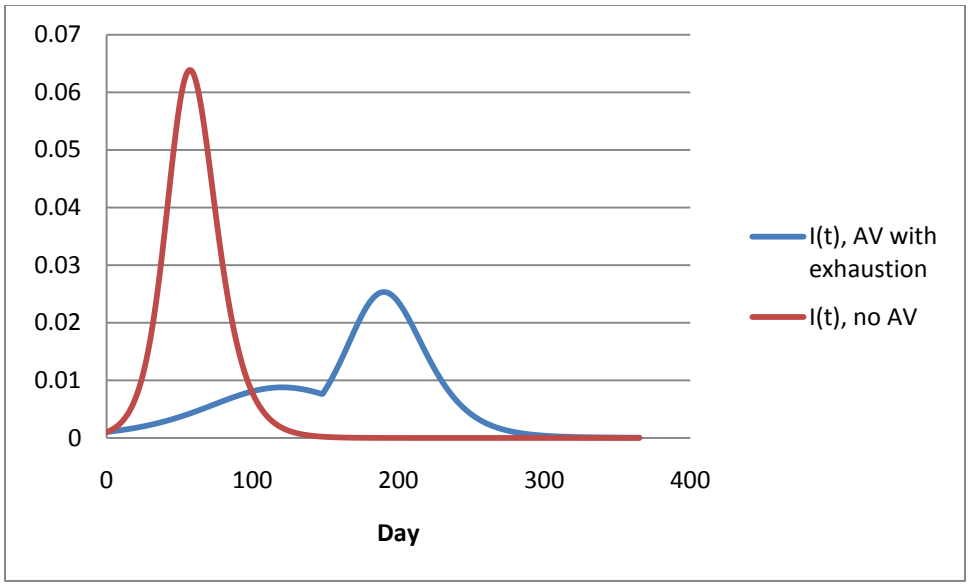


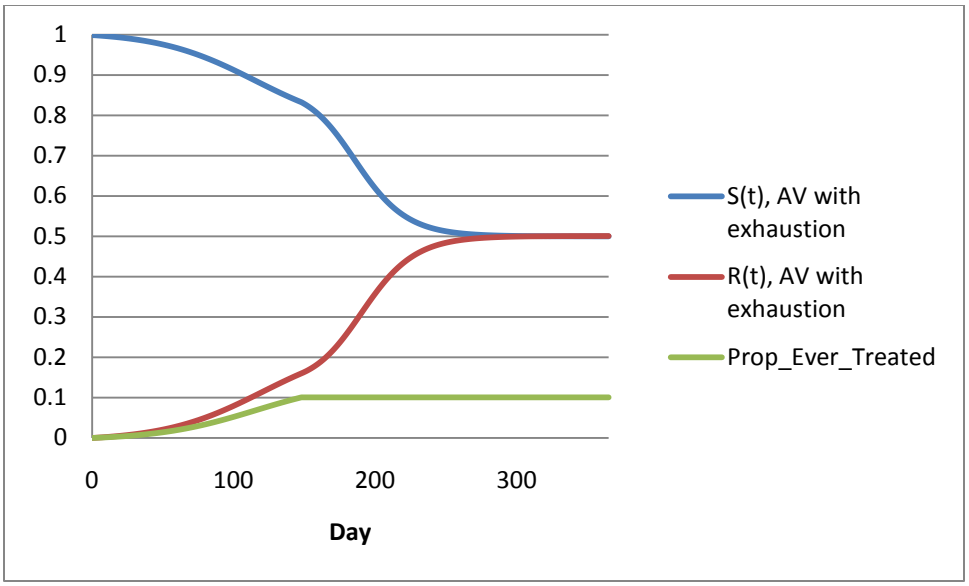
The attack rate here is much lower than in the case with no AV stockpile (24.1%, down from 58.4%), and the duration of the pandemic is longer (there are still new cases occurring near the end of the year). The proportion of people ever treated is 14.4%, which is less than the size of the 20% stockpile, and so the excess AV doses are never used. This highlights a key aspect of the choice of AV stockpile size; given any set of parameter inputs, there is an upper bound to the effective size of an AV stockpile - a city

cannot usefully deploy a stockpile larger than the number of people it can treat. But this number of treatable cases can be observed only by running the model with AV treatment included. Parameter inputs cannot be easily mapped directly into the implied maximum effective stockpile size.

Suppose now the same assumptions as used in Fig 3.4, but reduce the AV stockpile size from P* = 0.2 down to P* = 0.1. Figure 3.5 demonstrates the resulting dynamics.

Figure 3.5: One city model with AV stockpile exhaustion





The model follows an identical path to that of Figure 3.4 until day 151, when the AV stockpile is exhausted. On day 151, β instantly jumps from 0.228 up to the original value of 0.3, and so dI(t)/dt rises proportionally. The attack rate is 50.1%, as compared to 52.9% with no AV treatment, and 24.1% with an unexhausted stockpile.

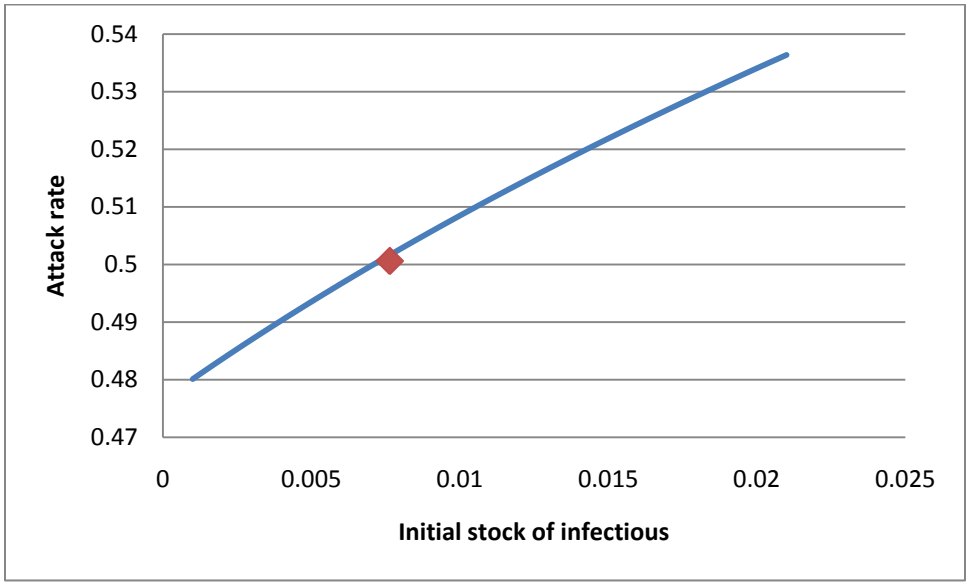
To understand the behavior, imagine splitting this epidemic with AV treatment into two components. The first component lasts from day 0 to 150, and is equivalent to a no-AV model with a β of 0.228. The second component starts on day 151, and behaves identically to a no-AV model with a β of 0.3 but with initial conditions equal to the those observed on day 150 of the first component; I(0) = 0.0076, S(0) = 0.8336 and R(0) = 0.1588. This means that changing the point where the stockpile is exhausted (by changing the stockpile size) in the original combined model is functionally identical to changing the initial conditions of this secondary component. The impact of changing the stockpile size can be examined either directly (by running the AV-treatment model with different stockpile sizes) or indirectly (by running the no-AV model with different initial

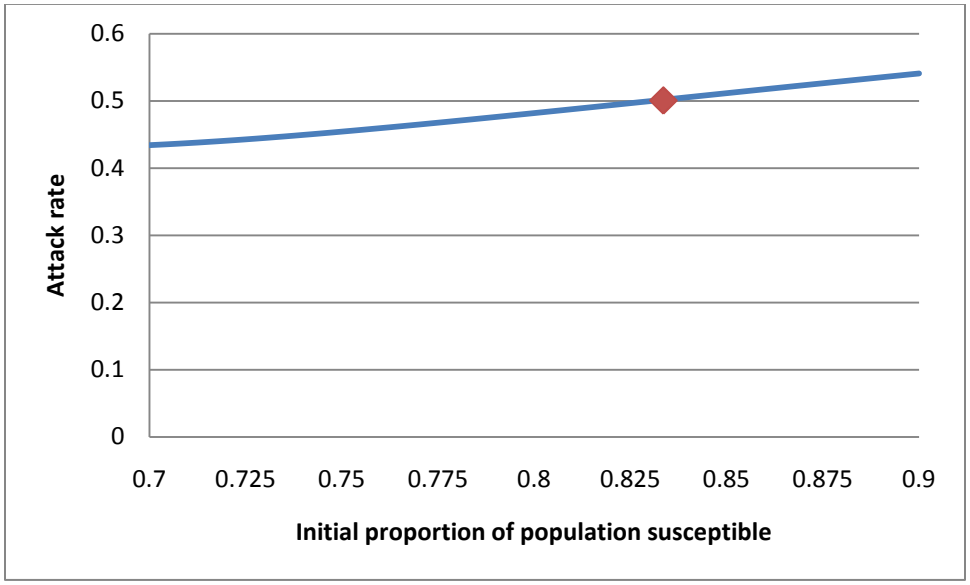
conditions). This does not allow running only the second part of the model (the full model is still needed in order to find the initial conditions for the second part), but it does explain how results can be mapped from the special case where we have a stockpile that is exhausted into the general case with no antivirus. Similarly, the case where antivirus does not expire is just another version of the general no antivirus model with a lower value of β.

Figure 3.6 demonstrates the impact of varying the initialization values I(0) or S(0) values around the day 150 values of the example above. Holding constant the pool of Susceptibles, the attack rate is increasing in the initial stock of infectious cases. Holding constant the stock of infectious cases, , the attack rate is increasing in size of the pool of Susceptibles.[10] Translating this back to AV stockpile size: increasing the stockpile size reduces the attack rate (as long as that larger stockpile is actually consumed), because it is equivalent to reducing the size of the stock of infectious cases present when the model reaches any value of S.

[10] This allows us to compare this model with antivirus to a model of vaccination, as in Boulier et al. (2007), where vaccination operates by directly removing individuals from the pool of Susceptibles.

Figure 3.6: One city model, sensitivity testing of initialization

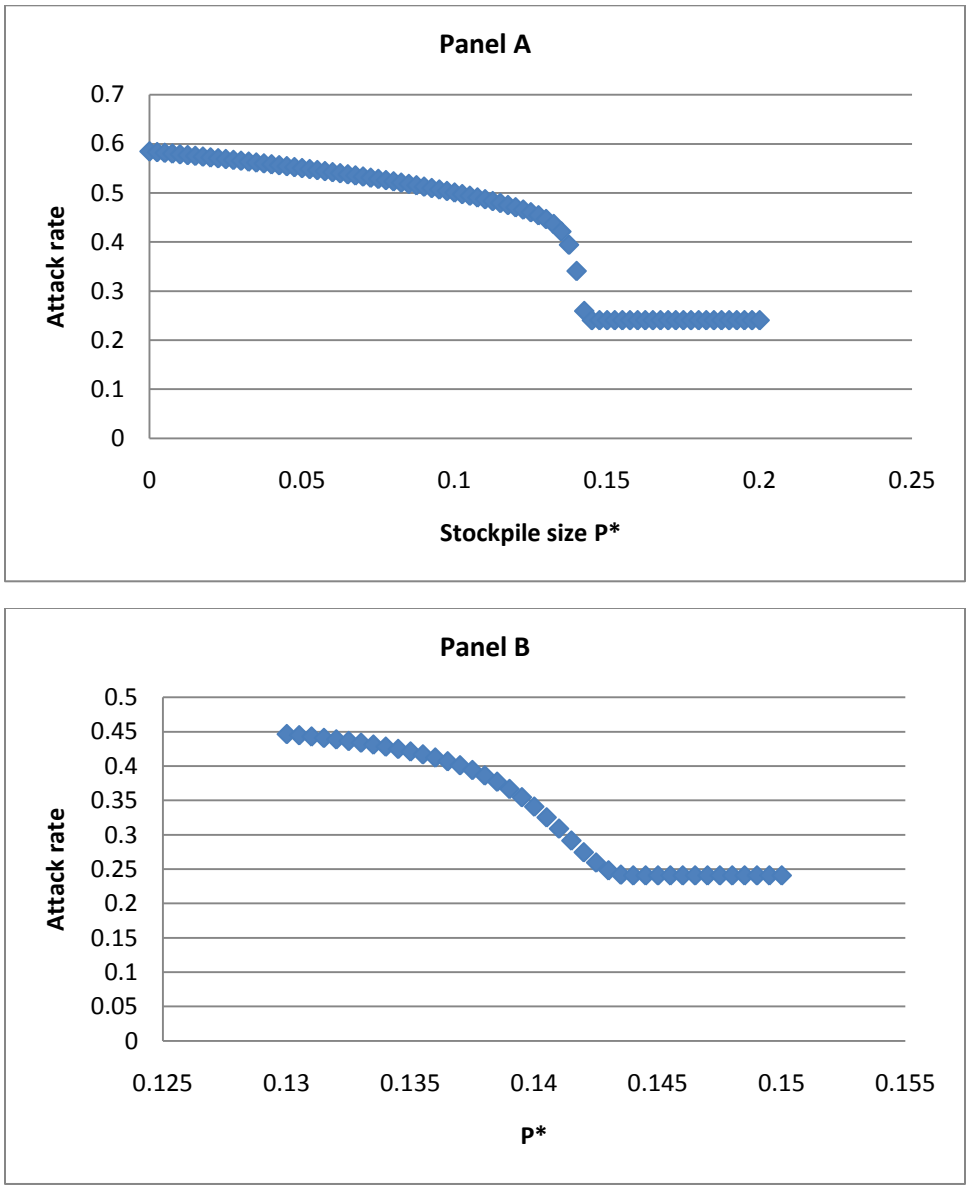




Looking at the impact of changing the stockpile directly highlights an important boundary condition: with any given parameter inputs, there is a maximum stockpile size that can be effectively consumed. Consider the previous one-city example, where β^o = 0.3, δ = 0.2, N = 1, I(0) = 0.001, p = 0.6, e = 0.4, and consider varying the stockpile size

P*. Figure 3.7 demonstrates the effect on the attack rate for various values of stockpile size P* (Panel B is a "zoomed" version of Panel A).

Figure 3.7: One city model, effect of changing stockpile size



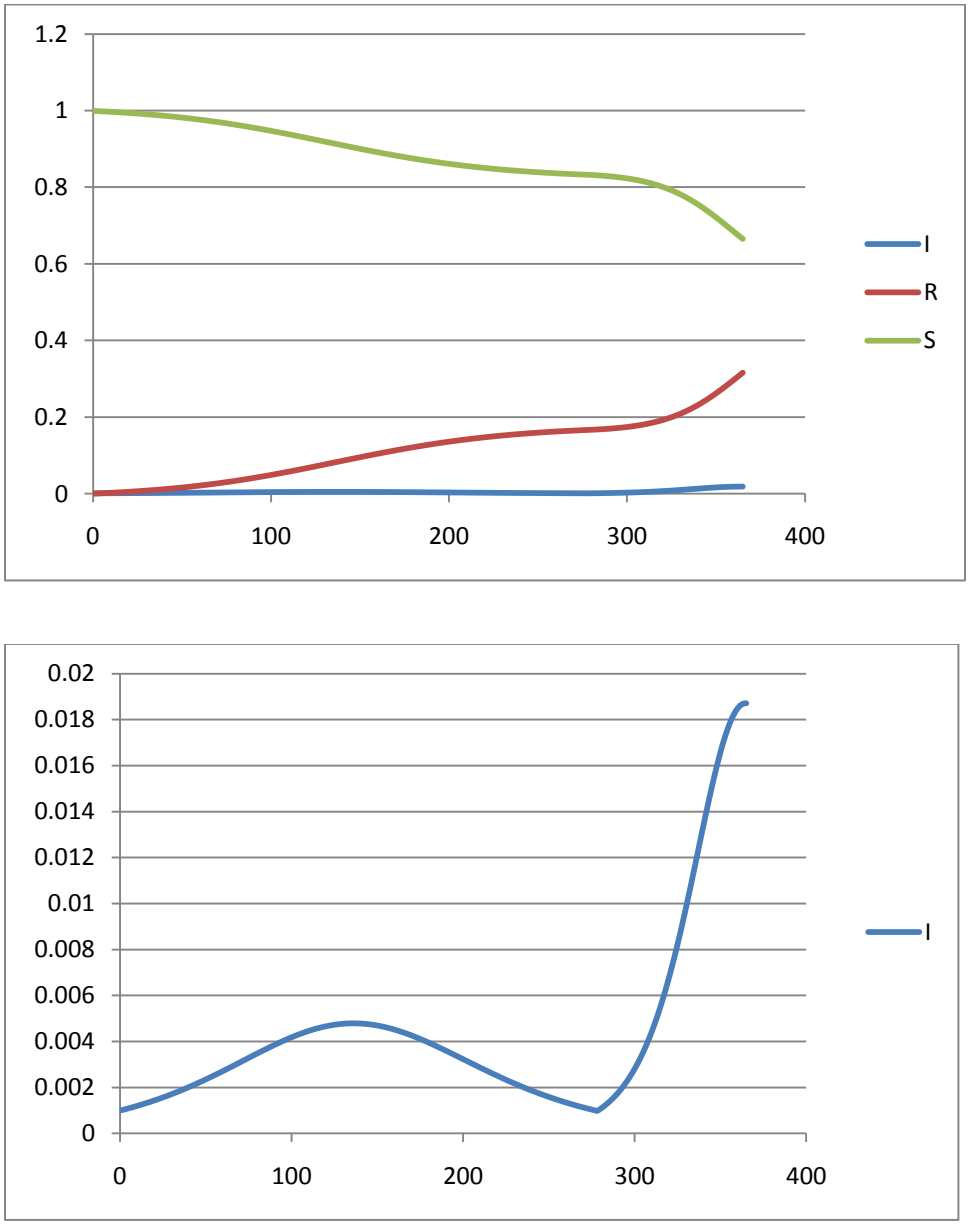
Increasing stockpile size reduces attack rate, and does so at an increasing rate, until the point is reached where the stockpile is not exhausted during the epidemic. This property is robust for any parameter inputs, and is critical in determining optimal behavior. Denote the maximum value of AV stockpile that can be consumed as $\hat{P}$. $\hat{P}$

depends on R_0 (which determine the attack rate), and on p, the proportion of people who can be treated. For example; if a particular set of parameter inputs lead to an attack rate of 0.2, then $\hat{P}$ is 0.2p. In Figure 3.7, $\hat{P} = 0.144$.

The length of an epidemic has implications for the dates at which an analyst may want to calculate the attack rate. A common approach is to examine attack rates after 1 year, but if the epidemic is slow to spread (from a low R_0 value, or a highly efficacious AV treatment policy) then it may not end within a year. Figure 3.8 shows model dynamics for P* = 0.10 but with e = 0.45 (as opposed to 0.4 in previous examples), slowing the course of the epidemic as long as there is any stockpile left (the stockpile is exhausted on day 271). The day attack rate after 365 days is 31.6%, but the final attack rate would clearly be higher if the model continued running past 1 year.

Considering only the attack rate on day 365 may be misleading. This issue becomes more important in the full GEM, where the large populations and limited travel between airports increase the pandemic duration. But it may be possible to develop, mass produce and distribute a strain-specific vaccine in roughly a year, so a one year window may be sufficient for examining the consequences of pandemic mitigation through antivirals, after which antiviral treatment would tend to be dominated by vaccine distribution in choice of control measure.

Figure 3.8: One city model, epidemic lasting more than 1 year



iv) One city model: optimal stockpile size

What size of AV stockpile should a city choose to purchase before an epidemic? The city prefers to have a lower attack rate - each person who becomes infectious suffers a welfare loss from disease symptoms (or death, if case fatality rate is positive) - but the city prefers to have a lower AV stockpile, since purchasing a stockpile is costly. So the

city solves a cost minimization problem, trying to minimize the sum of morbidity costs and stockpile/treatment costs:

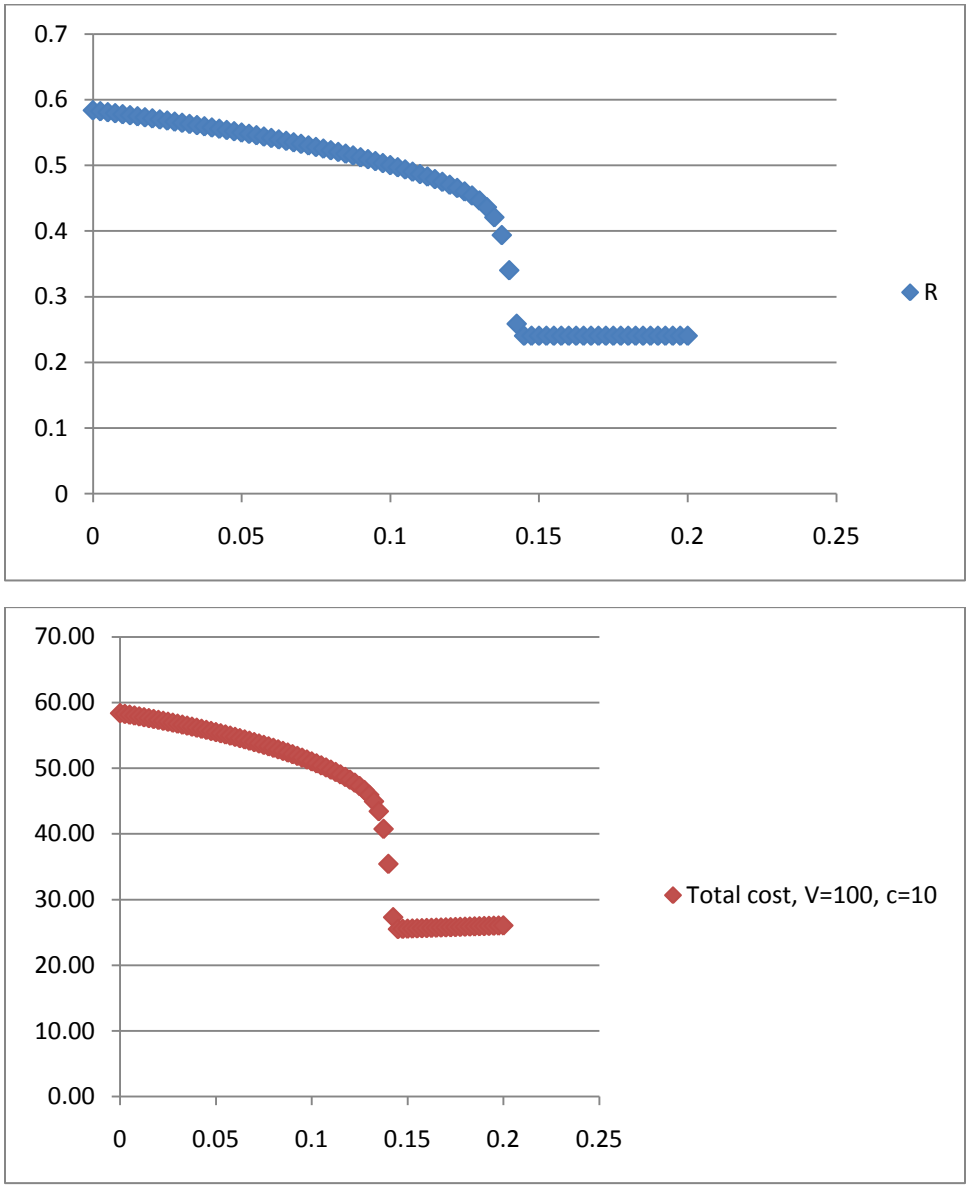
$$Min_{P*}: K\big(R(\infty, P^*)\big) + C(P^*) \qquad (12)$$

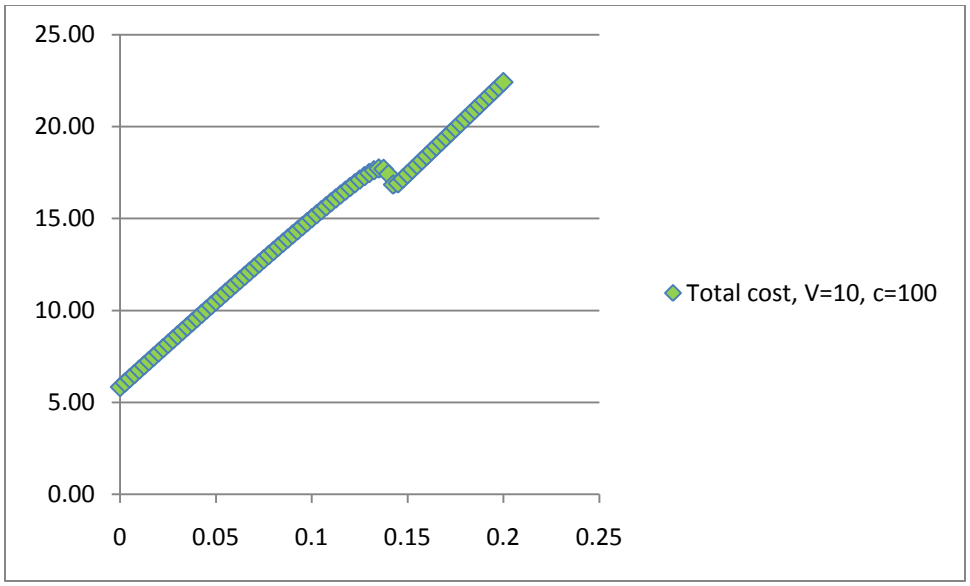
where R(∞,P*) is the total number of cases of the disease determined by the epidemic model[11], where K(.) is the disutility of morbidity (and is increasing in R(.)) and where C(.) is the cost of purchasing and disseminating the AV stockpile (and is increasing in P*). For simplicity, suppose that U(.) is linear, so U(.) = VR(∞,P*), where V is the welfare loss of an influenza case. Assume C(.) is linear, so C(P*) = cP*.

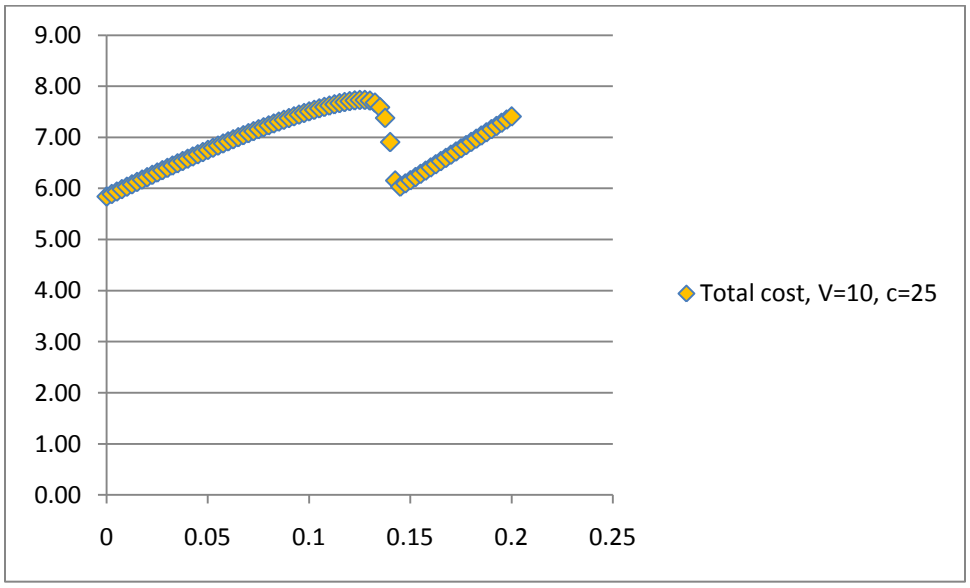
The concave shape of R(∞,P*) (see figure 3.7 and equation (5)) tends to drive the optimal policy towards the boundary cases: dR(∞,P*)/dP* is small in absolute value at low values of P*, but is large in absolute value as P* approaches $\hat{P}$. If V is high enough relative to c, then the solution is to choose P* = $\hat{P}$. If V is low enough relative to c, then the solution is to choose P* = 0. Figure 3.9 shows the payoff to the city in where β^o = 0.3, δ=0.2, e = 0.4, p = 0.6, for extreme cases [V=100,c=10] and [V=10,c=100] and an intermediate case [V=10,c=25].

[11] Since all infectious cases eventually recover and all Recovered cases were once infectious, it is simpler to track the value of R than the cumulative total of infectious cases.

Figure 3.9: One city model, Optimal stockpile size





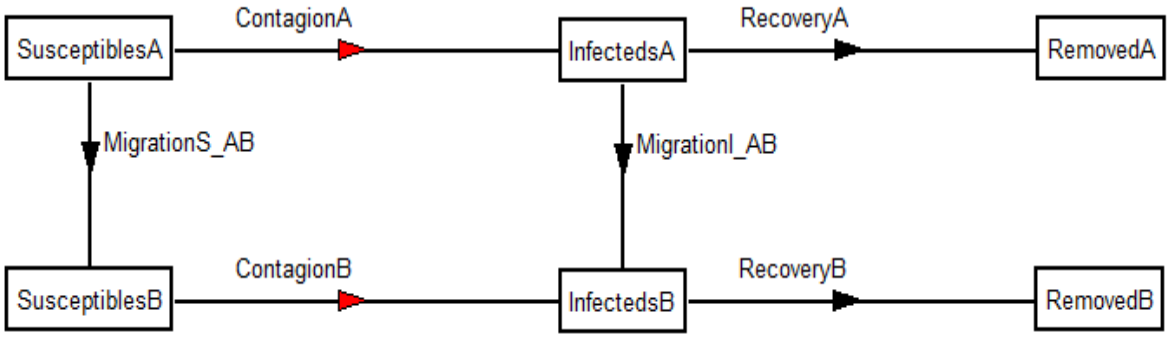


As expected, when V is high relative to c, the optimal choice is to pick P* = $\hat{P} \approx 0.145$. When c is high relative to V, the optimal choice is to pick P* = 0There is no interior solution; for any c large enough that utility is increased by choosing some value $\tilde{P}$ slightly less than $\hat{P}$, the planner is better off choosing P* = 0 instead of $\tilde{P}$. The fourth panel of Figure 3.9 also indicates a potential complication: small changes in parameter values can lead to a large change in the optimal value of P*. This means that best response functions can exhibit large discontinuities.

v) Two-city model

A one city model is useful for studying the course of an epidemic in a closed system, but to observe the interaction of choices across cities or countries a more complicated model is needed. Suppose that the simple model is expanded to encompass two cities A and B, linked by a travel connection, each with their own AV stockpile P_i*. Both cities are identical, except that the epidemic starts in city A and then spreads to B through this travel connection. Each period, a proportion α of the Infected cases in each city travel to the other city, and are replaced by an equal number of Susceptibles moving in the opposite direction.

Figure 3.10: Two city model



Migration $\Omega_A(t)$ is defined in terms of movement from A to B, but a negative value indicates a movement of people from B to A. Suppressing time subscripts (and using superscripts to indicate t+1), in city i :

$$S_i(t+1) = S_i(t) - \beta_i S_i(t) I_i(t) - \Omega_i(t) \qquad (13)$$

$$I_i(t+1) = I_i(t) + \beta_i S_i(t) I_i(t) + \Omega_i(t) - \delta I_i(t) \qquad (14)$$

$$R_i(t+1) = R_i(t) + \delta I_i(t) \quad (15)$$

$$\beta_i = (1-e)\beta_i^o \text{ if } P_i < P_i^* \quad (16)$$

$$\beta_i = \beta_i^o \text{if } P_i \geq P_i^*$$

$$P_i(t+1) = P_i(t) + p_i\beta_i I_i(t)S_i(t) \text{ if } P_i < P_i^* \quad (17)$$

$$P_i(t+1) = 0 \text{ if } P_i \geq P_i^*$$

where *i* and *j* subscripts denote cities A and B, where the daily travel rate $\Omega_i(t) = \alpha\left(I_j(t) - I_i(t)\right)$, where P_i^* is the AV stockpile in city *i*, where P_i is the number of AV doses remaining in city *i*, and where p_i is the proportion of people who become infectious in city i who receive city i.[12] Note that β_i is now time-varying, whereas β_i^* is a fixed parameter for each city.

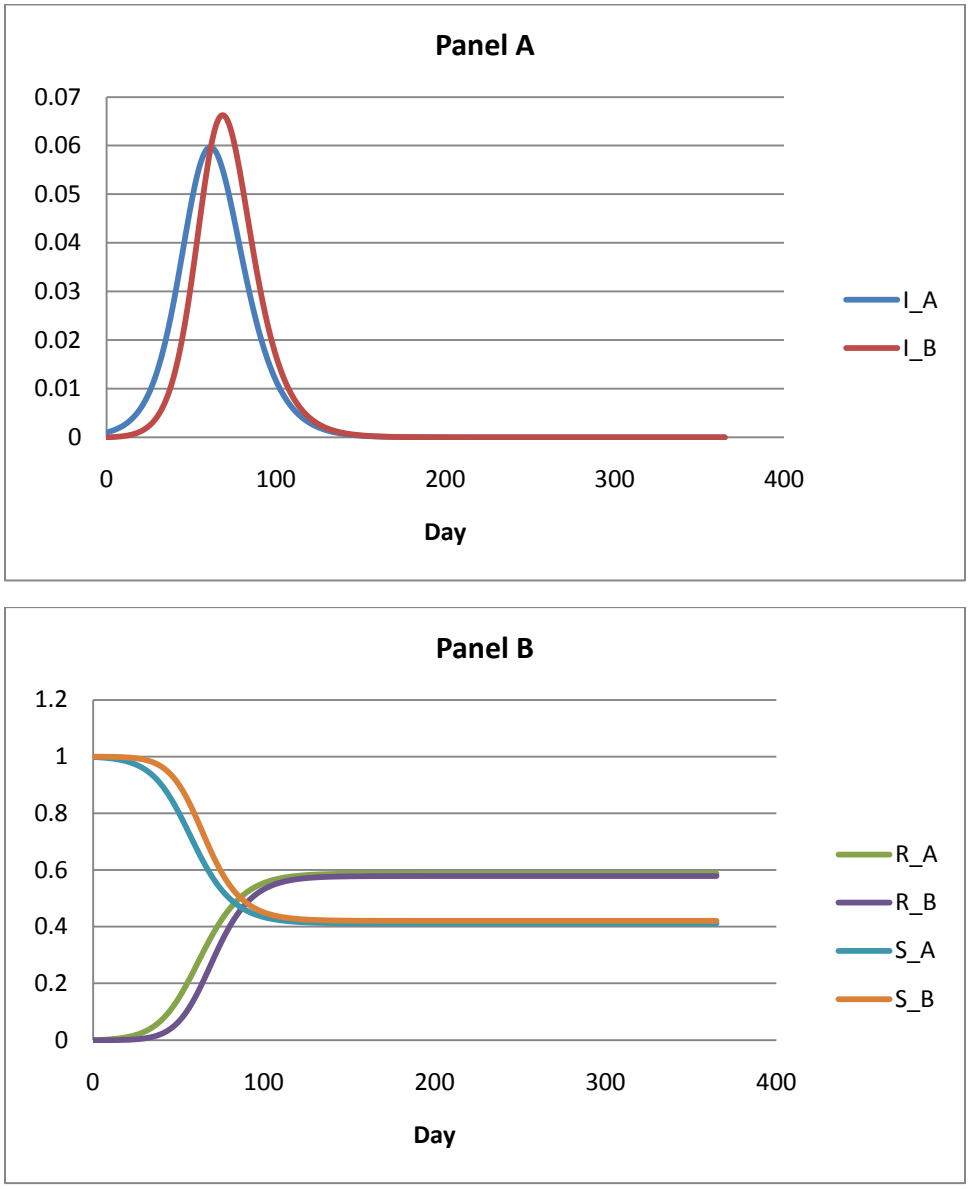
This model lets us examine the effect of changing the AV stockpile size in city *i* on the attack rate in city *j*. Assume for this example that:

- $\beta_A = \beta_B = 0.3$, $p = 0.6$, $e = 0.4$, $\delta = 0.2$ as before
- $\alpha = 0.01$ (i.e., 1% of the number of infectious cases travel between cities each day)

First, suppose that neither city has any AV stockpile. Figure 3.11 demonstrates model dynamics.

[12] Note that the change in the number of Susceptibles each day is a function of the difference in the number of Infected cases. This is necessary in order to ensure that an equal number of people move in each direction, to ensure that the population sizes stay constant over time. The model requires that we do not allow movement of Recovered individuals between cities (or that we interpret the Recovered category to be those who recovered while in city *i*), as doing so would cause the proportion of Recovered in each city to converge in the limit as $t \to \infty$ to the same vale in both cities.

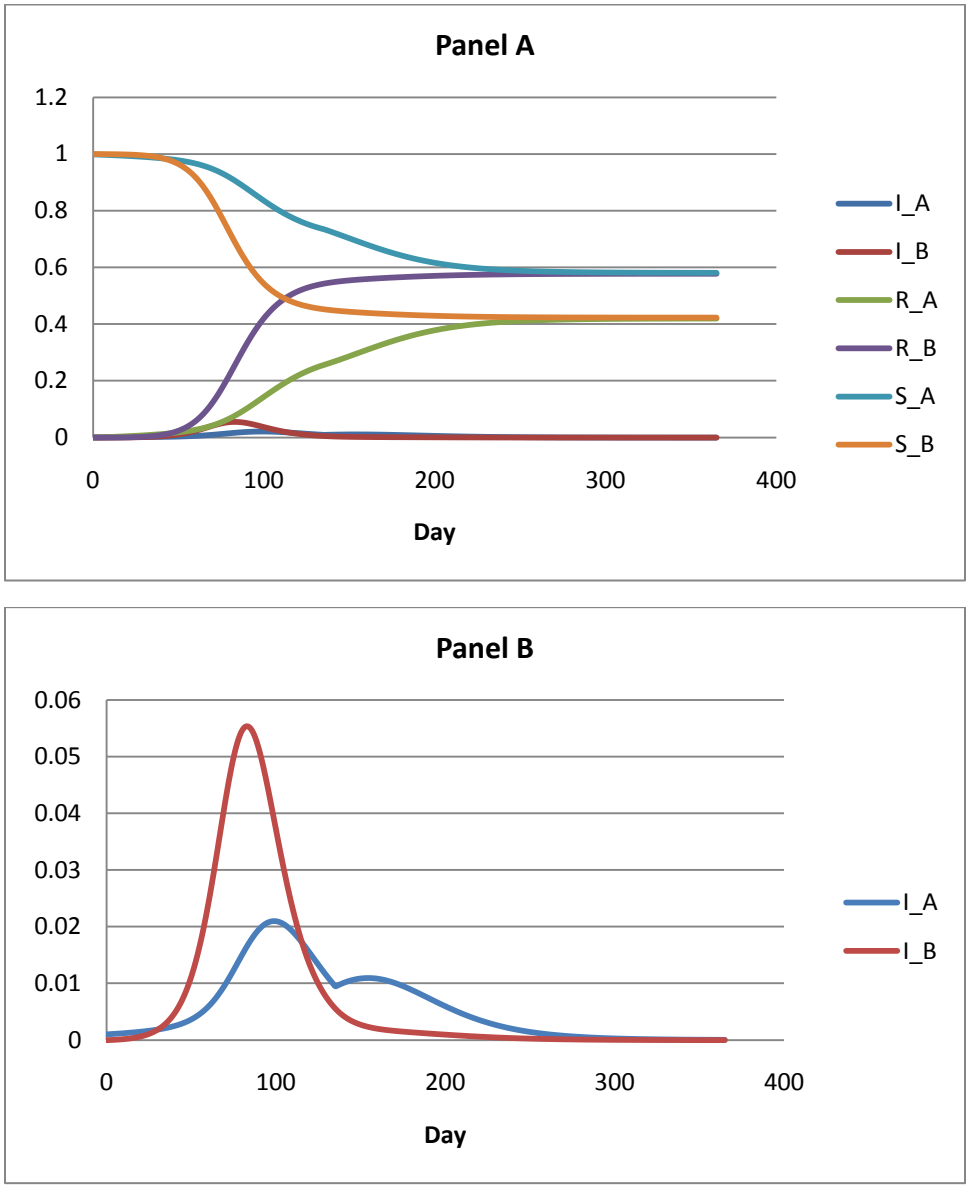
Figure 3.11: Two city model dynamics

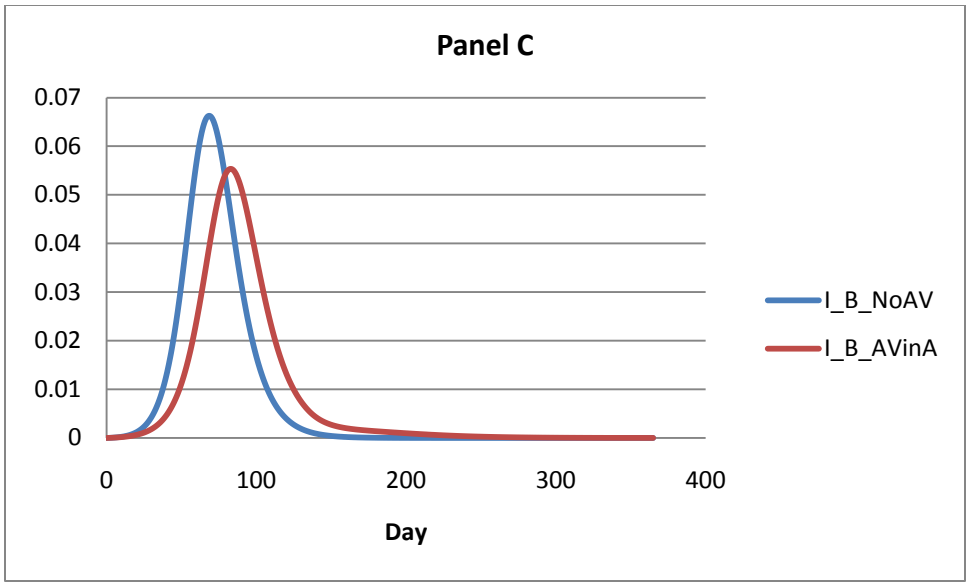


The epidemic starts in A, and then spreads to B after a lag. As long as $I_A > I_B$, City A is exporting infectious cases to B; when $I_A < I_B$ city A is importing infectious cases from B. The epidemic peak is higher in B (and the ascent is steeper) because as the epidemic progresses in City B, B is also importing additional infectious cases from A. However, both cities still end up with similar attack rates: 58.8% in City A, and 57.9% in City B.

Now the impact of adding AV treatment can be considered. Migration between the cities means that the impact of AV is not the same as in the one-city case, there are externalities and complementarities across cities. For example, suppose that only city A has a positive AV stockpile, while city B has no stockpile ($P_A^* > 0.150$, $P_B = 0$).

Figure 3.12: Two city model, City A AV stockpile, City B no AV

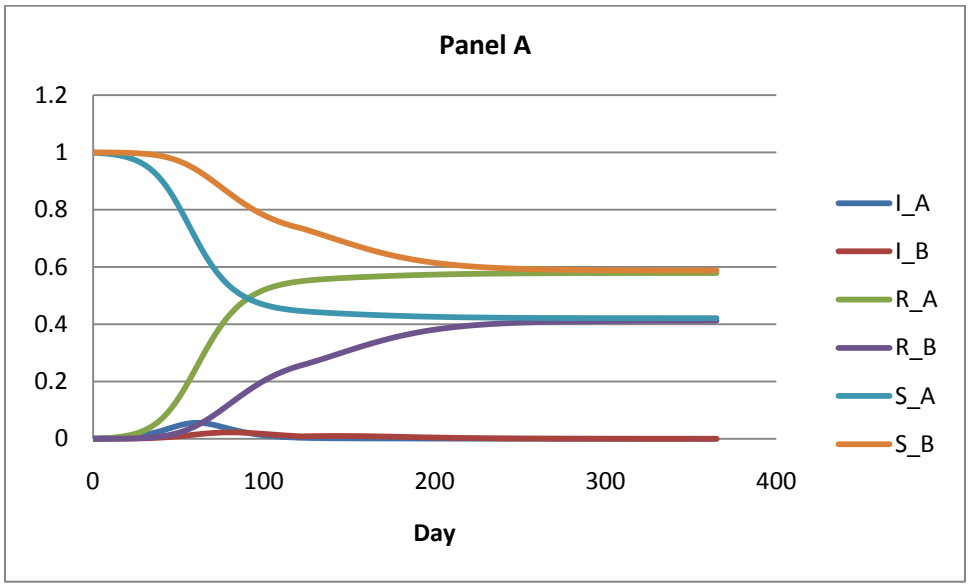


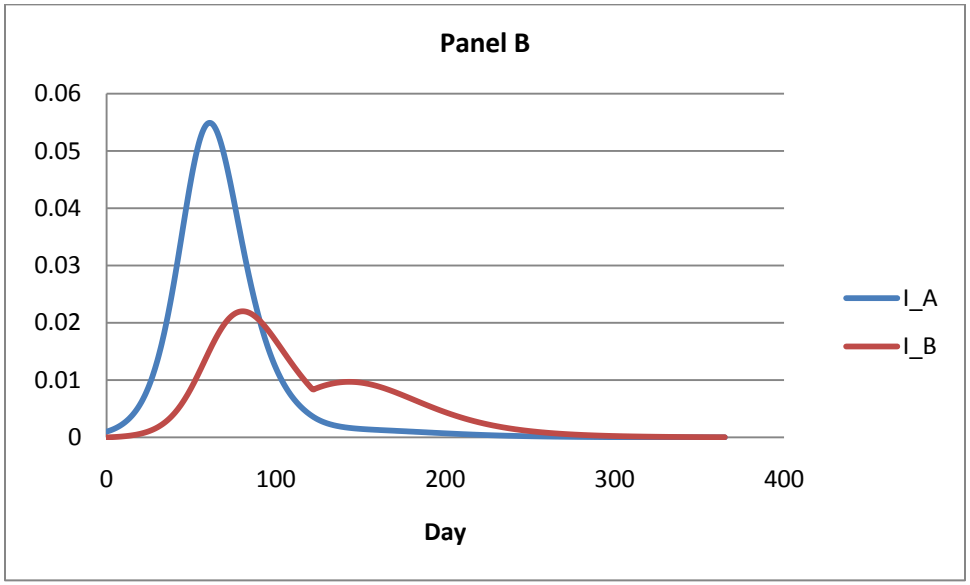


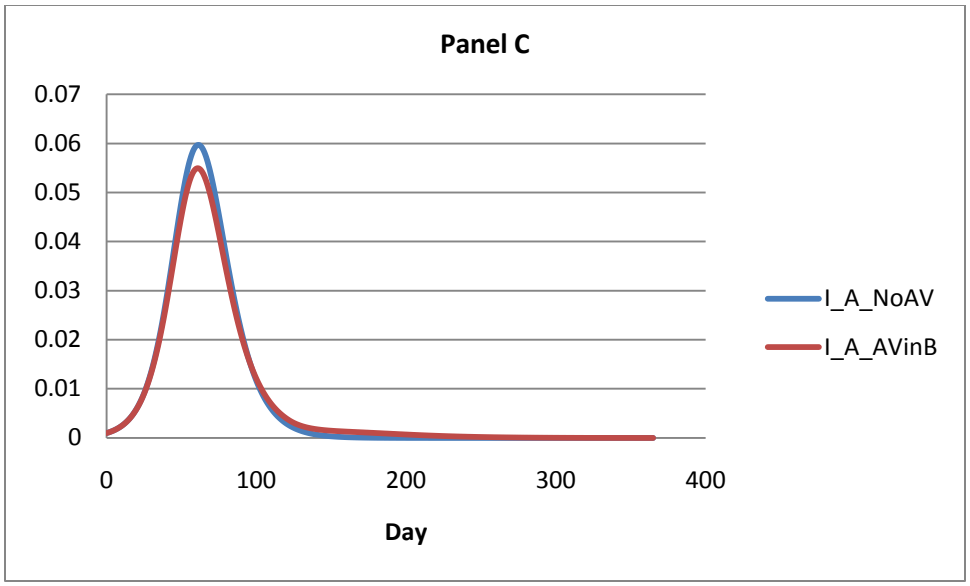
Panel A of Figure 3.12 shows the time path of the epidemic variables, Panel B shows the path of just $I_A(t)$ and $I_B(t)$, and Panel C compares the path of $I_B(t)$ in the case where neither city has any AV to the case where City A has a 15% stockpile. The epidemic starts in A and increases more gradually than in the no AV case, because of AV treatment. It spreads to B, where it spreads but still at a slower pace than in the no AV case, because there are fewer cases in A and so B is not importing as many infectious cases. When A's stockpile is exhausted dI(t)/dt instantly rises in city A, but the impact on B is small, because the epidemic in B in this case is largely over. The impact of A's stockpile exhaustion on B would be more significant if it occurred earlier relative to B's epidemic, either because A had a smaller stockpile or because the growth rate in B was lower. The attack rate in A is 41.9%, as opposed to 58.8% with no AV. There is a small positive externality to B from A's stockpile; the attack rate in B falls from 57.9% (when neither city has any AV) to 57.7% (when city A has a stockpile of 0.15). There is also a delay effect; the pandemic peak in B is delayed by 14 days.

Next, suppose that AV exists only in city B - suppose $P_A^* = 0$, $P_B^* = 0.15$. Panel A of Figure 3.13 shows the time path of the epidemic variables, the second shows the path of just I_A and I_B, and the third compares the path of I_A in the case where neither city has any AV to this case where City B has a 15% stockpile. The stockpile in B has a major impact in B (reducing attack rate from 57.9% to 41.2%), and a small positive externality on city A (attack rate falls from 58.8% to 57.9%), because A is importing fewer infectious cases from B.

Figure 3.13: Two city model, City B AV stockpile



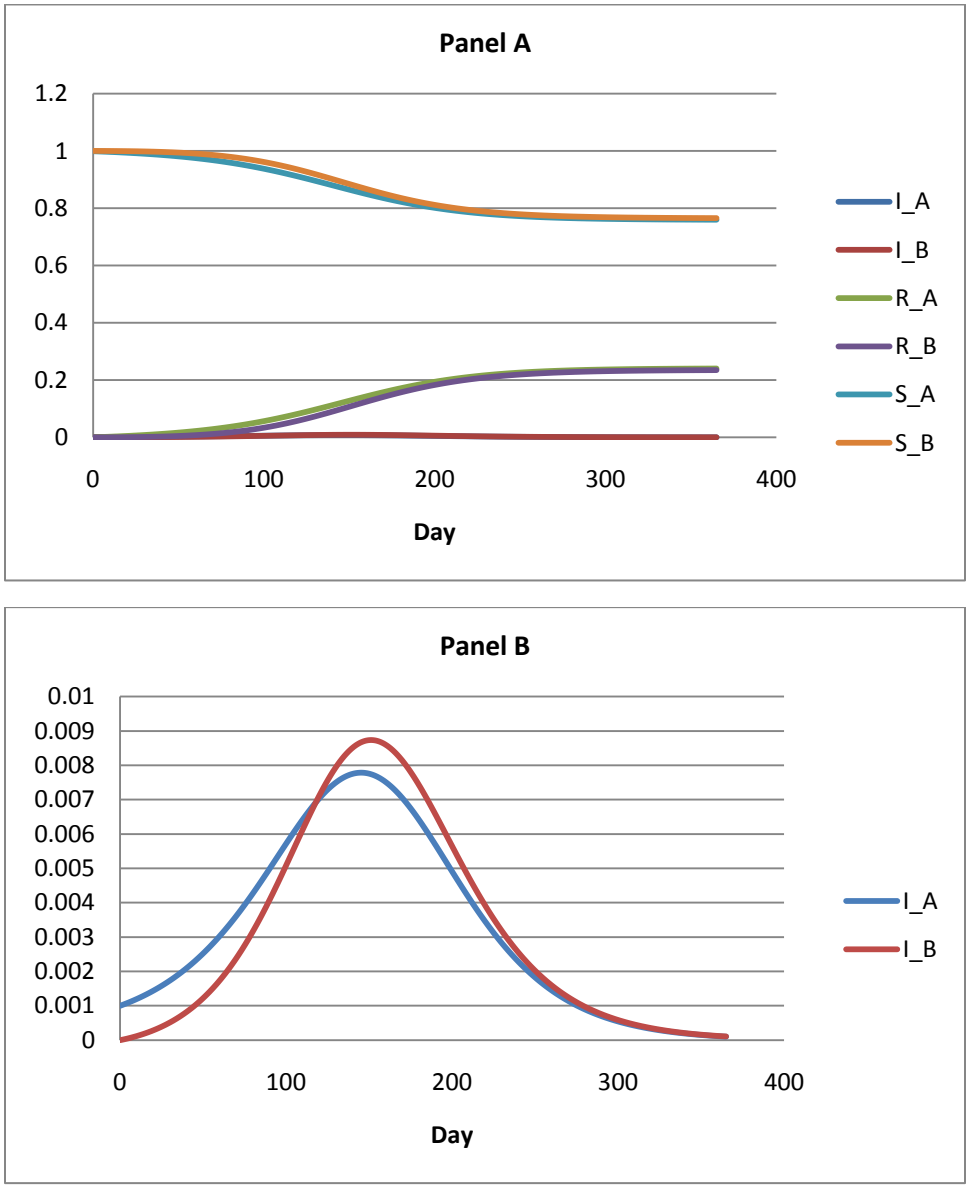


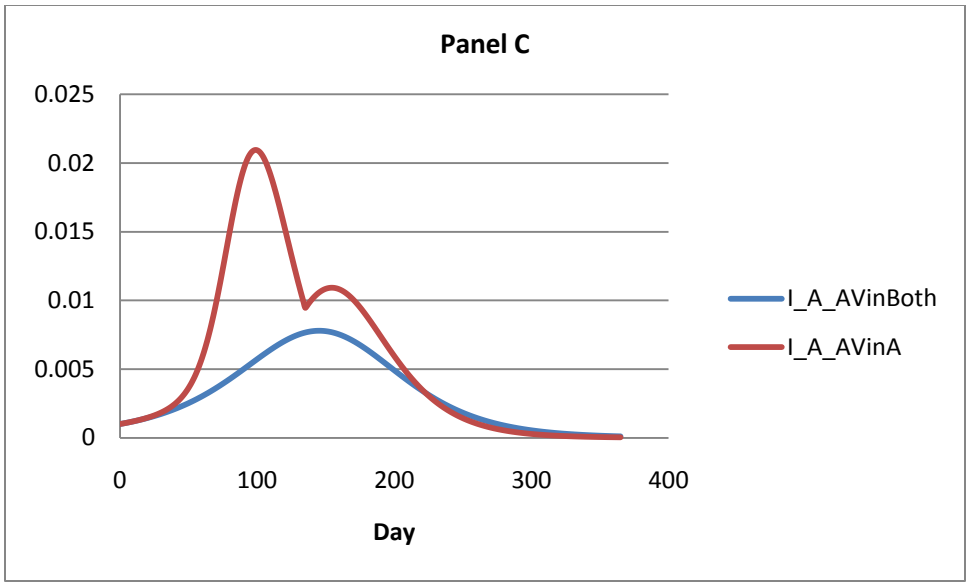


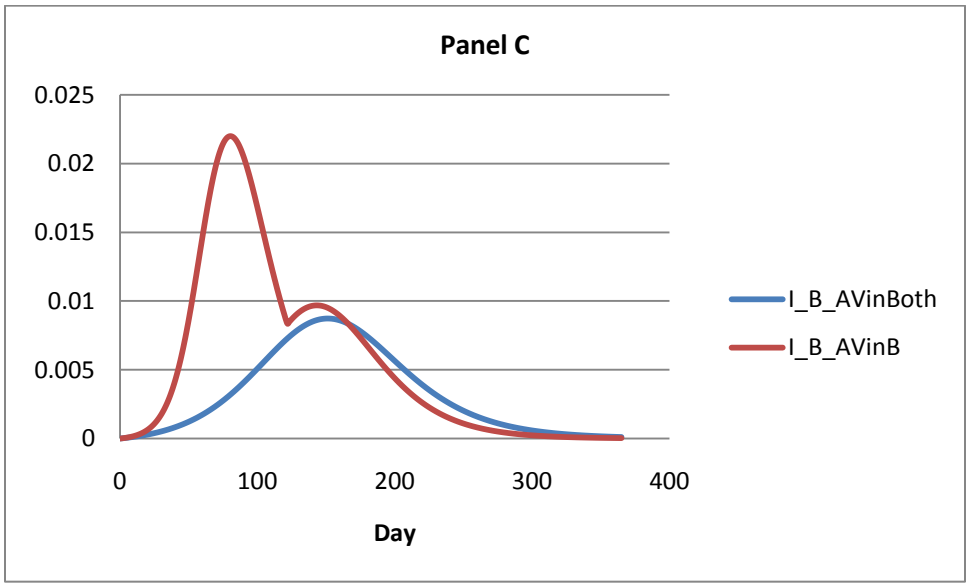
Now, suppose that both cities have some AV stockpile. Suppose $P_A^* = P_B^* = 0.15$. Figure 3.14 shows the model when both A and B have a stockpile of P*=0.15. Panels A and B show the epidemic dynamics in this scenario. Panels C and D show how I_A and I_B differ between the case where both cities have an AV stockpile and the cases in which only A or B has a stockpile. What is crucial to note is how large this difference is: when city *i* has a sizeable AV stockpile, the attack rate in *i* falls dramatically when city *j* also has a stockpile. This highlights a key property of the model: AV stockpile sizes are strategic complements (over at least some range[13]). City *i*'s marginal gain from choosing a higher value of P_i^* is increasing in P_j^* for at least some values of P_i^*, P_j^*.

[13] AV stockpile sizes will also be strategic substitutes for at least some range, because $\hat{P}_i$ is decreasing in P_j^*.

Figure 3.14: Two city model, AV in both cities







vi) Two city model, optimal stockpile sizes

Consider how cities choose an optimal stockpile size in this model. Each city has preferences over the number of cases caused by the epidemic, and over the size of their stockpile (a larger stockpile is more expensive). But the attack rate in each city is affected both by their own stockpile and by the stockpile size of the other city. So, City *i* solves:

$$Min_{P_i^*}: K_i\left(R_i\left(\infty, P_i^*, P_j^*\right)\right) + C_i(P_i^*) \quad (18)$$

where $R_i(\infty,.)$ is the terminal number of Recovered cases determined by the two-city epidemic model. Suppose as before that $K_i(.) = V_iR_i(\infty,P_i,P_j)$, where V_i is the morbidity cost of an influenza case in city *i*. If, for example, city *i* is poor while city *j* is rich, then the cities may have very different values of V. Assume $C_i(.)$ is linear and identical for both cities, so $C_i(P_i^*) = cP_i^*$.

The function R(.) is not well-behaved, so it is easier to analyze its behavior through simulation. Selecting parameter values of the model lets us run the simulation, and lets us trace out best response functions for each city, to find a Nash equilibrium. City *i*'s best response function $BR_i(P_j^*)$ is defined by:

$$Argmin_{P_i^*}: K_i\left(R_i\left(\infty, P_i^*, P_j^*\right)\right) + C_i(P_i^*) \quad (19)$$

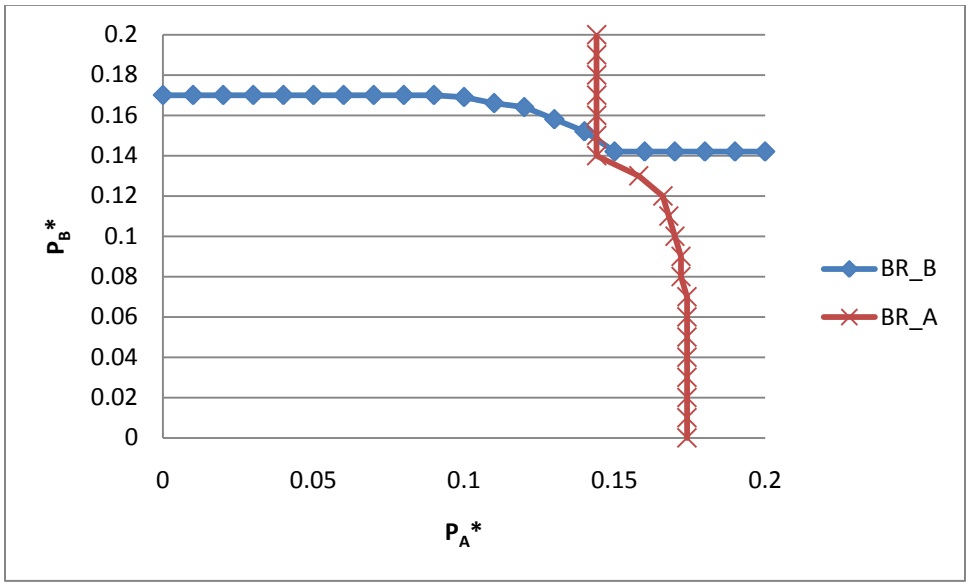
and a Nash equilibrium exists where each player chooses a value of P_i^* such that both cities are choosing a mutual best response. The shape of the best response functions and the properties of the Nash equilibrium depend crucially on the value of V_i relative to c. Several interesting cases arise. Assume $\beta_A = \beta_B = 0.3$, $\rho = 0.4$, $\delta = 0.2$, $\alpha = 0.01$ as before.

Case A: Mutual stockpiles.

Suppose $V_A = V_B = 100$, c = 10. This leads to each city to choose the largest stockpile that it can use in each case, $\hat{P}_i(P_j^*)$, and noting that $\hat{P}_i(.)$ is decreasing in P_j^*. Figure 3.15 sketches the best response functions for each city. The unique Nash equilibrium is at roughly [$P_A^* = 0.144$, $P_B^* = 0.142$]. The best response functions are

similar, but are not exactly symmetric, since the game is not exactly symmetric (the epidemic starts in city A). In this case, both players choose the highest feasible levels of stockpile, and this is socially optimal.

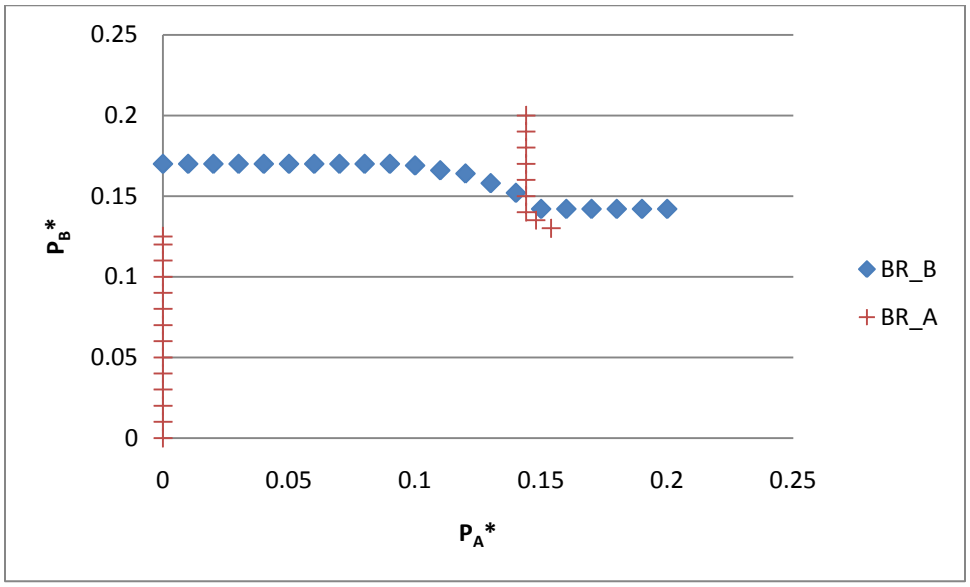
Figure 3.15: Two city model Nash equilibrium, mutual stockpiles



Case B: Discontinuous best response functions

Suppose V_A =5, V_B = 100, c = 10. Figure 3.16 shows how the best response functions can potentially be discontinuous. The best response function for city B is the same as in Case A, but now City A's best response is to choose P_A* = 0 for low values of P_B*. In this case, the Nash equilibrium is still roughly [0.144, 0.142] and this outcome remains socially efficient.

Figure 3.16: Two city model Nash equilibrium, discontinuous best response functions



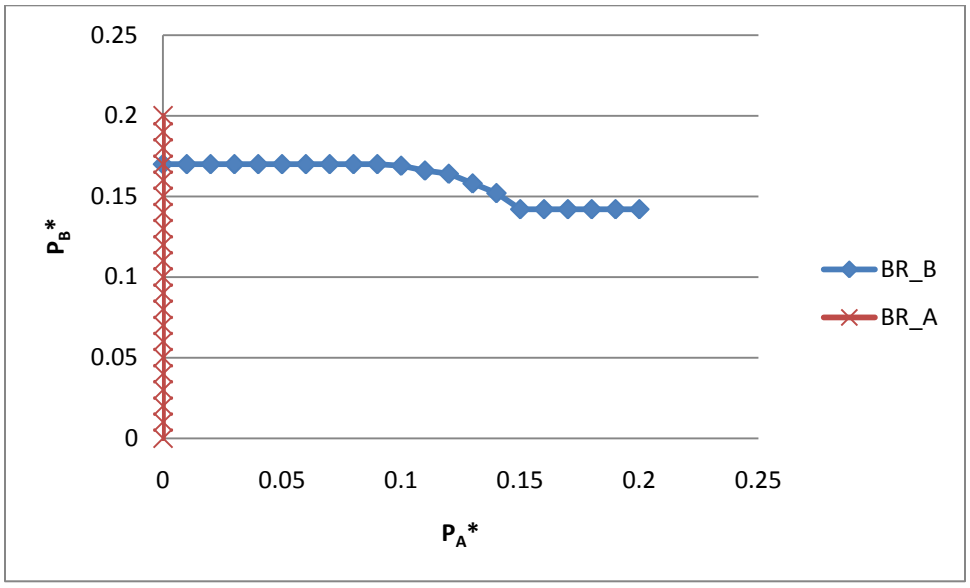
Case C: Inefficient outcome

Suppose $V_A = 3$, $V_B = 100$, $c = 10$. In figure 3.17, the Nash equilibrium occurs at roughly [0, 0.17]. However, this outcome is socially inefficient. At [0, 0.17], the total cost to City A is 1.69, the total cost to city B is 31.26, and so the total social cost is 32.97. However, at the [point 0.144,0.142] the total social cost is 27.05. The Nash equilibrium is inefficient; it gives a higher social cost than the point which would be chosen by the social planner. More importantly, there is scope for a Pareto improvement: if city B pays the entire treatment cost for the 0.144 stockpile in city A and the 0.142 stockpile in city B, the total costs borne by city B are only 26.36. City B would be strictly better off if it purchased stockpile and donated it to city A, because of the positive externality to B.

Ideally, this kind of analysis would be carried out in the full GEM. However, the model is simply too large for calculation of best response functions to be feasible (see

Chapter 5). Instead the focus of analysis will be on policies that individual countries are likely to pursue.

Figure 3.17: Two city Nash equilibrium, low value in city A



Analysis of these simple SIR models highlight several issues that warrant further investigation in a more realistic model. How would a pandemic spread in a descriptively realistic model using real world data and parameter values? How should stockpiles of antivirus be deployed in such a model? How large are the externalities in AV treatment? Is there a pure efficiency case for rich, downstream countries to pay for antiviral doses for poor areas near the source of a pandemic? How sensitive are answers to these questions to uncertainties in parameter values?

Chapter 4 A Global Epidemiological Model

While simple one- and two-city models can demonstrate the core qualitative properties of categorical epidemiological models, equilibrium outcomes in these models are highly dependent on particular parameter values. This makes it difficult to draw policy conclusions from such simple models. In order to address policy questions there is a need for a more complicated model that is sufficiently realistic to use real-world data for parameter value inputs.

This dissertation uses a version of the Global Epidemiological Model (GEM) developed by Goedecke and Bobashev and fully described in Hajdin et al. (2009), expanded and modified for the purposes of this research. The GEM is a discrete time stochastic SEIR model designed to simulate the spread of a pandemic throughout the entire world. The framework of the model was based on the work of Rvachev and Longini (1985) and on the epidemic model of Baroyan, Mironov, and Rvachev (1981). The GEM extends Rvachev and Longini by adding stochastic components, disease interventions, and a more detailed population structure. This dissertation further extends the network structure and parameterizes the model, but does not use the full capability of the GEM in modeling disease state transitions or treatment options. This chapter describes the mechanics of the model, and demonstrates the model dynamics in a baseline setting with no policy interventions.

i) Network structure

The GEM tracks the spread of a pandemic through a network of cities and rural areas that span the globe. The world is divided into 106 regions – each of which is either

an individual country or a group of countries[14] (see Appendix 2). Every region contains at least one city, and nearly all contain a rural area. There are 288 cities and 101 rural areas, for a total of 389 network points. The regions were selected using definitions from the Global Trade Analysis Project (GTAP). A rural area exists for every region except small islands and city states. Cities were selected by choosing the union of: one city for every region; the largest 130 cities in the world by population; the 130 cities with the highest average airport flows per day; sufficient cities to ensure that every city is connected to a city in another region, and that the network is fully connected.[15] Rural areas are defined as everything that is not explicitly urban, so the "rural" areas contain the population of all cities in the world that are not explicitly included in the model. Rural areas are treated the same as cities in the model, except that they are connected only to the cities within the same region, and not to cities in other regions. Total world population is roughly 6.4 billion, of which roughly 84% is allocated to a rural area.

Cities and rural areas are connected by a daily flow between every city and its associated rural area, and by airline flows between cities. It is assumed that 1% of the population of each urban area travels to its associated rural area each day, and an equal number travel from the rural area to the urban area[16]. This assumption is somewhat

[14] Roughly 86% of the world's population live in a country that is its own region, the rest live in aggregate regions like "Caribbean", "Rest of Eastern Africa", "Rest of Western Asia".

[15] This method means that we end up with many cities in rich countries, and relatively fewer cities in poorer countries. Running the model with many rich country cities excluded had no significant impact on model results.

[16] Attempts to estimate real world land travel rates using traffic count data encountered several problems. First, while data was available in many rich countries, data in developing countries was sparse. Second, even the high quality developing data could be extremely misleading, as it was difficult to distinguish between commuters and travelers, leading to daily traffic flow estimates that could be as high as 15% of the value of the urban population. Third, it was not possible to find reliable data for non-car based travel numbers.

arbitrary, but model results are not very sensitive to changes in travel rates[17]. Cities are connected though 7,668 city-pair connections with a total global daily air movement of 5,035,664 people per day. Travelers are randomly selected from the available pool of Susceptible, Exposed, asymptomatic-Infectious and Recovered cases; symptomatic Infectious and Dead cases do not travel (these terms are described in section ii) below). This means that the pool of travelers (on average) is exposed or asymptomatic-infectious in the same proportions as the general population (once symptomatic-infectious cases and dead cases are excluded).

Data on airline flows is drawn from Official Airline Guide (OAG) statistics on flight schedules provided by Guimerà, Mossa, Turtschi, and Amaral (2005). Air travel numbers are based on the daily seat capacity between each pair of cities for all cities in the model. The raw data only tracks direct flights, but a proportion of travelers is assumed to take 2-leg indirect flights, following Epstein et al. (2007)[18].

Data on regional populations is drawn from the World Bank development data portal. Data on urban metropolitan area populations is taken from several sources, primarily the United Nations World Urbanization Prospects (2007), and uses broad

[17] Running the model with doubled and halved travel rates had no significant impact on attack rates by the end of the pandemic.

[18] A dataset of 10% of domestic US airline ticket coupons for 2004 is used to estimate the proportion of travelers on any city pair who travel through a connecting airport, rather than directly. These estimates are then applied to the cities in the GEM, taking the OAG data for international travel and applying estimates that split the daily travel numbers on each city pair into 1-leg travelers and 2-leg travelers. For example for flights from Jakarta to Singapore and Singapore to New Delhi, a proportion of Jakarta to Singapore travelers are assumed to actually be 2-leg travelers moving between Jakarta and New Delhi (and similarly a proportion are assumed to be traveling from Jakarta to every other city connected to Singapore). On average, 68.2% of tickets are assumed to be 1-leg, and 31.8% are assumed to be 2-leg (though this varies across each city pair in the model).

definitions of urban areas to try to capture the greater metropolitan population[19]. Rural area populations are calculated by deducting urban populations from regional populations.

The GEM does not track individuals – it tracks only the size of each SEIR category in each city in each period**.** This means round-trip travelers cannot be distinguished from one-way travelers. In the real world, if a person flies from City A to City B or drives from city A to Rural Area C, then it is likely that they will return. But in the GEM, once that person travels to B or C, they enter the uniform mixing pool of their destination, and are no more likely to travel elsewhere than any other member of their new city.

ii) Disease spread mechanics

In each city or rural area, the population is divided into Susceptible, Exposed (infected but not yet Infectious), Infectious, Recovered and Dead disease state categories. Infectious cases may be either symptomatic or asymptomatic; asymptomatic cases are less infectious, will not be treated, and will still travel between cities. At time zero, the entire population is Susceptible, except for a small number of Exposed cases in a single outbreak city. A Suspectible individual who is contacted by an Infectious individual may be infected and become Exposed. Every Exposed case transitions to an Infectious case, and then every Infectious cases either recovers or dies.

Within each region the population is divided into 4 age categories; under 5 years, 5-14 years, 15-64 years and 65+ years, denoted with indexes a = 0,1,2,3, respectively. This allows incorporation of age-specific contact rates and to observe the incidence of the

[19] Some cities are urban agglomerations of multiple cities, like Washington DC-Baltimore.

pandemic by age-group. Data on population age structure is drawn from United Nations World Population Prospects for 2007. Within each city or rural area, the population is assumed to mix uniformly; every person in city i with the same characteristics (disease state, age, treatment status, infectiousness duration) is assumed to be identical.

For city i on day t define a set of variables[20] : *S(i,a,t), E(i,τ,a,t), I(i,s,θ,a,t), R(i,a,t)* and *D(i,a,t)* for those individuals who are Susceptible, Exposed, Infectious, Recovered or Dead, respectively. Define τ as the number of days a since a person became infected, so a new Exposed individual who has just become infected has τ = 0. Define $\bar{\tau}$ as the maximum number of days that a person may remain exposed before becoming Infectious, and $\bar{\theta}$ as the maximum number of days a person may remain infected before becoming Recovered or Dead. Those who are Exposed on day t were infected with the disease τ days earlier, on day t - τ. Those who are Infectious on day *t* became infectious θ days earlier, on day t – θ. Infectious individuals may be either Asymptomatic $(s=0)$ or Symptomatic $(s=1)$. From this point forward the city index *i* is suppressed for notational simplicity, but it is important to remember that each variable is city specific.

Given this notation, population of a city can be defined as the sum of this set of mutually exclusive disease states:

[20] It is convenient to describe the model in terms of city *i*, but note that this can refer to either a city or a rural area. In the notation that follows, each of these variables is particular to city *i*, but this notation is suppressed in order to improve readability.

$$N(a,t) = S(a,t) + \sum_{\tau=0}^{\bar{\tau}} E(\tau,a,t) + \sum_{\theta=0}^{\bar{\theta}} \sum_{s=0}^{1} I(s,\theta,a,t) + R(a,t) + D(a,t) \quad (20)$$

Variables can be aggregated across age groups, such that:

$$N(t) = \sum_{a=0}^{3} N(a,t) \;\; S(t) = \sum_{a=0}^{3} S(a,t) \quad (21)$$

$$E(\tau,t) = \sum_{a=0}^{3} E(\tau,a,t) \qquad I(s,\theta,t) = \sum_{a=0}^{3} I(s,\theta,a,t)$$

$$R(t) = \sum_{a=0}^{3} R(a,t) \;\; D(t) = \sum_{a=0}^{3} D(a,t)$$

A new Exposed case is the result of contact between a Susceptible person and an Infectious person. The daily infectious contact rate between age-groups j and k is defined as λ_{jk} (see below). This parameter λ_{jk} is equivalent to the parameter β in the simple SIR models in the preceding chapter.

Define $\gamma(\tau)$ as the probability that an Exposed case becomes Infectious, and $\sigma(\theta)$ as the probability that an Infectious person recovers. Assume for simplicity (following Longini 2005) that 80% of Exposed cases transition to Infectious after 1 day and 20% after 2 days; they will become symptomatic-Infectious with probability 2/3 and asymptomatic-Infectious with probability 1/3. Infectious cases either recover or die after 3-6 days since they became infectious, with 30% recovering after 3 days, 40% after 5 days, 20% after 5 days, and 10% after 6 days. This means $\bar{\tau} = 1$ and $\bar{\theta} = 5$. Thus:

$$\gamma(\tau) = \begin{cases} 0.8 \; if \; \tau = 0 \\ 1 \; if \; \tau = 1 \\ 0 \; otherwise \end{cases} \quad (22)$$

$$\sigma(\theta) = \begin{cases} 0.3 \; if \; \theta = 2 \\ (0.4)/(0.7) \; if \; \theta = 3 \\ (0.2)/(0.3) \; if \; \theta = 4 \\ 1 \; if \; \theta = 5 \\ 0 \; otherwise \end{cases} \tag{23}$$

Define *r(s)* as the relative infectiousness of symptomatic state s, and f(t) as a seasonality factor. Based on Ferguson, et al. (2005) and Longini, et al. (2005), it is assumed that r(s) = 1 if a person is symptomatic (s=1) and r(s) = 0.5 if a person is asymptomatic (s=0). Seasonality is discussed in part iv) below. Groups within a city mix uniformly, so the average number of new Exposed cases caused by one Infectious person in city *i* on day *t* is proportional to the number of Susceptible individuals in that city and is equal to λ·S(t)/N. Ignoring travel between cities for the moment, the total number of Exposed cases of age *k* in city *i* evolves by:

(24)

$$E(\tau, k, t+1) = \begin{cases} f(t) \sum_{j=0}^{3} \left[\lambda_{jk} \frac{S(k,t)}{N(j,t) - D(j,t)} \right] \sum_{\theta=0}^{\bar{\theta}} \sum_{s=0}^{1} [I(s, \tau, j, t) \cdot r(s)], for \; \tau = 0 \\ E(\tau - 1, k, t)\big(1 - \gamma(\tau - 1)\big), for \; \tau = 1, \ldots \bar{\tau} \end{cases}$$

and the number of infectious cases of age k in city *i* evolves by:

$$I(1, \theta, k, t+1) = \begin{cases} \left(\frac{2}{3}\right) \sum_{\tau=0}^{\bar{\tau}} [E(\tau, k, t)\gamma(\tau)], for \; \theta = 0 \\ I(1, \theta - 1, k, t)\big(1 - \sigma(\theta - 1)\big), for \; \theta = 1, \ldots \bar{\theta} \end{cases} \tag{25}$$

$$I(0, \theta, k, t+1) = \begin{cases} \left(\frac{1}{3}\right) \sum_{\tau=0}^{\bar{\tau}} [E(\tau, k, t)\gamma(m)], for \; \theta = 0 \\ I(0, \theta - 1, k, t)\big(1 - \sigma(\theta - 1)\big), for \; \theta = 1, \ldots \bar{\theta} \end{cases} \tag{26}$$

for symptomatic and asymptomatic cases, respectively. It is assumed 97.5% of symptomatic cases recover and 2.5% die, while 98.75% of asymptomatic cases recover and 1.25% die. [21] So the number of Recovered cases evolves by:

$$R(k,t+1) = R(k,t) + \sum_{\theta=0}^{\bar{\theta}} \sum_{s=0}^{1} \left[\left(0.975s + 0.9875(1-s)\right) I(s,\tau,k,t)\sigma(\tau) \right] \tag{27}$$

and the number of Dead cases of age k evolves by:

$$D(k,t+1) = D(k,t) + \sum_{\theta=0}^{\bar{\theta}} \sum_{s=0}^{1} (0.025s + 0.0125(1-s)) I(s,\tau,k,t)\sigma(\tau) \tag{28}$$

The preceding describes an S-E-I-R model with no travel. In the model, the evolution of S(.), E(.), I(.), R(.) and D(.) are determined by both these equations and by travel between cities. There is a fixed number of seats available X_{ij} for travelers moving between two cities *i* and *j* determined either from the airline data (for city to city travel) or 1% of the urban population (for city to rural travel). Each Susceptible, Exposed, Asymptomatic Infectious and Recovered individual has a probability of travel calculated by:

[21] This assumption is highly speculative, but gross attack rates would be similar so long as case fatality rates remained small. Slight differences would occur because deceased are not contacted by others. At a lower case fatality the proportion of proportion of Recovered people contacted by an infectious individual would be slightly higher, and the proportion of Susceptibles contacted would be slightly lower. The combined effect would be for the pandemic to spread slightly more slowly, but the effect would be minimal for low case fatality rates (for example, 0-3%). Increasing the case fatality rate further would have the opposite effect. Since the case fatality rate is highly speculative, results analysis focuses on attack rates from the total number of cases, rather than separately considering deaths vs recoveries.

$$pTravel(i,j) = \frac{X(i,j)}{N(i,t=0)} \quad (29)$$

The expected number of travelers of a particular subgroup who move from *i* to *j* on a particular day is the product of the number of persons in the subgroup multiplied by $pTravel(i,j)$.[22]

iii) Contact rates

The key parameter determining the spread of the pandemic is λ_{jk}, the (city-specific) daily infectious contact rate. This parameter defines the number of new exposed cases of age k generated by an infectious individual of age j. The rate is determined by two components: the number of people contacted daily by an infectious individual,[23] and the probability of transmission given contact. The contact rates are described by a matrix C_i whose elements C_{ijk} describe the number of contacts per day that a person of age j has with people of age k in city i. The P(T|C) is chosen by the modeler to determine the reproductive rate of the virus (determined by the particular biological characteristics of the specific influenza strain), and is constant in every city and age-group.

A key aspect of the modification of the GEM for this research has been an effort to incorporate realistic information on contact rates, so as to be able to partially explain observed variation in attack rates by age-group and to model variation in contact rates across urban/rural status and across countries. This dissertation derives its assumptions on age-specific contact rates from survey data from European countries as reported in

[22] This means that the expected number of travelers moving from i to j will be less than X_{ij} , as long as there are any cases who are symptomatic infectious or dead, since these people do not travel.

[23] A contact is assumed to be a physical contact, where two individuals encounter each other closely enough that there is a chance of influenza transmission. The precise definition of contact can be unclear in the literature, and this can complicate use of data where individuals report the number of contacts they undertake per day.

Mossong et al. (2009). The method by which the contact rate matrix (and thus λ_{jk}) is derived is described in Appendix 3.

Contact rates vary by age-group; in general each age-group is more likely to contact people of their own age than other ages; school children contact mostly other school-children, adults contact mostly other adults, and so forth. School-age children have the highest contact rates, while elderly people have the lowest contact rates. This pattern influences the incidence of influenza cases across age-groups. However, it is assumed that the total absolute number of contacts per day does not vary across cities based on age structure, so countries with younger populations will not have higher attack rates.[24]

Contact rates are adjusted across cities by urban density (for cities) and by the urbanization rate of the "rural" areas (see Appendix 2 for details). This has the effect of increasing contact rates (and thus attack rates) for cities in lower income countries (which tend to have high density cities), and for rural areas in higher income countries (which have higher urbanization rates).

A model extension could allow the P(T|C) to vary by age-groups, so that for example a child or elderly individual might be more likely to contract the virus than a healthy adult. But this dissertation retains the assumption of a constant P(T|C) for two reasons. First, reliable data on relative vulnerabilities is not available, and the relationship between age and vulnerability is not clear. While the elderly may have poorer health and weaker overall immune systems, older individuals are also more likely

[24] An alternative treatment that held absolute contact rates constant for each age-group would increase average contact rates and thus transmission rates and attack rates in countries with younger populations.

to have encountered related influenza strains in the past, and so may have developed some natural resistance. In the 2009 H1N1 pandemic, evidence suggests that healthy young adults may have been more vulnerable than older adults or children. Second, this research does not examine non-pharmaceutical interventions, and so the epidemiological consequences of allowing the P(T|C) to vary by age-group would be very similar to the consequences from incorporating variation in contact rates by age-group - both have the affect of introducing age-dependent transmissibility. For example, doubling the P(T|C) for some particular age-group pair would have a similar affect as doubling the number of daily contacts. If the model were to examine policies that affected contact rates, then it would be more important to separate age-specific contacts from age-specific transmissibility.

iv) Seasonality

One of the key stylized facts of influenza is that infection rates are much higher in winter months than in summer in temperate zones, but that there is less seasonal variation in tropical or arctic conditions. The mechanism for this is not well understood. Theories focus on both biological mechanisms and behavioral mechanisms. Biological mechanisms include the possibility that the influenza virus is more fragile in warmer temperatures and has lower transmission at higher humidity (confirmed in controlled environment tests, e.g. Lowen et al. (2008)), perhaps because virus particles are coated in moisture and less able to infect new hosts. Shaman and Kohn (2009) find that absolute humidity (rather than relative humidity) explains the seasonal variation of influenza in temperate zones, with lower transmissibility at higher humidity. Behavioral mechanisms suggest that in winter months (particularly in higher latitudes) people spend more time indoors in close proximity, and so have higher contact rates and thus a higher daily

infectious contact rate. In this dissertation, seasonality is incorporated by adjusting infectiousness rates rather than contact rates.

Existing global influenza models attempt to account for seasonality; most global pandemic models use some kind of step function as an approximation, where seasonality is constant for tropical regions and is lower during summer months in the temperate zones. For example, Colizza et al. (2007) divides the world into three zones; northern hemisphere, tropical, southern hemisphere. They assume constant seasonal transmissibility in the tropical zone, and use a seasonal transmissibility factor that varies between 0.1 in summer and 1.1 in winter in the temperate zones, using a monthly stepwise function such that the average temperate zone transmissibility is 68.3% of that in the tropical zone. The GEM takes a similar approach, but incorporates seasonality through a continuous seasonality infectiousness factor multiplier f(t) that varies by day of the year and by latitude of the city.[25] Infectiousness is assumed to be constant in the tropics, and elsewhere to follow a sinusoidal pattern where it is at maximum value on January 1 in the northern hemisphere (July 1 in the southern hemisphere) and minimum value on July 1 in the northern hemisphere (January 1 in the southern hemisphere). Specifically, for city *i* the seasonality factor is given by:

$$f(L_i, t) = (1 - \Lambda_i) + (0.55)\Lambda_i + (0.45)\Lambda_i \sin(L_i) \cos\left(\frac{2\pi(t + t_0)}{365.25}\right) \qquad (30)$$

L_i is the latitude of city i (interpreting southern hemisphere as negative latitude), where t_0

[25] The GEM uses the population-weighted mean latitude of the modeled cities in a region for the latitude of that region's rural area.

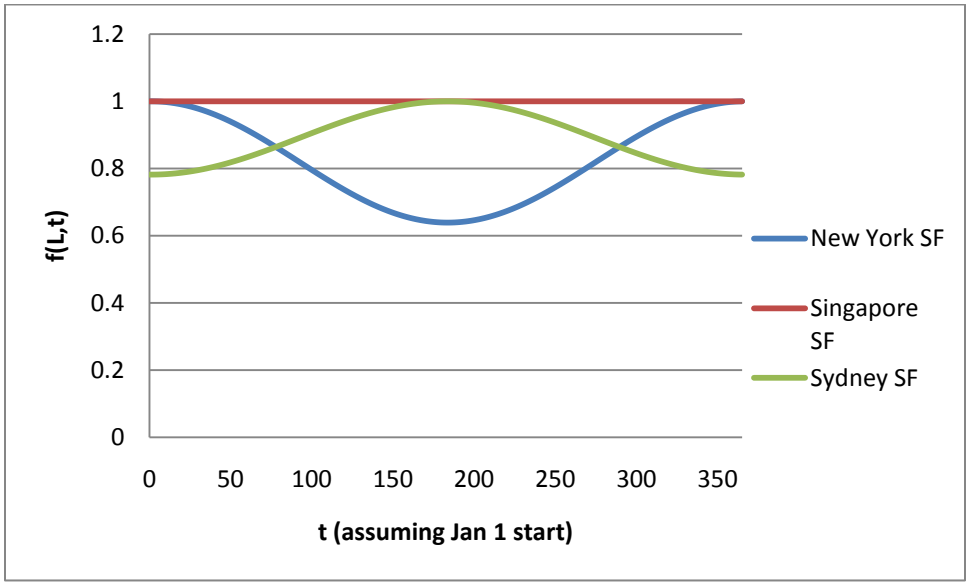
is the day of the year corresponding to the start of the pandemic, and where Λ_i scales L_i to a value between 0 and 1 according to:

$$\Lambda_i = min\left(1, max\left(0, \frac{|L_i| - 23.5}{66.5 - 23.5}\right)\right) \quad (31)$$

where 23.5 is the absolute value latitude of the tropics and 66.5 is the absolute latitude of the Arctic/Antarctic circles. The Λ_i function thus implies that the seasonality factor is constant in the tropics, has increasing amplitude through the temperate zone, and does not increase in amplitude further in the polar zones. The maximum value is constant for all latitudes, so the increasing amplitude means that in higher latitudes the average seasonality factor is lower - and so disease transmission (and thus attack rates) attack rates are highest in the tropics, and are lower the further a country is from the equator (up to the polar boundary).

Figure 4.1 shows the seasonality factors for Singapore (in the tropics), New York (temperate, northern hemisphere) and Sydney (temperate, southern hemisphere) with a pandemic starting on January 1 (winter in the northern hemisphere). Note that New York is further north than Sydney is south, and so New York's seasonality factor has higher amplitude and a lower mean value. These imply that all else equal, Singapore will have a higher attack rate than Sydney, which will have a higher attack rate than New York.

Figure 4.1: Seasonality factor for selected cities



Unfortunately, there is little empirical data on which to base this assumption. Cooper et al. (2006) compare a constant-maximum infectiousness model (as latitude changes) to a constant-mean model and finds the former performs slightly better when trying to fit observed attack rates from the 1968/69 influenza pandemic. They also find that a sine wave functional form fits the data better than a square wave or no seasonal variation.[26] In contrast, evidence from Shaman and Kahn (2009) suggests that influenza seasonality is driven largely by absolute humidity. Absolute humidity is always high in the tropics but is high in temperate zones only in summer, which suggests that a constant minimum model may be more accurate than a constant-maximum model. But, further evidence is lacking. This is unfortunate, since this seasonality assumption has a huge impact on pandemic dynamics and on the incidence of attack rates across countries.

[26] While this functional form matches the single peaked seasonal pattern observed in many countries, it does not matched the bimodal seasonality observed in some cases, such as Hong Kong.

v) Performance of the GEM with no policy interventions

Before moving on to consider policy interventions in the GEM, it is useful to first describe the auxiliary assumptions made in running the GEM, and to demonstrate the performance of the model in the baseline scenario.

Though the GEM is stochastic, the main simulations are run using a deterministic version of the model where each random variable[27] is assumed to take the mean value of its distribution (and so the model allows for fractional values of S(t), E(t), I(t), R(t) and D(t)). This reduces the number of runs that needs to be performed by an order of magnitude. While there is some variation across runs in the early stages of the stochastic model, the terminal values of key variables are nearly identical across runs of the stochastic model[28]. This means the model loses the ability to consider variation in early spread patterns, but also that the model run-time is drastically reduced, and so many more scenarios can be examined given the limited computing time available.

The pandemic is assumed to begin with 100 infectious cases on January 1 in Jakarta.[29] While the state of the model is observed for every day over the 3-year period for which the simulations run, analysis is focused on the attack rate after 1 year. This is a somewhat arbitrary cutoff (though it is common in the literature), but in part it represents the approximate amount of time that it might take to develop, test and mass produce a pandemic-strain-specific vaccine. Though any influenza vaccine remains exogenous, it is

[27] The probability of transmission given contact, the probability of recovery and the probability of travel are the key random variables in the stochastic version of the model.
[28] The global attack rate is identical to five significant figures across stochastic runs.
[29] There is no particular reason to pre-suppose that Jakarta is particularly likely to be a start point. A pandemic influenza strain that mutates from an avian or other animal strain is more likely to occur in a country where people in rural areas live in close proximity to livestock, but depending on the movement of the first few cases an outbreak could occur virtually anywhere. Jakarta was selected as a city in a large poor country that is well-connected to the international airline network.

an important motivating factor behind why consideration of attack rates after particular duration, rather than considering only the long-run steady state condition where the pandemic would be completely over without any further intervention. Many of the gains from antivirus policy come from delaying the spread of the virus, and delay is valuable in part because it buys time for vaccination development.

Every region in the model is assigned to an Income Group[30]; countries either Poor, Lower Middle, Upper Middle or Rich (see Appendix 2). This classification is useful for defining antiviral scenarios where the size of the stockpile differs depending on the purchasing power and public health system of the country in question (such as a scenario where rich countries have a large AV stockpile but poor countries do not), and is a useful means of showing the distributional impact of policies[31].

Figure 4.2: Distribution of population by income group and age-group

	Total	age 0	age 1	age 2	age 3
Poor	2,839,998,811	352,152,534	647,655,318	1,716,456,775	123,734,183
Lower_Middle	2,156,161,724	165,680,600	361,106,698	1,473,910,542	155,463,884
Upper_Middle	515,385,759	36,733,909	79,318,428	349,989,705	49,343,717
Rich	914,861,650	53,135,831	109,758,061	610,674,513	141,293,245
TOTAL	6,426,407,944	607,702,874	1,197,838,506	4,151,031,535	469,835,029

[30] Income groups classifications are based on International Monetary Fund data on nominal 2008 GDP per capita, at 2008 exchange rates. Poor countries are those with GDP per capita less than $USD 3,000, Lower Income are those with GDP per capita up to $10,000, Upper Middle have income up to $20,000 and Rich countries are those with incomes exceeding $20,000. For regions with multiple countries, the median income country is used to classify the region. We use this classification (rather than say the World Bank income group classifications) because we feel it is more representative of public health sector infrastructure.

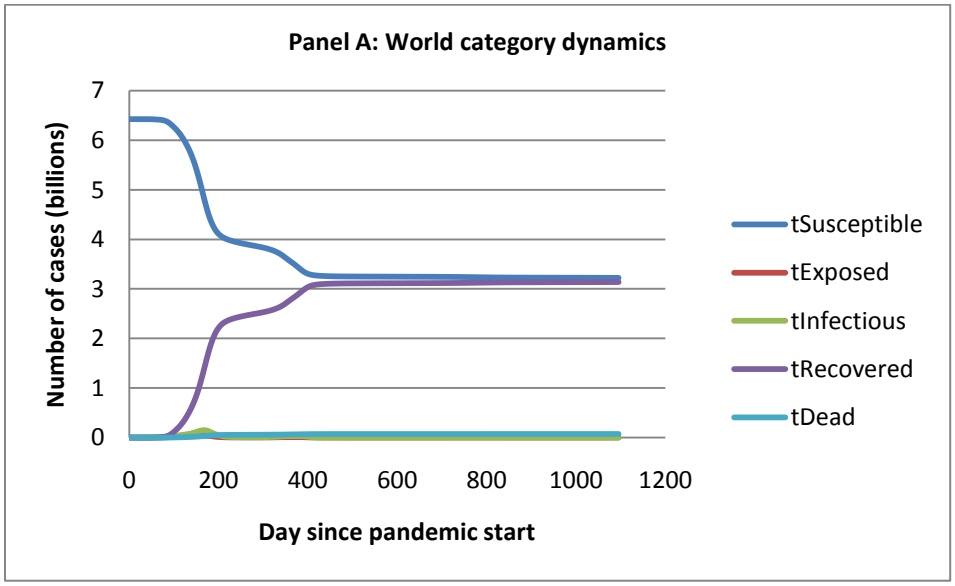
[31] The identical nature of agents within a city or rural area mean that we cannot capture distributional consequences within a country, though it is likely that higher income people will have better access to healthcare services including antivirals and may have lower contact rates than lower income people (eg by driving rather than using public transport).

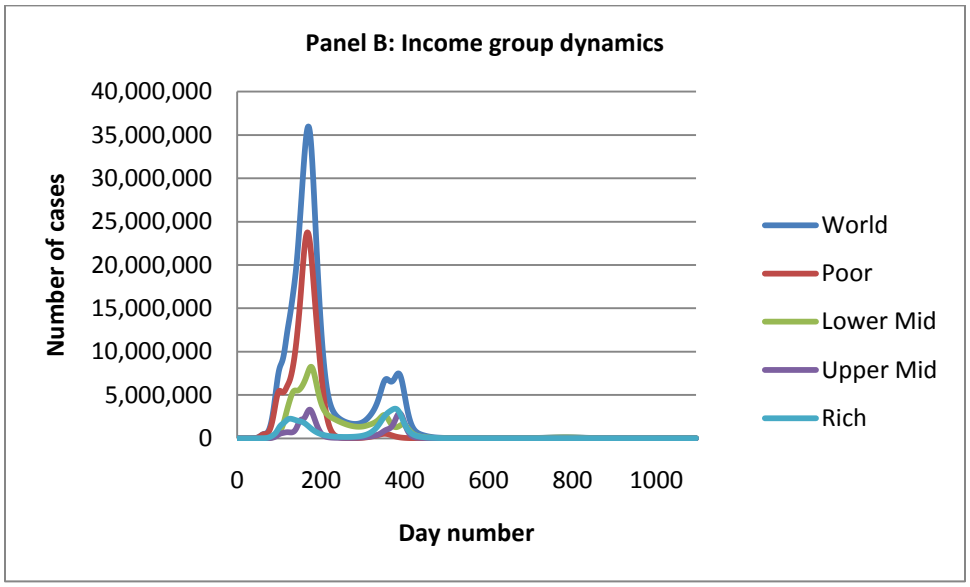
The baseline scenario assumes a "moderate" P(T|C) value of 0.05333, which implies a global average R_0 value of 1.48 (and an R_0 in the outbreak source Jakarta of 1.91). This value is somewhat arbitrary, but is designed to demonstrate the effectiveness of antivirus while remaining above the range where a small amount of anti-virus is able to prevent occurrence of a pandemic. "Low" and "High" values of P(T|C) are assumed to be P(T|C) = 0.045 and 0.06, respectively, which imply global average R_0 values of 1.25 and 1.67.

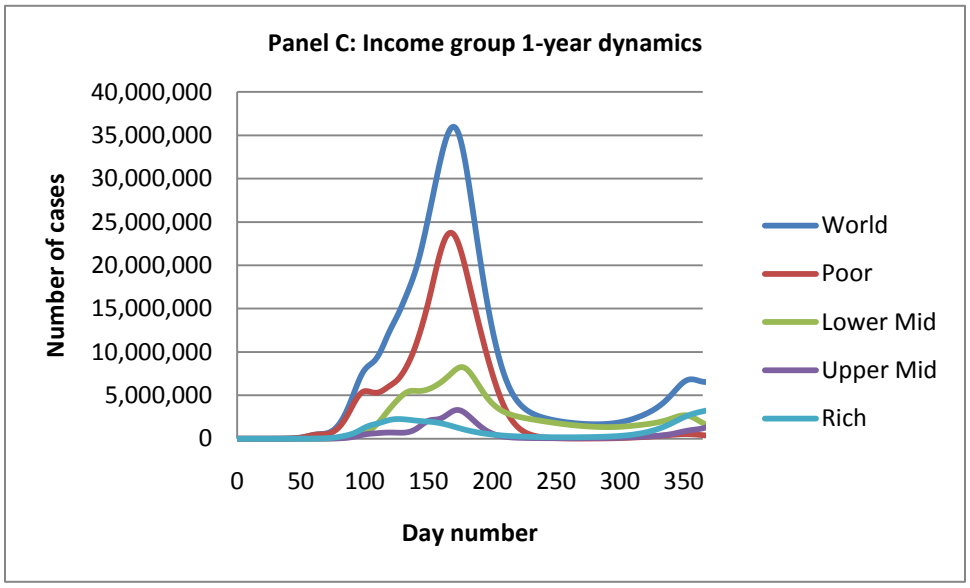
Figure 4.3 shows how the pandemic progresses in this baseline scenario. Panel A shows behavior of the global categorical variables (from aggregating each state across all cities and rural areas) for the entire 3-year run of the simulation model. Panel B shows the number of new infectious cases on each day, broken down across the 4 income groups, while Panel C shows the data but concentrates on the first year of the pandemic. Panel D shows the incidence of attack rates (total proportion of people ever infected) by country income group[32], while Panel E shows the incidence by age-group.

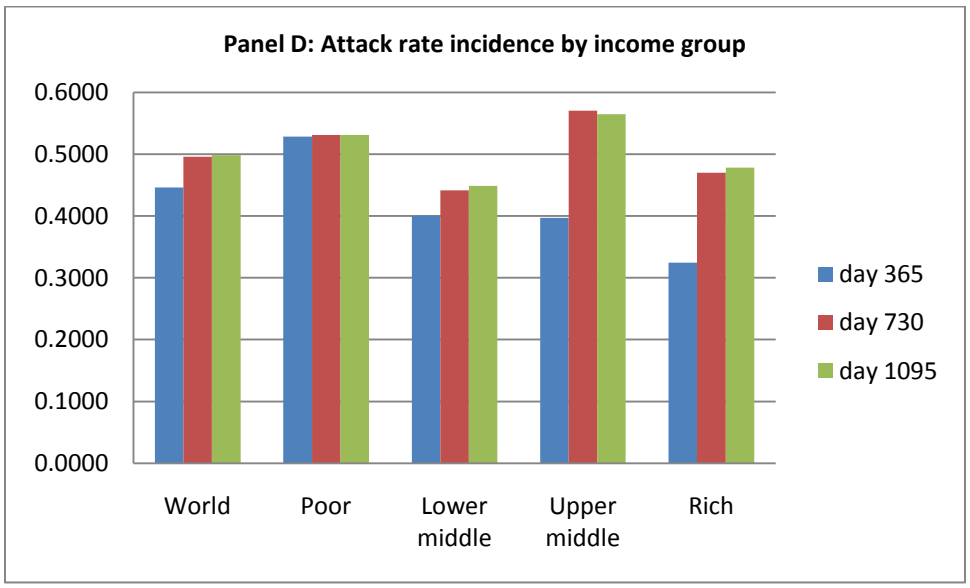
[32] Recall that this is incidence across countries by the average income of the country; the model says nothing about incidence varying by income within a particular country.

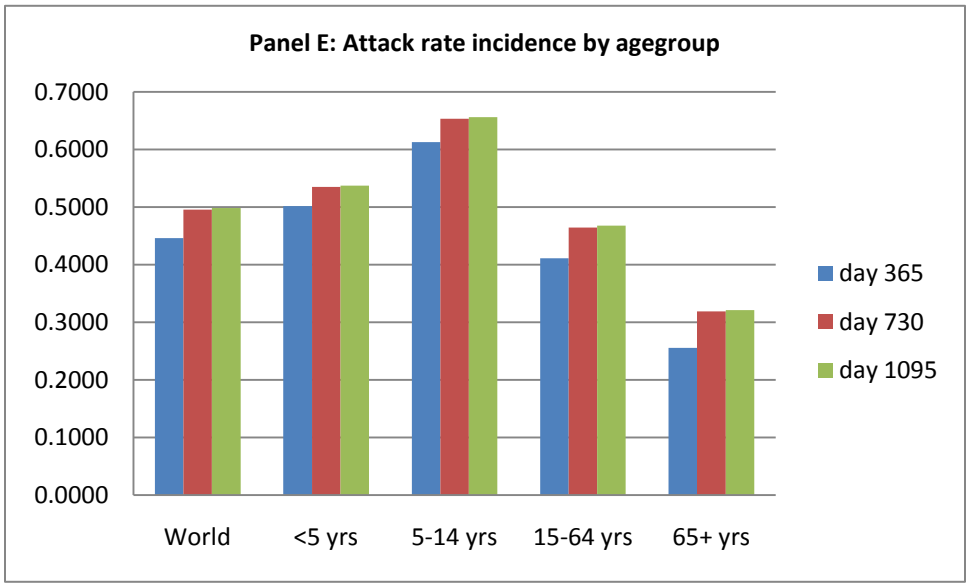
Figure 4.3: Pandemic time path, baseline






Panel C: Income group 1-year dynamics
Number of cases
40,000,000
35,000,000
30,000,000
25,000,000
20,000,000
15,000,000
10,000,000
5,000,000
0
0
50
100
150
200
250
300
350
Day number
World
Poor
Lower Mid
Upper Mid
Rich


Panel D: Attack rate incidence by income group
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
World
Poor
Lower middle
Upper middle
Rich
day 365
day 730
day 1095

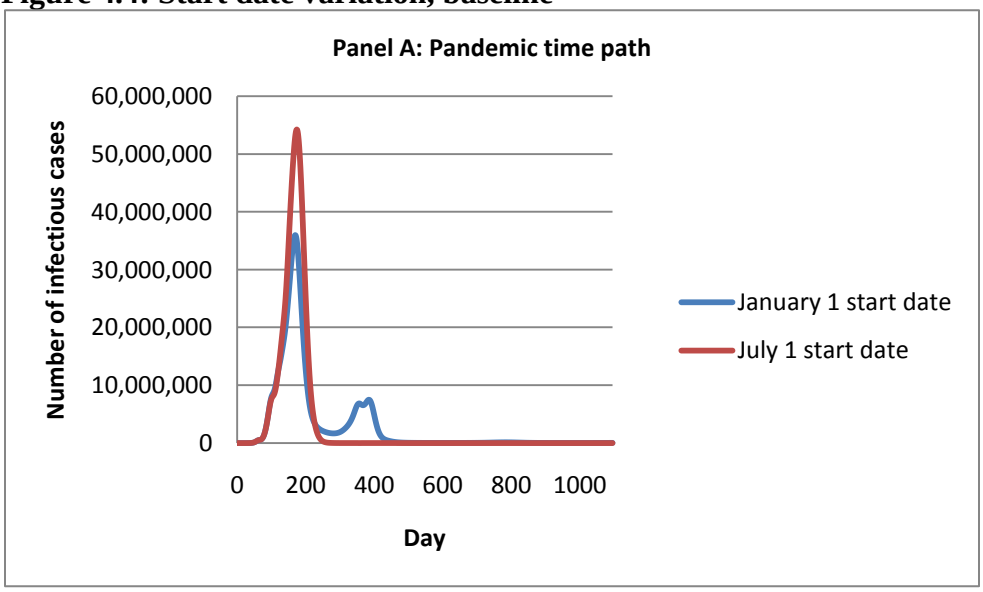


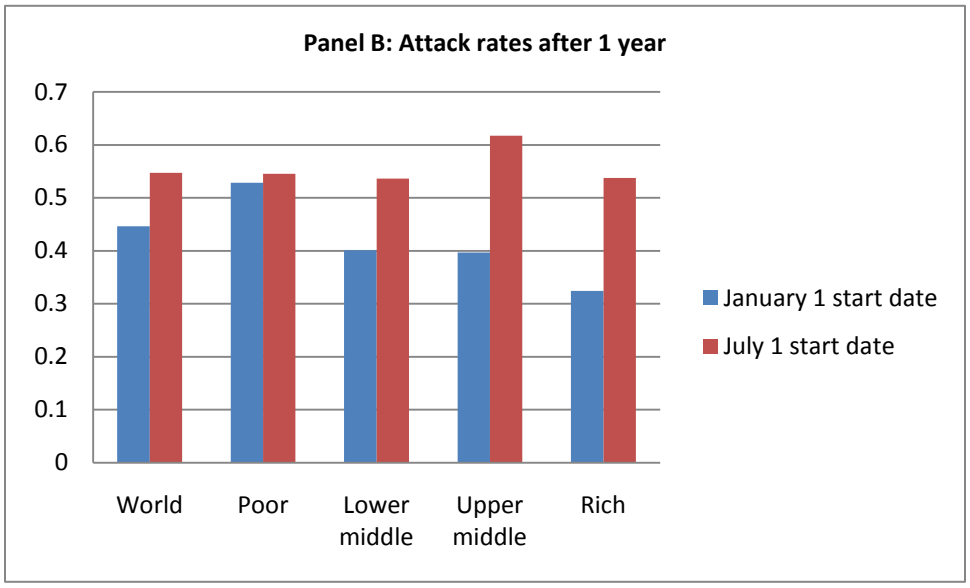
On day 1, the entire world is susceptible (except 100 infectious cases in Jakarta). Over time some people become Exposed, Infectious and then Recover or Die, first at an increasing rate though to the primary pandemic peak and then at a decreasing rate as the pandemic passes its peak (and the reproductive rate falls below 1 in most countries). A secondary wave occurs after roughly 300 days, as the northern hemisphere (where most of the population lives) re-enters winter. Most cases occur within a year; the 1-year attack rate is 44.6%, rising to 49.9% after 3 years.

Attack rates are higher in poor countries and lower in rich countries, due primarily to the positive correlation of income and latitude (see below. Rich countries have a larger increase in attack rate after the first year because they are at higher latitude, and so are more affected by seasonality, and so have a larger number of cases from the second pandemic wave. School-age children suffer higher attack rates than other age-groups, because of their high contact rates; the elderly suffer lower attack rates because of their relatively low contact rates.

Attack rates and attack rate incidence are highly sensitive to the pandemic start date, due to the interaction of start date and seasonality assumptions. Figure 4.4 compares the pandemic time path and attack rates for a January 1 start date (where the pandemic peaks during northern hemisphere summer) as compared to a July 1 start date (where the pandemic peaks during northern hemisphere winter). The secondary infection wave occurs in the Jan1 scenario only because the pandemic is choked off prematurely by rising temperatures as the northern hemisphere enters summer, while in the July1 scenario northern hemisphere temperatures are cooling as the disease spreads and so the pandemic reaches a single, more intense peak. The increase in attack rate after 1-year is largest for rich countries, due to the truncation of the full pandemic impact created by looking only at a 1-year attack rate.

Figure 4.4: Start date variation, baseline



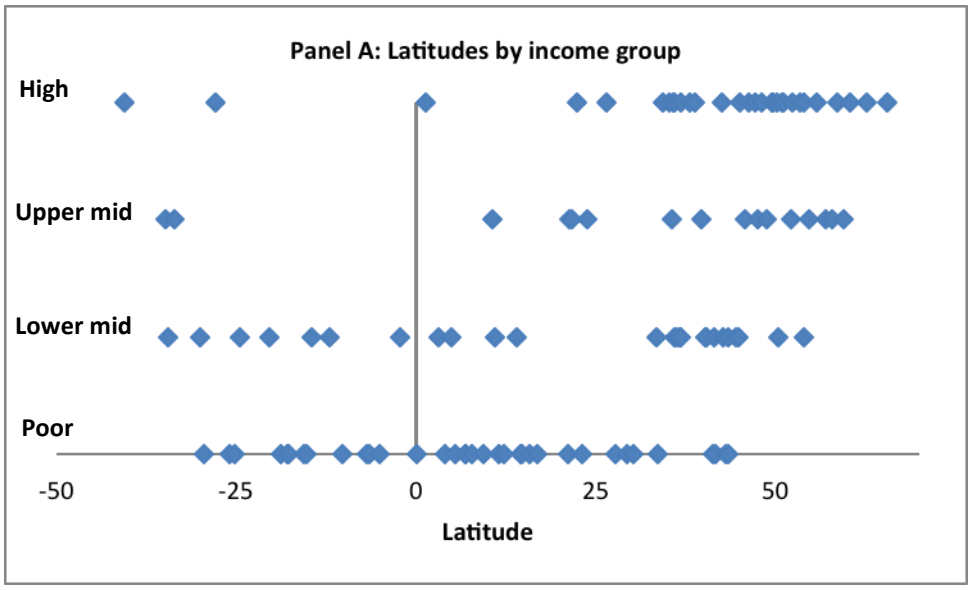


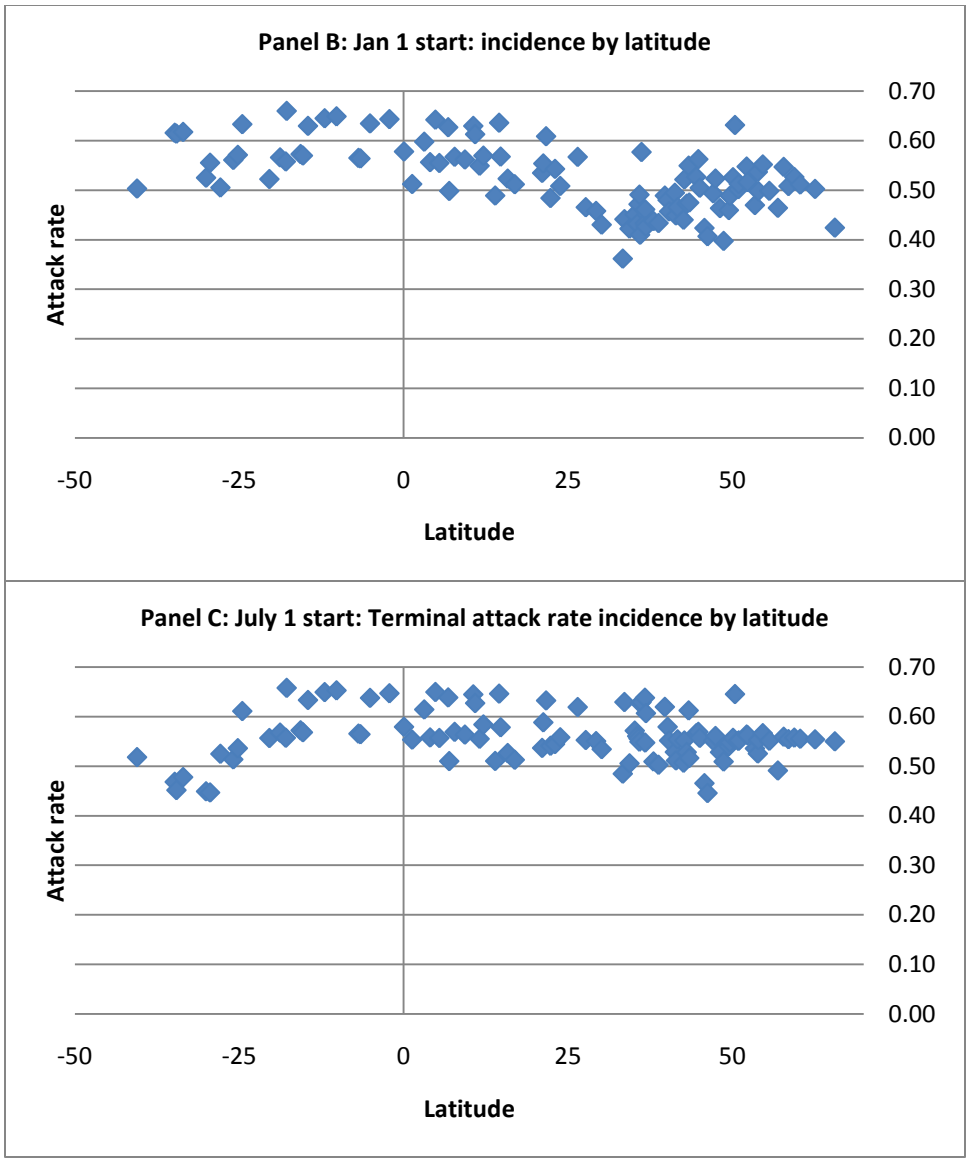
The incidence of attack rates across income groups is driven in large part by the combination of the seasonality assumptions and the real world correlation between income and latitude; countries more distant from the equator have lower average seasonality factors, and so lower attack rates. Figure 4.5 displays the distribution of attack rates (after 3 years) by latitude across the GEM's 106 regions. Panel A shows the distribution of regions across latitude by income group, and demonstrates the stylized fact that higher income countries lie further from the equator. Panel B shows attack rates under a January 1 start date, while Panel C shows attack rates under a July 1 start date.

Recall that the seasonality factor is constant in the tropics (between positive and negative 23.5 degrees latitude) and then average infectiousness decreases with distance from the equator. This explains why attack rates are roughly constant in the tropics, and then generally decreasing away from the tropics. The decrease is larger for the northern hemisphere in panel B and the southern hemisphere in panel C, because of the timing of the pandemic peak relative to summer in each case (seasonality factor is at minimum in

summer). Variation in attack rates not due to latitude (the upward slope above ~mid 30s latitude in Panel B for example) are largely due to the structure of the international airline network and regional populations (and so the number and timing of exposed cases imported) and to variation in density and urbanization rates across countries. Cities in poorer countries are more likely to be high density (increasing contact and attack rates), but poorer countries tend to have lower urbanization rates(decreasing contact and attack rates).

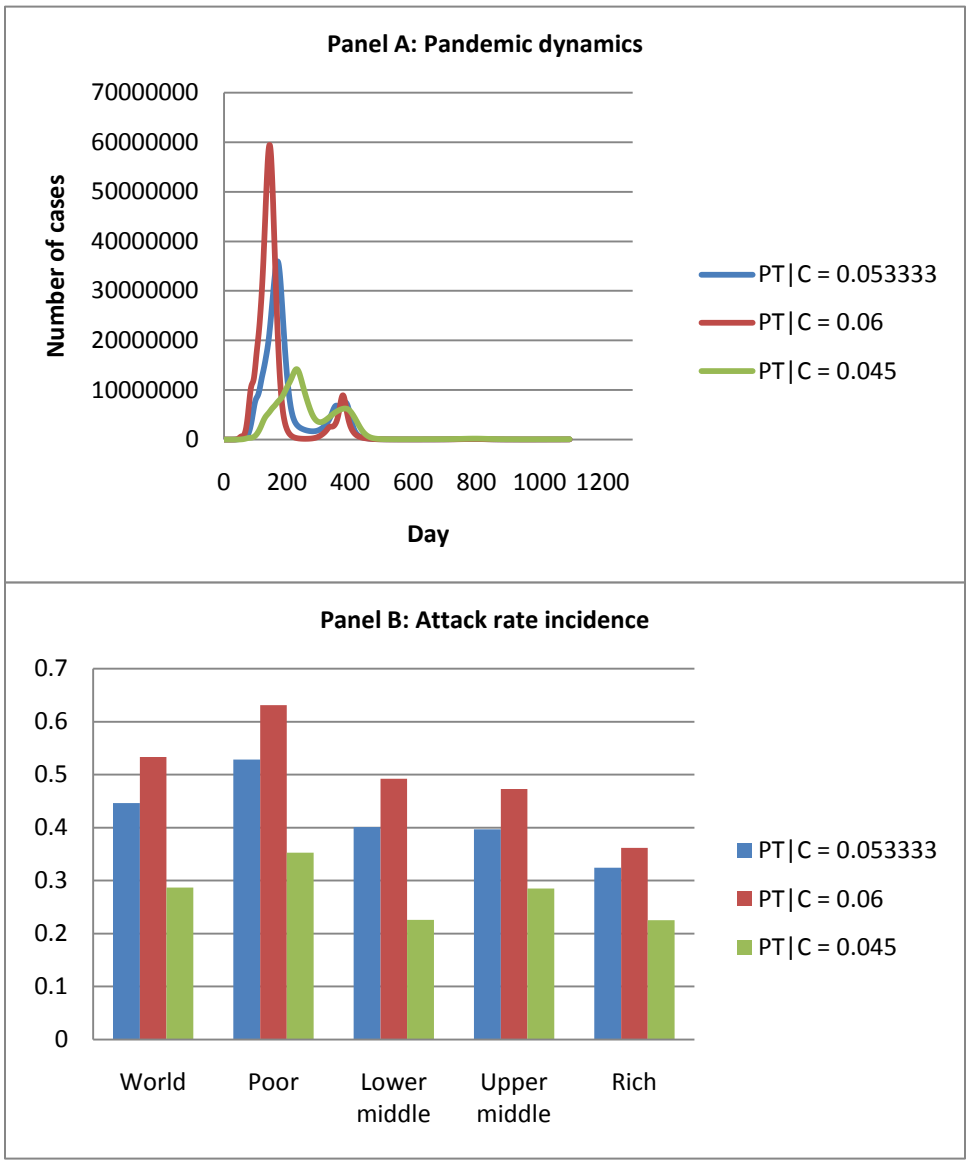
Figure 4.5: Effect of latitude in baseline model





A higher value of P(T|C) results in a faster pandemic and a higher attack rate (and the opposite for a lower P(T|C) value) while retaining similar qualitative properties (see Figure 4.6). In a "severe" pandemic with P(T|C) = 0.06 the global 1-year attack rate rises to 53.3%, while in a "minor" pandemic with P(T|C) = 0.045 the global 1-year attack rate falls to 28.7%.

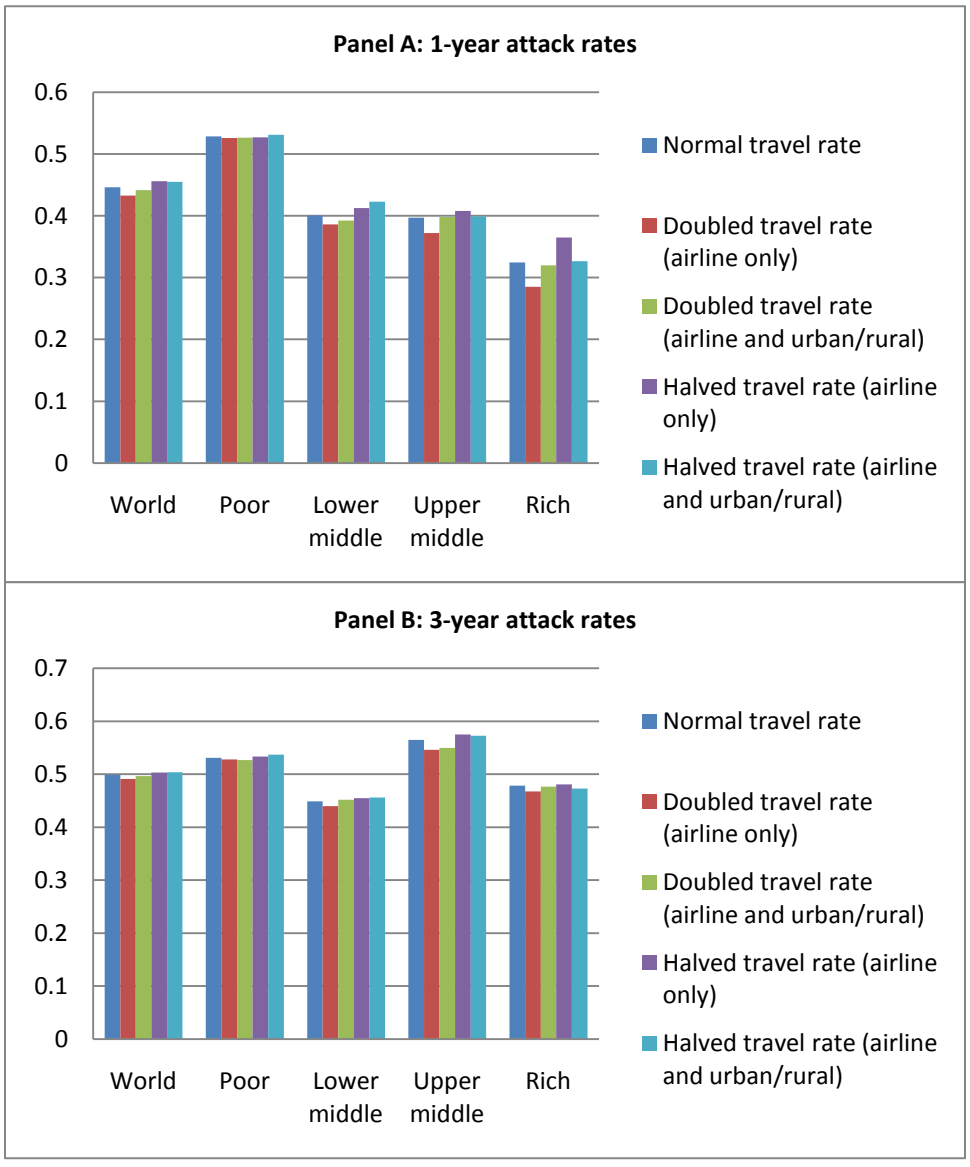
Figure 4.6: Effect of latitude in baseline model



Results are generally insensitive to travel rate assumptions. Figure 4.7 shows that 1-year attack rates for wealthier countries vary slightly with a doubling or halving of travel rates, and the impact on terminal attack rates after 3 years is minimal. Adjusting travel rates has two effects; a higher travel rate means a slightly faster pandemic spread (which might be expected to lead to higher attack rates), but a higher attack rate also shifts the season in which the pandemic peak occurs (which may increase or decrease

attack rates depending on whether it shifts the pandemic peak towards or away from summer in the northern hemisphere). It appears that the second effect dominates; attack rates are marginally lower with higher travel rates. The apparent insensitivity of attack rates to travel rates implies that airline travel restrictions would be relatively ineffective (as in Epstein et al. 2007).

Figure 4.7: Effect of latitude in baseline model



While the GEM has not been formally validated against modern pandemics, earlier versions of the GEM have been compared to those of earlier pandemic models. Rvachev and Longini (1985) test their model against the 1968-9 pandemic starting in Hong Kong, and found an approximate fit to the first wave of the pandemic, though they noted that there was often little empirical data with which to compare their model - either the data were not collected, or there was no actual outbreak in some cities. Grais et al. (2003) expand the model, and validate their model by reproducing the Rvachev and Longini results. In unpublished background work, Goedecke and Bobashev find that the GEM is able to replicate the Rvachev and Logini results when running the GEM with 1968 parameters.

Chapter 5 Policy interventions in the Global Epidemiological Model

Single-city and two-city epidemic models are unable are to give policy results that are robust to any input parameters. To progress further some idea of what is a "reasonable" input parameter is needed, but it is difficult to make judgments about what inputs are realistic in such highly stylized models. Adopting the GEM allows policy questions to be approached using a model with a high degree of descriptive realism, and allows use of data inputs and assumptions based on observable or semi-observable real world parameter values. This chapter describes how and why the GEM is adapted to focus on antivirus stockpiles, and presents results on: the value of antivirus in reducing the number of influenza cases, the incentive for wealthy countries to pay for antivirus in poor countries, the effectiveness of various antivirus stockpile allocation rules, the value of improving health infrastructure and the cost effectiveness of antivirus policy options.

i) The GEM and Antivirals

The generalized version of the GEM is capable of modeling several policy interventions; it can model vaccine use (which reduces the probability that a vaccinated Susceptible case will be infected if contacted), it can model travel restrictions (where cities cut air travel links), it can model non-pharmaceutical interventions such as physical distancing measures or partial quarantine (which reduce contact rates) or it can model treatment with antivirals (which reduces the probability that an infectious individual will infect others when contacted). This research focuses on antivirals.

The GEM is not very helpful in changing our understanding of vaccine policy issues. As in the 2009 influenza pandemic, by the time a pandemic influenza strain has

been detected and a strain-specific vaccine has been developed, mass produced and distributed, it is highly likely that a severe pandemic will have already run its course (though a milder pandemic that spreads more slowly may still be susceptible to significant mitigation through vaccine). The externalities of vaccine are relatively well understood (see for example Boulier et al. 2007); there is a strong "herd immunity" effect from preventative vaccination, and so mass prophylaxis with vaccine can be highly effective.

Travel restrictions have been studied using an earlier version of the GEM in Epstein et al. (2007). In general, travel restrictions are not very effective for most cities unless they are absolute and can be implemented extremely early (which is difficult to do when a pandemic may not be detected for weeks or months), because after a relatively small number of infectious cases exist in a city most of the additional contagion comes from natural increase rather than by importing infected cases. In the modern global economy absolute travel restrictions are extremely expensive, which limits their realistic effectiveness as a policy measure.

The GEM is capable of modeling physical distancing measures, such as school closures or workplace closures. These policies reduce the contact rates of individuals in particular age-groups (though they might also increase contact rates between other age-groups); for example, a school closure policy might reduce the rate at which children contact other children, but might increase the rate at which children contact adults. However, the effect of physical distancing measures in an epidemiological model is very similar to the effect of antivirals. What matters is the daily infectious contact rate, and so the effect of physical distancing measures that reduce the number of contacts will be

almost identical to antivirus treatment policies that reduce the probability of transmission given contact by an equivalent amount. There is also less policy value from considering external benefits from physical distancing, since it would be difficult for one country to pay for another country to undertake physical distancing measures (in contrast to antivirals, where one country could pay for doses to be provided to and distributed in another country).[33]

The GEM is prescriptive, rather than agent-based. Though governments implement policy actions, these policies are imposed as part of the model design, rather than policies that are selected optimally[34]. Individuals in the model make no decisions. They do not change their travel patterns in response to the pandemic, either at a global level (by reducing or changing the pattern of their movement between cities to avoid areas where the contact risk is high) or at a local level (by reducing their contact rates per day through self-sequestration when they are infectious, or self-isolation when there are many other infectious cases present in their city). Instead, people are assumed to act in a way dictated by parameter estimates calculated from behavior observed in non-pandemic settings. This will tend to lead to over-estimates of disease spread and attack rates from the GEM, relative to a real world pandemic.

There are agent-based epidemiological models that incorporate prevalence-elastic behavior endogenously, but not at the global scale of the GEM. The sheer scale of the GEM makes individual optimization infeasible; the computational power needed to

[33] However, it might be possible for a donor country to partially subsidize physical distancing measures by providing masks or gloves.

[34] The GEM is motivated by the idea that there is an initial first step where governments are simultaneously choosing an optimal stockpile size for their region, but this is not explicit in the model.

calculate optimizing decisions by individuals or even representative agents of every type in every city in every day would be extremely high. Such models are more effective when used in a smaller setting or a model with less detail.

The GEM is a model of perfect information. Even if the model included a step where governments chose an optimal stockpile size, it would still have to assume that the decision was made with perfect knowledge that a pandemic influenza strain with particular properties (start date, start location, biological properties such as transmission rate and recovery rate) will occur with probability 1. In reality, governments have to make pandemic plans and decisions about AV stockpiles without knowing whether or when a pandemic will occur.

The GEM is a large and complex model – the runtime of the model is significant even when using considerable computing power. This limits the number of model runs that can realistically be undertaken in a research project, and so limits the ability to explore the parameter space. In particular, it rules out the feasibility of doing the kind of optimization calculations that would be needed in order to calculate Nash equilibria in a formal strategic version of the model, with Regions or Super-Regions choosing stockpile sizes to maximize an objective function. This limits the ability to conduct formal economic analysis, but some important insights can be gained from examining particular parts of the parameter space to inform policy questions.

The GEM is not an effective model for tracking containment strategies**.** In some cases it may be possible to contain a virulent new influenza strain, and prevent a pandemic from occurring. In practice, this requires public health officials to perform a

focused contact-tracing procedure, where they target pharmaceutical interventions (such as antivirus) and physical distancing measures (such as quarantine) to specific individuals that may have had contact with early infectious individuals. This kind of process can be modeled, but it requires a hierarchical social network model that can differentiate between individuals of varying probabilities of having contacted an infectious individual, rather than the homogenous mixing procedures calculated in the GEM.

ii) Effects of Antivirus treatment

Following Longini et al. (2005), it is assumed that treating infectious cases with antivirus has two effects[35]: treatment reduces the probability of transmission given contact by a multiplicative factor (1 – e) where e is the efficacy of the antivirus,[36] and it reduces the average duration spent Infectious by one day, by uniformly shifting the probability distribution downwards. This means that an Infectious individual treated with AV will recover after 2-5 days of being Infectious (rather than 3-6 days for untreated individuals), with 30% recovering after 2 days, 40% after 3 days, 20% after 4 days, and 10% after 5 days, and so:

$$\sigma(\theta, treated) = \begin{cases} 0.3 \; if \; \theta = 1 \\ (0.4)/(0.7) \; if \; \theta = 2 \\ (0.2)/(0.3) \; if \; \theta = 3 \\ 1 \; if \; \theta = 4 \\ 0 \; otherwise \end{cases} \quad (32)$$

[35] Another possibility would be for treatment to reduce the case fatality rate, but here we focus on attack rates rather than mortality rates and so we do not consider this possibility. A modest reduction in case fatality rates will have little impact on the model dynamics or on the total number of people ever infected, but it could dramatically increase the value of avoiding a case.

[36] It is assumed that this efficacy is constant throughout the pandemic. The model does not allow for an AV-resistant strain to develop during the pandemic, which might be possible in the real world. An AV-resistant influenza strain (ie one where efficacy e was very low) would dramatically reduce the value of all AV interventions.

Each region has an AV stockpile size *P*(r),* and an amount of stockpile remaining *P(r,t)*. AV held by region *r* will be used to treat people in all cities (including the rural area) within that region. It is assumed that AV stockpile distribution begins in city *i* only when 1,000 people have become infectious in city *i,* [37] and then continues until there is no AV remaining in the region. Note however that in many cases a large stockpile in a region will not be exhausted. Any doses not consumed are wasted.

It is assume that a constant proportion *p* of symptomatic-Infectious cases are treated with AV (asymptomatic cases are not treated)[38]. This is a highly unrealistic assumption; real world treatment rates will vary depending on a number of factors including health system infrastructure and the proportion of the population infected. Unfortunately, there is little reliable data in which to build a plausible location-specific treatment number. In most model runs it is assumed that $p = 0.5$, so one-third of infectious cases are treated (while one-third are symptomatic but untreated and one-third are asymptomatic but untreated). This treatment occurs only after 1 day, so newly Infectious cases have 1 full day to infect others before they receive treatment to reduce their infectiousness.

Assume that there are no type 1 errors in treatment – AV doses are never used to treat non-infectious cases. While unrealistic, this assumption is relatively harmless, since

[37] This assumption is designed to model the difficulty in detecting the pandemic strain of influenza particularly against the "background noise" of regular seasonal influenza cases. Sensitivity tests show that final attack rates are not significantly affected by changing the number of cases needed before distribution begins, except in the special case of low P(T|C) where early detection will significantly reduce the number of cases, particularly in the outbreak source country where early detection can mean the outbreak is contained and no pandemic occurs.

[38] Treatment of influenza with AV involves a 5-day course of AV being consumed at a rate of 2 pills per day, so we assume each "dose" of AV is enough for the 5 day course.

the stockpile size could simply be increased in every city by some factor and receive the same qualitative results (treating x cases with no type 1 error is similar to treating 1.25x cases with a 20% false positive rate).

Antivirus is not used as prophylaxis. Though prophylactic use does have some medical benefit (a Susceptible individual treated with AV will have a lower probability of being infected if contacted), preliminary investigations with the GEM indicated that prophylactic use is relatively ineffective at influencing the pandemic spread. This occurs in part because the effect of being treated wears off rapidly after treatment ceases, and so prophylactic treatment is effective on a Susceptible person who is contacted during the brief duration in which the dose is effective. So a vast numbers of doses would be needed to sustain any significant reduction in the population at large. Those doses could be more effectively used for treating actual infectious cases rather than trying to blanket the population.

How will countries choose their antivirus stockpile? Ideally I could consider the possibility of locally and globally optimal allocation rules, assuming some welfare function that compared the social value of preventing cases across different countries.[39] But in practice, the GEM is too complicated for the optimal allocation to be calculated. An input parameter for the GEM with a stockpile allocation is a vector with 106 variables (and so 105 degrees of freedom for any given stockpile size) and so the number of possible combinations is enormous. It is not possible to predict the outcome of attack

[39] Such a welfare function might value all cases equally across all countries, or might value cases as a function of the dollar cost of cases in different countries (and so cases prevented in rich countries might be treated as more valuable than cases prevented in poor countries). Different beliefs about an appropriate global welfare function would lead to radically different outcomes.

rates across country without actually running the GEM, which takes over 30 minutes per run, and so it is infeasible to run the model enough times for any kind of optimization algorithm to converge. Instead, I examine the effects of antivirus by concentrating on a set of "plausible" antivirus stockpile combinations, where wealthier countries have larger AV stockpiles while poorer countries have smaller or no access to AV (unless it is provided to them by donors).

This notation indicates that all regions in the Poor income group have a stockpile size of a% of their population, Lower middle income regions have a stockpile size of b%, and so forth.[40] The four scenarios are (A) 0/0/5/10; (B) 0/1/5/10; (C) 1/1/5/10 and (D) 0/1/5/10 + a 4.2% stockpile for the outbreak country, Indonesia, which is in the Poor income group. The combined stockpile size of all rich regions is 91.5 and 25.8 million doses for all upper middle income regions (10% and 5% of their respective populations) in all scenarios.[41] The difference between scenario A and scenario B could be interpreted as a gift from wealthier countries of 21.5 million doses to Lower Middle income countries (equal 1% of their population), while the movement from B to C is equivalent to rich countries paying for 28.4 million doses in Poor countries (equal to 1% of their population). In scenario D Lower Middle income countries have stockpiles equal to 1%

[40] Although all regions in a particular income group receive the same stockpile, as a percent of their population, it is possible that the stockpile may be exhausted in some regions but not in others. This complicates interpretation of a change in stockpile size across an income group, since such a change may lead to additional people being treated in some regions but not in others.

[41] However, note that the entire 91.5 million doses are not always consumed, or are not always consumed within the first year, as some regions will not exhaust a stockpile equal to 10% of their population size. Since 1/3 of infectious cases are treated, a region will not exhaust a 10% stockpile any time its attack rate is less than 0.3.

of their populations and Indonesia has a stockpile of 9.2 million doses[42], which equals 4.2% of its population.

ii) The value of antivirus

Treatment with antivirus is highly effective in reducing pandemic attack rates. Figure 5.1 shows the effect on the global attack rate from AV treatment under a range of scenarios where AV stockpiles are concentrated in wealthier countries. The impact of each scenario on attack rates at the end of one year depends on the infectiousness of the flu (i.e., *P(T|C)*), the timing of the flu (whether the flu starts on January 1 or July 1) and the effectiveness of antivirals (*e*).

In all cases, there is a sizeable reduction in attack rate; recall that a reduction in global attack rate of 0.01 is equal to roughly 64 million fewer influenza cases by the end of the first year. Larger stockpile sizes lead to lower attack rates. Lower AV efficacy leads to a smaller reduction in attack rate. A high P(T|C) value reduces the effectiveness of AV. A low P(T|C) value means that the 1/1/5/10 scenario is able to contain the pandemic completely, because this scenario allocates AV to the outbreak source (Indonesia) and the low P(T|C) value implies that treatment in Indonesia is able to reduce the R_0 value below 1, and prevent a pandemic. But containment is not possible in other scenarios.

Figures 5.2-5.4 show the impact of varying these parameters on the cumulative number of influenza cases at the end of one year. In the standard case (medium infectiousness, January 1 start date and reduction in infectiousness of 60%), Scenario A

[42] The 9.2 million doses are 1% of the Rich income group population, so Scenario D is the equivalent of Rich countries purchasing an additional 1% stockpile, but distributing it to Indonesia.

reduces the attack rate by 5.2 percentage points (334 million cases); scenario B reduces the attack rate by 13.6 percentage points (877 million cases) and scenario C reduces the attack rate by 19.3 percentage points (1,237 million cases), all relative to the no-AV baseline. Scenario A leads to an average of 2.85 fewer cases per dose, 7.16 fewer cases per dose for Scenario B, and 8.60 fewer cases per dose for Scenario C.

Increasing the infectious of the flu ($P(T|C) = 0.06$) reduces the effectiveness of AVs, while a lower value of $P(T|C)$ makes containment of the flu possible under scenario C. Scenario C allocates AV to all poor countries, including the outbreak source (Indonesia), and the low $P(T|C)$ value implies that treatment in Indonesia is able to reduce the R_0 value below 1, and prevent a pandemic. But containment is not possible in other scenarios. Reducing the effectiveness of the AV dose from $e = 0.6$ to $e = 0.5$ reduces the number cases reduced per dose of AV from 2.85 to 2.55 in scenario A, from 7.16 to 6.36 in scenario B, and from 8.60 to 6.10 in scenario C.

Figure 5.1: Effect of AV on global 1-year attack rate

Scenario ID	AV stockpile size (% population, by income group)	P(T|C)	Start date	AV Efficacy	Baseline attack rate	Attack rate with AV	Difference
A1	0/0/5/10	Low	1-Jan	0.6	0.287	0.235	0.052
A2	**0/0/5/10**	**Medium**	**1-Jan**	**0.6**	**0.446**	**0.394**	**0.052**
A3	0/0/5/10	Medium	1-Jul	0.6	0.547	0.456	0.091
A4	0/0/5/10	Medium	1-Jan	0.5	0.446	0.400	0.047
A5	0/0/5/10	High	1-Jan	0.6	0.533	0.480	0.053
B1	0/1/5/10	Low	1-Jan	0.6	0.235	0.154	0.081
B2	**0/1/5/10**	**Medium**	**1-Jan**	**0.6**	**0.446**	**0.310**	**0.136**
B3	0/1/5/10	Medium	1-Jul	0.6	0.547	0.385	0.162
B4	0/1/5/10	Medium	1-Jan	0.5	0.446	0.325	0.121
B5	0/1/5/10	High	1-Jan	0.6	0.533	0.499	0.034
C1	1/1/5/10	Low	1-Jan	0.6	0.235	0.000	0.234
C2	**1/1/5/10**	**Medium**	**1-Jan**	**0.6**	**0.446**	**0.254**	**0.193**
C3	1/1/5/10	Medium	1-Jul	0.6	0.547	0.301	0.246
C4	1/1/5/10	Medium	1-Jan	0.5	0.446	0.309	0.137
C5	1/1/5/10	High	1-Jan	0.6	0.533	0.461	0.072

Note: "Standard" assumption scenarios are in bold. AV stockpile sizes a/b/c/d refer to Poor/LowerMid/UpperMid/Rich regions. Low P(T|C) = 0.045, Medium P(T|C) = 0.05333, High P(T|C) = 0.06.

Figure 5.2: Reduction in global influenza cases from AV, sensitivity to virulence

	P(T|C) = 0.045			P(T|C) = 0.05333			P(T|C) = 0.06		
Scenario	Δ Attack rate	Δ Cases (billions)	Cases reduced per AV dose	Δ Attack rate	Δ Cases (billions)	Cases reduced per AV dose	Δ Attack rate	Δ Cases (billions)	Cases reduced per AV dose
Baseline (no AV)	0	0		0	0		0	0	
Scenario A	0.052	0.333	2.836	0.052	0.334	2.848	0.053	0.341	2.906
Scenario B	0.132	0.850	6.946	0.136	0.877	7.163	0.034	0.219	1.793
Scenario C	0.286	1.839	12.773	0.193	1.237	8.594	0.072	0.466	3.235

Note: Change in attack rate and number of cases refer to the reduction relative to the no-AV baseline (so a positive value indicates fewer people infected). All figures are for the point 1 year after the pandemic commences.

Figure 5.3: Reduction in global influenza cases from AV, sensitivity to start date

	January 1 start			July 1 start		
Scenario	Δ Attack rate	Δ Cases (billions)	Cases reduced per AV dose	Δ Attack rate	Δ Cases (billions)	Cases reduced per AV dose
Baseline (no AV)	0	0		0	0	
Scenario A	0.052	0.334	2.848	0.091	0.585	4.993
Scenario B	0.136	0.877	7.163	0.162	1.041	8.508
Scenario C	0.193	1.237	8.594	0.246	1.581	10.984

Note: Change in attack rate and number of cases refer to the reduction relative to the no-AV baseline (so a positive value indicates fewer people infected). All figures are for the point 1 year after the pandemic commences.

Figure 5.4: Reduction in global influenza cases from AV, sensitivity to AV efficacy

	$\gamma = 0.6$			$\gamma = 0.5$		
Scenario	Δ Attack rate	Δ Cases (billions)	Cases reduced per AV dose	Δ Attack rate	Δ Cases (billions)	Cases reduced per AV dose
Baseline (no AV)	0	0		0	0	
Scenario A	0.052	0.334	2.848	0.047	0.299	2.550
Scenario B	0.136	0.877	7.163	0.121	0.779	6.360
Scenario C	0.193	1.237	8.594	0.137	0.879	6.104

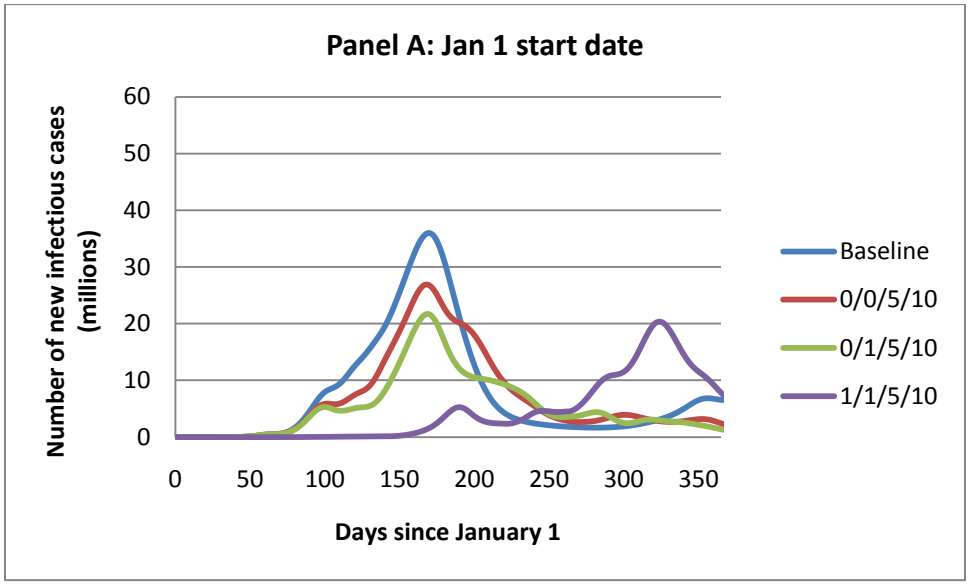
Note: Change in attack rate and number of cases refer to the reduction relative to the no-AV baseline (so a positive value indicates fewer people infected). All figures are for the point 1 year after the pandemic commences.

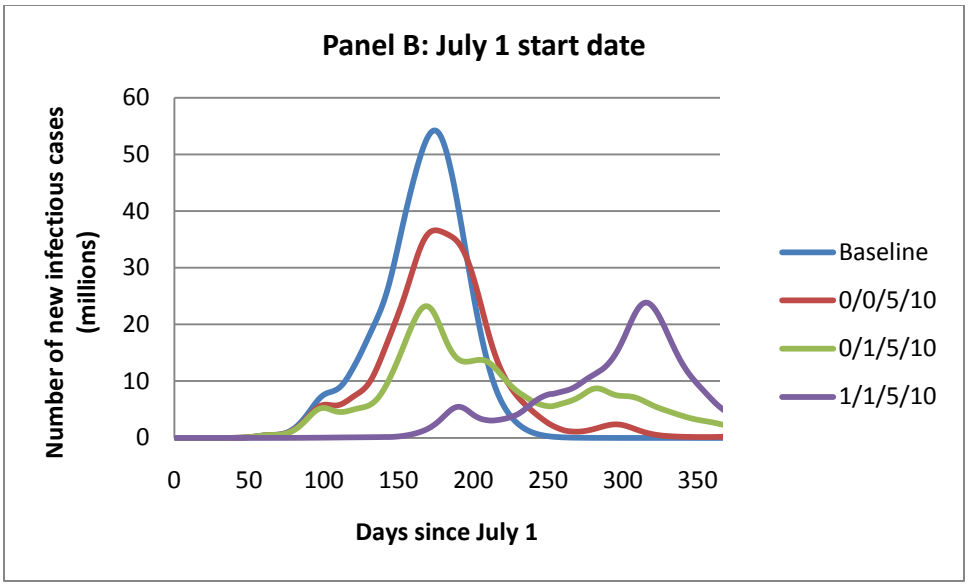
The impact of AV on attack rate is and is also affected by the start date of the pandemic, due to the seasonality assumptions. Recall that infectiousness in non-tropical regions is highest in winter and lowest in summer. This means that AV distribution, which slows the pandemic, will also have an indirect effect on infectiousness by changing the time of year during which the pandemic peak occurs. As shown in Chapter 4, a pandemic that starts on January 1 peaks during northern hemisphere summer given a "medium" P(T|C), but peaks during northern hemisphere winter for a pandemic starting on July 1. AV policies that slow the pandemic will tend to increase the average

seasonality factor for a Jan 1 start date, and decrease the average seasonality factor for a July start date, because most of the population lives in the northern hemisphere. The net impact of an AV policy is a combination of both its direct effect from reducing infectiousness, and its indirect effect via the change in seasonality factor.

Figure 5.5 shows how AV treatment delays the pandemic (shifting the distribution rightwards). With a January 1 start date, delay blunts the effectiveness of AV in reducing the attack rate, because it pushes the pandemic peak towards northern hemisphere winter. With a July 1 start date, delay augments the effectiveness of AV in reducing the attack rate, because it pushes the pandemic peak away from northern hemisphere winter.

Figure 5.5: GEM pandemic dynamics with AV treatment





Note: Results are for a P(T|C) of 0.05333, and AV efficacy 0.6.

iii) Antivirus, externalities and complementarities

In one and two-city SIR models, most of the gains from AV treatment accrued to the region in which the treatment was taken, but there were some positive spillover effects. These results also hold in the full GEM. Figure 5.6 shows the reduction in attack rates from AV treatment across income groups for the standard set of assumptions. Moving from no AV to the 0/0/5/10 scenario, there is a very large reduction in attack rate in Rich regions, a large attack rate reduction in Upper middle income regions, and negligible impacts in lower income regions[43]. Moving to 0/1/5/10 adds a stockpile in lower middle income regions, which significantly reduces the attack rate there while causing a small but positive impact elsewhere. Moving then to 1/1/5/10 adds an AV stockpile in poor countries, leading to a large attack rate reduction in poor countries and moderate benefits elsewhere.

[43] In fact, there is a very small increase in attack rate in Poor and Lower Middle income regions. This is due to interaction with seasonality effects; AV in wealthier countries delays the pandemic slightly, which means that northern hemisphere countries (including India and China) have slightly higher average seasonality factors.

Figure 5.6: Impact of AV on 1-year attack rate, incidence by income group, standard parameters

AV scenario	#AV doses	World	Poor	Lower middle	Upper middle	Rich
Jan 1 start date, Medium P(T|C)						
Baseline (no AV)	0.0	0.446	0.528	0.401	0.397	0.325
Scenario A	117.3	0.394	0.529	0.403	0.210	0.059
Scenario B	138.8	0.310	0.525	0.169	0.166	0.052
Scenario C	167.2	0.254	0.458	0.121	0.051	0.045
Jan 1 start date, Low P(T|C)						
Baseline (no AV)	0.0	0.287	0.353	0.226	0.285	0.225
Scenario A	117.3	0.235	0.351	0.223	0.010	0.027
Scenario B	138.8	0.154	0.343	0.003	0.003	0.010
Scenario C	167.2	0.000	0.001	0.000	0.000	0.000
Jan 1 start date, High P(T|C)						
Baseline (no AV)	0.0	0.533	0.631	0.492	0.473	0.362
Scenario A	117.3	0.480	0.630	0.487	0.299	0.102
Scenario B	138.8	0.499	0.627	0.547	0.299	0.102
Scenario C	167.2	0.461	0.606	0.458	0.302	0.108
July 1 start date, Medium P(T|C)						
Baseline (no AV)	0.0	0.547	0.546	0.536	0.617	0.538
Scenario A	117.3	0.456	0.544	0.532	0.268	0.110
Scenario B	138.8	0.358	0.542	0.331	0.249	0.102
Scenario C	167.2	0.301	0.484	0.183	0.183	0.077

Note: Results are for a January 1 start date, and AV efficacy 0.6

The small external benefits to Rich countries of Lower Middle and Poor countries having stockpiles still holds when the flu starts on July 1, assuming it is moderately transmissible. A highly transmissible flu, however, sharply reduces the benefits to Rich countries of stockpiles in Lower Middle income and Poor countries, because a small stockpile is rapidly exhausted under a virus with a high reproductive rate. Indeed, in the January 1, High Transmissibility case Rich countries are slightly worse off in Scenario C compared to Scenario B. This effect is driven by seasonality; in the high transmissibility scenario, AV in poor countries is ineffective in reducing attack rates by the slight delay it

causes pushes the pandemic peak in the northern hemisphere towards winter. The fact that Lower Middle income countries themselves are worse off when using AVs (118 million additional cases relative to no treatment) in the High transmissibility scenario with the January 1 start date follows largely from a higher attack rate in China, where the slight delay from a 1% stockpile pushes the pandemic peak towards winter (and so increases China's average seasonality factor).

It seems likely that in reality poor and lower middle income countries will not purchase and maintain their own stockpiles of antivirus, for use in the course of a pandemic. But the external benefits to rich countries suggest that there may be scope for a pareto-improvement; rich countries may find it in their interest to pay the costs of acquiring and distributing antivirals in poor countries. As discussed more fully below, Table 5.7 implies that the reduction in number of influenza cases in rich countries per AV dose administered in poor or lower middle income countries is 1/3 or more when the transmissibility of the flu is low to moderate.

An alternative possibility is for rich countries to donate AV directly to the outbreak source. Consider a scenario where rich countries collectively provide AV doses equal to 1% of their population directly to Indonesia. Figure 5.7 shows the effect of this policy when combined with a 0/1/5/10 scenario. In nearly every case, the donation policy reduces the global attack rate and the attack rate in rich countries. In a few cases attack rates are increased, due to the seasonality effect with the January 1 start date, where slowing the pandemic spread increases the average seasonality factor in the northern hemisphere. The gain from donation is largest in the Low P(T|C) scenarios, because in these cases donation to the outbreak source reduces the R_0 below 1 and the pandemic is

contained. the gains are also large in the July 1 start date scenario, where delay is more valuable due to the seasonality interaction. The external benefits of providing AV are lower when the AV efficacy is lower, which contributes to the negative impact of donation in the low AV efficacy scenario.

Figure 5.7: Impact of AV on 1-year attack rate: Scenario D

P(T|C)	Start date	AV efficacy	Attack rate for:	Policy: No donation	Policy: Donation to Indonesia	Difference
Low	1-Jan	0.6	World	0.178	0.001	0.176
Low	1-Jan	0.6	Rich	0.015	0.001	0.014
Medium	1-Jan	0.6	World	0.310	0.289	0.021
Medium	1-Jan	0.6	Rich	0.052	0.047	0.005
Medium	1-Jan	0.5	World	0.310	0.319	-0.009
Medium	1-Jan	0.5	Rich	0.052	0.068	-0.016
Medium	1-Jul	0.6	World	0.450	0.328	0.122
Medium	1-Jul	0.6	Rich	0.111	0.081	0.031
High	1-Jan	0.6	World	0.489	0.468	0.021
High	1-Jan	0.6	Rich	0.103	0.108	-0.005

Which strategy is more effective for rich countries, paying for AV doses to be spread amongst poor countries in general, or paying for AV doses to be targeted to the outbreak source? Under what circumstances will welfare in rich countries be increased by undertaking this strategy? Figure 5.8 compares the effectiveness of purchasing doses for low income countries in general to the effectiveness of purchasing doses for the outbreak source country. Neither strategy is effective in reducing the number of cases in rich countries for a high virulence pandemic (P(T|C) = 0.06). But in all other cases, targeting the outbreak source is dramatically more cost effective than spreading doses throughout Poor or Lower middle income regions.

Figure 5.8: Rich country cases reduced per dose purchased for low income countries

Transition:	P(T|C)	Start date	Rich country cases reduced per extra AV dose purchased
Scenario A1 → B1	Low	1-Jan	0.74
Scenario A2 → B2	Medium	1-Jan	0.31
Scenario A3 → B3	Medium	1-Jul	0.33
Scenario A4 → B4	High	1-Jan	0.00
Scenario B1 → C1	Low	1-Jan	0.29
Scenario B2 → C2	Medium	1-Jan	0.22
Scenario B3 → C3	Medium	1-Jul	0.83
Scenario B4 → C4	High	1-Jan	-0.17
Scenario B1 → D1	Low	1-Jan	1.42
Scenario B2 → D2	Medium	1-Jan	0.52
Scenario B3 → D3	Medium	1-Jul	3.06
Scenario B4 → D4	High	1-Jan	-0.51

Note: AV efficacy is 0.6 in all cases.

Figure 5.8 suggests that providing AVs to developing countries in the event of a pandemic may pass a benefit-cost test. Although the percentage reduction in cases from providing AVs is small, millions of cases of the flu in rich countries would thereby be avoided, at a cost of 3-4 doses of antivirals per case avoided. Even at a cost of $25-$30 per course of treatment (including distribution costs), this would likely pass a benefit-cost test, even without any fatalities.[44] Sander et al. (2009) predict the economic cost of an

[44] Evidence on costs of AV is difficult to come by. Lokuge et al. (2006) report average stockpile acquisition costs per course for the USA, UK, Australia and Canada of $17-$50. Sander et al. (2009) assume a 20% markup to cover distribution costs. It is also unclear which costs should be considered; from a particular country's perspective we may wish to consider wholesale prices, but from a global welfare perspective we

influenza epidemic in the U.S. at $187 per person, based on a 50% attack rate. Keogh-Brown et al. (2010) suggest that a mild influenza pandemic in the UK (similar to the 1957 or 1968 pandemics) would reduce GDP by 0.58% over the course of year; a more severe pandemic (with a case fatality rate of 1%) would reduce GDP by 4.5% over the course of a year. Typical estimates of the value of a statistical life in rich countries are in the millions of dollars, and so even a very low but positive case fatality rate would lead to a large value from reducing cases.

v) The value of health infrastructure

The effectiveness of using AV to mitigate pandemics is a function of the effectiveness of the public health system infrastructure in each country, and its ability to rapidly and accurately identify infectious individuals and provide them with AV treatment. In the primary simulation runs, it is assumed that 50% of symptomatic cases (and thus 33.35% of infectious cases, since 33.3% of cases are asymptomatic) could be reached a single day after symptoms presented, in any country that had a stockpile. These assumptions are arguably too optimistic, particularly for developing countries, though they are more pessimistic than the assumptions used in some other global influenza papers[45]. By weakening these assumptions and reducing the proportion of symptomatic infectious cases who receive treatment, I can examine the sensitivity to these assumptions – and investigate the value of strengthening public health systems in poor countries.

---

should be concerned only with pure manufacturing and distribution costs, as the wholesale markup is a welfare transfer to drug manufacturers rather than a cost.

[45] For example Colizza et al. (2007) assume that 30, 50, or 70% of new infectious cases, can be treated.

Consider a weak health infrastructure scenario (Figure 5.9), where only 30% of symptomatic infectious cases can be treated in poor and lower middle income countries, 40% in upper middle income countries, and (as before) 50% in rich countries.

Figure 5.9 presents simulations based on a weak health infrastructure scenario in which only 30% of symptomatic infectious cases can be treated in poor and lower middle income countries, 40% in upper middle income countries, and (as before) 50% in rich countries. The effect of reducing the proportion who are treated is dramatic, both on the direct value of antiviral treatment in countries that have a stockpile, and on the external benefits to rich countries from providing antiviral doses to poorer countries. In terms of direct effects, the marginal benefit (in terms of world cases reduced) from an AV stockpile in upper middle and rich countries is reduced from 5.2 percentage points to 4.8 percentage points. The marginal benefit from adding a 1% stockpile to lower middle income countries is reduced from 8.4 percentage points to 2.5 percentage points. The marginal benefit from adding a 1% stockpile to poor countries is reduced from 5.6 percentage points to 1.3 percentage points.

Figure 5.9: Attack rate after 1 year, standard vs. reduced proportion of infectious treated, January 1 start date

	Scenario	**Strong health infrastructure**	**Weak health infrastructure**
World	Baseline (no AV)	0.446	0.446
	Scenario A	0.394	0.398
	Scenario B	0.310	0.373
	Scenario C	0.254	0.360
Poor	Baseline (no AV)	0.528	0.528
	Scenario A	0.529	0.529
	Scenario B	0.525	0.525
	Scenario C	0.458	0.518
Lower middle	Baseline (no AV)	0.401	0.401
	Scenario A	0.403	0.403
	Scenario B	0.169	0.336
	Scenario C	0.121	0.314
Upper middle	Baseline (no AV)	0.397	0.397
	Scenario A	0.210	0.254
	Scenario B	0.166	0.248
	Scenario C	0.051	0.224
Rich	Baseline (no AV)	0.325	0.325
	Scenario A	0.059	0.061
	Scenario B	0.052	0.058
	Scenario C	0.045	0.055

Note: Strong health infrastructure means 50% of new symptomatic infectious cases are treated (in regions that have an AV stockpile); weak health infrastructure means that 30% of cases are treated in poor and lower middle income regions, 40% in upper middle income regions, and 50% in rich regions.

The effect of reducing the proportion who are treated is dramatic, both on the direct value of antiviral treatment in countries that have a stockpile, and on the external benefits to rich countries from providing antiviral doses to poorer countries. In terms of direct effects, the marginal benefit (in terms of world cases reduced) from an AV stockpile in upper middle and rich countries is reduced from 5.2 percentage points to 4.8 percentage points. The marginal benefit from adding a 1% stockpile to lower middle income countries is reduced from 8.4 percentage points to 2.5 percentage points. The marginal benefit from adding a 1% stockpile to poor countries is reduced from 5.6 percentage points to 1.3 percentage points. The indirect effects are also significant. The marginal reduction in the attack rate in rich countries from adding a 1% stockpile for lower middle income countries falls from 0.7 percentage points to 0.3 percentage points. Similarly for adding a 1% stockpile to poor countries, the benefit falls from 0.7 to 0.3. These results follow from the basic "increasing returns" property of the core SIR model; with a lower percentage of people treated, the effect of AV on reducing the reproductive rate is diminished proportionally, which has a greater than proportional impact on the benefits of AV from reducing attack rtes.

These results suggest that if the "Reduced" proportion treated parameters are a more accurate description of the real world, then investments that increase the number of infectious people than can be treated in poorer countries will have large benefits, and will be complementary with policies that provide antiviral doses to poor countries.

Figure 5.10: Rich country cases reduced per AV dose purchased in poor countries, sensitivity to weak health infrastructure

Scenario Transition:	P(T|C)	Start date	Rich country cases reduced per extra AV dose purchased	
			Strong health infrastructure	Weak health infrastructure
Scenario A → B	Medium	1-Jan	0.31	0.12
Scenario B → C	Medium	1-Jan	0.22	0.09
Scenario B → D	Medium	1-Jan	0.52	0.21

Note: "4.2 in Indonesia" refers to the fact that the 9.14 million doses donated to Indonesia constitute 4.2% of Indonesia's population. AV efficacy is 0.6 in all cases. "Strong health infrastructure" means 50% of symptomatic infectious cases are treated in all income groups. "Weak health infrastructure" means 30% of symptomatic infectious cases are treated in Poor and Lower middle income group regions, 40% in Upper middle income group regions, and 50% in Rich regions.

vi) Allocating antivirals across countries

Discussions of global public health are often motivated by different perspectives on altruism and global fairness, as much as economic efficiency. But what are the epidemiological consequences of different views of fairness? Suppose that a global agency such as the World Health Organization was able to purchase a stockpile and distribute it over the course of a pandemic. If it did so, how should the agency distribute doses so as to maximize the number of cases prevented? Consider two extreme allocation rules that might be adopted under different conceptions of global fairness. One allocation would be to simply have a single global stockpile, that uses doses to treat infectious cases as they develop symptoms, no matter where in the world they are. This "patient-oriented" rule could be viewed as fair under a perspective that sees people as having an equal right to treatment. A second possibility is to evenly divide up the stockpile among regions in proportion to population size, so each region uses its share of

the stockpile to treat only its own citizens. This "nation-oriented" rule could be seen as fair under a moral view that each country has equal right to access to treatment, regardless of the whether the pandemic reaches them quickly or slowly.

What are the implications of these allocation rules? Consider each allocation rule for a global stockpile size equal to 1% or 2% of the world population (roughly 64 million doses or 128 million doses, respectively). The main practical differences between the allocation rules are the speed at which they disburse doses and the number of doses allocated to each country. Doses are disbursed faster in the early stage of the pandemic under the “patient-oriented” rule than under the “nation-oriented” rule, because people in regions that suffer many cases early on keep treating patients even after they would have consumed a stockpile equal to 1% or 2% of their population size. However, this also means that the global stockpile is exhausted more rapidly under the “patient-oriented” rule, and so this rule effectively transfers doses from countries where the pandemic is late to arrive to countries where the pandemic is early to arrive.

Figure 5.11 compares global attack rates across the two rules for the standard assumptions (P(T|C) = 0.05333, Jan 1 start date), the severe pandemic (P(T|C) = 0.06), the mild pandemic (P(T|C) = 0.045) and the July 1 start date (P(T|C) = 0.05333).

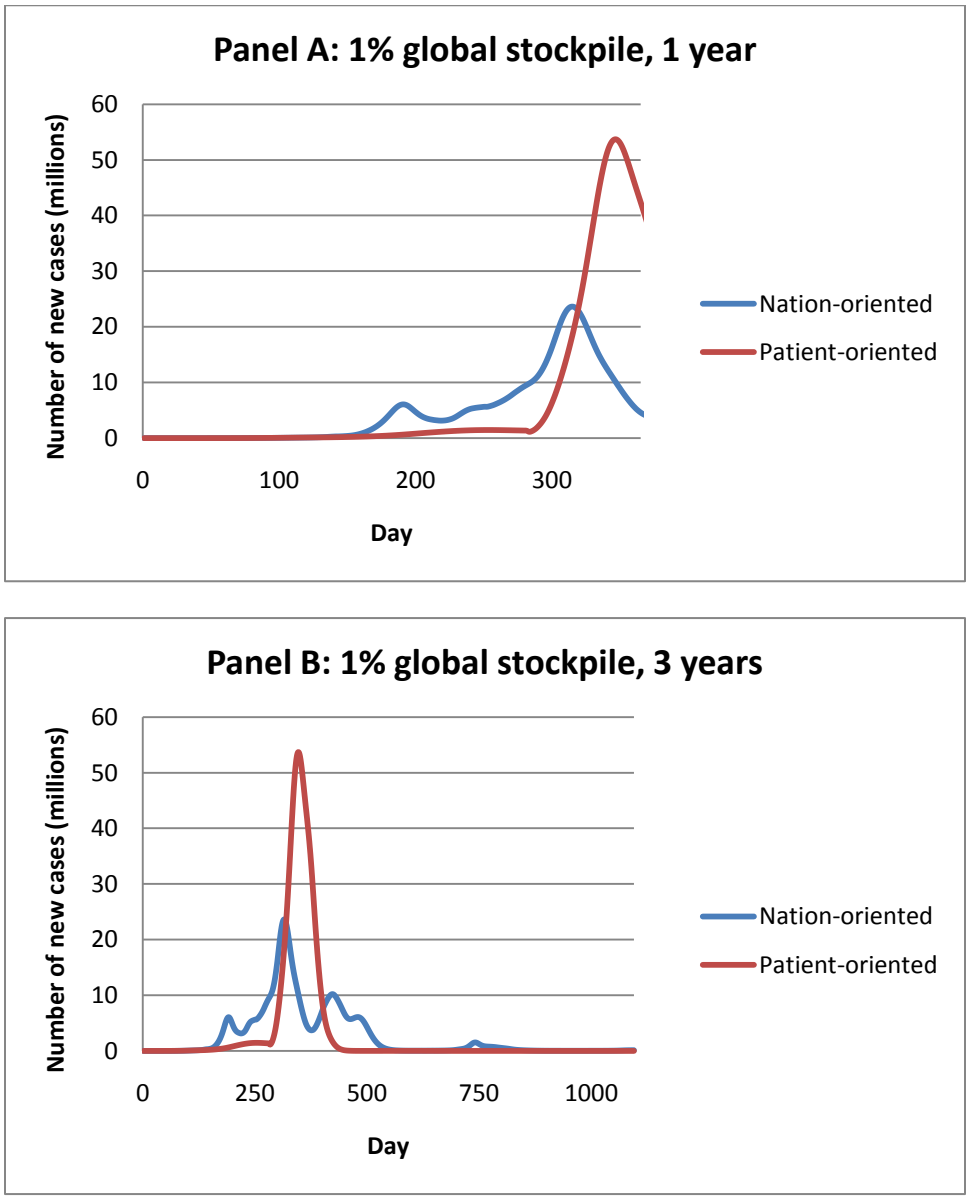
Figure 5.11: World attack rates under varying allocation rules

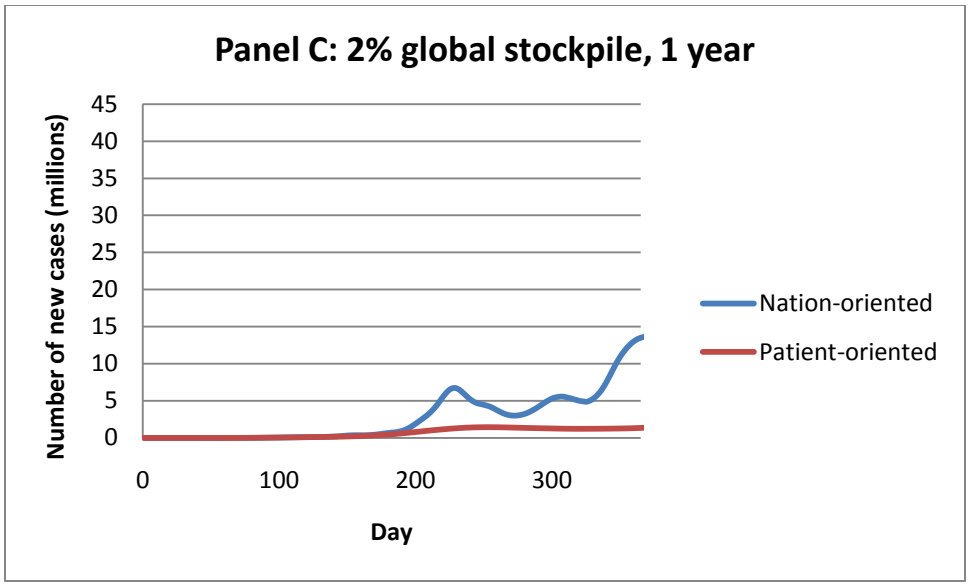
		1% global stockpile		2% global stockpile	
	Baseline (no AV)	Patient-oriented	Nation-oriented	Patient-oriented	Nation-oriented
Standard	0.446	0.389	0.281	0.037	0.152
High P(T|C)	0.533	0.589	0.509	0.601	0.409
Low P(T|C)	0.287	0.000	0.001	0.000	0.000
July 1 start date	0.547	0.412	0.369	0.210	0.227

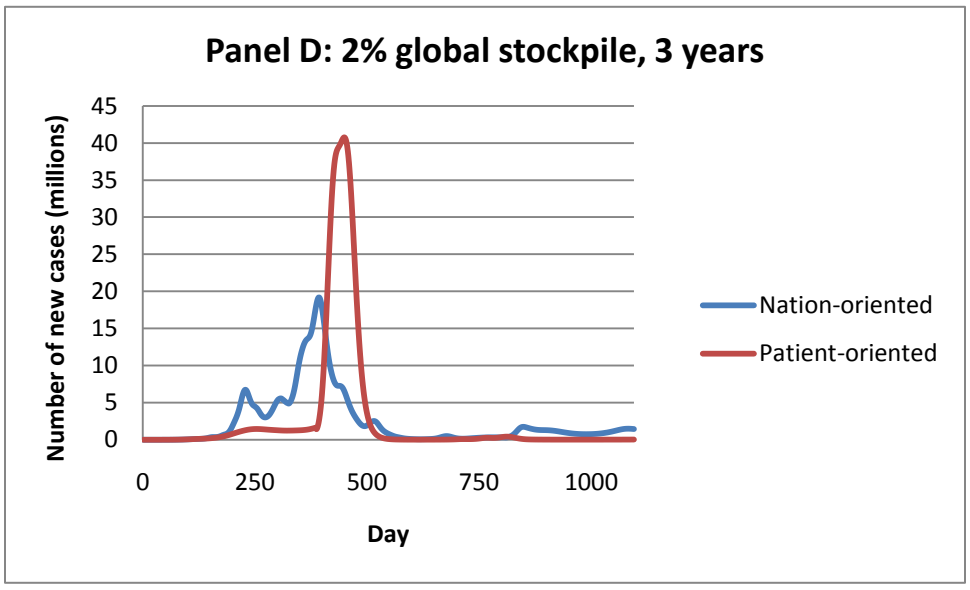
In the "Low" scenario, the pandemic is contained because AV doses are provided to the outbreak source under both rules, and this is sufficient to push the R_0 below 1. In most other cases, the nation-oriented rule performs better (ie leads to a lower attack rate after 1 year) than the patient-oriented rule. The reason for this is straightforward; under a patient-oriented rule, doses continue to be allocated to the outbreak country even after the pandemic has spread beyond the outbreak country. Those additional doses in the outbreak country have a negligible impact on the rest of the world, but they exhaust the stockpile rapidly, so there are many fewer doses to be used in other countries, and the number of global cases increases sharply as the global stockpile expires.

The patient-oriented method appears to give a better result for the standard case with the 2% stockpile, but this is highly sensitive to the 1-year cutoff date. Figure 5.12 demonstrates the impact of the 1-year cutoff under the two rules. The 1% stockpile is exhausted after 280 days under the patient-based rule. The 2% stockpile is not exhausted until day 370, which makes it superior for reducing the attack rate after 1 year, but once the stockpile is exhausted the pandemic spreads very rapidly, and beyond day 455 the attack rate is higher under the patient-based rule than under the nation-based rule.

Figure 5.12: World pandemic dynamics for Patient- vs. Nation-oriented allocation rules







The optimal policy depends on how long it will take to develop, manufacture, distribute, and administer an effective vaccine, and on the size of the stockpile (relative to the pandemic reproductive rate. If the global stockpile is large, then the rapid disbursement property of the patient-oriented method makes it superior; when exhaustion of the stockpile is less of a risk, then the faster disbursement rate is valuable. Fast disbursement is similarly advantageous if the appropriate time horizon (after which

vaccine is available) is short – for example, the attack rate after 9 months is lower under the patient-oriented rule for both 1% or 2% global stockpiles. But if the global stockpile is small or the relevant time horizon is longer, then the nation-oriented method is more desirable, because it ensures that limited AV uses are used to treat the early cases in every country. This is more effective in slowing the pandemic spread and reducing the one-year attack rate than continue to spend doses on new cases in the outbreak country even once the pandemic has spread throughout the world.

Chapter 6 Conclusions

Treating infectious individuals with antiviral drugs may be an effective method for mitigating the consequences of an influenza pandemic. While most of the benefits from a country choosing to treat its population with antivirals will accrue to that country, there are some positive externalities. This implies that, in a world in which poor countries are unlikely to hold stockpiles of antiviral drugs, it may be in the self-interest of wealthy countries to (collectively) pay for purchase and distribute antivirals to poor countries, even without any altruistic or humanitarian motivations.

The payoff to providing antiviral doses to poor countries is illustrated by a simple, two-country SIR model. In that model there are significant externalities and complementarities in antiviral treatment: when one country treats more of its population this both reduces the attack rate in the other country and increases the marginal benefit from additional treatment in the other country.

It would, however, be misleading to draw policy conclusions from a simple two-country model. In reality, influenza spreads through a complex network of air- and land-based travel. The spread of the flu and the effectiveness of treatment policies depend on seasonal factors—e.g., whether the flu peaks in Northern Hemisphere winter or Northern Hemisphere summer. And, the number and distribution of poor v. rich countries differs significantly from the symmetric two-region SIR model. The spread of the flu is simulated in a more descriptively realistic global epidemiological model in order to capture the impact of these features on the effectiveness of treatment policies.

Under the base case assumptions of moderate transmissibility of the flu, the distribution of antiviral stockpiles from rich countries to poor and lower middle income countries may indeed pay for itself: providing a stockpile equal to 1% of the population of poor countries will reduce cases in rich countries after 1 year by about 6.13 million cases at a cost of 4.62 doses per rich-country case avoided. Concentrating doses on the outbreak country is, however, even more cost-effective: in the base case it reduces the number of influenza cases by 4.76 million cases, at the cost of roughly 1.92 doses per case avoided.

These results are, however, dependent on the transmissibility of the flu, its effectiveness in reducing infection and on the proportion of infectious who can realistically be identified and treated. Simulations reveal that reducing the proportion of symptomatic infectious that can be treated from 50% (our base case assumption) to 40% in lower middle income countries and 30% in poor countries more than doubles the number of doses required to reduce a case of the flu in rich countries. Providing stockpiles to poor countries may still pass the "selfish" benefit-cost test. But the results suggest that improving the delivery of health services in poor countries will complement policies to treat pandemic flu, in addition to yielding other health benefits.

These results also focus on the use of antivirals alone. In practice, a variety of interventions may be adopted, including non-strain specific vaccination or non-pharmaceutical interventions such as face masks and gloves or school or workplace closures. These policies will tend to complement antiviral treatment, and should not be ignored when considering pandemic mitigation.

Appendix 1: One and Two-City SIR models reframed

Chapter 3 considered one and two city SIR models with antiviral policy interventions where the choice variable was the stockpile size. While this characterization is useful in providing background and intuition for the policy interventions applied in the full Global Epidemiological Model, it is difficult to characterize in an analytic fashion. An alternative characterization that is easier to analyze is to instead consider a model where the choice variable is the proportion of infectious cases who receive treatment p, and to assume that the planner has perfectly information and so purchases exactly enough doses to treat proportion p of infectious cases for the duration of the epidemic. Hence, there is no concept of a "stockpile" that can be exhausted, and the effect of applying antivirus is merely to reduce the effective reproductive rate for the full epidemic duration. The advantage of this characterization model is that the terminal values S(∞) and R(∞) can be observed from the transcendental equation, and so there is no need to run a simulation model.

The equations of motion for S(t), I(t) and R(t) remain unchanged from equations (7) through (9), but now rather than choosing a stockpile size P* a city chooses a proportion of people to treat $p \in [0, p^*]$, where p* is the maximum possible proportion of infectious cases that can be treated, determined by the level of health infrastructure. This means that the effective beta for the full duration of the pandemic is $\beta = (1 - pe)\beta^o$, and the number of people treated is $p[1-S(\infty)]N$. A city chooses p to maximize:

$$F(p) = V[S(\infty)]N - cp[1-S(\infty)]N \qquad \text{(A1)}$$

where S(∞,p) is defined by the transcendental equation $ln(S(\infty)) = R_0(S(\infty)-1)$,where $R_0 = \beta^o(1-pe)/\delta$.

For values of R_0 that characterize pandemic flu, the p that maximizes equation F(p) is a corner solution: either p=0 or p=1. First-order conditions to the maximization problem call for p to be chosen to the point where the marginal benefits of increasing $S(\infty)$ both in terms of reducing the number never infected and the number of doses needed to treat the infectious, equal or exceed the cost of an additional dose,

$$dF(p)/dp = (V + cp)[dS(\infty)/dp] - c[1- S(\infty)] \geq 0 \qquad \text{(A2)}$$

The second-order conditions for an interior maximum, $d^2F(p)/dp^2 < 0$, are, however, not satisfied for pandemic flu.[46] Equation (6) implies that the proportion of people never infected, $S(\infty)$, increases as R_0 falls (i.e., as p is increased); however, for values of $R_0 >$ 1.05, increasing p increases $S(\infty)$ at an increasing rate.[47] Influenza pandemics have generally been characterized by reproductive rates greater than 1.5.[48] This implies that p = 0 maximizes (5) if V is low relative to c, but p = 1 maximizes (5) if V is high relative to c. To illustrate, when $\beta° = 0.3$, $\delta = 0.2$ and $e = 0.2$, setting $p = 1$ is optimal if $V/c > 1.2$.[49]

[46] $d^2F(p)/dp^2 = (V + cp)[d^2S(\infty)/dp^2] + c[dS(\infty)/dp]$ which need not be negative.

[47] A sufficient condition for $d^2F(p)/dp^2 > 0$ is $d^2S(\infty)/dR_0^2 > 0$. Formally, $dS(\infty)/dR_0 = [S(\infty)-1]/[S(\infty)^{-1} - R_0]$ and
$d^2S(\infty)/dR_0^2 = [S(\infty)^{-1} - R_0]^{-1} [dS(\infty)/dR_0] [2 + S(\infty)^{-2}[dS(\infty)/dR_0]]$.
This second derivative $d^2S(\infty)/dR_0^2 > 0$ provided $2 + S(\infty)^{-2}[dS(\infty)/dR_0] < 0$. Equation (6) implies that this condition is satisfied if $R_0 > 1.05$.

[48] The R_0 for the 1957 Asian flu has been estimated at 1.8 (Vynnycky and Edmunds 2008). Vynnycky et al. (2007) estimate an R_0 of 2.4-4.3 for the 1918 Spanish flu.

[49] If we considered the case where e = 0.4 then p = 1 would nearly always be optimal, because p = 1 and e = 0.4 imply $R_0^{eff} = 0.9$, and so no pandemic occurs.

This model with choice variable p can be expanded to the two-city case in a similar fashion to the stockpile size model. The qualitative results are similar, and are dependent on the particular parameter values. Appendix table 1 demonstrates the attack rates that result from this model for various input parameters.

Three points about Appendix Table 1 deserve emphasis: (1) when only one city treats its infectious, the effectiveness of its treatment is less than in the one-city case where there is no travel; (2) when only one city treats its infectious, the external benefits to the no-treatment city are small; (3) when both cities treat, the effectiveness of each city's stockpile is greater than when only one city employs a treatment strategy. When city A treats 60% of its infectious in isolation, the attack rate is reduced from 58% to 24%. In the two-city case, the attack rate is reduced from 59% to 30%. People from city B continue to re-infect city A, making city A's treatment of its infectious less effective. The benefits to city B of A's stockpile are, however, small: when city B has no stockpile, its attack rate is reduced by only 1.6 percentage points by virtue of the fact that A treats its infectious. When both cities treat their infectious, the benefits to both cities increase, compared to the case in which each city acts in isolation. When both cities treat, the attack rate is reduced from about 58% to 24% in each city, as in the single city case. (In effect, the two cities have become a single city, employing the same treatment as in the single-city case.)

The magnitude of the externalities and complementarities observed in Appendix Table 1 are affected by the transmissibility of the flu (β^0/δ), the proportion of infectious who can be treated (p^*) and by the effectiveness of treatment (e). The benefits to the no-treatment city of antivirals in the treatment city are greater the less transmissible is the

flu, as are the complementarities of treatment (see Table 1). Treatment complementarities are also greater the higher the proportion of infectious who can be treated (p^*) and the greater the efficacy of treatment (the higher is e).

With appropriate choice of V and c, it is possible to generate results where the Nash equilibrium is ($p_A = 0$, $p_B = p^*$) while the socially optimal solution is ($p_A = p^*$, $p_B = p^*$) and yet there exists a pareto improvement where city B is made better off by paying for the costs of increasing p_A from zero to p*. For example, consider the case when $V_A = 3$, $V_B = 100$ and $c = 10$. The Nash equilibrium is $p_A = 0$, $p_B = 0.6$, whereas the treatment strategy that maximizes the sum of cases avoided minus treatment costs for the two cities combined is $p_A = p_B = 0.6$. It is also the case that it pays city B to pay for antivirals sufficient to treat 60% of infectious in city A. The cost of the treatment ($10*(0.6)*0.235N$) is less than the value of the resulting reduction in city B's influenza cases ($100*0.061N$).

Appendix Table 1: Results from 2-city model with proportion treated as choice variable

β°	e	**α**	p_A	p_B	Attack rate in City A	Attack rate in City B
0.3	0.4	0.01	0	0	0.588	0.579
			0.6	0	0.301	0.563
			0	0.6	0.565	0.296
			0.6	0.6	0.240	0.235
0.35	0.4	0.01	0	0	0.719	0.707
			0.6	0	0.475	0.705
			0	0.6	0.709	0.468
			0.6	0.6	0.456	0.449
0.25	0.4	0.01	0	0	0.376	0.370
			0.6	0	0.111	0.326
			0	0.6	0.329	0.107
			0.6	0.6	0.012	0.005
0.3	0.3	0.01	0	0	0.588	0.579
			0.6	0	0.381	0.569
			0	0.6	0.573	0.376
			0.6	0.6	0.353	0.348
0.3	0.5	0.01	0	0	0.588	0.579
			0.6	0	0.224	0.555
			0	0.6	0.557	0.220
			0.6	0.6	0.090	0.083
0.3	0.4	0.01	0	0	0.588	0.579
			0.7	0	0.249	0.558
			0	0.7	0.559	0.244
			0.7	0.7	0.145	0.138
0.3	0.4	0.01	0	0	0.588	0.579
			0.5	0	0.355	0.567
			0	0.5	0.570	0.349
			0.5	0.5	0.318	0.313
0.3	0.5	0.015	0	0	0.587	0.580
			0.6	0	0.318	0.555
			0	0.6	0.557	0.314
			0.6	0.6	0.239	0.235
0.25	0.5	0.005	0	0	0.588	0.579
			0.6	0	0.278	0.572
			0	0.6	0.574	0.271
			0.6	0.6	0.240	0.234

Appendix 2: List of Regions in the GEM, by Income Group

Regions in the Poor Income Group:

Bangladesh, Bolivia, Cambodia, Egypt, Georgia, India, Indonesia, Kyrgyz Republic, Madagascar, Malawi, Morocco, Mozambique, Nicaragua, Nigeria, Pakistan, Paraguay, Philippines, Senegal, Sri Lanka, Tanzania, Uganda, Viet Nam, Zambia, Zimbabwe, Rest of Central Africa, Rest of East Asia, Rest of Eastern Africa, Rest of Former Soviet Union, Rest of Oceania, Rest of South African Customs Union, Rest of South America, Rest of South Asia, Rest of South-Central Africa, Rest of Southeast Asia, Rest of Western Africa,

Regions in the Lower Middle Income Group:

Albania, Argentina, Armenia, Azerbaijan, Botswana, Brazil, Bulgaria, China, Colombia, Ecuador, Iran, Kazakhstan, Malaysia, Mauritius, Peru, Romania, South Africa, Thailand, Tunisia, Ukraine, Rest of Central America, Rest of Eastern Europe, Rest of Europe, Rest of North Africa, Rest of Western Asia

Regions in the Upper Middle Income Group:

Chile, Croatia, Estonia, Hungary, Latvia, Lithuania, Mexico, Poland, Republic of Korea, Russian Federation, Slovak Republic, Taiwan, Turkey, Uruguay, Venezuela, Rest of the Caribbean

Regions in the Rich Income Group:

Australia, Austria, Belgium, Canada, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong S.A.R. of China, Iceland, Ireland, Israel and Arabia, Italy, Japan, Luxembourg, Malta, Netherlands, New Zealand, Portugal, Singapore, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States of America, Rest of EFTA, Rest of North America

Appendix 3: Deriving the daily infectious contact rate

This appendix describes the derivation of the matrix of contact rates, and hence the daily infectious contact rates λ. The goal is to derive contact rates for each age-group in each city based on real world data from Europe, and attempt to adjust the European data to fit other cities by accounting for differences in age structure and density.

First define the 4x4 contact rate matrix C_i. The elements C_{ijk} describe the number of contacts per day that an Infectious person of age-group *j* has per day with individuals in age-group *k* in city *i*. Following Hethcote and Yorke (1984) and Over and Piot (1993), decompose this matrix into two parts: a 4x4 mixing matrix M_i and a 4x1 contact vector A_i, such that :

$$C_i = M_i \cdot I_4 \cdot A_i \qquad \text{(A3)}$$

where I_4 is the 4x4 identity matrix.

The vector A_i gives the absolute number of contacts per day for each agegroup; so for example $A_{Jakarta}' = [7.29, 9.45, 6.41, 3.98]$ means that in Jakarta, each Infectious case with a=0 has 7.29 contacts per day, each Infectious case with a=1 has 9.45 contacts per day, and so forth. Denote the 4x1 vector n_i as the proportion of the population that are of each age category, so the elements of n_i sum to 1. The total number of contacts per person per day in city *i* is then the product $n_i'A_i$. For example, in Indonesia, 9.64% of the population are age0, 18.75% are age1, 66.09% are age2 and 5.52% are age3, so the total contacts per day is the product $n_{Jakarta}'A_{Jakarta} = 6.94$.

The mixing matrix M_i describes the relative mixing rates between age-groups. Each row of the M matrix sums to 1, so the element M_{ijk} describes the proportion of an individual of age-group *j*'s contacts that they have with age-group *k*. For example, if the *j*th row of M_i were [0.25, 0.25, 0.25, 0.25] then age-group *j* individuals in city *i* would contact equal numbers of all four age-groups each day.

C, M and A have subscripts *i* because these vary across cities, though all cities within a region are assumed to have the same values of C, M, A and n. Values for A and M are generated based on mixing data from Mossong et al. (2008) and from data on population density and urbanization rates. Mossong et al. measure self-reported physical contact rate matrices for eight European countries for ten age categories from self-reported contacts across 7,290 individuals.[50] Their data provides the equivalent of a C-matrix for each of the 8 countries. The goal is to use these to extract an underlying set of "core behavior" that can be assumed to hold in all countries, and will allow reconstruction of a C-matrix for every region in the GEM.

The Mossong data is used directly to estimate the absolute number of contacts in each city, and the relative mixing rates of one age-group with another. A-vectors are defined for each region such that n'A = 6.94 in every region – where 6.94 is the weighted-average absolute contact per day across the 8 Mossong countries and across all age-groups.[51]

[50] Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, Netherlands and Poland.

[51] This crucial assumption, made for simplicity, implies that there is no variation in total contacts per day by age-structure; it will not be the case that the pandemic spreads faster in countries with younger populations merely because young people have higher contact rates. An alternative structure that would have this property could be generated by fixing the individual elements A_i across countries, but that would lead to very large variation in contact rates (and thus attack rates) across countries.

But I cannot simply work directly with the C or M matrices from the Mossong data; the C and M matrices are in part a product of the particular age structure of the country they are observed in, and age-structures vary across countries. For example, consider the hypothetical example above, where the jth row of M_i = [0.25, 0.25, 0.25, 0.25], and suppose that all 4 age-groups in city i are of equal size (ie n_i' = [0.25,0.25,0.25,0.25]). This represents fully random mixing, every person has an equal probability of contacting any other person, regardless of age-category. Now, consider some other city y with the same underlying behavior (fully random mixing), but with a younger population; suppose n_y' = [0.35, 0.25, 0.25, 0.15]. Clearly, it cannot be the case that the jth row of M_y =[0.25,0.25,0.25,0.25], because that would imply that a person of age j was contacting the same number age1 and age4 people in both cities, despite the fact that there are many fewer age3 people available to contact in city *y* and many more age1 people. So the confounding effect of the relative age-group sizes needs to be separated out.

An M matrix for each region by is estimated by assuming a structural form for the matrix, and then using the 8 Mossong matrices to estimate an underlying mixing matrix stripped of the country-specific n-vectors. Following Hethcote and Yorke, define a 4x1 vector B_i:

$$B_{ij} = \frac{A_{ij} n_{ij}}{\sum_{m=1}^{4} A_{im} n_{im}} \qquad \text{(A4)}$$

And a 4x4 matrix G and G whose elements are defined by:

$$B_{ij}\left(1 - G_{ijk}\right) = M_{ijk} \qquad \text{(A5)}$$

where G_{ijk} is the (j,k) element of the B matrix for city *i*, and similarly for M. B_{ij} is the jth element of the B-vector in city *i*, and can be interpreted as the proportion of contacts that would be with age-group j city *i* under random mixing. The G-matrix is what allows us to move away from random mixing, and captures the more complex structure where people can have differential mixing rates across age-groups.

Assume a structural form for the G-matrix where $G_{ijk} = G_{ij}$, i.e. all elements of the same row have the same value.[52] That is, each age-group is assumed to have the same relative preference for mixing with its own age-group as for all other age-groups. Estimate the G-matrix that minimizes the sum of squared residuals across the 8 Mossong countries to find an "optimal" G-matrix G*. M, and hence C matrices, for every region in the model are then reconstructed by substituting G* and the region-specific age structures from real world population data into the equations above.

Contact rates are adjusted by adding a semi-arbitrary adjustment for population density, recognizing that areas with higher population density have higher contact rates. Define the "residual urbanization rate" of a region as the urbanization rate of the "rural" population that is not a member of the 289 cities explicitly included included in the model. For example, Indonesia has an urbanization rate of 48.1%, but when the populations of Bandung, Denpasar, Jakarta, Medan and Surabaya (the Indonesian cities in the model) are excluded, the residual urbanization rate is 41.5%.

Every city and rural area is assigned to one of four categories, each with a relative density factor d_i; d = 1 for rural areas with residual urbanization <40%, d=1.1 for rural

[52] Tested alternative forms showed that moving from a 16-unique-element G-matrix to a 4-unique element G-matrix had very little impact on the fit to the Mossong data, but that moving from a 4-element G-matrix to a 1-element G-matrix significantly worsened the fit.

areas with residual urbanization >40%, d = 1.3 for low density urban areas (cities with density < 1,000 people per square kilometer), and d = 1.4 for urban area low density (for cities with density > 1,000 per square kilometer). In general, "rural" areas in high income countries have high residual urbanization while those in low income countries have low residual urbanization, and cities in high income countries have low population densities while those in low income countries have high population densities.

Finally, the daily infectious contact rate is λ given by:

$$\lambda_{ijk} = d_i \cdot \sum_{a=0}^{3} C_{iak} \cdot P(T|C) \qquad \text{(A6)}$$

where the values of the C-matrix for each city are derived from the Mossong data as described above. This leaves us with the infectious contact rate dependent on one free parameter input (the probability of transmission given contact), which can then be varied to explore moderate, mild and severe pandemic strains.

References

Baroyan, O. V., Mironov, G. A., and L. A. Rvachev (1981). An algorithm modeling global epidemics of mutant origin. *Programming and Computer Software* 6(5), 272--277.

Barrett, S. (2003). Global Disease Eradication. *Journal of the European Economics Association,* 1, 2-3, pp.591-600

Boulier, B.L., et al. (2007). Vaccination Externalities. *B.E. Journal of Economic Analysis & Policy*, 7, 1, 23

Colizza, V., et al. (2007). Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Med* 4(1), e13.

Cooper, B. S., et al. (2006). Delaying the international spread of pandemic influenza. *Public Library of Science Medicine*. 3(6), e212

Epstein, J.M., et al. (2007). Controlling pandemic flu: The value of international air travel restrictions. *PLoS ONE* 2(5), e401.

Ferguson, N. M., et al. (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*. 437, 209-214.

Ferguson, N. M., et al. (2006). Strategies for mitigating an influenza pandemic. *Nature*, advance online publication,26 April 2006, (doi: 10.1038/nature04795).

Francis, P.J. (2004). Optimal tax/subsidy combinations for the flu season, *Journal of Economic Dynamics and Control,* 28, pp. 2037-2054.

Germann, T. C., et al. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*. 103(15), 5935-5940.

Glezen W. P., (1996). Emerging infections: pandemic influenza. *Epidemiology Review,* 18, 64–76.

Grais, R. F., Ellis, J. H., and Glass, G. E. (2003). Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology,* 18, 1065-1072.

Guimerà**,** R. et al. (2005). *The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proceedings of the National Academy of Sciences.* 102, 7794-7799.

Hajdin, C. et al. (2009). Stochastic Equation-Based Model of a Global Epidemic (Version 3.0). RTI International, May 2009.

Hethcote, H. W. and Yorke, J. A. (1984). Gonorrhea transmission dynamics and control. *Lecture Notes in Biomathematics*, vol. 56. Berlin, Germany: Springer

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

Keogh-Brown, M. R. et al. (2010). The Possible Macroeconomic Impact on the UK of an Influenza Pandemic. *Health Economics.* 19,1345-1360.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond. A* 115, 700-721.

Lokuge, B., Drahos, P. and Neville, W. (2006). Pandemics, antiviral stockpiles and biosecurity in Australia: what about the generic option? *Medical Journal of Australia*, 184(1), pp.16-20

Longini, I. M. Jr., et al., (2005). Containing pandemic influenza at the source. *Science,* 309, 1083-1087.

Lowen, A.C., et al. (2008). High temperature (30°C) blocks aerosol but not contact transmission of influenza virus. *Journal of Virology*, 82 (11), 5650-5652

Mossong, J. et al. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine.* 5, 381-91.

Over, M., and Piot, P. (1993). HIV Infection and Sexually Transmitted Disease, in D.T. Jamison and others (eds.), *Disease Control Priorities in Developing Countries*. New York: Oxford University Press.

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

Sander, B., et al. (2009). Economic evaluation of influenza pandemic mitigation strategies in the United States using a stochastic microsimulation transmission model. *Value in Health*, 12(2), 226-233.

Shaman, J. and Kohn, M. (2009). "Absolute humidity modulates influenza survival, transmission, and seasonality", *Proceedings of the National Academy of Sciences*, 106, pp. 3243-3248

United Nations, (2007). World Urbanization Prospects: The 2007 Revision, New York.

Vynnycky, E., et al. (2007). Estimates of the reproduction numbers of Spanish influenza using morbidity data. *International Journal of Epidemiology. 36, 881-889*

Vynnycky, E. and Edmunds, W.J,. (2008). *Epidemiol.Infect.*, 136, 166-179

Yang, Y. et al. (2009). The Transmissibility and Control of Pandemic Influenza A (H1N1) Virus, *Science*, 326, 5953, pp. 729-733