Models for managing surge capacity in the face of an influenza epidemic

Ana Cecilia Zenteno Langle

Submitted in partial fulfillment of the

Requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Daniel Bienstock, Sponsor
(Industrial Engineering and Operations Research)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Garud Iyengar, Chair
(Industrial Engineering and Operations Research)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Van-Anh Truong, 2nd reader
(Industrial Engineering and Operations Research)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Jay Sethuraman, 3rd reader
(Industrial Engineering and Operations Research)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Raul Rabadan, Outside Examiner
(Biomedical Informatics)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____
Y. Claire Wang, Outside Examiner
(Health Policy and Management)

Approved for the Department of Industrial Engineering and Operations Research:

_____
Clifford Stein
Department Chair

ABSTRACT

Models for managing surge capacity in the face of an influenza epidemic

Ana Cecilia Zenteno Langle

Influenza pandemics pose an imminent risk to society. Yearly outbreaks already represent heavy social and economic burdens to society. A pandemic could severely affect infrastructure and commerce through high absenteeism, supply chain disruptions, and other effects over an extended and uncertain period of time. Governmental institutions such as the Center for Disease Prevention and Control (CDC) and the U.S. Department of Health and Human Services (HHS) have issued guidelines on how to prepare for a potential pandemic, however much work still needs to be done in order to meet them. from a planner's perspective, the complexity of outlining plans to manage future resources during an epidemic stems from the uncertainty of how severe the epidemic will be. Uncertainty in parameters such as the contagion rate (how fast the disease spreads) makes the course and severity of the epidemic unforeseeable, exposing any planning strategy to a potentially wasteful allocation of resources.

In this thesis we consider robust models of surge capacity planning. We focus on surge staff deployment strategies that aim to mitigate the impact of an influenza epidemic on an organization's operations.

Our approach involves the use of additional resources in response to a robust model of the evolution of the epidemic as to hedge against the uncertainty in its evolution and intensity. Under existing plans, large cities would make use of networks of volunteers, students, and recent retirees, or "borrow" staff from neighboring communities. Taking into account that such additional resources are likely to be significantly constrained (e.g. in quantity and duration), we seek to produce robust emergency staff commitment levels that work *well* under different trajectories and degrees of severity of the pandemic.

Our methodology combines Robust Optimization techniques with Epidemiology (SEIR models) and system performance modeling. We describe cutting-plane algorithms analogous to generalized Benders' decomposition that prove fast and numerically accurate. Our results yield insights on the structure of optimal robust strategies and on practical rules-of-thumb that can be deployed during the epidemic. To assess the efficacy of our solutions, we study their performance under different scenarios and compare them against other seemingly good strategies through numerical experiments.

This work would be particularly valuable for institutions that provide public services, whose operations continuity is critical for a community, especially in view of an event of this caliber. As far as we know, this is the first time this problem is addressed in a rigorous way; particularly we are not aware of any other robust optimization applications in epidemiology.

Table of Contents

Table of Contents

List of Figures

Acknowledgments

I would like to start by thanking my advisor, Prof. Daniel Bienstock, whose support was a key element in the development of this thesis. Not only is he smart, he is very encouraging, funny, extremely patient, and he truly cares about his students' wellbeing.

Other professors to whom I am indebted for their contribution to my personal and professional development are Professors Garud Iyengar, Ward Whitt, and Don Goldfarb. I would also like to thank Professor Maria Chudnovsky; I deeply enjoyed working as her Teaching Assistant for four consecutive years.

I am also grateful to the department staff, who was always happy to answer my numerous questions. Donella, Maria, Jaya, Jenny, Risa, Darbi, Aysha, Shi Yee, Adina; they were always helpful and responsive!

My most sincere appreciation goes for all the good friends I met these past five years. They made the good times merrier and the hard times easier to overcome, becoming an essential part of my life in New York. It is impossible to quantify how much I have learned from them. Special thanks go to my officemates at Mudd 313A and at Schapiro 821, with whom I shared so many hours of my Ph.D. Running the terrible risk of forgetting someone, I want to mention Yori, Romain, Rodrigo, Dani, Silvi,

Acknowledgments

Vero, Majo, Jing Dong, Song-hee, María de la Garza, Elia, Colin; I will always cherish all the moments we shared. Very special thanks go to Angel Almada and Raúl González, for being always with me, no matter the distance. I am indebted to Angel for his constant push to go beyond what I think it is possible. I thank Yixitín for watching my back for 4 precious years (literally and figuratively). I am ever grateful to Bar Ifrach, for the long-lasting friendship we built from the first recitation we attended; his caring and honest advice will be with me always. I am especially beholden to Rouba Ibrahim for her constant kindness and most encouraging support; she must be the most warm hearted person I know.

I owe my deepest gratitude to my family in Mexico, particularly to my mom, my dad, Dany, and Raúl. Being far away from them continues to be one of the toughest challenges I have faced. Their love and support have been pivotal during the whole of my graduate studies. Their hard work and generosity will forever be a source of inspiration.

This thesis would have never been possible without Filippo Balestrieri, to whom I owe more than any words can describe. I love him with all my being. I dedicate this work to him.

1

Introduction

In this paper we consider robust models for emergency staff deployment in the event of a flu pandemic. We focus on managing critical staff levels at organizations that must remain operational during such an event, and develop methodologies for managing emergency resources with the goal of minimizing the impact of the pandemic. We present numerical experiments using realistic data to study the effectiveness of our approach.

A serious flu epidemic or pandemic, particularly one characterized by high contagion rate, would have extremely damaging impact on a large, dense population center. The 1918 influenza pandemic is often seen as a worst-case scenario as it arguably represents the most devastating pandemic in recent history, having killed more than 20 million people worldwide **[20, 53, 71]**. However, even a much milder epidemic would have vast social impact as services such as health care, police and utilities became severely hampered by staff shortages. Workplace absenteeism might also become a serious

concern [72]. In the United States, the Implementation Plan for the National Strategy for Pandemic Influenza foresee absenteeism levels as high as 40% at the height of the pandemic wave ([40]). In the health care setting, there are mixed views: While some preparedness plans project high absenteeism due to illness or need to care for family members ([57]), the opposite may also take place: health care workers may reportedly avoid calling in sick during an emergency as observed during the last H1N1 outbreak [27, 16].

In this study we focus on managing the inevitable staff shortfall that will take place in the case of a severe epidemic. We take the viewpoint of an organization that seeks to diminish decreased performance in its operations as the epidemic unfolds, by appropriately deploying available resources, but which is not directly attempting to control the number of people that become infected. Some examples of infrastructure of critical social value we are interested in are hospitals, police departments, power plants and supply chains; these are entities that must remain operative even as staff levels become low. In cases such as police departments, staff would likely be more exposed to the epidemic than the general public and (particularly if vaccines are in short supply, or apply to the wrong virus mutation) shortfalls may take place just when there is greatest demand for services. Power plants and waste water treatment plants are examples of facilities whose operation will be degraded as staff falls short and which probably require minimum staff levels to operate at all [39]. Supply chains would very likely be significantly slowed down as their staff is depleted, resulting, for example, in food shortages [66]. In all these cases, organizations cannot implement "work from home" strategies as urged for private businesses by the Centers of Disease Control and Prevention (CDC) and the United States Department of Labor [63].

In contrast to our focus, much valuable research has been directed at addressing the epidemic itself.

Such work studies public health measures that would reduce the epidemic's severity and its direct impact, for example by managing the supply of vaccines and antivirals (see **[50, 38, 51]**). While we do not address this topic, it seems plausible that robust optimization techniques could be applied in these settings, as well. To the best of our knowledge this is the first time that these methodologies are introduced into this research area.

A pandemic contingency plan for a large organization (such as a city government) would include resorting to emergency (or "surge") sources for additional staff: for example by temporarily relying on personnel from outlying, less dense, communities. Such additional resources are likely to be significantly constrained in quantity, duration and rate, among other factors. Such emergency staff deployment plans would entail some complexity in design, calibration and implementation, but as a result of other disasters such, as Hurricane Katrina in 2005 and the anthrax attacks in 2001, it is now agreed that there is a compelling need for emergency planning **[30]**.

From a planner's perspective, the task of managing future resource levels during an epidemic is complex, partly because of uncertainties regarding the behavior of the epidemic, in particular, uncertainty in the contagion rate. The evolution of the contagion rate is a function of poorly understood dynamics in the mutations of the different strains of the flu virus and environmental agents such as weather **[20, 52]**. Additionally, non-pharmaceutical public health interventions such as quarantine and social distancing could impose significant changes in social contact patterns that would in turn affect the contagion rate. In addition to uncertainty, a decision-maker will also likely be constrained by logistics. In particular, it may prove impossible to carry out large changes to staff deployment plans on short notice, particularly if such staff is also in demand by other organizations (as might be the case during a severe epidemic). We will return to these issues in Sections 5.2.2 and 8. As a consequence of

the two factors (uncertainty, and logistical constraints) a decision-maker may commit too few or too many resources - in this case emergency staff- or perhaps at the wrong time, if there turns out to be a mismatch between the anticipated level of staff shortage and what actually transpires (after the resources have been committed).

We present models and methodology for developing emergency staff deployment levels which optimally hedge against the uncertainty in the evolution an the epidemic while accounting for operational constraints. Our approach overlays adversarial models on the classical SEIR model for describing epidemics to characterize the potentially wide variability of the contagion rate. The resulting robust optimization models are non-convex and large-scale; we present convex approximations and algorithms that empirically prove numerically accurate and efficient, and we study their behavior and the policies they produce under a range of scenarios.

This thesis is organized as follows. Chapter 2 contains preliminaries to the main topics addressed in this work. Issues related to surge capacity planning are described in Chapter 3; Chapter 4 describes the classical SEIR model and makes the case for robustness. Chapter 5 contains the description of our robust model and Chapter 6 presents the results of our numerical experiments; details of the robust algorithm are presented in Chapter 7. We discuss extensions and give final remarks in Chapters 8 and 9, respectively.

2

Preliminaries

2.1   Operations Research and Public Health

Operations Research is the science of better decision making. It provides a structured framework to analyze complex systems by capturing their main uncertainties and interactions. While the field originally had military applications, it is now prevalent in supply chain management, transportation, services and, more recently, homeland security and health care management.

Public Health studies how to protect and improve the health of communities through education, promotion of healthy lifestyles, and research for disease and injury prevention [81]. Typical Public Health programs include disease screening and surveillance (such as HIV and influenza), vaccination, outbreak investigation (SARS), inspection and standard enforcement at public establishments, environmental monitoring, and vector control (mosquitoes, ticks that transmit diseases) [41].

Like any other services, Public Health programs require adequate design and effective implementation to have the desired impact. Herein lie many opportunities to apply Operations Research principles and techniques: the delineation, evaluation and operation of these services may benefit significantly from interweaving theoretical insight with practical experience which no health worker can afford to overlook [1]. Operations Research has proved to be successful in all steps of this process; it is significantly helpful in decision making with limited resources under uncertainty and in gauging the potential impact of various programs. These are all common elements of Public Health policy design.

In the case of epidemiology, public health concerns are focused on disease surveillance, prevention services, on the design and delivery of health programs and on the evaluation of such interventions. Questions of interest include how to reduce the final size of an outbreak of an infectious disease; what is the best combination of prevention strategies - vaccination, quarantine, social-distancing - and how to implement such measures. On the health economics side, an important matter is how to reduce the monetary and societal costs of these events due to the loss of productivity and the related business and community disruptions [41].

2.2 Influenza virus and pandemics

Influenza is an acute, highly contagious respiratory disease caused by a number of different virus strains. According to the CDC, annual outbreaks cause an average of 23,600 deaths and more than 200,000 hospitalizations in the U.S. [77]. These outbreaks are considerably costly as well: based on 2003 US population data, Molinari et al. estimate the total economic burden of annual influenza epidemics to be $87.1 billion [54]. It is most often a mild viral infection from which people usually recover within one or two weeks without requiring medical treatment; however, it may evolve into

lethal complications like pneumonia due to secondary bacterial infections, imposing a heavy medical burden. For certain virus strains this is especially pronounced in the case of susceptible groups such as young children, the elderly or people with certain medical preconditions.

In nature, the influenza virus can be found in wild aquatic birds, who are typically not harmed by it. However, it can jump from wild to domesticated ducks and then to chickens, from where it can infect pigs. Pigs can be infected by avian flu and the types that infect humans. In rural settings where humans, chickens, and pigs are all in close contact, pigs act as an influenza virus mixing bowl. Such virus can sometimes make a further jump from swine to people **[64]**.

There are three types of influenza viruses based upon their protein composition: A, B, and C **[64]**. Type A viruses are found in humans and in many kinds of animals including ducks, chickens, pigs, and whales. Type B mainly circulates in humans. Type C has been found in humans and animals like pigs and dogs, but it does not spread as fast as to cause an epidemic. Type A influenza subtypes have been catalogued according to two different protein components, also known as antigens, that are found on the virus surface: haemagglutinin (H) and neuraminidase (N). There are no type B or C subtypes.

Viral genomes are constantly mutating, producing new forms of these antigens. Whenever the mutation is significantly different, the human immune system can no longer recognize the virus, making people who have had influenza in the past lose their immunity to the new strain. Needless to say, vaccines against the original virus will also become less effective.

Two processes drive the antigens to change: antigenic drift and antigenic shift**[77, 18]**. Antigenic drift involves small, gradual, unpredictable changes in the genetic content of the same virus strain, and thus in the antigens H and N. This leads to loss of immunity and vaccine mismatch. On the other hand, antigenic shift refers to the process by which a new subtype of the virus is created by the combination

of two or more different strains of a virus, or strains of two or more different viruses. While antigenic drift occurs in all types of influenza, antigenic shift occurs only in influenza virus A because it is capable of infecting other animals asides from humans, creating the opportunity to reassort its genetic content dramatically. Depending on the reassortment of bird-type flu proteins, if it makes it to the human population, the flue may be more or less severe.

A pandemic has been defined by the World Health Organization as the worldwide spread of a new disease [82]. There were three flu pandemics in the twentieth century, the worst of which occurred in 1918; known as the "Spanish flu", it killed 20-40 million people worldwide. Milder pandemics occurred in 1957 (Asian Flu) and 1968 (Hong Kong Flu). Researchers think type A influenza is responsible for all of them [64]. In 1977, it was found that the avian flu was transmitted to humans directly for the first time [64]. The virus did not pass easily between humans, and a pandemic did not take place.

The most recent pandemic occurred in 2009 - 2010 with the surge of the H1N1 virus (also known as swine flu). Although we learned after the epidemic was over that it was the least lethal of the modern pandemics (it appeared to kill one of every 2,000 people who get it), health authorities around the world took extraordinary measures to combat its spread [73]. The outbreak caused concern because officials had never seen this particular strain of flu passing among humans before.

Currently, fears are that an antigenic shift between avian influenza and human influenza will result in a new highly virulent strain for which humans have little or no immunity resulting in a pandemic: the disease would rapidly spread worldwide, possibly with high mortality rates. These worries are well founded: bird flu has killed 60% of the 570 known cases since 2003 [2]. So far, the virus lacks of sustained human-human transmission; however, a single mutation could make this transmission not only possible, but efficient [10].

To anticipate nature, in late 2011 researchers tweaked the virus's genes to produce a strain that could be passed from person to person through air. A debate about whether the results of this investigation should be published or not (due to fears that sensitive information could fall into the wrong hands) ensued. After months of deliberation involving world's experts on many fields, on April 20th of 2012, American officials decided that the benefits of publishing such results outweighed the risks. Quoting The Economist [2]: "The reason is that, as bioterrorists go, humans pale in comparison with nature. [...] From the Black Death via Spanish flu to AIDS, bacteria and viruses have killed on a scale that terrorists and dictators can only dream of." The main take-away we draw from these studies is that experts still don't know how to predict the virus' mutations. Indeed, they don't know how likely it is that H5N1 (avian flu) will follow the mutations presented in the papers or a different one [74].

2.2.1   Pandemic Preparedness

While new, better vaccines are being studied, there is great interest in evaluating possible emergency management strategies (the focus of this thesis) due to the social and economic costs that would arise with a big influenza outbreak. During an epidemic, particularly a long one, public services such as health care facilities, police, fire fighting services and refuse collection would, in all probability, experience staff reduction due to illness or fear of infection. Public utilities such as power and water plants may require a minimum level of staff below which they must shut down [39].

It is worth mentioning that preparing for an epidemic is quite different from planning for "regular" catastrophes. While they are both "emergency" settings, catastrophes such as earthquakes, hurricanes or nuclear plant failures are sharp, extant events that have materialized. During these events municipalities would need to resort to large numbers of first responders. These would be members

of possibly distant communities that would be brought in to the emergency site. An important sociology question concerns whether these people would actually participate in the relief effort, not only abandoning their roles in their communities, but potentially risking their own lives. Research shows that this indeed happens. First-responder corps with as many as three or fourfold additional staff are normally prescreened and trained in emergency protocols **[76]**.

Epidemics pose a different challenge in that they evolve in multiple locations (vs a single site) and over a potentially protracted time frame, with extensive uncertainty. The main difficulty lies not in the proclivity of staff participation, but in the scarcity of staff and its effective management during a potentially extended period of time, under significant incertitude (a point indirectly alluded to in **[76]**).

In the process of designing contingency plans organizations must design frameworks to make the best use of the available resources and look to extend their capacity in case of an emergency, including surge staff. This will require the local planners to solve major logistical problems. We note that federal agencies such as the CDC and the United States Department of Labor have recommended businesses to implement flexible human resources strategies - such as "work from home" - in the event of an epidemic. However, organizations that provide public services like the ones described above cannot afford such plans and are encouraged to build sources of additional staff, a strategy that could require careful preplanning. Credentialing and legal preparations should be made in case personnel needs to be brought in from other states or recalled from retirement, especially in the case of health care workers **[10, 11]**.

Thus, a virulent influenza epidemic that would develop over time and geography, characterized by uncertainty and noise, will have deep social impact. Given the immediacy of events once the epidemic starts, a significant amount of preplanning is needed to build an adequate response, from a staffing

and resource perspective, that will allocate resources as effectively and efficiently as possible. While

we focus on influenza mainly motivated by the amount of attention it has received in recent years, we

expect our models to be useful for other diseases that could represent public health concerns that could

meet this characteristics. In this thesis we focus on building contingency plans that maintain critical

staff levels required for the operations continuity of organizations of interest. We follow algorithmic

and data-driven forecasts that hedge against the inherent uncertainty of the epidemic.

2.3   Robust Optimization and Benders' Decomposition

A generic mathematical programming problem is of the form

$$\min_{x_0 \in \mathbb{R}, x \in \mathbb{R}^n} \{x_0 : f_0(x, \zeta) - x_0 \leq 0, f_i(x, \zeta) \leq 0, i = 1, ..., m\} \tag{2.1}$$

where x is the decision variable, f_0, the objective function, and f_i, the constraints, are structural

elements of the problem; ζ stands for the *data* specifying a particular problem instance.

Optimization problems posed to solve real-world problems are usually presented with the following

challenges [12]:

1. The data are uncertain/unexact;

2. The optimal solution may be difficult to implement;

3. The constraints must remain feasible for *all* meaningful realizations of the data;

4. Problems are *large-scale* (the number of constraints and/or variables is large);

5. It's common to have solutions which, deemed to be optimal, behave *badly* in the face of small changes to the input data[1].

Robust Optimization is a modeling methodology, combined with a suite of computational tools, which is aimed at accomplishing the above requirements. Thus, the robust counterpart of (2.1) is

$$\min_{x_0 \in \mathbb{R}, x \in \mathbb{R}^n} \{x_0 : f_0(x, \zeta) - x_0 \leq 0, f_i(x, \zeta) \leq 0, i = 1, ..., m, \forall (\zeta \in \mathcal{U})\}. \tag{2.2}$$

It's important to stress that any candidate solution to this problem must satisfy a *large* system of constraints dictated by all $\zeta \in \mathcal{U}$, where $\mathcal{U}$, known as uncertainty set, represents the collection of possible values that the data could attain. Many simple instances ($\mathcal{U}$ being an interval in $\mathbb{R}$) already make (2.2) into a semi-infinite mathematical program.

Formulating a problem like (2.2) faces two major difficulties: determining its computational tractability (even if just approximately), and specifying $\mathcal{U}$. Once solved, the optimal solution to a robust problem will have the desirable property of being insensitive to perturbations of the data within set $\mathcal{U}$. At the same time, the robust solution is a worst-case solution, and thus, be deemed too conservative. It is up to the modeler to evaluate this trade-off.

Methodologies to tackle robust problems vary according to the characteristics of the objective f_0, the constraints f_i, and the structure of uncertainty set $\mathcal{U}$ (see **[12]** for a survey). In this work we focus on Robust Linear Programs and, in particular, in a cutting plane method known as Benders' Decomposition to solve them.

Benders' Decomposition follows the concept of delayed constraint generation. In a problem with an

---

[1]Used as the solution of the same problem but with small changes of the input data yields a very distant value from the optimal one.

excessive number of constraints, the idea is to add them iteratively to a relaxed version of the original problem known as the master problem. A constraint is explicitly considered only when it is violated by an optimal solution to the master problem. To "discover" it, instead of individually checking all of the constraints, auxiliary subproblems need to be solved efficiently. A detailed description of how we construct the master problem and the corresponding subproblems is given in Chapter 7.

It is worth pointing out that for our purposes, the utility of Benders' Decomposition becomes clear from the "constraint-wise" formulation of robust problems as described above [12]. Our uncertainty set will be characterized by the different intensity levels that an epidemic can take; our goal will be to design a contingency plan that is as insensitive as possible to these (potentially many) scenarios. This methodology is also used in solving multi-stage decision problems and in Stochastic Programming problems, among other applications[24].

3

Contingency Planning

3.1   Planning for workforce shortfall

An influenza pandemic could severely stress the operational continuity of social and business structures through staff shortages. Altogether, public health and utility professionals predict **[39, 40]** that the direct and indirect staff shortfall caused by an epidemic, in a worst-case scenario, could result in 20 to 40% of the workforce absenteeism for an extended period of time. Even though the outlook is dire, organizations that provide critical infrastructure services such as health care, utilities, transportation, and telecommunications, should clearly continue operations and are required to plan accordingly **[77]**. (See **[55]** for additional background on emergency staff planning.)

Staff shortfall directly resulting from individuals becoming sick could be intensified by policy or absenteeism. For example, during the last H1N1 influenza outbreak, CDC recommended that people with

influenza-like symptoms remain at home until at least 24 hours after they are or appear to be free of fever. In the particular case of health care workers it is advised that they refrain from work for at least 7 days after symptoms first appear (see [18] for additional details). Moreover, staff shortages may occur not only due to actual illness, but also from illness among family members, quarantines, school closures (combined with lack of child care), public transportation disruptions, low morale, or because workers could be summoned to comply with public service obligations [72, 31]. Indeed, employees who have been exposed to the disease (especially those coming into contact with an ill person at home) may also be asked to stay at home and monitor their own symptoms.

U.S. authorities, acknowledging these facts, have released documentations such as the Implementation Plan for the National Strategy for Pandemic Influenza [40] at the federal level and the Pandemic Influenza Response Standard Operating Guide in the state of Georgia [29], promoting guidelines to coordinate careful planning. The latter, for example, promotes county planning committee kits with three objectives:

1. Educate community members on the pandemic threat: how to prepare for it and what to expect from authorities.

2. **Planning for continuity of services in the face of high absenteeism and possible closures.**

3. Understand how members can contribute to their community's response.

It considers a special planning kit for urgent care facilities and health clinics to project for possible demand increases and the need for surge capacity. Planning is considered at Federal, State, and District levels. The first have the responsibility to appropriately disseminate and coordinate regulation plans with state authorities; the second should gather this information and disseminate it to the Districts,

as well as sharing any additional information with their federal partners. Public health districts are responsible for "local" level activities; they develop and execute plans upon request or as required.

At the federal level, the Department of Health and Human Services (HHS) and the CDC have also provided directions to help organizations and their employees in such planning. In addition to recommendations addressing the spread of disease and antiviral drug stockpiling, there is a focus on staff planning, which is the subject of our study **[77, 18]**. Additionally, there are federal and state programs such as the New York Medical Reserve Corps whose mission is to organize volunteer networks prepare for and respond to public health emergencies, among other duties **[62]**.

Multiple federal agencies have done some work directed to this end. HHS and CDC urge organizations to identify critical staff requirements needed to maintain operations during a pandemic and develop detailed emergency staff deployment plans to maintain operations **[78]**. In particular, organizations should develop detailed criteria to determine when to trigger the implementation of an emergency staffing plan. Most significantly, organizations should identify the minimum number of staff needed to perform vital operations. For example, in the case of water treatment plants, approximately 90% of the personnel is critical for keeping the utility running; for refineries, losing 30% of their staff would force a shutdown **[39]**.

In the particular case of hospitals, an effective *contingency staffing* plan should incorporate information from health departments and emergency management authorities at all levels, and would build a data base for alternative staffing sources (e.g., medical students). For additional details, see **[77, 37]**. To help with these tasks, the CDC has also developed the software package FluSurge **[17]** aimed to help hospital administrators and public health officials to estimate the surge in demand for hospital-based services (such as number of hospitalizations and persons require ICU care) during an influenza

epidemic. The software is meant to provide a starting point in planning activities since the estimates presented are based on a given scenario. The Department of Transportation has also developed a preparedness strategy calling for development of surge staffing and response capabilities under general emergencies [65].

In New York City, for example, after the events of 9/11, Columbia University created a database of volunteers to be recruited and trained both in basic emergency preparedness and their disaster functional roles [55]. At the city level the NYC Medical Reserve Corps ensures that a group of health professionals ranging from physicians to social workers is ready to respond to health emergencies. The group is pre-identified, pre-credentialed, and pre-trained to be better prepared in the wake of a crisis. Similar emergency staff backup plans could be implemented in all other cases of utilities and social infrastructure [62]. Moreover, for specific infrastructure needs, commercial sector providers of fully trained and certified surge staff are available to operate in vital command, operations, and emergency response centers at a cost [8].

In spite of all these efforts, it seems clear that much remains to be done and that a severe epidemic would place extreme strain on infrastructure. A good example is provided by the 2009 Swine Flu epidemic. Even though the virus mutation caused few fatalities, and a successful vaccine became available, New York hospitals were severely stressed [73]: "The outbreak highlighted many national weaknesses: old, slow vaccine technology; too much reliance on foreign vaccine factories; **some major hospitals pushed to their limits by a relatively mild epidemic**" (our emphasis).

It is important to mention that other organizations besides those devoted to public service could be affected in this sense. A supermarket chain, due to decreased staff, may experience increased spoilage, requiring more frequent restocking. A manufacturer managing a supply chain may see its production

yield decrease due to low staff, causing a change to alternate manufacturing methods which require, in turn, additional resource allocation. Such disruptions could manifest themselves in events similar to the so-called "Bullwhip effect" **[48]**.

From our perspective, the uncertainty concerning the time line and severity of the pandemic brings substantial complexity to the problem of deploying replacement staff. This problem, which will be the core issue that we address, is relevant because significant preplanning must take place and it is unlikely that major quantities of additional workforce can be summoned on a day-by-day basis.

3.2   Declaring an epidemic

A technical point that we will return to below concerns when, precisely, an epidemic is "declared". In the case of an infectious disease such as influenza, an initially slow accumulation of cases followed by a more rapid increase in incidence **[34]** is viewed by epidemiologists as an epidemic. However, this definition is too general for planning purposes. This is a significant issue since emergency action plans would be activated when the epidemic is declared.

In the United States, the CDC declares an influenza epidemic when death rates from pneumonia and influenza exceed a certain threshold **[26]**. Each week, vital statistics of 122 cities report the total number of death certificates received and the proportion of which are listed to be due to pneumonia or influenza. This percentage is compared against a seasonal baseline, which in turn was computed using a regression model based on historic data. There is a different baseline for each week of the year to capture the different seasonal patterns of influenza-like illnesses (ILI) (the epidemic threshold sits 1.645 standard deviations above the seasonal baseline). This type of measure is not specific for the United States. In **[49]**, the authors' base case assumed that the duration of an influenza pandemic

in Singapore was defined as the period during which incident symptomatic cases exceeded 10% of baseline ILI cases.

Motivated by this discussion, we will use the convention that an epidemic is known to be present as soon as the number of (new) infected individuals on a given time period exceeds a small percentage of the overall population, e.g. 0.93%, which corresponds to the national epidemic threshold of 6.5% for week 40[26].

From the point of view of an organization, of course, action need not wait until an "official" epidemic declaration and would instead rely on its own guidelines to possibly implement a preparedness plan at an earlier point. However, we expect that the mechanism underlying declaration will be the similar.

4

Modeling Influenza

In this chapter we describe a classic compartmental model that follows the spread of influenza in big populations. We discuss the benefits and pitfalls of using such models as underlying elements of policy design tools and make the case for their use incorporating robustness.

4.1 Modeling literature

The mechanism of transmission of most communicable diseases such as influenza or tuberculosis is now known. As in most other fields, the degree of complexity of the mathematical modeling of disease transmission varies with the desired accuracy of short-term specific calculations and the ability to derive broad, general principles that are good to establish theoretical principles.

During late 1920's and early 1930's, public health physicians McKendrick and Kermack laid the foundations of the study of epidemiology based on compartmental models in a sequence of three papers

[45, 43, 44]. Their model divides large populations into classes or compartments that reflect different states of the disease and describes rates of deterministic flow between them under relatively simple assumptions. Individuals in a large population are classified into compartments depending on their status with regard to the infection under study. The disease is tracked at a population level, not individual. Because of their relative simplicity, they have become widely used in Mathematical Epidemiology and have increasingly incorporated complexity into their structure[1].

In recent years, agent-based modeling has emerged as an alternative to model disease spreading **[3]**. These models simulate the actions and interactions of autonomous agents at the individual level. As such, they are very effective at incorporating population heterogeneity. At the same time, they are very computation-intensive.

Compartmental models have proved to be effective at fitting epidemic curves (see, for example **[75]**) and are still the most common modeling tool. Thus, we have opted for the use of these models for our algorithms.

The most basic model (and maybe most common) of an influenza epidemic is the Susceptible-Infected-Removed (*SIR*) model. However, as more information about the disease has been gathered, additional compartments have been added. In particular, influenza is characterized by an incubation period between infection and the appearance of symptoms, accounted for by the Susceptible-Exposed-Infected-Removed (*SEIR*) model. Secondly, a significant fraction of people who are infected never develop symptoms, but go through an asymptomatic period, during which they are still infectious at a lesser degree, and then recover. The model Susceptible-Exposed-Infective-Asymptomatic-Removed takes these assumptions into account **[15]**.

---

[1]For a general reference on Mathematical Epidemiology see, for example, **[15, 7]**.

The critical difficulty in modeling the impact of a future influenza pandemic is our inability to accurately predict the spread of disease on a given population. Specifically, knowledge of the rates at which individuals flow between compartments becomes problematic when looking to forecast the behavior of a new virus strain. In the following section we provide a description of the *SEIR* model which we will modify in order to follow the evolution of the disease and its spread among members of a given population and within a workforce group of interest. We will later discuss its weaknesses and discuss the importance of incorporating robustness.

4.2   SEIR model

The *SEIR* model describes, in a deterministic fashion, the spread of the disease in a given population. It divides the host population into a small number of groups (or compartments) that correspond to different stages of the disease in question. In the *SEIR* model there are 4 compartments:

- *Susceptible*: holds individuals who have no immunity to the infectious agent and so, can become infected if exposed.

- *Exposed*: also known as Latent compartment, contains individuals who are incubating the disease. Individuals are infected, but they are not yet infectious.

- *Infectious*: describes infected individuals who can transmit the disease to those susceptibles with whom they have contact.

- *Removed*: has individuals that are immune to the disease. They don't affect the transmission dynamics in any way.

The number of individuals in each compartment is traditionally denoted as S, E, I, and R, respectively. The total host population size is given by the sum of these totals, and it is denoted by N. *SEIR* models thus describe the size the compartments at each time period via a set of equations that model the transition between them. Before giving them out, we state the following standard assumptions [15]:

1. There is a small number I_0 of initial infectives relative to the size of the total population, which we denote by N.

2. The rate at which individuals become infected is given by the product of the probability that at time t a contact is made with an infectious person, β_t, the average constant social contact rate λ, and the likelihood of infection, p, given that a social contact with an infectious person has taken place. The probability β_t changes with time because we assume it depends on the number of infectious agents in the population.

3. We assume exposed individuals proceed to the I compartment with rate μ_E.

4. Infectives leave the infectious compartment at rate μ_{RR}.

5. We assume there is only one epidemic wave. Thus, people who recover are conferred immunity.

6. The fraction of members that do not die from the disease, when removed from the infectious class, is given by $0 \leq f < 1$.

7. We do not include births and natural deaths because influenza epidemics usually last few months. We also omit any migration. In other words, excluding deaths by disease, the total population remains constant.
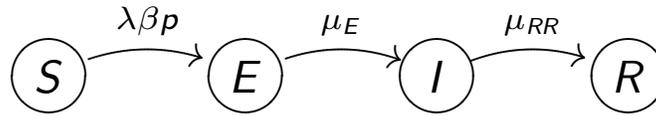
Figure 4.1: Basic *SEIR* model.

SEIR models are usually described as a system of nonlinear ordinary differential equations (see for example [15, 60]). For our purposes, we use a discrete-time Markov chain type approximation along the lines of [47] and similar to those found in [4, 5, 6]:

$$S_{t+1} = S_t(e^{-\lambda \beta_t p})$$

$$E_{t+1} = E_t(e^{-\mu_E}) + S_t(1 - e^{-\lambda \beta_t p})$$

$$I_{t+1} = I_t f(e^{-\mu_{RR}}) + E_t(1 - e^{-\mu_E})$$

$$R_{t+1} = R_t + I_t(1 - e^{-\mu_{RR}}).$$

(4.1)

At each time, a fraction of the susceptible population becomes infected and transitions to the exposed compartment. Such fraction depends, among other factors, on the number of infectious individuals in the population at that time, which is captured by $beta_t$, described above. Exposed members of the population may remain in the compartment or progress to the infectious group, where they similarly either stay (if they survive) or move on to the Removed compartment.

Compartmental models incorporate a number of assumptions to describe social contact dynamics. First, the number of social contacts with infectious people for an arbitrary person is thought of as a Poisson random variable with rate $\lambda \beta_t$; we use one day as time unit. Second, the models assume homogeneous mixing, that is, all individuals have a fixed average number of contact rates per unit of time and are all equally likely to meet each other. Thus, the probability that a contact is made with

an infectious person, β_t is given by I_t/N. The daily infectious contact rate λp is usually written as λ

[15, 5] and taken as a constant throughout the epidemic. We will remove this assumption later when

we make the transmissibility parameter p explicit.

If the number of initial infectives is relatively very small compared to the whole population ($S_0 \sim N$),

then a newly introduced contagious individual is expected to infect people at the rate λp during the

expected infectious period $1/\mu_{RR}$. Thus, each initial infective individual is expected to transmit the

disease to an average of

$$R_0 = \frac{\lambda p}{\mu_{RR}} \tag{4.2}$$

individuals. R_0 is called the *basic reproduction number* (also known as *basic reproduction ratio* or

basic reproductive rate.) It is without doubt the most important quantity epidemiologists consider

when analyzing the behavior of infectious diseases [15]. Its relevance derives mainly from its threshold

property: when $R_0 < 1$, the disease does not spread fast enough and there is no epidemic; when

an epidemic does take place - i.e. $R_0 > 1$ - the magnitude of R_0 is a parameter of great interest.

Considering that we assume *a priori* that an epidemic will take place, R_0 is not of particular interest

to us; however, it is presented here for completeness.

Keeping track of staff availability

We are interested in tracking workforce availability at a particular organization during the epidemic;

following previous work [9, 28] we divide the population into two groups: (1) the general population

and (2) the workforce under consideration. Individuals from the latter group could have a very different

exposure to the epidemic. For example, people working at a water plant could have lower contact

rates than average (by virtue of having contact with few individuals during the workday) while the staff at a health clinic may not only have higher contact rates, but may also have easier access to antiviral medicines that reduce their infectiousness and the length of the infectious period.
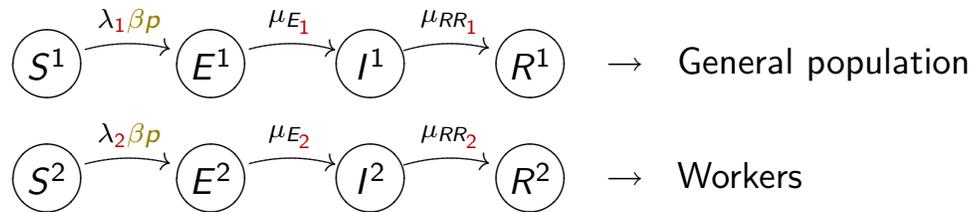
Figure 4.2: Keeping track of staff availability: Two parallel *SEIR* models.

For ease of notation, we use superscript 1 to refer to the general population and 2 to refer to the group of workers of interest. For $j = 1, 2$, define compartments S^j, E^j, I^j, R^j corresponding to group j. Following the discussion above, we allow the groups to have different contact, incubation, and recovery rates. The probability that a random contact is one with an infected person at time t, β_t, is now defined as

$$\beta_t = \frac{\lambda' I_t}{\lambda' N_t},\tag{4.3}$$

where $I_t = [I_t^1, I_t^2]$ is the vector of infectious individuals at time t, $\lambda' = [\lambda^1, \lambda^2]$ is the vector of contact rates, and $N_t = [N_t^1, N_t^2]$ denotes the size of each group at time t. We note that β_t is constant across groups. We now have two parallel thinned Poisson process approximations, each with rate $(\lambda_j \beta_t p)$.

The set of equations that correspond to group j $(= 1, 2)$ is

$$
\begin{aligned}
S_{t+1}^j &= S_t^j e^{-\lambda_j \beta_t p} \\
E_{t+1}^j &= E_t^j e^{-\mu_{E_j}} + S_t^j (1 - e^{-\lambda_j \beta_t p}) \\
I_{t+1}^j &= I_t^j f e^{-\mu_{RR_j}} + E_t^j (1 - e^{-\mu_{E_j}}) \\
R_{t+1}^j &= R_t^j + I_t^j (1 - e^{-\mu_{RR_j}}).
\end{aligned}
\tag{4.4}
$$

We now give an expression for R_0 for the case in which $\mu_{RR_1} = \mu_{RR_2}$ [15, 23]. First we note that the average contact rate for subgroup j at time 0 is given by

$$
\frac{\lambda_j N^j}{\lambda_1 N^1 + \lambda_2 N^2}.
$$

Thus, the average number of new infections caused by a newly infective person introduced into an otherwise susceptible population is given by

$$
\begin{aligned}
R_0 &= \left[\lambda_1 \frac{\lambda_1 N^1}{\lambda_1 N^1 + \lambda_2 N^2} + \lambda_2 \frac{\lambda_2 N^2}{\lambda_1 N^1 + \lambda_2 N^2} \right] \frac{p}{\mu_{RR}} \\
&= \frac{\lambda_1^2 N^1 + \lambda_2^2 N^2}{\lambda_1 N^1 + \lambda_2 N^2} \cdot \frac{p}{\mu_{RR}}.
\end{aligned}
\tag{4.5}
$$

We note that when $\lambda_1 = \lambda_2$ (4.5) reduces to (4.2).

Nonhomogeneous mixing and social distancing

We initially assumed that the population mixed homogeneously, that is the contact rates λ_1 and λ_2 remained constant throughout. However, the homogeneous mixing and mass action incidence

assumptions have clear pitfalls; individuals from each compartment are hardly indistinguishable in terms of their social patterns and likeliness of infection. It has been suggested that people reactively reduced their contact rates in response to high levels of mortality during the 1918 pandemic [14]. Following [5], we make the assumption that this could also be the case for number of infectious members of the population. This makes the contact rates to be reexpressed as

$$\lambda_t^j = \Lambda_j \frac{S_t^j + E_t^j + R_t^j}{N_t^j}, \quad j = 1, 2, \tag{4.6}$$

where Λ_j are fixed constants, $j = 1, 2$. Using this definition, λ_t^j decreases whenever there is a high number of infectious agents in the population. As mentioned in [5], other functional forms are possible; we rely on (4.6) because it provides a simple way to capture changes in contact rates as a function of severity of the epidemic.

Additionally, we also consider a scenario in which authorities impose social distancing measures as soon as the epidemic is declared. This kind of public health intervention was used in some cities of the United States during the 1918 epidemic with different degrees of success. San Francisco, St. Louis, Milwaukee, and Kansas had the most effective social distancing bans, reducing transmission rates by up to 30 - 50% [14]. A similar situation took place in Mexico during the last 2009 H1N1 epidemic, where venues such as schools, movie theaters, and restaurants were forced to close temporarily. It is estimated that the transmission of the disease was diminished by 29% to 37% [19]. We incorporate this element by multiplying the contact rates by an additional dampening factor when the epidemic is considered declared and until the rate of growth of daily infectives is below some threshold (we refer the reader to section 3.2). Effectively, the contact rate for group j at time t becomes $\lambda_t^j = \theta \Lambda_j \left(\frac{S_t^j + E_t^j + R_t^j}{N_t^j} \right)$ $(0 < \theta < 1)$ when the epidemic is officially ongoing; otherwise, it remains as per equation (4.6).

4.3 Robustness in planning

In planning the response to a future or impending epidemic, one would need to rely at least partly on an epidemiological model, and such a model would have to undergo careful calibration in order to be put to practical use. This is especially the case for the SEIR model presented above, which is rich in parameters that need to be estimated to fully define the trajectory of the epidemic. In the case of a flu pandemic caused by an unknown virus strain, new parameter values would need to be promptly estimated as the epidemic emerges. Ideally, robust statistical inference would provide information on all parameters, though data paucity would present a challenge.

On the positive side, the infectious and incubation periods can usually be independently estimated via clinical monitoring of infected agents, either by observation of transmission events or by the use of more detailed techniques [42].

On the other hand, it is not clear how to accurately estimate the *transmission rate* $\lambda_j p$. One approach is to approximate the basic reproductive ratio R_0 and the mean infectious period, and then use equation (4.2). There are multiple ways of estimating R_0; see, for example, [20]. Given that the definition of R_0 is based on the early stages of the epidemic, one should examine its early behavior. However, the progression of the epidemic in its initial stages could fluctuate widely because of the very small number of initial cases, making the fitting process more difficult [42].

Direct estimation of the transmission rate gives rise to a number of challenges [79, 83]. First, pandemic influenza (along with smallpox and pneumonic plague) has not been present in modern times frequently enough so as to gather sufficient data for accurate estimation. Second, for existing age-specific transmission models, there are more parameters than observations on risk of infection for

each age class. The infectivity of the disease can be approximated using serologic data and contact rates are usually estimated from census and transportation data after making assumptions about the contact processes that reduce the number of unknowns to the number of age classes. However, both contact or infection rates estimates can serve only as a baseline; a population could change its behavior significantly during a severe epidemic (due to school closures, for example); further, environmental changes (e.g. weather changes) could also have a significant impact on the virus transmissibility **[52]**.

The infectivity parameter p is particularly hard to estimate because (at least partly) it reflects the characteristics of the virus; it is usually estimated after the epidemic has taken place. It is very difficult to predict the evolution of new virus strains. In fact, as far as we know **[67, 56]** research that relates mutations of influenza virus to infectivity is inconclusive. Previous work has proposed different upper and lower bound values for p, depending, among other factors, on the geographical location of the study and the pandemic wave of interest. For example, Walling et al. use the interval $[0.025, 0.5]$ in one work **[79]** and $[0.02, 0.16]$ in another **[80]**. Both studies use these intervals to conduct sensitivity analysis. In another study, Larson **[47]** classifies the population into three groups according to social activity levels; the three groups have p value 0.07, 0.09 and 0.12 and corresponding λ values 50, 10 and 2, respectively. In summary, the precise estimation of p values appears quite challenging, especially prior to or even during an epidemic.

Given this uncertainty, it is likely that basing the response to an epidemic on a fixed estimate for p is incorrect. To illustrate the impact of such a decision we refer the reader to Figure 4.3. It shows the availability of staff as a function of time, for different two values of p, from two different perspectives. Figure 4.3a shows how the workforce becomes ill at the actual time it happens. In contrast, 4.3b presents these curves all starting from the time the epidemic is declared. Indeed, a planner deploying
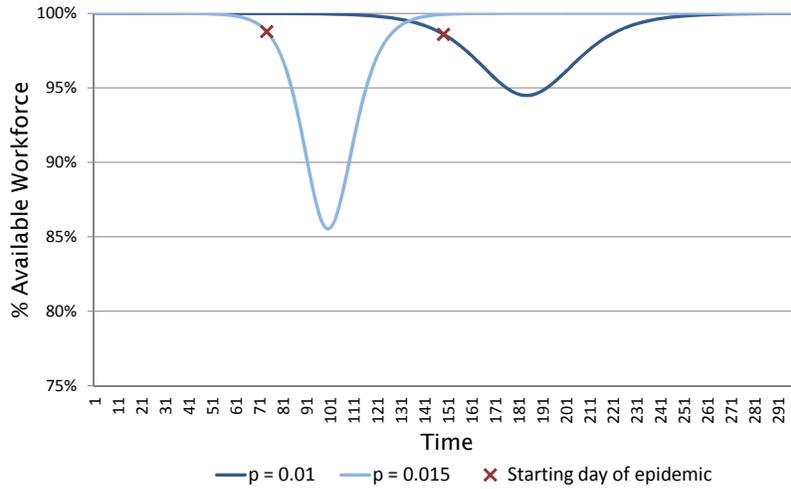
(a) An absolute perspective in time: Curves start from the moment infectives are introduced into the susceptible population.
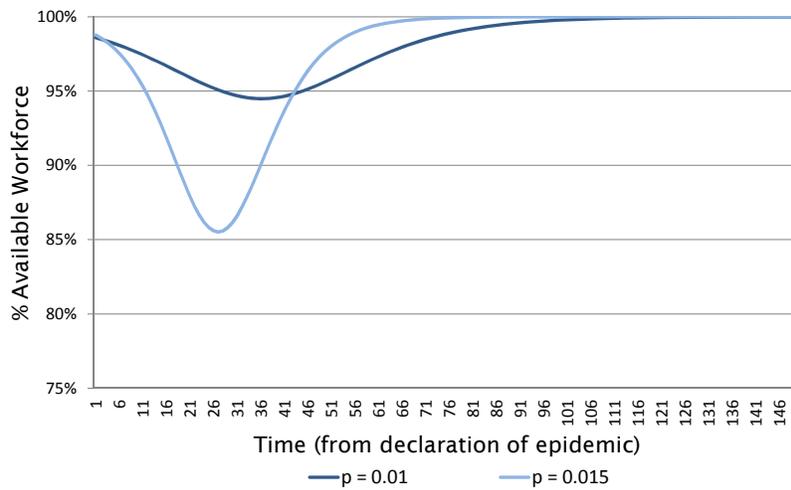
(b) A planner's perspective: Curves for *both epidemics* start from the moment an epidemic is declared.

Figure 4.3: Availability of workforce as epidemic progresses for different values of *p*.

a contingency plan at the moment an epidemic is declared would be interested in the preparing for different scenarios according to what is shown in 4.3b, rather than 4.3a. This point will be discussed again in Section 5.2.2.

Consider a baseline threshold of 90%, i.e. the system is considered to be performing poorly if fewer than 90% of the staff is available. For $p = 0.015$ this period spans days 18 through 37 (from the

declaration of the epidemic), whereas for $p = 0.01$ the baseline is not reached. As expected, the epidemic is more severe for higher values of p; however, somewhat lower values of p result in longer-lasting epidemics. In particular, $p = 0.01$ results in an extended period of time (days 13 through 69) where even though staff availability is above the baseline, it is still significantly below 100%. If, in the above example, p were unknown, a planner would have to carefully ration scarce resources over a nearly month-long period. If the planner were to assume a *fixed* value for p throughout the epidemic, then the example suggests allocating comparatively higher levels of surge staff to earlier periods of time, to overcome the higher shortfalls to be expected in the case of higher values for p. Of course, this higher level of initial allocation needs to be carefully chosen to obtain maximum reward.

Consider now Figure 4.4, which shows the impact of a *change* in p in the midst of an epidemic. Here, p changes from 0.012 to 0.03 on day 130, 28 days after the epidemic has been declared. Thus, for more than half of the epidemic, the disease spreads slowly and even though the 90% baseline is approached, it is not reached. However, after the change in p the epidemic becomes severe and a significant shortfall arises. This type of variability would be especially problematic when resources are limited. The epidemic, on the basis of observations of its initial progression, would likely be classified as relatively mild, and, perhaps, action might be taken to at least partially abandon the surge staff buildup. However, if the a change in p as shown in Figure 4.4 were plausible, then a careful planner would have to hedge by holding back staff so as to handle the potentially critical situation in later periods. Should the change in p not take place, the chart indicates that any held back staff would essentially be wasted. And should the change take place somewhat earlier, then the opportunity cost is significantly higher. The critical question, of course, is whether this type of virus behavior is possible. As far as we can tell, current knowledge of the influenza virus cannot categorically reject a change in

p as shown, although planning for such a change might be construed as an overly conservative action.

Thus, it remains to be seen if a robust strategy that can protect against a change in *p* entails higher

cost or other compromises.
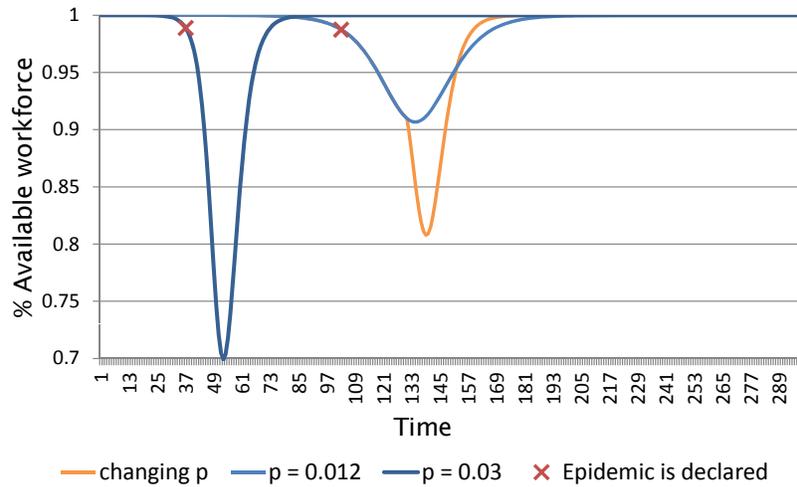
(a) An absolute perspective in time: Curves start from the moment infectives are introduced into the susceptible population.
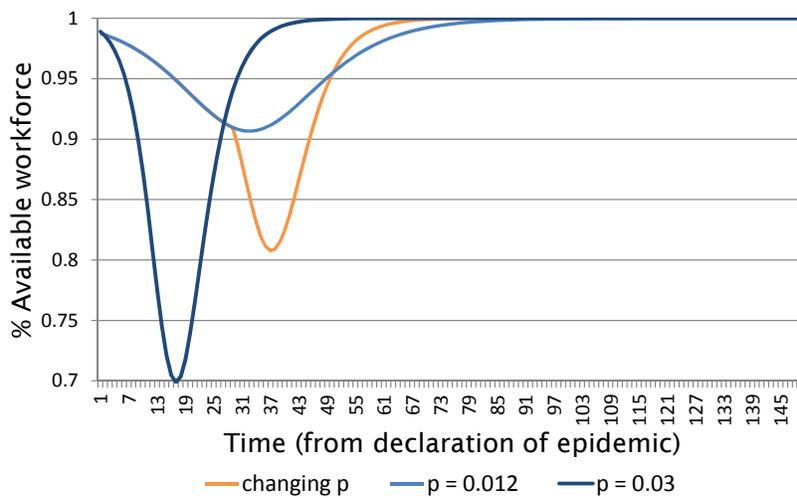
(b) A planner's perspective: Curves for *all three epidemics* start from the moment an epidemic is declared.

Figure 4.4: Availability of workforce as epidemic progresses when probability of contagion *p* is allowed to change *during* the epidemic.

In this paper we model the variability of the product $\lambda_j p$ using techniques derived from robust opti-

mization. For ease of exposition, we assume that the infectivity probability *p* is uncertain, and that

the values for the contact rates λ_j are fixed (and known); thus the contact probabilities β_t are also

fixed. Our algorithm's flexibility would allow us to easily incorporate uncertainty in other parameters;

however, as explained above, we advocate that the biggest source of uncertainty relies on the con-

tagion rate. Our goal is to produce surge staff deployment strategies that are robust with respect

to variability of p in a number of models of uncertainty. Robust optimization provides an agnostic

methodology for assessing competing allocation plans and for computing good plans according to

various criteria. Above, we described two such criteria. Our optimization procedures make it possible

to efficiently evaluate multiple plans and different levels of conservatism.

5

Robust Models

5.1 Performance Measures

Personnel shortage during an epidemic may jeopardize the continuity of operations of critical infrastructure organizations. To gauge the impact of this shortfall, we consider cost functions that model two scenarios. In the first one the organization requires at least certain percentage of personnel present to operate. This could be the case of water and energy plants, as mentioned in [39]. The second scenario uses classic queueing theory to measure the simultaneous impact of the decrease in number of available workers and the change in demand for the service that the organization provides. Health care institutions would be clear examples in this context.

5.1.1 Threshold functions

Here we model an organization that requires a minimum number of staff in order to operate under normal conditions. We further assume that this "minimum" is a *soft* constraint in the sense that if the available staff should fall below the threshold, the organization will still manage to operate, but at very large cost. This could be the case, for example, if the organization is able to purchase output or hire staff from another source (such as a competitor) or if it is able to reduce its level of service or output, at cost.

To model this behavior, we assume that operating costs at each point in time are given by a convex piecewise linear function. This function is represented as the maximum of L linear functions, with slopes $\sigma_L < \sigma_{L-1} < ... < ... \sigma_1 = 0$ and intercepts $k_L > k_{L-1} > ... > k_1 = 0$. Thus, denoting by ω_t the work force level at time t and z_t the cost associated with time period t, we have

$$z_t = \max_{1 \le i \le L} \{\sigma_i \omega_t + k_i\}. \tag{5.1}$$

5.1.2 Queueing models

We are interested in modeling basic queueing systems where both the service rate and the incoming demand rate are affected by the evolution of the epidemic. For each time period t, denote the average incoming demand rate by ζ_t and let s_t denote the number of available servers (staff). The system utilization at time t, using an M/M/s_t queueing model, is given by

$$\rho_t = \frac{\zeta_t}{s_t \mu}. \tag{5.2}$$

As it is well known, ρ_t and related quantities are good indicators of system performance. In our simulations of severe epidemics, ρ_t can grow larger than 1, a situation that depicts a system that is catastrophically saturated (a situation observed during 2010's H1N1 pandemic **[73, 27]**) and consequently system performance will drastically suffer. Accordingly, we choose as a reasonable representative of "cost" incurred at time t an exponentially increasing function of ρ_t. In particular we consider $e^{\rho_t - \delta}$ (for small $\delta > 0$) which we approximate with a piecewise-linear function, much as in Section 5.1.1. Other alternatives are possible. For example, one could use a traditional measure of system performance, such as average queue length for an M/M/s_t system, computed using a shifted ρ_t, i.e. a quantity $\hat{\rho}_t = \rho_t - \delta$ for appropriate δ.

In a regime where ρ_t is close to (but smaller than) 1, one could use one of the popular measures of queueing system performance as "cost" – or use ρ_t itself as the cost.

5.1.3 Other cost functions

Other situations of practical relevance besides the two considered above are likely to arise. Our robust planning methodology, described below, is flexible and rapid enough that many alternate models could be accommodated. Moreover, we would argue that in the context of the cost associated with staff shortfall, any reasonable cost function would be, broadly speaking, increasing as a function of the shortfall. The approach described above approximates cost, in the two cases we listed, using piecewise-linear convex functions, and we postulate that many cost functions of practical relevance can be successfully approximated this way.

5.2 Robust Models

We now describe our methodology for the robust surge staffing problem. We will first describe our uncertainty model, then the deployment policies we consider, and finally the robust optimization model itself. In what follows, we make the assumption that the total quantity of surge staff is small enough that their deployment does not affect the evolution of the epidemic in the SEIR model, in particular the values β_t.

5.2.1 Uncertainty models

Let p_t denote the probability of contagion at time t (time measured relative to the declaration of the epidemic). The model for uncertainty in p that we consider embodies the notion of increasing uncertainty in later stages of the epidemic. We will first describe the model and then justify its structure. We assume that there are four given parameters $p^L, p^U, \hat{p}^L$, and $\hat{p}^U$. The model behaves as follows:

There is a time period $\check{t}$, unknown to the planner, such that for $1 \leq t \leq \check{t}$, p_t assumes a constant (but unknown to the planner) value in $[p^L, p^U]$, and for $\check{t} < t$, p_t assumes a constant (and also unknown to the planner) value in $[\hat{p}^L, \hat{p}^U]$.

We stress that $\check{t}$ is not known to the planner. We are interested in cases where $\hat{p}^L \leq p^L$ and $p^U \leq \hat{p}^U$, that is to say, the period following day $\hat{t}$ exhibits more uncertainty than the initial period. We can justify our model as follows. In the event of an epidemic, a planner would be able to obtain some information on the spread of the epidemic in the period leading to the actual declaration of the epidemic, resulting

in a (perhaps tight) range of values $[p^L, p^U]$. As we will discuss below (Section 5.2.2) we are interested in surge staff deployment strategies with limited flexibility, i.e. requiring staff commitment levels that are prearranged.

In this setting, at the point when the epidemic is declared, the planner would deploy a staff deployment strategy. A prudent planner, however, would not simply accept the range $[p^L, p^U]$ as fixed. In particular, if at first the epidemic is mild (small p^U) the planner might worry that a change, such as sudden drop in temperature, could effectively increase p beyond p^U. Note that a change in weather would not simply change the contact rates, λ; it might produce other environmental changes (such as decreased ventilation); further, it is known that the influenza virus has higher transmissibility in colder and drier conditions [67, 52]. we model such changes, collectively, through changes in p. Should the change take place, a staff surge strategy that consumed most available staff in the earlier part of the epidemic would be ineffective.

As a means to avoid this overcommitment of resources to early phases of the epidemic, we can assume that a second and more virulent regime of the epidemic *could* be manifested at an unknown later time. We can parameterize this later regime by assuming a range $[\hat{p}^L, \hat{p}^U]$ with, for example, $\hat{p}^L = p^L$ and $\hat{p}^U > p^U$. The exact relationship between the two values $\hat{p}^U$ and p^U is a measure of the conservativeness or risk-aversion of the decision-maker. We will touch upon this issue later.

Of course, exactly the reverse could happen: the epidemic could become *milder* in the second stage. This could take place as a function of changes in public behavior due to non-pharmaceutical interventions (see [16, 32]) resulting in a (difficult to accurately predict) decrease in infectivity. In that case, a surge strategy that defers most staff deployment to the second stage of the epidemic could also be ineffective. The challenge is how to properly hedge in view of these extreme situations, and other

intermediate situations that could also arise.

5.2.2 Deployment strategies

We make several assumptions that constrain the feasible deployment patterns. First, we assume that a surge staff member, when deployed, will be available for up to τ time periods, provided that he or she does not get infected first - if that happens, this person will be removed from the system and will not available for deployment in the future. We also assume that the pool of available surge staff over the planning horizon is finite. Finally, there is a maximum quantity of surge staff that can be summoned on any given time period. More detailed models are possible and easy to incorporate in our optimization framework.

In this paper we will focus on *off-line*, or *fixed* strategies. More precisely, we assume that a *deployment vector* is computed immediately after an epidemic is declared, on the basis of the available information. From a formal perspective, the deployment vector is obtained as follows:

(1) First, the epidemic must be declared. As indicated in Section 3.2, this takes place as soon as the percentage of infected individuals exceeds some (small) threshold.

(2) Based on data available at that point, estimates are constructed for the various SEIR parameters. In particular, the ranges $[p^L, p^U]$ and $[\hat{p}^L, \hat{p}^U]$ for p (see Section 5.2.1) are constructed.

(3) Using as inputs the population statistics, the cost function, the nominal SEIR model parameters, and the uncertainty model for p, we compute the deployment vector $h = [h_1, h_2, \ldots, h_T]$, where h_t indicates the number of staff procured t time periods **after** the declaration of the epidemic. The parameter T is chosen large enough to encompass one epidemic wave and will be discussed

later.

We next address some points implicit in (1)-(3). First, recall that in the SEIR framework the formal epidemic will have a starting point which precedes the time period when an epidemic is actually declared. The exact magnitude of this run-up period is *not* known to a decision maker. In all our notation below (as in (2) above), "time period 1" refers to the time period where the epidemic was declared, that is to say, we always label time periods by the amount elapsed since the declaration of the epidemic. When using the SEIR machinery to *simulate* an epidemic, of course, we always proceed as in the formal model, and we merely relabel as "time = 1" that period where the declaration condition is first reached (since that would be the start of the epidemic as far as a planner is concerned).

Also, we assume that the epidemic is *correctly* declared, that is to say, the first day in which the criterion in Section 3.2 applies is correctly observed. In practice, this observation would include noise (for example, due to individuals infected by a different strain) but we make the assumption that the resulting decrease in strategy robustness is small.

Item (2), the estimation of the SEIR model and in particular the construction of the initial range $[p^L, p^U]$, gives rise to a host of other issues that are outside the scope of this paper (but are nonetheless important). We assume that at the time the epidemic is declared there is enough data (however incomplete, and noisy) to enable the application of *robust* least squares methods (see e.g. **[25]**) to fit an SEIR model, and that the model is sufficiently accurate to permit point estimates for all parameters other than p. Having constructed the initial range $[p^L, p^U]$, the later range $[\hat{p}^L, \hat{p}^U]$ is constructed on the basis of (a) risk aversion, and (b) environmental considerations. The role of (a) is clear -for example, we could obtain $\hat{p}^U$ by adding to p^U a multiple (or fraction) of the standard deviation of p in the observed data. As an example for (b), an epidemic that starts in late-Autumn might possibly

become more infectious as colder weather develops.

Another point concerns the time horizon, T, for the SEIR model (see Section **??**) which to remind the reader is measured from the point that the epidemic is declared. It is assumed that T is large enough to handle an epidemic of interest. If we assume a fixed (but unknown) p, then the milder (i.e., less infective) epidemics will tend to run longer, but, significantly, will also be less disruptive. Higher values of p give rise to sharper epidemics in the near term. From a technical standpoint, the parameters p_L and $\hat{p}_L$ can be used to construct estimates for T. In order to be sure that we capture as much of an epidemic as possible, in this study we will use values for T up to 150, which according to our experiments seems more than adequate.

One could consider models of epidemics spanning, for example, a one-year period, with multiple epidemic "waves". If some of these waves are very prolonged then they will necessarily be mild waves for long periods of time. As a result, during such long, mild epidemic periods (1) the social cost will be low, and (2) a significant pool of the population will become sick and as a result become immune to the virus. Consequently, future epidemic waves will necessarily be milder and cause decreased social cost (so surge staff will be less critical). On the other hand, over a long period we could have well-separated strong epidemic waves[1]. However, we would argue that from the perspective of surge staff deployment, each such wave should be handled as a separate event, with its own surge staff deployment strategy.

A significant element in our approach is that it produces a fixed deployment vector $h_1, \ldots, h_T$ which, from a formal standpoint, will be applied regardless of observed conditions. Of course, the strategy

---

[1]This was the case of H1N1: the first wave hit in the spring into the summer, followed by another outbreak in the fall, both of 2009.

should be interpreted as a template and a planner would apply small deviations from the planned surge levels, as needed. In any case, we have focused on this assumption for practical reasons. Having a pre-arranged schedule greatly simplifies the logistics of deploying possibly large numbers of staff, especially if many of the staff originate from geographically distant sources.

A broader class of models would allow for on-the-fly revision of a strategy in the midst of an epidemic, which would allow a planner to dynamically react to changes in the epidemic.

However, a significant underlying issue concerns the actual flexibility that would be possible under a virulent epidemic. We expect that large, sudden changes in deployment plans may be difficult to implement. In particular, if a significant and unplanned *increase* in surge staff is needed, it may prove impossible to rapidly attain this increase; and by the time the surge staff is available a severe peak of the epidemic may already have taken place. See Section 6.1.3 for some experimental validation of these views.

A different type of dynamic strategy would implement many, but small corrections as conditions change. We will discuss examples of such strategies (and appropriate modifications to our methodology) in Section 8; an issue in this context is the calibration of our statement "as conditions change" in an environment that will be characterized by noisy, partial and late data.

Last, but not least, while implicitly assumed in our description of the *SEIR* model, we reiterate that we consider that the deployment of surge staff does not alter the course of the epidemic. We assume that the number of additional workers is too small to change significantly the social contact patterns during the epidemic. Additionally, in the case of hospitals or clinics we take the position that surge staff is brought in to preserve operations continuity as much as possible, but they do not have any direct incidence on the outcome of patients infected with the flu. This is mainly because it is not

clear whether existing medications would be effective with a new virus strain - proper vaccines are not likely to be ready for a pandemic. If antiviral medications were available, they are usually expensive and should be carefully administered for it to be the most effective. Recommendations point towards focusing on the treatment of ill health care workers, and not the general population [21, 50, 61]. A model for the potential impact of these medications - such as a decrease in the infectious or recovery rates - would be needed and is considered as a possible refinement to this work.

5.2.3 Robust problem

We can now formally state the problem of computing an optimal robust pre-planned staff deployment strategy. Let h denote a deployment vector and $\mathcal{H}$, the set of allowable deployment vectors. Let $\vec{p} = (p_1, p_2, \ldots, p_T)$ be a vector indicating a value of p for each time period. Define

$$V(h \,|\, \vec{p}) := \text{cost incurred by deployment vector } h, \text{ if the infection probability equals } p_t \text{ at time } t.$$

$$(5.3)$$

[Here we remind the reader that time is measured relative to the declaration of the epidemic.] Let $\mathcal{P}$ indicate a set of vectors $\vec{p}$ of interest; our uncertainty set. Our robust problem can now be formally stated, as follows:

Robust Optimization Problem

$$V^* := \min_{h \in \mathcal{H}} \max_{\vec{p} \in \mathcal{P}} \; V(h \,|\, \vec{p}). \qquad (5.4)$$

Problem (5.4) explicitly embodies the adversarial nature of our model. Given a choice of vector h, a fictitious adversary chooses that realization of the problem data (the contagion probabilities) that maximizes the ensuing cost; the task for a planner is to minimize this worst-case cost. The set $\mathcal{H}$

describes how staff deployment plans are constrained. In our implementations and testing, we assume that any deployed staff person works for a given number τ of time periods, or until he or she gets sick; following this period of service this person will not be available for re-deployment. We also assume that whenever surge-staff is called up, there is a one-period lag before actual deployment (this assumption is easily modified to handle other time lags).

Finally, we model the impact of the epidemic on surge staff using the SEIR model with parameters as for the workforce group (λ_t^2 and μ_{E_2}), and we assume that the total quantity of available surge staff is small enough so as to not modify the outcome of the epidemic.

In Chapter 7 we will present an efficient procedure, based on linear programming, for solving problem (5.4) to very tight numerical tolerance.

6

Numerical Experiments

In this chapter we investigate the structure of the optimal robust policies by numerical experiments.

First, it is worth raising a point concerning "approximate" vs "exact" solutions. Our algorithms construct numerically optimal solutions to the formal problems they solve; nevertheless, as argued above, our models approximate real problems. Again, we postulate that given the context (i.e. the behavior of social entities) an approximation is the best we can hope for; moreover we expect that robust strategies computed for the approximate models will translate into actual robustness in practice. This last feature can at least be experimentally tested through simulation, as we present via examples in this section.

In order to assign numerical values to the various parameters in our models, we followed the existing literature and consulted with various experts **[67, 56]**. The numbers in the literature may vary significantly depending on various factors such as the location where the study took place, the under-

lying model for which the parameters were calibrated, and the epidemic wave of study. For example, published values for the influenza R_0, the number of secondary transmissions caused by an infected individual in a susceptible population, range from 1.68 to 20 [53]. Our main sources are listed in Table 6.1.

6.1 Examples in a health care setting

We now present examples modeling a hospital of the size of New York-Presbyterian Hospital in New York City under slightly different scenarios. Staff at this hospital amounted to $19,376$ in 2010 (we have rounded this up to $20,000$) and served a population of approximately $900,000$ people [59]. We used these numbers as a baseline to obtain different scenarios with different initial conditions.

In the event of an epidemic, hospitals will be likely to experiment surge demand trailing the incidence curve. We use the queueing-based cost function to capture both shifts in demand and workforce availability while the epidemic ensues. This increase in demand for service is modeled by adding a proportional factor of I_t^1 to the average arrival rate of patients when we compute the system's occupation rate, ρ_t:

$$\rho_t = \frac{\bar{\zeta} + \delta \, I_t^1}{s_t \mu}. \tag{6.1}$$

6.1.1 Example 1

For this example we assume 3,000 surge staff members are available, each of which will be called up for at most one period of service lasting one week. If the member becomes infectious while on call, then service is terminated with no further future availability.

Table 6.1: List of references for different parameter values.

Parameter	Description	Value(s)	Reference
Epidemic related parameters			
p	Probability of contagion	$[0.02, 0.16]^a$	[80]
		$\{0.07, 0.09, 0.12\}^b$	[47]
λ	Contact rate	47.48 (per day)	[79]
		35 (per day)^c	[47]
μ_E^{-1}	Latent period	1 day	[36]
		1.9 days^d	[50]
		1.25 days	[28]
		1.48 ± 0.48 days	[28] and within
μ_R^{-1}	Infectious period	2 days	[36]
		4.1 days	[50, 49]
		4.1 days^e	[49]
$1 - f$	Mortality rate	0.02	[50]
		0.002 per day^f	[28]
		1.2%^{g}	[19]
		0.05^g	[49]
R_0	R_0	1.68 to 20^h	[53] and within
		1.73^i	[79]
		2.5^j	[49]
	Length of pandemic wave	$8 - 12$ weeks	[28]
Hospital related measures			
ζ	Patient arrival rate	2,800 ILI cases/day^k	[49]
μ	Service rate	$24 - 30$min/patient^l	[33]
ρ	Occupancy rate	Average of 85%^{m}	[22]
δ	Increase in demand for health care	up to 50%	[58]
Miscellaneous			
	Seasonal threshold	$[0.06, 0.076]$ weekly^n	[26]
	Epidemic threshold	$[0.064, 0.08]$ weekly^n	[26]
	ILI National Baseline	2.4% weekly^o	[26]
θ	Transmission reduction factor	29% to 37%^{p}	[19]

^{a} 1957 wave in the Netherlands
^{b} Values of p for different socially active groups
^{c} Weighted average of different age groups presented in the paper
^{d} 1957 pandemic wave
^{e} Untreated symptomatic
^{f} Treated symptomatic
^{g} Overall ILI case-fatality ratio during 2009 H1N1 pandemic in Mexico
^{g} Spanish Flu fatality rate
^{h} From multiple studies
^{i} 1957 wave
^{j} Min 1.5, Max 6
^{k} Data for Singapore, 2005. Population then was about 4.35 million people
^{l} Emergency Department, New York Presbyterian Hospital
^{m} Maintained beds capacity for New Jersey Hospitals, 2005
^{n} The threshold is for percentage of deaths caused by pneumonia and influenza.
It is different for each week throughout the year.
^{o} Percentage of influenza-like illness (ILI) patient visits reported through the U.S. Outpatient ILI Surveillance Network.
^{p} Estimated effect of social distancing measures during H1N1 epidemic in Mexico, 2009

We used the uncertainty model in Section 5.2.1 which allows p_t to change once. The change is further constrained to take place within a fixed range of time periods.

The social contact model assumed the nonhomogeneous-mixing rule described in section 4.2. Additionally, we assumed that social contact rates were dampened by a factor of 30% while the epidemic is declared; that is, during the days in which the growth of infectives is higher than a given threshold. Additionally, we assumed an epidemic is declared only once. In other words, once the rate of infectives slow down below the epidemic threshold, the epidemic is considered to be terminated.

The parameter values for this example are summarized in Table 6.2; the range of days where p is allowed to change are measured from the *start* of the epidemic (rather than the day the epidemic is declared).

In what follows, the staff deployment plan computed by our algorithm will be called the *Robust Policy*. To gauge its usefulness, we study its performance under different scenarios, each of which is defined by a tuple (p_1, p_2, d) representing the initial probability of contagion (p_1), the time period where it changes (d) and the value to which it changes (p_2).

A natural alternative to the Robust Policy is that which best responds to what could be construed as a worst-case scenario - the scenario which yields the highest cost when no contingency plan is implemented. Such a strategy can be obtained by solving a linear program as described in (7.5) once the tuple that implies the evolution of p_t in this scenario is identified. In this example such tuple is given by $(0.01092, 0.0135, 140)$; we will term it the *No-Action-Max-Cost tuple*, and the deployment strategy which achieves minimum cost under this tuple the *Naïve-worst-case Policy*. We compare the Naïve-worst-case and Robust Policies under the following scenarios:

Table 6.2: Parameter values for example 1

Parameter	Description	Value(s)
Epidemic related parameters		
P	Uncertainty set	$[0.01, 0.012] \times [0.0125, 0.0135]$
	Possible days of change	$\{140, \dots, 160\}$
(Λ_1, Λ_2)	Contact rates	$(30, 35)$ per day
μ_E^{-1}	Latent period	1.9 days
μ_R^{-1}	Infectious period	4.1 days
$1 - f$	Mortality rate	0
R_0	R_0	$[1.24, 1.48]$
	Deployment threshold	2.4% weekly
Population parameters		
N_1	General population size	900,000
N_2	High risk population size	20,000
$[I_1, I_2]$	Initial infectives	[5,0]
System Utilization (Queueing setting)		
$\bar{\zeta}$	Patient arrival rate	500 per day
μ	Service rate per person	30 patients per day
ρ_0	Initial occupation rate	0.875
δ	Daily increase in demand for health care	0.07%
Deployment parameters		
	Total number of available volunteers	$3,000$
τ	Length of stay (w/o sickness)	7
	Deployment lag	1

1. No-Action-Max-Cost tuple $(0.01092, 0.0135, 140)$;

2. The tuple that achieves highest cost if the *Naïve-worst-case Policy* is implemented; and

3. The tuple that achieves highest cost if the *Robust Policy* is implemented.

Results are summarized in Table 6.3. In Scenario 1 the Robust Policy presents a cost improvement of almost 99% over the no-intervention policy. While it is not completely able to prevent ρ from exceeding 1, it reduces the number of critical days from 28 to 8, and the maximum ρ from 1.05 to below 1.002. Indeed, the undesirable instability in which a system is put through when $\rho > 1$ is heavily penalized with the use of the exponential function in our objective function (see 5.1.2). On the other

Table 6.3: Results. Example 1.

Scenario / Strategy / tuple		No Intervention	Robust Policy	Naïve-worst-case Policy
1	Cost	4.5812	0.0495	0.0000
No intervention	Maximum ρ	1.0481	1.0018	0.9999
(0.01092,0.0135,140)	# days $\rho \geq 1$	28	8	0
2	Cost	1.4300	0.0500	0.7100
Naïve-worst-case Policy	Maximum ρ	1.0210	1.0020	1.0185
(0.01172, 0.0135, 140)	# days $\rho \geq 1$	20	8	13
3	Cost	1.6938	0.0521	0.6862
Robust Policy	Maximum ρ	1.0236	1.0027	1.0170
(0.01168, 0.0135, 140)	# days $\rho \geq 1$	21	7	12

hand, under Scenario 1 the Naïve-worst-case Policy reduces the cost to 0; that is, there is no day in which ρ is greater than 1. Without doubt, it represents a better solution than the Robust Policy *for this particular case*. This result is expected; the solution derived from that single linear program is *specialized* in this scenario, while the Robust Policy, having to hedge against other possible bad cases, does not perform as well in this instance.

The situation is considerably different, however, when we look at Scenario 2, which achieves the highest cost when the Naïve-worst-case Policy is being implemented. In this case, the Robust Policy performs much better than the Naïve-worst-case Policy: the former reduces the cost of this scenario by 96.5%, while the latter does so by only 50%; the Robust Policy reduces the number of critical days from 20 to 8 days, while this number jumps to 13 for the Naïve-worst-case Policy. The maximum value of ρ is 1.002 for Robust Policy and 1.018 for scenario 2, compared to 1.021 with no intervention. The third scenario presents similar characteristics; what is worth noticing is that even though this is the *highest-cost* scenario for the Robust Policy, its cost is still the same as in Scenario 2.

This fact is further illustrated in Figure 6.1, where we compare the behavior of ρ_t with and without
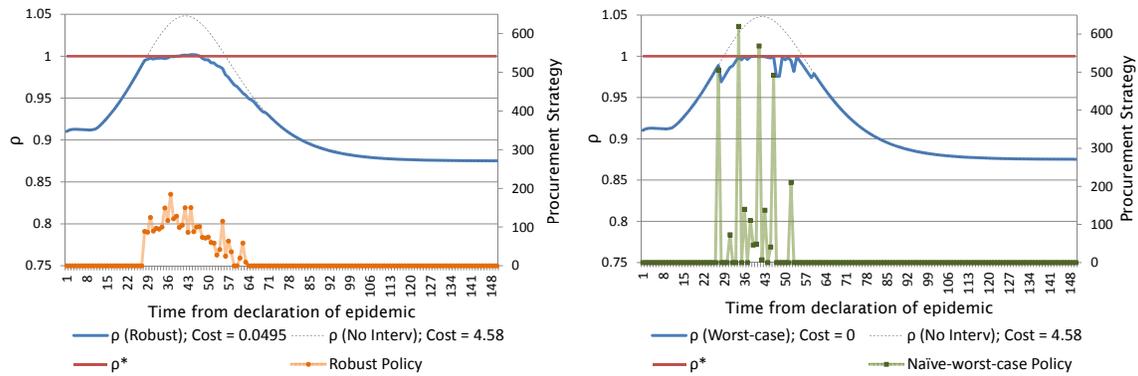
(a) **Scenario 1**:Effectiveness of Robust Policy

(b) **Scenario 1**:Effectiveness of Naïve-worst-case Policy
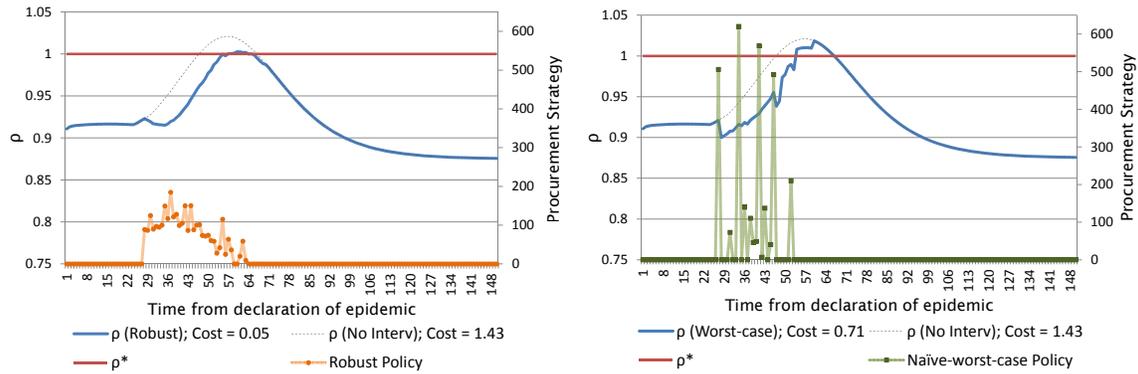
(c) **Scenario 2**:Effectiveness of Robust Policy

(d) **Scenario 2**:Effectiveness of Naïve-worst-case Policy
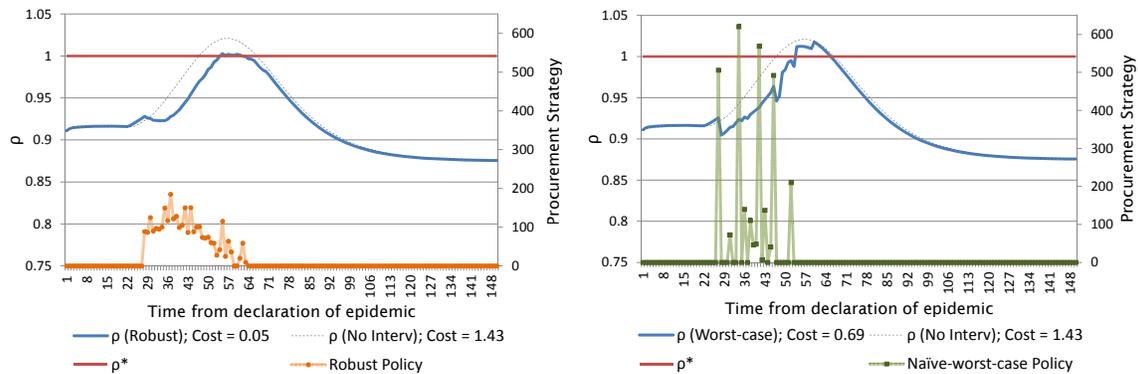
(e) **Scenario 3**:Effectiveness of Robust Policy

(f) **Scenario 3**:Effectiveness of Naïve-worst-case Policy

Figure 6.1: Reduction in ρ obtained by the Robust and Naïve-worst-case deployment strategies under three scenarios: 1) Highest-cost scenario given that no policy is implemented, 2) Highest-cost scenario given that the Naïve-worst-case Policy is implemented, and 3) Highest-cost scenario given that the Robust Policy is deployed.

interventions for the three scenarios, together with the deployment patterns. We note that the Naïve-worst-case Policy deploys staff around in large, sharp bursts between days 27 and 52. The Robust Policy deployment, on the other hand, trails ρ in Scenario 1 in a smoother way by calling in personnel in smaller batches, and also over a longer period of time: 36 days. This shows the importance of Scenario 1 as the costliest case and, at the same time, how the strategy hedges against scenarios where the epidemic has its peak later (than under the No-Action-Max-Cost tuple), even if not so aggressively. In particular, the last wave of staff deployed by the Robust Policy (on days $54, 58$, and 63) could seem somehow wasteful under Scenario 1; however their utility is displayed under Scenario 2, where the Robust Policy tackles a delayed, weaker ρ peak much more effectively than the Naïve-worst-case Policy. Because an unstable system builds up a queue exponentially fast, even this small changes in ρ could translate in significant difference in the load experienced by the system.

A similar situation is presented in Scenario 3, as illustrated in both Figure 6.1 and Table 6.3. The Naïve-worst-case Policy is not as effective as handling what would have been a milder epidemic, while the Robust Policy performs consistently well. In fact, this is the scenario with highest cost for the Robust Policy; nevertheless it manages to significantly outperform the Naïve-worst-case Policy here as well. Indeed, a feature of the Robust Policy is its near-uniform behavior under all scenarios; this epitomizes the use of the term 'robust'.

Other strategies such as the Naïve-worst-case Policy, on the other hand, may be very sensitive to changes in p depending on how they are constructed. In this example we make the case that preparing only for the highest-cost scenario (under the No-Action Policy) is not necessarily enough to be prepared for other milder cases: This is critical from the point of view of a decision maker: as seen above, a policy built only to tackle such case does not allocate enough resources to other time periods in

which milder cases may need them the most. At the same time, being able to cut down the number of days in which the organization is completely overloaded should represent considerable economic and operational advantages for the institution implementing the contingency plan. One estimate of the costs saved in a given scenario could be derived as a function of the area between the ρ curves corresponding to the No-Action Policy and the Robust Policy, whenever $\rho > 1$.

Smoothing out the solution

An additional restriction on the implementation of the proposed Robust Strategy could be for it to be as regular as possible. In other words, it could be deemed more desirable to have a contingency plan with fewer fluctuations in the surge staff calls in terms of its applicability. There are several ways of imposing these constraints. We chose to incorporate the constraints

$$h_t > 0 \Rightarrow h_{t+1} \leq (1+r)h_t \quad \forall t, \tag{6.2}$$

where $0 < r < 1$ is a given tolerance. These constraints allow the policy to consider more even deployment blocks, avoiding drastic increases in the number of surge staff called in. It also discourages long sequences of days in which relatively few people are deployed, for the following days would be very constrained if a strong increase is needed.

At the same time, it is important to mention that these additional restrictions imply a loss in robustness: constraining the variability of the hiring policies hinders the structure of the strategy's best response. The obvious question is how much of this robustness would be lost to favor an easier implementation.

To answer this question and to illustrate the new policies, we repeated Exercise 1 for 3 values of
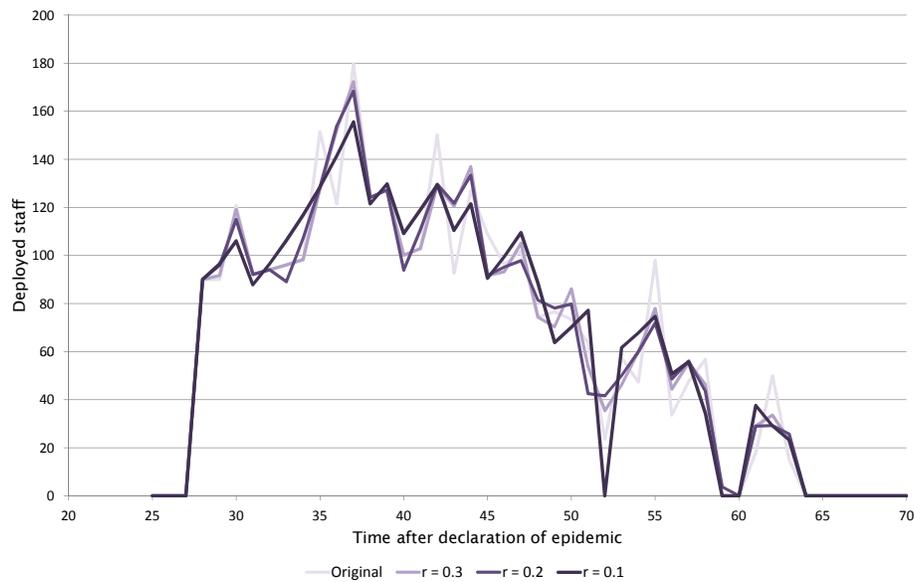
Figure 6.2: Example 1. Deployment strategies for different smoothing tolerances.

r : 0.1, 0.2, and 0.3. The changes in the structure of the deployment strategies can be observed in Figure 6.2 and the change in the highest cost in Table 6.4. The loss in robustness is modest, but it would be up to the decision maker to evaluate the trade off.

Table 6.4: Change in highest cost due to smoothing out solution.

	Original	r = 0.3	r = 0.2	r = 0.1
Highest cost	0.050767	0.051025	0.051308	0.052327
% increase		0.51%	1.07%	3.07%

In algorithmic considerations, note that these are so-called "If-then" constraints and they require the use of binary variables. However, this didn't present any computational difficulty: each iteration's running time did increase slightly, but the number of iterations didn't, making the change in total running time negligible[1].

---

[1]The variables and constraints are added to the master problem – See Chapter 7 for its definition.

Sensitivity of the grid's resolution

To compute the tuple that yields the highest cost at each iteration we make use of grid search over the interval(s) p_t belongs to. In this section we inspect the sensitivity of our algorithms' solution to different grid resolutions of our uncertainty set. We compare the obtained highest costs under the discretization they were computed with and under finer grid resolutions to gauge any loss in robustness. In particular, we present a finer discretization of the grid we used as a base example and three other coarser resolutions. We also present the tuples associated with such costs and the average CPU time required to compute them. Results are summarized in Table 6.5.

This exercise shows the importance of using the correct resolution to compute the final policies. Extremely fine grids may provide somehow more accurate approximations. At the same time, they can be computationally much more expensive. Comparing the Base case with the Super Fine case, we can see that the difference in the highest cost is minimal while the time to find the desired tuple for the finer grid is more than 30 times more than the time it takes for the Base case grid. As expected, even coarser grid searches provide faster but less robust solutions. While our algorithm performs well, this analysis suggests that one could make use of coarse discretizations for the first iterations and finer grids in subsequent iterations.

The resulting deployment strategies for the different grid resolutions are also quite dissimilar as shown in Figure 6.3. The deployment strategies resulting from coarser grids are represented in dashed, light-colored lines, while those obtained from finer discretizations are shown in darker, solid lines. The first thing to notice is that the differences between the strategies obtained from the Original and the Fine discretizations are negligible. On the other hand, as the resolution of the grid gets coarser, the staff is deployed in bigger, more sprawled bursts. The fact that the smoother strategies are more robust (see

Table 6.5: Change in highest cost for different discretizations of the uncertainty set. In this example, it is given by $P = [0.01, 0.012] \times [0.0125, 0.0135] \times \{140, ..., 160\}$. Different resolutions were taken for the first 2 intervals.

Discretization Step$_1$ × Step$_2$	cost tuple	Highest cost with given resolution	Highest cost with Base resolution	Highest cost with Fine resolution	CPU time worst-tuple	Size of Uncertainty Set
0.00001 x 0.00001 (Fine)		0.050950 (0.01111,0.0135,140)	-	-	125s	420K
0.00001 x 0.0005 (Base)		0.050768 (0.01172,0.0135,140)	-	0.050768 (0.01172,0.0135,140)	3.75s	16.8K
0.00005 x 0.00005 (Coarse 1)		0.004059 (0.01105,0.0135,156)	0.2515 (0.01172,0.0135,140)	0.2515 (0.01172,0.0135,140)	5.234s	33.6K
0.0001 x 0.0001 (Coarse 2)		0.000104 (0.0113,0.0135,140)	0.3141 (0.01172,0.0135,140)	0.3141 (0.01172,0.0135,140)	1.484s	8.4K
0.0005 x 0.0005 (Coarse 3)		0.002302 (0.011,0.0135,143)	0.2031 (0.01168,0.0135,140)	0.2031 (0.01168,0.0135,140)	0.094s	0.336K
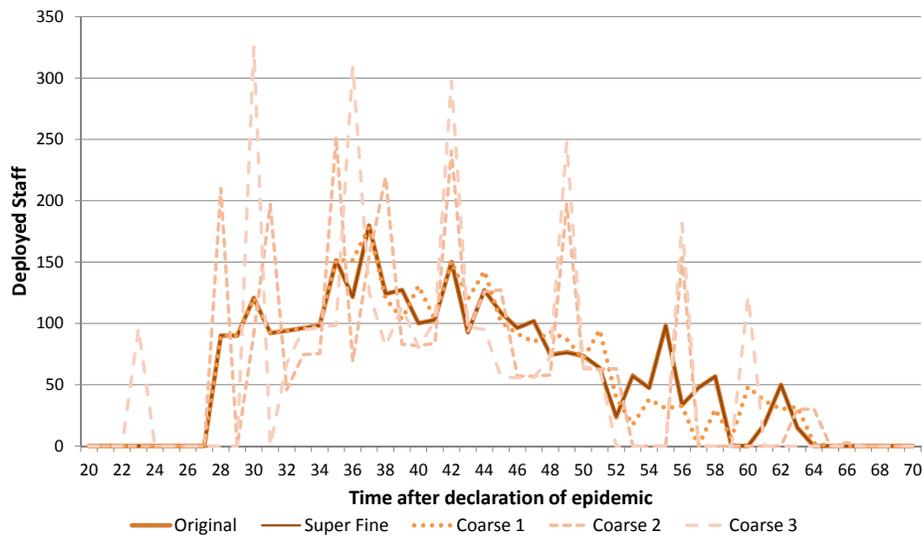
Figure 6.3: Example 1. Deployment strategies for different discretizations of the uncertainty set.

Table 6.5) shows that, in this particular example, a more even deployment of the surge staff is better capable of dealing with different epidemic scenarios. It is important to mention that this assumes that a certain amount of surge staff is available. We modify this assumption in the following example (Section 6.1.2).

Threshold objective function

In this section we present analogous results, now considering a threshold objective function in which cost increases as staff levels decrease (see 5.1.1 for more details). For this example we take 95% and 99% as the breakpoints of the piecewise linear function: whenever staff levels are below 95%, costs increase by 1 unit for every percentage point workforce availability is reduced; if staff levels are between 95% and 99% costs increase by 1/4 of a unit for each percentage point, and there is no cost whenever there is 99% or more of the workforce available. In mathematical terms, given a percentage of available staff at time t, $\omega_t \in [0, 1]$, the daily cost is given by

$$z_t = \max\{-100\omega_t + 96, -25\omega_t + 24.75, 0\}. \tag{6.3}$$

All other parameters remain as listed in Table 6.2.

As with the queueing model-based objective, we compare the Robust solution with the Naïve-worst-case Policy under different scenarios: 1) the No-Action-Max-Cost tuple $(0.01092, 0.0135, 140)$ and 2) the highest-cost tuple if the Naïve-worst-case Policy is deployed, $(0.01168, 0.0135, 140)$. In this case Scenario 1) coincides with Scenario 3); that is, the tuple that achieves the highest cost if the Robust Policy is implemented is the same as the No-Action-Max-Cost tuple. As with the queueing model-based objective function, while the Naïve-worst-case Policy fares better than the Robust Policy in Scenario 1), its cost can be up to 28% higher than the Robust Policy for other tuples. Moreover, given the number of surge staff available, the Robust Policy avoids going under the preestablished critical threshold of 95% in both scenarios, while the Naïve-worst-case Policy presents as many as 8 of these days. These statistics are exhibited in Table 6.6.

Table 6.6: Results with threshold function. Example 1.

Scenario / Strategy / tuple		No Intervention	Robust Policy	Naïve-worst-case Policy
1	Cost	83.1445	32.4503	32.2351
No intervention &	Max. absenteeism	6.90%	4.96%	5.00%
Robust Policy	# days AvWF < 95%	28	0	0
(0.01092,0.0135,140)				
2	Cost	64.3800	29.0090	37.3449
Naïve-worst-case Policy	Max. absenteeism	5.97%	4.27%	5.95%
(0.01168,0.0135,140)	# days AvWF < 95%	22	0	8

The structure of the deployment strategies together with its impact on the workforce availability (measured in percentage) for Scenarios 1 and 2 is displayed in Figure 6.4. It is worth nothing that if Scenario 2 took place, it could be thought that the Robust Strategy is somehow wasteful: staff levels surpass 100% around day 36 (Figure 6.4c). However, one must remember that the strategy must be resilient to all situations, including one such as Scenario 1, where there is no underutilization of the surge workers (Figure 6.4a). In the same scenario, the Naïve-worst-case strategy not only would deploy too many extra workers (around day 25) for many more days, it would allow the absenteeism to grow above 5% for 8 days as we mentioned above.

6.1.2 Example 2

To further illustrate the structure of the Robust Policy under different scenarios, we now consider the previous example with some modifications. Here, the uncertainty set is p lies in the interval $[0.01, 0.0125]$ throughout, but can change values once on any day in the range $\{100, \ldots, 115\}$. R_0 is now between 1.24 and 1.54. The total number of available volunteers is $2,000$ and the contact rates are not dampened while the epidemic is being declared. All other data remains as in the first example.

Following the same logic as in the previous example, we present the Robust Policy and compare it

(a) **Scenario 1**:Effectiveness of Robust Policy

(b) **Scenario 1**:Effectiveness of Naïve-worst-case Policy

(c) **Scenario 2**:Effectiveness of Robust Policy

(d) **Scenario 2**:Effectiveness of Naïve-worst-case Policy
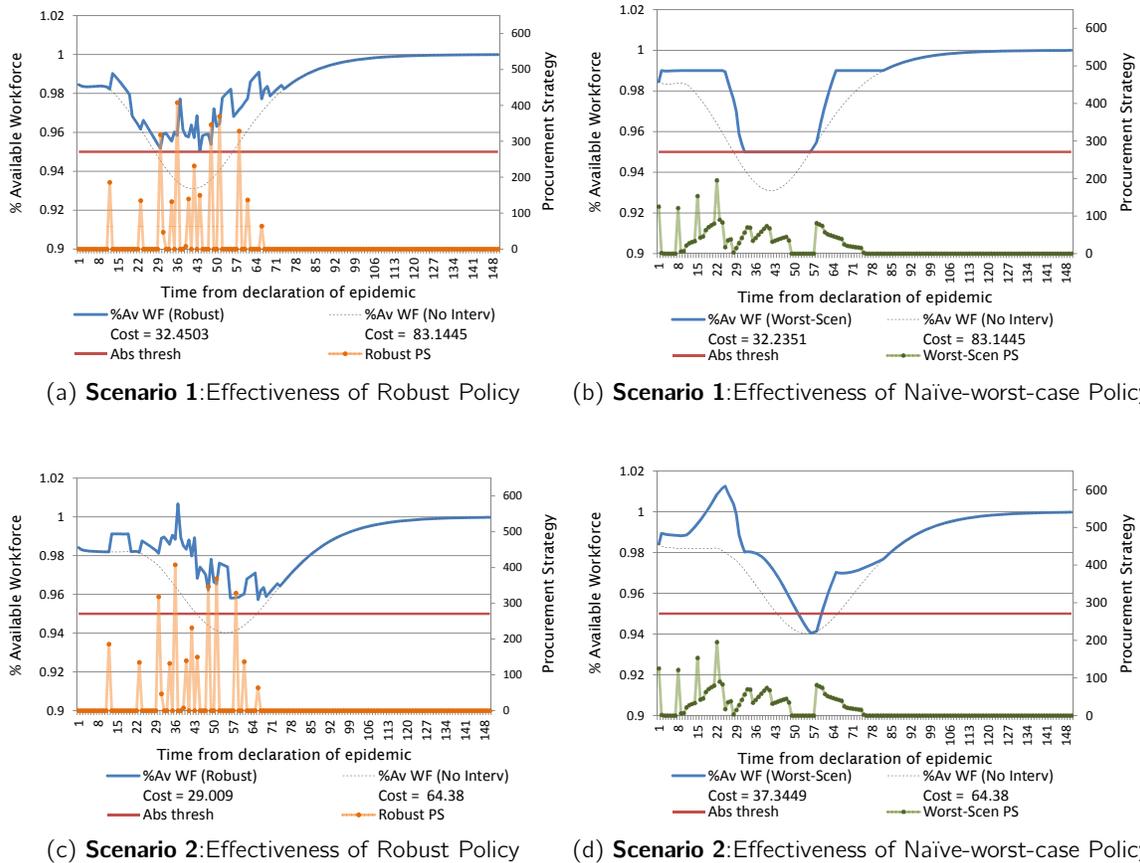
Figure 6.4: Changes in the workforce availability after the deployment of the Robust and Naïve-worst-case strategies under two scenarios: 1) Highest- cost scenario given that no policy is implemented, 2) Highest- cost scenario given that the Naïve-worst-case Policy is implemented.

to the Naïve-worst-case Policy (obtained as before by solving the linear program associated with the highest-cost tuple of our uncertainty set). We present results for the same three scenarios: the No-Action-Max-Cost tuple, the costliest tuple for the Naïve-worst-case Policy, and the costliest tuple for the Robust Policy, respectively. This fact hints on the non-convexity of the problem, a topic we will touch upon later in this section.

Table 6.7 displays the results. The Robust Policy is worse than the Naïve-worst-case Policy in the first instance, while keeping a fairly consistent and low cost for the other scenarios; where the Naïve-worst-case Policy does not perform as well. What is worth noting is that, unlike in the previous example,

Table 6.7: Results. Example 2.

| Scenario / Intervention / Tuple		No Intervention	Robust Policy	Naïve-worst-case Policy
1	Cost	3.8332	0.0280	0.0066
No intervention	Maximum ρ	1.0410	1.0012	1.0003
(0.0125, 0.0125, 100)	# days $\rho \geq 1$	27	10	6
2	Cost	3.5906	0.0280	0.0542
Naïve-worst-case Policy	Maximum ρ	1.0393	1.0019	1.0025
(0.01235, 0.0125, 112)	# days $\rho \geq 1$	26	7	8
3	Cost	3.7320	0.0295	0.0379
Robust Policy	Maximum ρ	1.0403	1.0010	1.0017
(0.01015, 0.0125, 101)	# days $\rho \geq 1$	26	10	10

Figure 6.5: Example 2. Structure of the deployment strategies.

the structure of the deployment strategies is much more similar (see Figure 6.5); nevertheless, the apparently small differences make the Robust Policy more resilient to other scenarios. A second observation regards the structure of the highest-cost tuples evinced by the policies. In this example the No-Action-Max-Cost tuple corresponds to having an epidemic with a *fixed* probability of contagion $p = 0.0125$, the highest in the interval of the uncertainty set. However, the costliest tuples for the Robust and Naïve-worst-case policies are quite different.

Cost-benefit analysis

We now present a cost-benefit analysis on the availability of surge staff. We took this example as a base case and recomputed the optimal Robust and the Naïve-worst-case policies assuming $1,000$, $1,500$, $2,500$, and $3,000$ surge staff on hand. Table 6.8 displays their corresponding worst-scenario tuples along with the associated costs for when the corresponding policy is being implemented and for when no contingency plan was put in place.

We point out that the difference in the highest-cost tuples explains the difference in the costs when no intervention takes place: the highest-cost scenarios are obtained assuming that a corresponding policy is being implemented. In fact, it is interesting to see how the structure of the tuples change with the different policies. When there is not much staff available, the worst scenario for both policies is given by the tuple $(0.0125, 0.0125, 100)$. The highest costs for the Naïve and the Robust Policies are very close, although the deployment strategies not as much. When there are $1,000$ and $1,500$ staff available, the Robust Strategies are more concentrated towards the middle, while the Naïve-worst-case Policies are more spread out (see Figure 6.6).

Table 6.8 also presents the change in the maximum value of ρ and the number of days that ρ is at or above 1. The last row of the table shows the ratio between the change in the highest cost and the change in the number of available surge staff. For example, when there are $1,500$ volunteers available, there is a change in the cost of 0.156% for each of the 500 that are now not at hand. At the same time, the benefit obtained per additional worker available decreases for both the Robust and the Naïve-worst-Case Policy, but more so for the latter. That is, the Robust Strategies make better use (cost-wise) of extra staff in terms of the highest cost. This is again because the Robust Policy "knows" how to allocate the additional staff to cover up for different bad cases, while the Naïve-worst-case

Policy has only one scenario to focus on. This fact is particularly noticeable when the Naïve Policy chooses to hire only 2,579 emergent workers out of the 3,000 that are available. While this amount of staff is enough to tackle the highest-cost scenario (under no intervention), it is not distributed well enough to tackle milder cases, whereas the Robust Policy has a maximum cost of 0 when so much staff is on hand.

Figure 6.6 illustrates the structure of the above mentioned strategies. To make the strategies comparable, we display the percentage of the staff that is called in on each day; percentages are stacked on top of each other. The plots further support our claims: for the cases in which the Robust Policy has more surge staff available, more people are hired towards the end of the planning horizon, while the Naïve-worst-case Policies look to tackle the system congestion early after the epidemic has been declared.

Table 6.8: Cost-Benefit Analysis for different quantities of available surge staff.

	Robust Policy				
Total staff deployed	1,000	1,500	2,000	2,500	3,000
Highest-cost tuple (p_1, p_2, d)	(0.0125,0.0125,100)	(0.0125,0.0125,100)	(0.01015,0.0125,101)	(-,-,-)	(-,-,-)
Highest cost	1.755971	0.818334	0.029545	0	0
Highest cost (No Int)	3.8332	3.8332	3.732	3.8332	3.8332
Maximum ρ	1.02	1.02	1.001	0.999	0.998
Maximum ρ (No Int)	1.04	1.04	1.04	1.04	1.04
# days $\rho \geq 1$	27	27	10	0	0
# days $\rho \geq 1$ (No Int)	27	27	26	27	27
Cost Change / Extra Surge Staff	0.345%	0.158%	-	-0.006%	-0.003%
	Naïve-worst-case Policy				
Total staff deployed	1,000	1,500	2,000	2,500	2,579
Highest-cost tuple	(0.0125,0.0125,100)	(0.0125,0.0125,100)	(0.01235,0.0125,112)	(0.01235,0.0125,105)	(0.01235,0.0125,113)
Highest cost	1.75351	0.811355	0.05418	0.018956	0.015555
Highest cost (No Int)	3.8332	3.8332	3.6077	3.6997	3.5906
Maximum ρ	1.02	1.01	1.002	1.002	1.002
Maximum ρ (No Int)	1.04	1.04	1.04	1.04	1.04
# days $\rho \geq 1$	27	27	8	3	2
# days $\rho \geq 1$ (No Int)	27	27	26	26	26
Cost Change / Extra Surge Staff	0.345%	0.156%	-	-0.002%	-0.002%

(a) Robust Policies.

(b) Naïve-worst-case Policies.
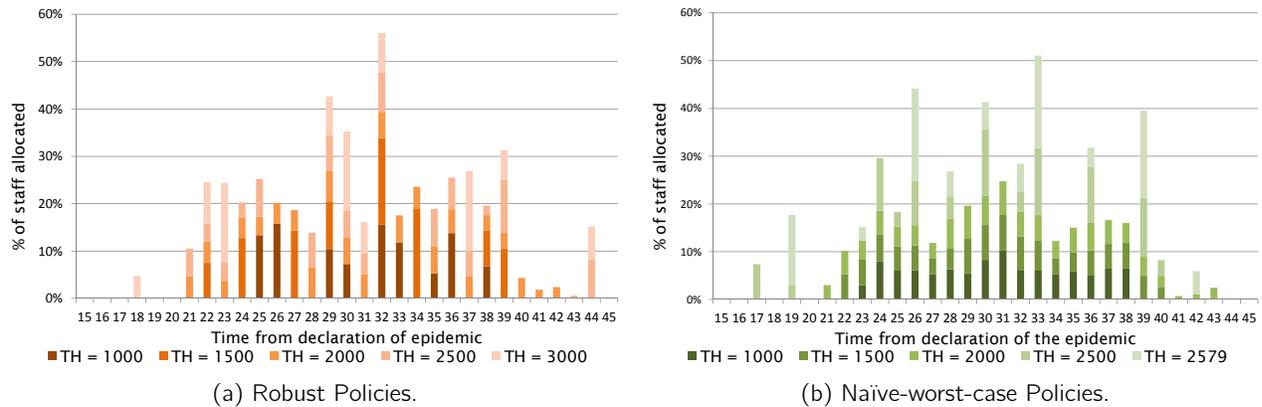
Figure 6.6: Changes in the distribution of the deployment strategies for different amounts of available surge staff.

Non-convexity

As the examples above make clear, the systems we study can be strongly dependent on the choice of p, with clear nonlinearities and non-convexities. This justifies the use of robustness in the design of response strategies.

To further study the dependence on p, we considered the cost of a set of strategies when p is constant throughout the epidemic. Figure 6.7a plots the objective function as a function of p in the interval $[0.0115, 0.0125]$ when there is no intervention; clearly it is monotone increasing in p. Figure 6.7b, on the other hand, shows the cost of three strategies for computed by our algorithms[2]. These curves are certainly not monotonic and, not surprisingly, particular values of p are able to exploit relative "gaps" in deployment. This fact has algorithmic consequences; we use grid search to look for the highest-cost tuple at each iteration of our algorithm. We delve into more details in Chapter 7.

---

[2]These strategies were obtained from intermediate iterations of our Robust algorithms. These are fully described in Chapter 7.

(a) Total cost as a function of p under no intervention.(b) Total cost as a function of p under 3 different strategies.
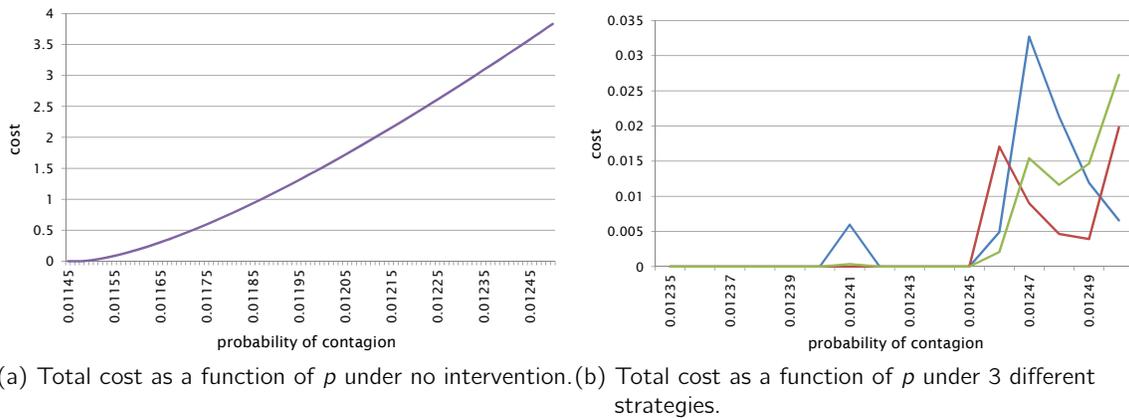
Figure 6.7: Total cost of an epidemic as a function of p under a)No intervention, and b) Three different deployment policies.

6.1.3 Out-of-sample tests

The set of experiments we present below amount to out-of-sample testing. Its purpose is to evaluate the robustness of our proposed strategies *outside* the uncertainty set they were built upon. We first build on Example 1 (Section 6.1.1) in which the initial uncertainty interval for p was $[0.01, 0.012]$, and the second interval was $[0.0125, 0.0135]$; p was allowed to change between days 140 and 160. We also recall that the contact rates were diminished by 30% while the epidemic is declared.

The first test goes about gauging the impact of modifying the tuple $(\tilde{p}_1, \tilde{p}_2, \tilde{d})$ that corresponds to the highest-cost scenario should the Robust Policy be implemented (Scenario 3), which is given by $(0.01168, 0.0135, 140)$. For this experiments the epidemic is declared on day 113 with deployment under the Robust Policy beginning on day 140, i.e. 28 days after the declaration (thus, p changes on the day of deployment).The experiments go out-of-sample in two ways: we made changes both in $\tilde{p}_2$ and $\tilde{d}$. $\tilde{p}_2$ was allowed to take values in the set $\{0.0135, 0.014, 0.015\}$ and $\tilde{d}$ in $\{125, 130, 135, 140, 150, 155, 160, 165\}$.

We refer the reader to Figure 6.8a to exemplify the results. Each solid curve represents the cost of

a fixed pair of the form $(0.01168, \tilde{p}_2)$; the different days in which the change in p occurs are indexed in the x axis. For example, the darkest colored curve corresponds to the cost of tuples of the form $(0.01168, 0.015, \tilde{d})$, with $\tilde{d}$ varying in the x axis. As benchmark, the dotted lines represent the cost of the No-Intervention Policy for the same tuples.

The first thing to point out is that any change before day 130 translates into zero cost for both, the Robust Policy and the No-Intervention strategy. This is because of the contact rate dampening factor. Given the specified parameters in this experiment, early changes in the probability of contagion p result in extended periods of time in which social distancing is implemented. We admit this could be too strong an assumption; social distancing measures that include the closure of public spaces such as restaurants, theaters, and stadiums could be prohibitively expensive. Further numerical experiments could be done limiting the number of days the social distancing measures are applied. At the same time, it highlights the clout of such measures should they effectively diminish the contact rates by that amount. This phenomenon does not occur in the case of a late changes in p, wherein the epidemic does not last for more than 20 days.

The second fact to point out is that for any fixed day (after 130) the cost of the Robust Policy does increase over its highest cost for higher values of $\tilde{p}_2$ — however, the increase is modest relative to the cost of not intervening. Most importantly, the cost of the Robust Policy decreases rapidly as $\tilde{d}$ moves out of scope of the original model. In fact, the same is true even under the No-Intervention Policy.

This is the salient fact that we want to point out. Its explanation is simple: even though p is increasing to a larger value than originally envisioned, if $\tilde{d}$ is large, the impact of the increase is negligible — the increase is taking place so late, that by then the epidemic (under the initial p value) has largely run its course. In other words, we have a long-running epidemic which, *under the Robust Policy*, ends up
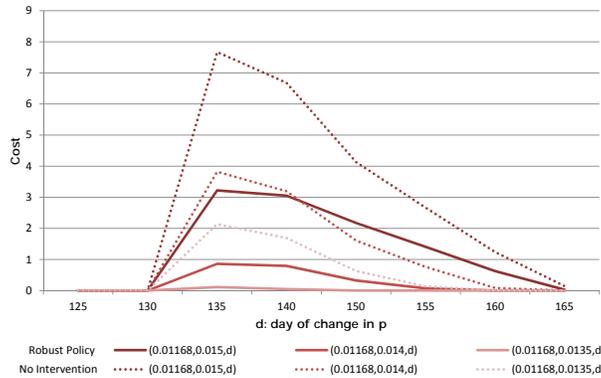
(a) Out-of-Sample Testing 1. Change in the cost of Robust Strategy due to changes its highest-cost tuple.
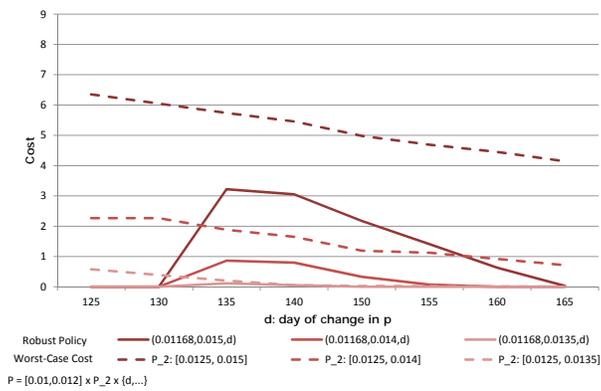
(b) Out-of-Sample Testing 2. Change in the cost of Robust Strategy due to changes its highest-cost tuple.
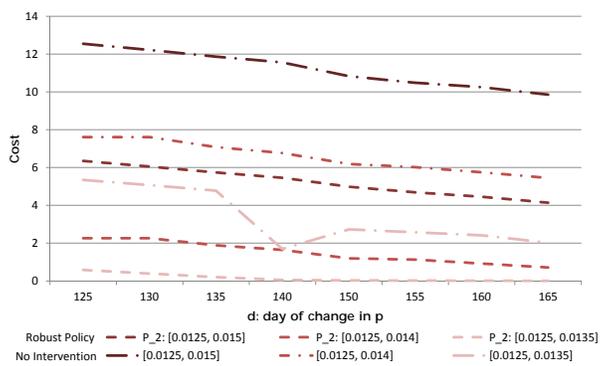
(c) Out-of-Sample Testing 3. Change in the cost of Robust Strategy due to changes its highest-cost tuple.

Figure 6.8: Out-of-sample testing: Changes in the cost of the Robust Strategy for scenarios that are not in the uncertainty set it was built upon.

having no impact. Numerical details of the change in cost given a change in p after day 140 for both, the Robust and the No-Intervention Policies, are reported in Table 6.9.

Table 6.9: Effects of a late change in p on Scenario 3

Scenarios	Original	Hypothetical 1				Hypothetical 2			
p_1	0.01168	0.01168				0.01168			
p_2	0.0135	0.014				0.015			
day epidemic declared	113	113				113			
day deployment starts	140	140				140			
day change p	140	150	155	160	165 -	150	155	160	165
cost Robust Policy	0.0508	0.3268	0.0729	0	0	2.1737	1.4068	0.6282	0.0294
cost No Intervention	4.58	1.6087	0.7619	0.087	0	4.1327	2.6688	1.2427	0.1462

Along the same lines, it is also worthwhile to note that in all cases where the out-of-sample increase in p actually does have a cost impact, the change takes place shortly after the epidemic declaration and very shortly after deployment of the robust strategy begins. These facts indicate that there is a critical period (in this example of duration less than forty days or even shorter) starting from the declaration of the epidemic during which *correct* action is important. To some degree this supports our model of rolling-out a fixed strategy that deals with the immediate future; should the epidemic "slow down" only to restart much later, a completely new plan would then be deployed.

A second, more general, related out-of-sample exercise modifies the uncertainty set directly (vs changing the highest-cost tuple only, as in the previous test). With the use of Robust Optimization, it is natural to inspect what the highest cost would be should the uncertainty set change as a whole. Again, the tests were constructed in two ways: The uncertainty sets are of the form $[0.01, 0.012] \times [0.0125, \tilde{p}_2] \times \{\tilde{d}, ...\}$[3]. Once again, we let $\tilde{p}_2$ be either $0.0135, 0.014$ or 0.015 and $\tilde{d}$ to be within the set $\{125, 130, 135, 140, 150, 155, 160, 165\}$. The Robust Policy's highest cost for each of these uncertainty sets are represented by the dashed lines in Figure 6.8b. The corresponding

---

[3]The second interval, $[0.0125, \tilde{p}_2]$ is referred to as P_2 in Figure 6.8b
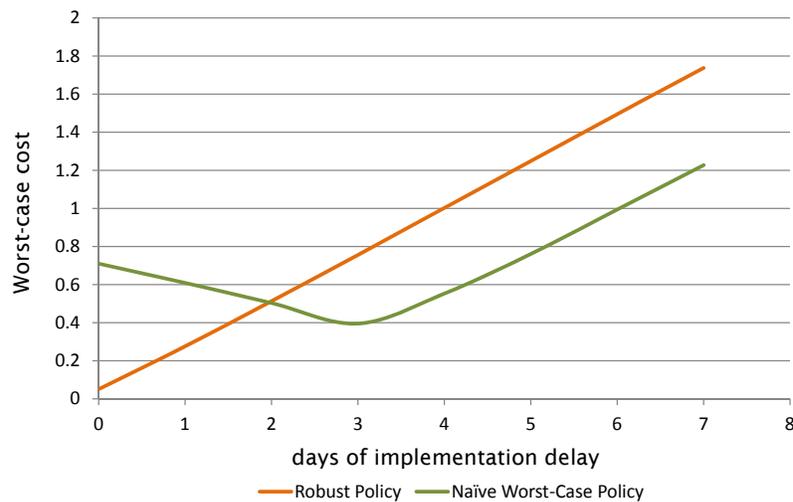
Figure 6.9: Example 2. Out-of-sample testing: Implementing the Robust and Naïve strategies *n* days late.

costs for the No-Intervention policy are shown in dotted lines in Figure 6.8c. These two charts reinforce what we had previously argued: while there is an increase in the total cost for out-of-sample scenarios, it is moderate, especially if compared against the No-Intervention Policy. Moreover, late changes in the probability of contagion have a milder impact in the total cost because the epidemic has largely run its course by then. The implementation of the Robust Policy helps to bring these costs further down.

To finish this section we present a last out-of-sample test, also built upon Example 1. Its purpose is to measure the impact of a delay in the implementation of the Robust Policy. That is, we are interested in the change in the highest cost if the policy is shifted *n* days late. This could happen if the decision maker fails to declare the epidemic on time, for example. We compare its performance by making the same computations for the Naïve Policy. Results are presented in Figure 6.9. It is interesting to see how the Naïve Policy outperforms the Robust Policy after 3 days of delay. This is coherent with our previous results. The test just checks how sensitive policies are when they are *shifted to the right*. The Naïve Policy was specialized in tackling an early, very aggressive epidemic. As a consequence, when

adjourned a few days it becomes resilient to more bad cases, as the unshifted Robust Policy is. At the same time, if postponed too many days, the Naïve Worst-Case Strategy also loses its effectiveness as expected. While this could translate in the Naïve Worst-Case Strategy being more attractive to implement than the Robust Strategy, this could also imply that we should incorporate this parameter into the uncertainty set. We leave that exercise for future work.

7

Solving the Robust Problem

In this chapter we describe an algorithm for solving the robust optimization problem (5.4), which is worth restating:

$$V^* := \min_{h \in \mathcal{H}} \max_{\vec{p} \in \mathcal{P}} \ V(h \,|\, \vec{p}). \tag{7.1}$$

We will formulate this problem as a semi-infinite linear program and use a variant of Benders' decomposition [13] to solve it. In this chapter we also present procedures to improve the numerical performance of an otherwise purely theoretical algorithm. We begin by characterizing the representation of the cost incurred by a given deployment vector under a given vector of contagion probabilities $\vec{p}$, $V(h \,|\, \vec{p})$ as a linear program, a critical component of our method.

7.1   The robust problem as an infinite linear program

Let h be a given surge staff deployment vector, and let $\vec{p}$ describe an evolution of the contagion probability. We recall that the i-th element of h dictates how many members of our surge staff are called i days after the epidemic is declared. $\vec{p}$ dictates the value of the probability of contagion at each point in time (since the epidemic starts) and determines the severity of the epidemic once *all other parameters are fixed.*

We now compute $V(h\,|\,\vec{p})$, the cost induced by h *given* $\vec{p}$, representing it as the value of a linear program. To this end, we observe that this quantity captures the impact on the operations performance of the total availability of the regular workforce together with that of the surge staff. This availability –for both groups– is given by the intensity of the epidemic, that is, by $\vec{p}$. All other parameters fixed, knowing $\vec{p}$ allows us to calculate the expected trajectory of the epidemic and thus, the total workforce availability at each time period. In particular, we have knowledge of what the absenteeism will be like for the regular workforce, and the conditions the surge staff would be subject to should they be called in (via h).

To this effect, let h and $\vec{p}$ be given, let $(S_0^j, E_0^j, I_0^j, R_0^j,\ j = 1, 2)$ denote the initial *SEIR* distribution, and let all the other *SEIR* parameters be fixed. As we just described, these conditions fully describe the predicted epidemic trajectory, particularly the values β_t for all t. These in turn fully define the disease dynamics of the deployed staff through equation (7.2) given below.

The next step is to obtain a linear inequality representation of the quantity of available surge staff at any given time t. Define:

- $S_{t,j}^3$, the number of surge staff deployed at time $t - j$ that remain susceptible on day t, and

- $E_{t,j}^3$, the number off exposed surge staff deployed at time $t - j$.

As before, T is the time horizon for the epidemic. Using this notation, we have that:

$$S_{t+1,1}^3 = h_t, \qquad\qquad t = 1, ..., T - \tau$$

$$S_{t+1,j+1}^3 = e^{-\lambda_t^2 \beta_t p_t} S_{t,j}^3, \qquad t = 2, ..., T - 1, \quad j = 1, ..., \min\{t - 1, \tau - 1\}$$

$$E_{t+1,2}^3 = (1 - e^{-\lambda_t^2 \beta_t p_t}) S_{t,1}^3, \qquad t = 2, ..., T - 1$$

$$E_{t+1,j+1}^3 = (1 - e^{-\lambda_t^2 \beta_t p_t}) S_{t,j}^3 + e^{-\mu_{E_2}} E_{t,j}^3, \quad t = 3, ..., T - 1, \quad j = 2, ..., \min\{t - 1, \tau - 1\}$$

(7.2)

We note that the parameters λ_t^2, μ_{E_2} correspond to those of the workforce subgroup (and not the general population). This follows the assumption that the surge staff will be in the same circumstances as the workforce of interest.

Given (7.2) and assuming that as soon as a staff member (regular or surge) becomes infectious he/she does not show up for work, the available *surge* staff at each point in time t is given by

$$\sum_{j=1}^{\min\{t,\tau\}} S_{t,j}^3 + \sum_{j=1}^{\min\{t,\tau\}} E_{t,j}^3, \quad \text{for } t = 2, ..., T - \tau, \tag{7.3}$$

where we have set $E_{2,j}^3 = 0$.

Similarly, the total available *regular* staff at time t is given by the sum

$$S_t^2 + E_t^2 + R_t^2. \tag{7.4}$$

Thus, the sum of (7.4) and (7.3) yields the total available workforce at time t, the quantity of interest

to compute the desired operational performance.

Now we return to the robust optimization problem (7.1). For each period t, let ω_t denote the total number of available workers at time t. Based on discussions above, we assume that the cost function to be optimized is given by a convex piecewise-linear function of the form $\max_{1 \leq i \leq L} \{\sigma_i \omega_t + k_i\}$, for appropriate L, σ and k. This holds for the threshold function case (section 5.1.1) and for the queueing case (section 5.1.2).

Using notation as in equation (5.3), we therefore have

$$V(h \mid \vec{p}) := \min \sum_{t=1}^{T} z_t$$

s.t.

$$S_{t+1,1}^3 = h_t, \qquad\qquad\qquad t = 1, ..., T - \tau$$

$$S_{t+1,j+1}^3 = e^{-\lambda_t^2 \beta_t p_t} S_{t,j}^3, \qquad\qquad t = 2, ..., T - 1,$$

$$j = 1, ..., \min\{t - 1, \tau - 1\}$$

$$\text{(7.5)}$$

$$E_{t+1,2}^3 = (1 - e^{-\lambda_t^2 \beta_t p_t}) S_{t,1}^3, \qquad\qquad t = 2, ..., T - 1$$

$$E_{t+1,j+1}^3 = (1 - e^{-\lambda_t^2 \beta_t p_t}) S_{t,j}^3 + e^{-\mu_{E_2}} E_{t,j}^3, \qquad t = 3, ..., T - 1,$$

$$j = 2, ..., \min\{t - 1, \tau - 1\}$$

$$\omega_t = S_t^2 + E_t^2 + R_t^2 + \sum_{j=1}^{\min\{t-1,\tau\}} S_{t,j}^3 + \sum_{j=1}^{\min\{t-1,\tau\}} E_{t,j}^3, \quad t = 1, ..., T$$

$$z_t \geq \sigma_i \omega_t + k_i, \qquad\qquad\qquad 1 \leq i \leq L, \ t = 1, ..., T$$

In this formulation S, E, ω and z are the variables (they should be indexed by $\vec{p}$, but we omit this for simplicity of notation). Notice that the quantity h_t appears explicitly in only the first set of

constraints. Also, by construction all variables are nonnegative. We can summarize this formulation in a more compact form. Using appropriate matrices $A_{\vec{p}}$ and $C_{\vec{p}}$ and vectors $\kappa_{\vec{p}}$ and $d_{\vec{p}}$,

$$V(h|\vec{p}) \ := \ \min_x \kappa'_{\vec{p}} x \tag{7.6}$$

$$\text{s.t.} \quad A_{\vec{p}} x = h \tag{7.7}$$

$$C_{\vec{p}} x \geq d_{\vec{p}} \tag{7.8}$$

Here, x denotes a vector of auxiliary variables comprising all S, E, w and z. We can now write:

$$V^* \ := \ \inf_{h,x} \quad \nu$$

$$\text{s.t.} \quad \nu \geq \kappa'_{\vec{p}} x, \quad \forall \vec{p} \in \mathcal{P}$$

$$A_{\vec{p}} x = h, \qquad \forall \vec{p} \in \mathcal{P} \tag{7.9}$$

$$C_{\vec{p}} x \geq d_{\vec{p}}, \qquad \forall \vec{p} \in \mathcal{P}$$

$$h \in \mathcal{H}.$$

The feasible set of problem is an intersection of closed, convex sets (provided $\mathcal{H}$ is of this form). Hence, we can substitute infimum for minimum. We note that even in the simple case in which $\mathcal{P}$ is an interval (7.9) is a *semi-infinite linear program*. This motivates the use of cutting-plane algorithms. A fine discretization of such an interval creates a prohibitively large number of constraints, and as we mentioned in Section 6.1.1, this can affect the quality of the solution.

7.2 Algorithm

We can now describe our procedure for solving optimization problem (7.9). We make use of the dual of (7.6)-(7.8) to construct cuts that will approximate the desired objective function.

Let π, α be *feasible* dual vectors for the LP (7.6)-(7.8), corresponding to (7.7) and (7.8), respectively. Then by weak duality,

$$V(h \mid \vec{p}) \geq \pi' h + \alpha' d_{\vec{p}}. \tag{7.10}$$

Furthermore, π and α are optimal for the dual if and only if

$$V(h \mid \vec{p}) = \pi' h + \alpha' d_{\vec{p}}. \tag{7.11}$$

Denoting by $D(\vec{p})$ the set of feasible duals to LP (7.6)-(7.8), it follows that

$$V(h \mid \vec{p}) = \max_{(\pi,\alpha) \in D(\vec{p})} \pi' h + \alpha' d_{\vec{p}}. \tag{7.12}$$

and so we can rewrite (7.1) as

$$V^* = \min_{h \in H} \max_{\vec{p} \in P} \max_{(\alpha,\pi) \in D(\vec{p})} \alpha' h + \pi' d_{\vec{p}}. \tag{7.13}$$

Now consider a finite family of vectors (π_k, α_k) $(k = 1, \ldots, K)$ such that for each k there is a vector $\vec{p}(k) \in \mathcal{P}$ with $(\pi_k, \alpha_k) \in D(p^k)$. Then, by (7.13), we have

$$V^* \geq \min_{h \in H} \max_{1 \leq k \leq K} \alpha_k' h + \pi_k' d_{\vec{p}(k)}. \tag{7.14}$$

This observation gives rise to Algorithm B.

<div align="center">

Algorithm B

</div>

---

0. Set $K = 0$ and $r = 1$.

1. Let h^r be an optimal solution for the LP

$$W^r := \min_{z,h} \quad z$$
$$\text{s.t.} \quad z \geq \alpha'_k h + \pi'_k d_{\vec{p}(k)}, \quad 1 \leq k \leq r - 1$$
$$h \in \mathcal{H}.$$

2. Let $\vec{p}(r) \in \mathcal{P}$ be such that

$$V(h^r \mid \vec{p}(r)) = \max_{\vec{p} \in \mathcal{P}} V(h^r \mid \vec{p}), \tag{7.15}$$

Note: $\vec{p}(r)$ is the contagion probability vector that attains the worst case should deployment vector h be used.

3. Using notation as in formulation (7.6)-(7.8), let (π_r, α_r) be optimal duals for the LP

$$\min_x \kappa'_{\vec{p}(r)} x \tag{7.16}$$
$$\text{s.t.} \quad A_{\vec{p}(r)} x = h^r \tag{7.17}$$
$$C_{\vec{p}(r)} x \geq d_{\vec{p}(r)} \tag{7.18}$$

Note: The value of this LP is $V(h^r | \vec{p}(r))$.

4. If $W^r \geq V(h^r \mid \vec{p}(r))$, **STOP** − algorithm has terminated.
Else, reset $r \leftarrow r + 1$ and go to **1**.

Remark: The above algorithm can be viewed as a special case of Benders' decomposition **[13]**. The linear program solved in Step 1 is called the *master problem*.

Lemma 7.2.1 *For any $r \geq 0$ we have (a) $W^r \leq W^{r+1}$ and (b) $W^r \leq V^* \leq V(h^r \mid \vec{p}(r))$.*

Proof. (a) This follows because in each iteration we add a new constraint to the problem solved in Step 1. (b) This follows as per equation (7.14). ∎

Corollary 7.2.2 *If the algorithm terminates in Step 4, it has correctly solved problem (7.1).*

In practice we would not stop the algorithm in Step 4, but rather use part (b) of Lemma 7.2.1 to stop when a desired optimality guarantee is reached. An issue of theoretical interest is the rate of convergence of Algorithm B. The following result indirectly addresses this question.

Lemma 7.2.3 *There is an algorithm that computes V^* by solving problem (7.15) a polynomial number of times.*

We stress that the algorithm in Lemma 7.2.3 is *not* Algorithm B; rather, it relies on the well-known equivalence between separation and optimization **[35]** and it is similar to Algorithm B except that (essentially) the computation of h^r in Step 1 is performed differently. The resulting algorithm, while theoretically efficient, may not be practical. Instead, in practice researchers in the nonconvex optimization community would rely on a classical cutting-plane algorithm such as Algorithm B. In the following section we give more details about its implementation.

7.2.1 Implementation

The above discussion highlights the pivotal elements of the implementation of the algorithm, particularly the ones that help build the correct cuts that are added to the master problem at each iteration:

SEIR **model:** Given a set of initial parameters[1], and a deployment strategy h, the algorithm heavily depends on a routine that computes and keeps track of the progression of the epidemic for both population subgroups and the surge staff. This is done by using equations 4.5. We remind the reader that the surge staff is subject to the epidemic as soon as they are called in and that we assume that the parameters of the *SEIR* model we'll use are those of the regular workforce (or subgroup 2).

The utility of such routine is mainly twofold: *i*) Combined with the appropriate objective function, it is used to compute the worst-case tuple given a deployment strategy h^r (Step 2). *ii*) Given a deployment vector h^r and its associated worst-case $\vec{p}(r)$, it generates the necessary data that will be fed as coefficients in the auxiliary linear program of our algorithm (Step 3).

Highest-cost tuple: The discussion in the previous section brings attention to the importance, both theoretical and practical, of having a fast way of computing $\text{argmax}_{\vec{p} \in \mathcal{P}} \ V(h|\vec{p})$ for a given deployment vector h (Step 2). Even in the case in which p does not change throughout an epidemic, a one-dimensional maximization problem, it is possibly non-concave (recall Figure 6.7b). Thus, we proceed to compute this quantity via grid search by discretizing $\mathcal{P}$. In this context, an important experimental observation is that the calculation of $V(h|\vec{p})$ (via the *SEIR* routine) is indeed extremely fast even for large T (long time horizons) – typically, $V(h|\vec{p})$ can be computed in approximately one *hundred thousandth* of a second on a modern computer[2]. More details about how this step could be furthered improved are described below.

Auxiliary LP: In Step 3 a linear program must be solved to compute $V(h^r|\vec{p}(r))$. (See equations (7.16-7.18) or (7.5) for a detailed description.) The structure of this linear program does not

---

[1]Initial number of infectives for both subgroups, I_0, and parameter values for all the rates of transition between compartments.

[2]We implemented this subroutine in C.

change with each iteration; only the input data does: for a given $\vec{p}(r)$, a deployment strategy h^r, and preset initial conditions, we need to run the *SEIR* model to generate the coefficient matrices (both of the *SEIR* model for the surge staff, and the objective function piece-wise linear approximation), the right-hand side and the objective vectors of this LP.

We used AMPL together with the Gurobi optimization solver to update such input and solve this LP at each iteration. This software readily gives us the optimal dual variables α_k and π_k that are used to construct the cuts that go into the master problem.

Master problem: While this is the Linear Program that provides the optimal solution, it is the step that requires the least effort to implement. Using the information obtained from the lastly-solved auxiliary LP, at each iteration one only needs to append a new cut to the file that carries the master problem and re-solve the LP. Such cut is easily constructed with the dual optimal variables of the auxiliary LP after it is solved at the end of Step 3.

In addition to these elements, a wrap-up routine synchronizes all the subroutines involved and, among other things, keeps track of the bounds that determine whether to stop the algorithm or not at a pre-specified level of accuracy.

Thus, the $r-$th iteration of the algorithm is implemented as follows:

1. Solve the master problem with the current cuts (empty if at first iteration). The optimal value of this linear program constitutes a lower bound on the optimal cost we are looking for, and is recorded for future reference.

2. Use obtained deployment vector h^r to compute its associated highest-cost tuple using grid search.

3. Take such tuple to construct corresponding vector $\vec{p}(r)$. Run the *SEIR* model subroutine with $\vec{p}(r)$ and h^r as inputs to construct the input data of the auxiliary LP. The optimal solution of this program is an upper bound to our optimal cost. Store optimal dual variables to append to Master Problem if necessary.

4. Check if algorithm should be stopped. If the lower and upper bound are within certain pre-specified tolerance, the algorithm is stopped. Otherwise, the optimal solution readily provides the optimal dual variables needed to construct a new cut that is appended to the master problem.

A smarter use of grid search

In Section 6.1.1 we discussed the importance of using the right discretization of the grid to compute the optimal robust solution. We also commented on the possibility of using coarser grids at the beginning of the algorithm to speed it up (in particular, to speed up Step 2). In this section we give theoretical details of this approach.

Recall that in our uncertainty model (Section 5.2.1), that for $t \leq \check{t}$, p_t takes a fixed value in $[p^L, p^U]$ whereas for $t > \check{t}$, p_t takes a fixed value in $[\hat{p}^L, \hat{p}^U]$. Let N be a large integer, and write

$$\mathcal{Q}_N = \left\{ p^L + \frac{p^U - p^L}{N} j, \ 0 \leq j \leq N \right\} \quad \text{and} \quad \hat{\mathcal{Q}}_N = \left\{ \hat{p}^L + \frac{\hat{p}^U - \hat{p}^L}{N} j, \ 0 \leq j \leq N \right\} \tag{7.19}$$

$\mathcal{Q}_N$ is a finite approximation to the interval $[p^L, p^U]$ and likewise with $\hat{\mathcal{Q}}_N$ and $[\hat{p}^L, \hat{p}^U]$. Define

$$\Pi_N = \left\{ (q, \hat{q}, \check{t}) \ : \ q \in \mathcal{Q}_N, \ 1 \leq \check{t} \leq T, \ \hat{q} \in \hat{\mathcal{Q}}_N \right\},$$

and for any $\pi = (q, \hat{q}, \check{t}) \in \Pi_N$, let $\vec{p}(\pi)$ be the vector of probabilities defined by

$$p_t(\pi) = q \text{ for } 1 \leq t \leq \check{t}, \text{ and } p_t(\pi) = \hat{q} \text{ for } \check{t} < t \leq T.$$

Finally, define

$$\mathcal{P}_N = \{ \vec{p}(\pi) : \pi \in \Pi_N \}.$$

Thus, $\mathcal{P}_N$ is a discrete approximation to $\mathcal{P}$, and, furthermore $V_N^* := \min_{h \in \mathcal{H}} \max_{\vec{p} \in \mathcal{P}_N} V(h \,|\, \vec{p})$ can be

computed using a (finite) linear program analogous to (7.9) which we include for completeness:

$$
\begin{aligned}
V_N^* := \min_{h,x} \quad & \nu \\
\text{s.t.} \quad & \nu \geq \kappa'_{\vec{p}} x^{\vec{p}}, \quad \forall \vec{p} \in \mathcal{P}_N \\
& A_{\vec{p}} x^{\vec{p}} = h, \quad \forall \vec{p} \in \mathcal{P}_N \\
& C_{\vec{p}} x^{\vec{p}} \geq d_{\vec{p}}, \quad \forall \vec{p} \in \mathcal{P}_N \\
& h \in \mathcal{H}.
\end{aligned}
\tag{7.20}
$$

Clearly, $V_N^* \leq V^*$ since $\mathcal{P}_N \subset \mathcal{P}$, and one can show that $V_N^* \to V^*$ as $N \to +\infty$. From a practical

perspective, a large enough value for N, such as $N = 1000$, should provide an excellent approximation

to our robust staffing problem. Note that, for any N we have $\mathcal{P}_N \subset \mathcal{P}_{2N}$. Hence, we can proceed

by computing V_1^*, followed by V_2^*, followed by V_4^*, and so on until the desired accuracy (in terms of

p) is attained in a logarithmic number of iterations. Notice that in this framework the cutting planes

(7.15) discovered in each iteration of Step 1 performed when computing each value V_N^* remain valid

for the computation of V_{2N}^* (precisely because $\mathcal{P}_N \subset \mathcal{P}_{2N}$) and thus each iteration is warm-started by

the preceding iteration. Additional strategies are possible.

7.3   Improved algorithm

Algorithm B described is theoretically efficient. However, it may sometimes require many iterations to achieve a desirable tolerance. The master problem starts *empty* and, given our computational experience, approximating the objective function could easily take more than 150 iterations (This was the case of Example 1 as discussed later in this section).

We started off the formulation of our Robust Problem with the description of a semi-infinite linear program that could be approximated with a prohibitively large number of constraints (7.20). The procedure outlined in this section is based on the view that we can partially use the structure of this problem to set up a non-empty master problem to start Algorithm B. Even though (7.20) does not offer a computationally viable solution, we combine it with the idea of looking for the *right* cuts of Algorithm B to solve our problem in much fewer iterations.

Making a slight abuse of notation, problem (7.20) can be rewritten as

$$
\begin{aligned}
V_N^* := \min_{h,x} \quad & \nu \\
\text{s.t.} \quad & \nu \geq V(h|\vec{p}) \quad \forall \vec{p} \in \mathcal{P}_N \\
& h \in \mathcal{H}.
\end{aligned}
\tag{7.21}
$$

Recall $V(h|\vec{p})$ represents the cost of a deployment vector h given $\vec{p}$ (7.6-7.8). The cuts of Algorithm B *approximate* $V(h|\vec{p})$ with each iteration for different values of $\vec{p}$, particularly of those dictated by Step 2. We construct a master problem that instead of starting off by adding these cuts, it appends

whole sets of constraints that define $V(h|\vec{p})$, each of them of the form:

$$\nu_{\vec{p}} \geq \kappa'_{\vec{p}} x^{\vec{p}} \tag{7.22}$$

$$A_{\vec{p}} x^{\vec{p}} = h,$$

$$C_{\vec{p}} x^{\vec{p}} \geq d_{\vec{p}}.$$

Let us call (**??**) the block (of constraints) associated with $\vec{p}$.

In other words, Procedure A mimics Algorithm B, but adds a block of constraints associated with $\vec{p}$ to the master problem instead of a cut with each iteration. Clearly, the master problem acquires a significantly large number of constraints with each of these blocks, but numerical experiments confirm that even adding a few of these blocks improve the numerical performance significantly. The reason is that instead of adding cuts that may loosely approximate $V(h|\vec{p})$ –especially during the first iterations –, adding a set of constraints such as (**??**) is paramount to having constraints that describe its exact cost. The careful choice of which blocks to add using Step 2 of Algorithm B counterbalances their computational cost. When the master problem becomes *too large,* we can switch to adding cuts as in Algorithm B. Formally, this approach is described as follows.

<div align="center">

Procedure A

</div>

---

0. Set $Q_0 = \emptyset$ and $r = 1$.

1. Let $\vec{p}(r) \in \mathcal{P}$ be the tuple such that

$$V(h^r \mid \vec{p}(r)) = \max_{\vec{p} \in \mathcal{P}} V(h^r \mid \vec{p}). \tag{7.23}$$

Let $\tilde{V}^r = \min\{V(h^k \mid \vec{p}(k))\}_{k=1}^r$ and update $Q_0 \leftarrow Q_0 \cup \{\vec{p}(r)\}$.

2. Let h^r be an optimal solution for the LP

$$\tilde{W}^r := \min_{h,x} \quad \nu$$

$$\text{s.t.} \quad \nu \geq \kappa'_{\vec{p}} x^{\vec{p}}, \quad \forall \vec{p} \in Q_0$$

$$A_{\vec{p}} x^{\vec{p}} = h, \quad \forall \vec{p} \in Q_0$$

$$C_{\vec{p}} x^{\vec{p}} \geq d_{\vec{p}}, \quad \forall \vec{p} \in Q_0$$

$$h \in \mathcal{H}$$

$$x \geq 0.$$

3. If $|\tilde{V}^r - \tilde{W}^r| < \epsilon \tilde{W}^r$ or $r > K$, **STOP** – algorithm has terminated.

Else, reset $r \leftarrow r + 1$ and go to **1**.

Parameters ϵ and K are fixed a priori; they represent a duality gap tolerance and a maximum number of iterations for Procedure A, respectively. As we mentioned above, the program acquires a significant number of constraints with each iteration: for a single tuple, under the queueing scenario, there are close to 2,800 constraints and more than 2,500 variables (after pre-solving). Thus, we expect K to be small (not larger than 10).

Once Procedure A terminates, we start Algorithm B, with the proviso that the master problem is initialized as:

$$W^r := \min_{z,h,x} \quad z$$

$$\text{s.t.} \quad z \geq \alpha'_k h + \pi'_k d_{\vec{p}(k)}, \quad 1 \leq k \leq r - 1$$

$$z \geq \kappa'_{\vec{p}} x^{\vec{p}}, \qquad \forall \vec{p} \in Q_0$$

$$A_{\vec{p}} x^{\vec{p}} = h, \qquad \forall \vec{p} \in Q_0$$

$$C_{\vec{p}} x^{\vec{p}} \geq d_{\vec{p}}, \qquad \forall \vec{p} \in Q_0$$

$$h \in \mathcal{H}$$

$$x \geq 0.$$

Algorithm B is then run as before. At each iteration r, Step 2 discovers a scenario $\vec{p}(r)$, and Step 3 produces a dual vector (π_r, α_r) which gives rise to the cut $z \geq \alpha'_r h + \pi'_r d_{\vec{p}(r)}$ that is added to the master. The effect of the enhancement provided by Procedure A is, typically, to drastically shortcut the number of iterations required by Algorithm B; essentially, the algorithm has been hot-started with a very good initial representation of the critical constraints needed to define problem (7.1).

7.3.1 Example 1 Revisited

To illustrate the performance of the enhanced Algorithm B, we present a number of statistics corresponding to Example 1 for both the queueing and the threshold cost functions (Section 6.1.1). For each iteration of the algorithm we first point out if it is part of Procedure A (Enh_i) or if it is a regular iteration of Algorithm B (Bend_i). We list its corresponding highest-cost tuple (p_1, p_2, d) and the lower and upper bounds on the cost of the robust optimization problem. Finally we also add the CPU

time per iteration (in seconds) used to compute the highest-cost tuple and to solve the corresponding linear programs with AMPL[3].

Tables 7.1 and 7.2 contain the statistics for the queueing and the threshold objective, respectively. In both cases, it is clear that the computation of the highest-cost tuple is the most time-consuming task at each iteration. As we briefly mentioned in Section 6.1.1, at present we have implemented this step as a search process which could be improved in a number of ways. At the same time, we note that for both objective-function types the time that it takes to solve the corresponding linear programs increases with each iteration, but remains at less or about 0.5 second. In summary, the algorithm appears effective: in each case the total CPU time does not exceed 35 seconds.

Tables 7.1 and 7.2 also show how fast the algorithm converged for this example. In terms of the parameters that correspond to Procedure A, we chose the maximum number of iterations to be $K = 10$ and $\epsilon = 0.05$. The procedure was used for the first 8 iterations, after which only one more cut was added to the master problem. The resulting duality gaps were slightly larger than 0.005% and 0.0168% for the queueing and the threshold objective, respectively.

Table 7.1: Bounds and CPU times per iteration for Example 1, queueing cost function (Section 6.1.1). The algorithm used the enhancement for the first 8 iterations and switched to our algorithm for one more iteration.

	Highest-cost tuple			Convergence Bounds		CPU time (seconds)	
Iter	p_1	p_2	**day ch**	**Lower**	**Upper**	**highest-cost tuple**	**AMPL**
Enh_1	0.01092	0.0135	140	0	4.581151	3.406	0.328
Enh_2	0.01172	0.0135	140	0	0.710181	3.453	0.344
Enh_3	0.01132	0.0135	140	0.007060	0.217431	3.453	0.343
Enh_4	0.01081	0.0135	142	0.031251	0.132066	3.468	0.391
Enh_5	0.01185	0.0135	140	0.048387	0.132367	3.453	0.391
Enh_6	0.01168	0.0135	140	0.050479	0.073479	3.469	0.453
Enh_7	0.01177	0.0135	140	0.050686	0.056242	3.562	0.422
Enh_8	0.01111	0.0135	140	0.050686	0.052368	3.453	0.453
Bend_1	0.01172	0.0135	140	0.050765	0.050768	3.437	0.453

---

[3]Time is measured using `_total_solve_elapsed_time`, which reflects the elapsed seconds used by the solve commands.

Table 7.2: Bounds and CPU times per iteration for Example 1, threshold cost function (Section 6.1.1). The algorithm used the enhancement for the first 8 iterations and switched to our algorithm for one more iteration.

	Highest-cost tuple			Convergence Bounds		CPU time (seconds)	
Iter	p_1	p_2	**day ch**	**Lower**	**Upper**	**worst p**	**AMPL**
Enh_1	0.01092	0.0135	140	32.23508	83.14448	2.891	0.343
Enh_2	0.01168	0.0135	140	32.24410	37.34494	2.906	0.359
Enh_3	0.01132	0.0135	140	32.25421	34.37022	2.875	0.406
Enh_4	0.01081	0.0135	142	32.25740	33.12586	2.875	0.406
Enh_5	0.01111	0.0135	140	32.25866	32.55180	2.891	0.437
Enh_6	0.0109	0.0135	140	32.25912	32.33714	2.953	0.484
Enh_7	0.01125	0.0135	140	32.25914	32.29835	2.922	0.485
Enh_8	0.01118	0.0135	140	32.25915	32.27235	2.891	0.5
Bend_1	0.01091	0.0135	140	32.25915	32.26458	2.907	0.531

We now compare the deployment strategies obtained with our Robust Algorithm as presented in Section 7.2 with the one from the improved version described in the present section.

In the case of the queueing objective, the Robust Algorithm has a worst-case cost of 0.05214 with a duality gap of close to 5%; in turn, the enhanced algorithm has 0.0507 and 0.0052%, respectively. As depicted in Figure 7.1, the strategies follow the same pattern for the first 15 days, which trail the worst-case scenario without intervention (as described in Section 6.1.1). They later differ slightly albeit maintaining the same general structure, which makes up for the difference in the duality gaps.
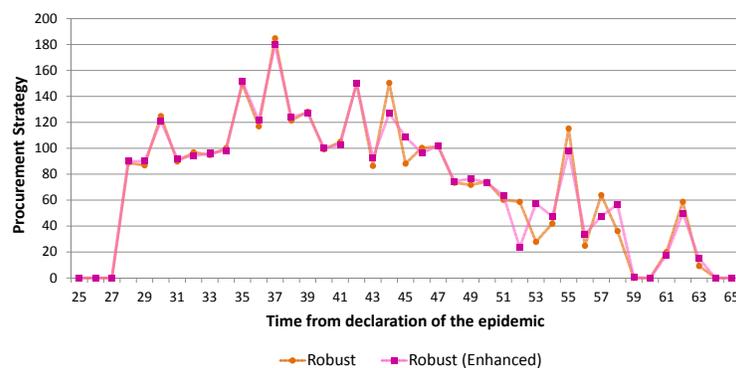

Figure 7.1: Resulting deployment strategies from our Robust Algorithm and its improved version.

On the other hand, the corresponding deployment strategies for the threshold cost function are quite

different (see Figure 7.2a). At the same time, their worst-case costs are less than 1% apart: the strategy obtained from the Robust Algorithm without improvements has a cost of 32.45 with a duality gap of 2.7% and with enhancements, 32.26 and 0.017%, respectively. We hypothesize that the strategy obtained from the Improved Algorithm is very influenced by the No-Action-Worst-Tuple, as evidenced by Figure 7.2b, while the Robust Strategy, obtained from an empty master problem, is void of such influence.
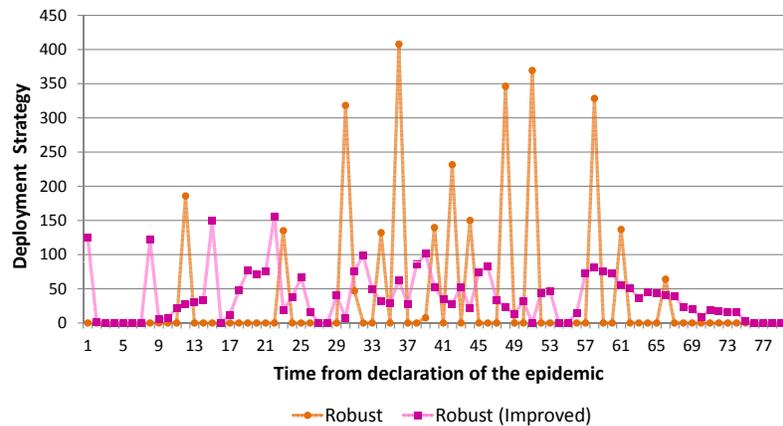
(a) Robust Strategy compared to Improved Robust Strategy
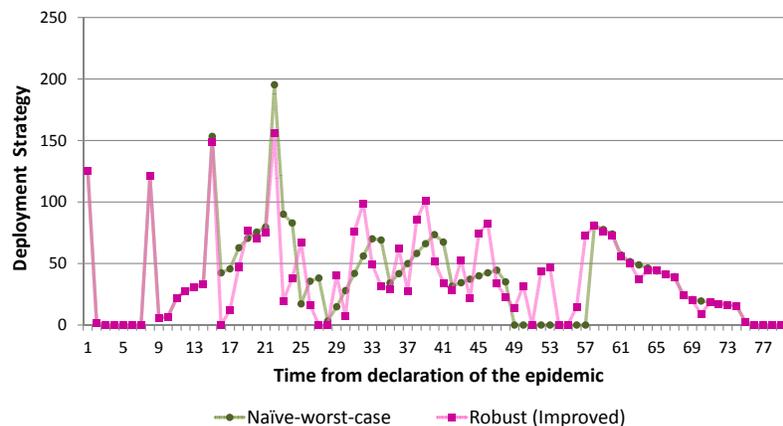
(b) Improved Robust Strategy compared to Naïve-worst-case Strategy

Figure 7.2: Comparison of deployment strategies obtained from our Robust Algorithm, our Improved Robust Algorithm, and the Naïve-worst-case strategy. Strategies correspond to Example 1, using the threshold cost function.

8

Extensions

In this chapter we outline two extensions to our approach where the decision-maker can apply recourse when conditions significantly deviate from predictions. From a broad optimization perspective such strategies clearly make sense – why stick with a rigid strategy?. However, we stress that in view of the experiments at the end of the last section, the likely benefit from a dynamic policy would be primarily realized during a period immediately following the epidemic declaration and of relatively short duration. As as a result, a decision maker would likely face severe logistical constraints if attempting to rapidly redeploy large numbers of staff. Thus, a dynamic policy might only be able to perform *small* changes on a day-to-day basis. From an operational perspective such small changes would of course make sense and would be applied; however the recourse might only amount to a set of small adjustments to a pre-computed strategy.

In both cases that we consider we assume that a decision-maker can only approximately observe the

behavior of p_t (the value of p at time t). The Benders-decomposition solution methodology that we described in Chapter 7 can be adapted to handle both models.

8.1   Robust optimization with recourse

Assume that uncertainty of p_t is modeled using m intervals, or *tranches*, denoted $I_1, I_2, \dots I_m$, for some integer $m > 0$. In addition, two time periods are given, J^{min} and J^{max}. A realization of the uncertain values p_t plays out as follows:

(i) A period $J^{min} \leq J \leq J^{max}$ is chosen, as well as a value $p^{(1)} \in I_1$, a tranche H, and a value $p^{(2)} \in H$.

(ii) For each $t < J$ we have $p_t = p^{(1)}$ whereas for each $t \geq J$, $p_t = p^{(2)}$.

In other words, p_t is allowed to change values, once. The decision-maker operates as follows. First, a fixed time period, S, in which the deployment strategy will be *revised* is chosen in advance. Then

(a) At time $t = 1$, the decision-maker knows that $p_1 \in I_1$, but does not know its precise value. The decision-maker then produces an initial, or "first-stage" deployment plan indicating the amount of staff to call-up at time t, for each $t < S$. Further, m alternative "second-stage" deployment plans (labeled $1, 2, \dots, m$) are announced, each of which specifies the level of staff to deploy at each period $t \geq S$. Each second-stage plan is compatible with the first-stage plan in terms of deployment constraints (such as total staff availability).

(b) At time $t = S$ the decision-maker observes the tranche that p_S belongs to, but not the actual value of p_S. If tranche h is observed, then second-stage plan h is rolled out on periods $t \geq S$.

Example. Consider a case with three tranches: $I_2 = [.1, .15]$, $I_1 = [.15, .18]$, and $I_3 = [.18, .19]$.

Further, $J^{min} = 18$ and $J^{max} = 60$. A realization of the data is the following: $J = 20$, $p^{(1)} = .155$,

$H = I_3$ and $p^{(2)} = .19$. Thus, for $t < 20$ we have $p_t = .155$ while for $t \geq 20$, $p_t = .19$. Suppose

$S = 30$. The decision-maker knows that $.15 \leq p_1 \leq .18$; at time $S = 30$ the decision-maker observes

the tranche I_3 and thus knows that for $t \geq 30$, $.18 \leq p_t \leq .19$, leading to an appropriate set of

deployments for future periods.

The model we just described is inherently adversarial: it assumes that, subject to the stated rules,

data will take on worst-case attributes. This feature can, of course, lead to overly conservative plans.

However, the proper statement to make is that the plans will be conservative only if the data (the

tranches, in particular) allow it to be so. A different drawback in this model is that an adversary could

simply "wait" to change p until just after the review period S. As the experiments in Section 6.1.3

indicate, if S is not too small the impact of such a late change in p is much decreased. If S is chosen

too large, then the adversary can change p much earlier, and then the impact will be felt before the

review can take place. Thus, care must be chosen in selecting S. A review strategy that proceeds

dynamically is given next.

8.2 Stochastic optimization

In contrast to the above, we also outline a model where the contagion probabilities p_t behave stochas-

tically. Assume:

(i) p_1 takes a random value drawn from a uniform distribution over a *known* interval Z.

(ii) For any t, $p_{t+1} = p_t + \epsilon_t$, where ϵ_t is normally distributed, with zero mean and small standard

 deviation (small compared to the width of Z).

Thus, p_t executes a random walk, starting from Z. Let $\bar{p}$ indicate the center (mean value) of Z. The

decision-maker's actions, in this model, are as follows.

(a) At time $t = 1$, the decision-maker produces a policy consisting of a triple $(\tilde{h}, \alpha, R)$ where $\alpha \geq 0$,

 and $\tilde{h}$ is an initial staff surge deployment plan indicating for each t the quantity $\tilde{h}_t$, the amount

 of staff to call-up at time t. Finally, the decision-maker also places an additional quantity $R \geq 0$

 of surge-staff on *reserve* (i.e., not yet part of the deployment plan). For notational purposes,

 write $h_t = \tilde{h}_t$, for each t.

(b) The decision-maker can revise the plan at any of several checkpoints $t_1 < t_2 < ... < t_r$.

(c) Consider checkpoint t_i $(1 \leq i \leq m)$. Let

$$\Theta(t_i, \bar{p}) \;=\; \text{the total number individuals infected by time } t_i, \text{ in the SEIR model where } p_t = \bar{p},$$
$$\text{for } 1 \leq t_i$$

$$\Gamma(t_i) \;=\; \text{the } \textit{actual} \text{ total number of individuals who are infected by time } t.$$

Thus, $\Theta(t_i, \bar{p})$ is an estimate of the random variable $\Gamma(t_i)$. We assume that the decision-maker

observes, as an estimate for $\Gamma(t_i)$, a quantity

$$\nu(i) \;=\; \Gamma(t_i) + \delta(t_i), \tag{8.1}$$

where $\delta(t_i)$ is normally distributed with zero mean and known variance. Then, the deployment

plan is revised as follows. Write

$$\Delta = \frac{\nu(i) \,-\, \Theta(t_i, \bar{p})}{t_i}.$$

Then, the decision-maker resets

$$h_t \;\leftarrow\; h_t \,+\, \alpha\,\Delta, \quad \text{for each } t_i \leq t < t_{i+1}, \text{ and} \tag{8.2}$$

$$R \;\leftarrow\; R \,-\, (t_{i+1} - t_i)\,\alpha\,\Delta. \tag{8.3}$$

Comment. The proposed scheme constitutes an example of affine control. The decision-maker computes an initial estimate of the deployment plan (the initial h_t) which are corrected on the basis of real-time observations, using the parameter α to modulate the corrections. The quantity Δ indicates the (per-period) deviation between the observed behavior of the epidemic (including "noise", given by δ) and the predicted behavior as given by $\Theta(t_h, \bar{p})$.

When $\Delta > 0$, the behavior of the epidemic is worse than initially estimated and Rule (8.2) will increase the near-future surge staff deployment. This increase is drawn from the surge staff held in reserve, as per equation (8.3). When $\Delta < 0$, (8.3) moves staff back into the reserve ranks.

Note that in order for Rule (8.3) to be feasibly applied, the resulting R must be *nonnegative*. For this to be the case when $\Delta > 0$ we should have α small enough, and, especially, that the initial estimate for R be "high enough." These are constraints to be maintained in our solution algorithms.

In order to consider algorithmic implications of this model, consider a fixed *sample path* specifying a value for each p_t and for each $\delta(t_i)$ (see equation (8.1)). Denote by T the length of the planning horizon. Recall that in the stochastic model the decision-maker announces, at $t = 1$, a triple $(\tilde{h}, \alpha, R)$.

We have:

Lemma 8.2.1 *Under the given sample path the cost incurred by the policy* $(\tilde{h}, \alpha, R)$ *(as per rules (b)-(c)) is given by a linear program whose right-hand side is an affine function of the* $\tilde{h}_t$ *and* α.

As a consequence, we expect that several well-known methodologies for stochastic optimization, such as sample average approximation [69], [70] and stochastic gradient schemes [68], [46], will be well-suited for our problem.

9

Discussion

The potential impact of a severe influenza epidemic on essential services is a matter of critical importance from a public health perspective. There are current concerns that an influenza virus might mutate into a highly contagious strain for which humans have little or no immunity, resulting in a rapid widespread of the disease. In the event of such an epidemic or pandemic, organizations that provide critical social infrastructure, such as health care clinics, police departments, public utilities, food markets, and supply chains are likely to face severe workforce shortage that would jeopardize their continuity of operations. In particular, at the most critical stage of the epidemic, health care clinics would be required to provide care for an extraordinary number of patients, and thus could ill-afford staff shortfalls.

Pandemic influenza preparedness would require an extensive set of challenging steps. It requires the involvement and commitment of many segments of the academic, industrial and governmental

communities. The planning process requires joint efforts to improve, among other things, the vaccine and anti-viral development and distribution, the surveillance systems, and the design of pragmatic and effective regional plans. These plans must include, among other important activities, methods to expand surge capacity and to maintain essential services, including the deployment of 'surge' staff to compensate for potential abnormal absenteeism rates. To address these issues, the CDC and the HHS Department together with State authorities have published suggested courses of action.

The intelligent design of such contingency plans would necessarily model the spread of the epidemic. In the case of a new strain of the influenza virus, such modeling entails incorporating significant uncertainty: as far as we know, there are no clear ways of accurately predicting the transmissibility of a new virus strain. Moreover, it has been suggested that the effective contagion rate may change under different environmental conditions, such as weather, which are likewise unpredictable. Additionally, social patterns may change during the epidemic, for example due to the implementation of public health measures such as quarantine or social distancing.

Pre-event planning is critical: once pandemic flu strikes a community, it could be over in few months, barely enough time to put together all the elements of emergency management that we have extensively described. We believe Robust Optimization is the best tool to tackle the problem from this perspective, particularly because of the sheer unpredictability of the epidemics' intensity and length. Although it could be deemed too conservative (in a way, it is optimizing against the worst-case scenario), this methodology effectively immunizes the solutions against a pre-specified uncertainty set. This could prove invaluable from a social cost perspective. Moreover, a decision-maker interested in using our tool could vary the degree of risk aversion by varying the structure of the uncertainty set: the size of the intervals that conform it is proportional to the degree of risk aversion. Several scenarios could be

considered until a balanced tradeoff is reached.

In this thesis we build a tool that helps to study how to design robust pre-planned surge staff deployment strategies that would aide an organization to cope with workforce shortfalls while optimally hedging against uncertainty. As far as we know, this is the first time that this methodology is used in this research area. As discussed in previous chapters, earlier work primarily focuses on conducting simple sensitivity analysis as a way to measure robustness. These works implicitly assume probability distributions on the parameters when these may not be readily available or are easily justifiable. Our models do not need to make such assumptions.

Taking advantage of a version of the generalized Benders' Decomposition, we propose fast and accurate algorithms that prove sufficiently flexible to incorporate intricate uncertainty sets that reflect the changeability of the environment. We bring insights on the structure of optimal robust strategies and on practical rules-of-thumb that can be deployed during the epidemic.

To benchmark the effectiveness of the obtained robust policies, we compared them to the policies that optimally solve the problem for the scenario with highest cost given that no action is taken (the Naïve-worst-case Policy). Our proposed solutions could also be compared to policies obtained from an equivalent Stochastic Optimization problem, which minimize the average cost over all possible scenarios within the uncertainty set. While this is a useful exercise, we use the Naïve-worst-case Policy as a first benchmark because preparedness plans often suggest decision makers to augment their surge capacity based on a scenario similar to the Spanish Flu. This is considered to be the *worst* influenza pandemic known by human kind (in terms of mortality and social impact). Our results show that following a strategy that just considers the highest-cost scenario may allocate resources inappropriately for milder cases, which are also likely to occur. This would translate in big savings for

a decision maker: surge capacity would be effectively used in *any* scenario within the preestablished uncertainty set avoiding wastage. We expect the comparison of our Robust Strategy to one obtained from a Stochastic Optimization framework to yield similar results and leave it as an extension of this thesis.

Our strategies also behave well in out-of-sample tests in which the probability of contagion was allowed to change to larger values outside the uncertainty set, and at least 10 days after the surge staff started to be deployed. A notable fact was that when p changes relatively late with respect to the strategy's deployment, the epidemic has largely run its course and the change doesn't have any meaningful impact. When the change occurs much earlier, social distancing prevents the epidemic from unfolding, highlighting the importance of these kind of interventions. Further out-of-sample tests suggested that it may be desirable to include a lateness factor in the uncertainty set.

A number of refinements and extensions of this thesis are possible. First, one could include a more specific compartmental model for influenza: adding more population subgroups to the model should be able to capture population dynamics better; similarly, incorporating more compartments would describe the disease more accurately (such as an asymptomatic compartment, as presented in Chapter 12 of [15]). These enhancements would have to handled with care, for they would require corresponding parameter values. Data paucity could represent a disadvantage more than a benefit.

A second extension would incorporate the potential impact that a staff contingency plan could have on the course of the epidemic (so far assumed to be negligible). This could be the case, for instance, at vaccination centers; their adequate staffing could be crucial to reach critical vaccination fractions in a timely matter. We did not venture into this endeavor because any such model would heavily rely

on the ability of the surge staff to change the epidemic dynamics. In the case of the vaccination centers, an effective vaccine would have to be readily available. For a new flu virus mutation it is unclear whether this is possible, especially considering that influenza vaccines take at least 6 months to manufacture. Thus, this extension would be useful should the model be used for seasonal influenza or for epidemics for which effective means of fighting it are accessible and their deployment depends on the effective implementation of a pre-specified contingency plan.

Another extension includes the use of cross-staffing, either within a hospital or a network of clinics and institutions. In this study we are considering the entire workforce of a particular organization as a whole; however, as part of the contingency plans proposed by the CDC, organizations are encouraged to identify critical staff positions and cross-train personnel accordingly. The model could also be further refined by modeling more than one employee group depending on levels of exposure to the disease or social contact patterns. This would, of course, vary from organization to organization and should be tailored judiciously.

Our methods could be easily used to manage any quantifiable scarce resource during an epidemic. Moreover, it is very likely that a decision-maker will face parallel resource problems during such an event - for example, a hospital would need to plan for bedding and other supplies on top of surge staffing. Any multiple-objective cost function could be incorporated provided it has a convex piecewise-linear shape. One such example involves antiviral inventory management. In the case of hospitals and clinics, rates of transmission for health care workers may be minimized during a severe pandemic by means of protective equipment or the use of antiviral medication. Given that stockpiles are limited, a hospital could be interested in managing the doses during the outbreak. Appropriate extensions to the *SEIR* model could be carefully made to accommodate the reduction on disease transmissibility, while

keeping track of the antiviral stockpile. Additionally, our model currently assumes that surge staff is not exposed to the epidemic until they are called into service. This relatively strong assumption could be remedied by incorporating proper *SEIR* constraints to the linear program defined by $V(h|p)$ (see Chapter 5); we also leave this for future projects.

Finally, from an operational perspective, organizations may be able to react to evolutions of the epidemic that significantly differ from initial estimations. While the space for action could be limited (from what we learned from our out-of-sample tests), the topic is clearly of great interest. How to incorporate optimal recourse decisions is a topic we briefly discussed in Chapter 8 and plan to address in future work.

Bibliography

[1] Operations research and public health. *American Journal of Public Health and the Nations Health*, 42(10):1306–1307, October 1952.

[2] The world's deadliest bioterrorist. *The Economist*, April 28th 2012.

[3] Dionne M. Aleman. Predicting the spread of pandemic disease. *OR/MS Today*, 39:24–28, February 2012.

[4] L.J.S. Allen and A.M. Burgin. Comparison of deterministic and stochastic sis and sir models in discrete time. *Mathematical biosciences*, 163(1):1–33, 2000.

[5] LJS Allen, MA Jones, and CF Martin. A discrete-time model with vaccination for a measles epidemic. *Mathematical biosciences*, 105(1):111–131, 1991.

[6] L.J.S. Allen and P. van den Driessche. The basic reproduction number in some discrete-time epidemic models. *Journal of Difference Equations and Applications*, 14(10):1127–1147, 2008.

[7] R.M. Anderson and R.M. May. *Infectious diseases of humans: dynamics and control*, volume 26. Wiley Online Library, 1992.

[8] Command center operations. http://apogeeengineering.net/index.cfm/ID/54/Command-Center-Operations/, 2012.

[9] J. Arino, F. Brauer, P. Van Den Driessche, J. Watmough, and J. Wu. Simple models for containment of a pandemic. *Journal of the Royal Society Interface*, 3(8):453, 2006.

[10] J.G. Bartlett. Planning for avian influenza. *Annals of internal medicine*, 145(2):141–144, 2006.

[11] J.G. Bartlett and L. Borio. The current status of planning for pandemic influenza and implications for health care planning in the united states. *Clinical Infectious Diseases*, 46(6):919, 2008.

[12] A. Ben-Tal and A. Nemirovski. Robust optimization–methodology and applications. *Mathematical Programming*, 92(3):453–480, 2002.

[13] J.F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962.

[14] M.C.J. Bootsma and N.M. Ferguson. The effect of public health measures on the 1918 influenza pandemic in US cities. *Proceedings of the National Academy of Sciences*, 104(18):7588, 2007.

[15] F. Brauer, P. Van den Driessche, J. Wu, and L.J.S. Allen. *Mathematical epidemiology*. Number 1945. Springer Verlag, 2008.

[16] C.B. Bridges, M.J. Kuehnert, and C.B. Hall. Transmission of influenza: implications for control in health care settings. *Clinical Infectious Diseases*, pages 1094–1101, 2003.

[17] Flu surge 2.0. http://www.cdc.gov/flu/tools/flusurge/, August 2011.

[18] Seasonal influenza (flu). Retrieved October, 2011. http://www.cdc.gov/flu/, 2011.

[19] G. Chowell, S. Echevarría-Zuno, C. Viboud, L. Simonsen, J. Tamerius, M.A. Miller, and V.H. Borja-Aburto. Characterizing the epidemiology of the 2009 influenza a/h1n1 pandemic in mexico. *PLoS Med*, 8[5]:e1000436, 2011. Available at doi:10.1371/journal.pmed.1000436.

[20] G. Chowell, H. Nishiura, and L. Bettencourt. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface*, 4(12):155, 2007.

[21] MD Cinti, Sandro, MD Chenoweth, Carol, and MD Monto, Arnold S. Preparing for pandemic influenza: Should hospitals stockpile oseltamivir? *Infection Control and Hospital Epidemiology*, 26(11):pp. 852–854, 2005.

[22] Derek DeLia. *Hospital Capacity, Patient Flow, and Emergency Department Use in New Jersey*. Rutgers Center for State Health Policy, 2007.

[23] O. Dieckmann and JP Heesterbeek. *Mathematical epidemiology of infectious diseases*. Wiley, New York, 2000.

[24] John N. Tsitsiklis Dimitris Bertsimas. *Introduction to Linear Optimization*. Athena Scientific Belmont, MA, 1997.

[25] L. ElGhaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. App.*, 18(4):1035–1064, 1997.

[26] Centers for Disease Control and Prevention. Flu view. a weekly influneza surveillance report prepared by the influenza division. Reviewed on October 2011. http://www.cdc.gov/flu/weekly/.

[27] E. Yoko Furuya, May 2011. Assistant Professor of Clinical Medicine, Division of Infectious Diseases, Columbia University Medical Center. Personal communication.

[28] M. Gardam, D. Liang, S.M. Moghadas, J. Wu, Q. Zeng, and H. Zhu. The impact of prophylaxis of healthcare workers on influenza pandemic burden. *Journal of the Royal Society Interface*, 4(15):727, 2007.

[29] Pandemic influenza preparedness and response standard operating guide. Available at http://health.state.ga.us/pandemicflu, June 2006.

[30] R.R.M. Gershon, K.A. Qureshi, P.W. Stone, M. Pogorzelska, A. Silver, M.R. Damsky, C. Burdette, K.M. Gebbie, and V.H. Raveis. Home health care challenges and avian influenza. *Home Health Care Management & Practice*, 20(1):58, 2007.

[31] C.D. Gomersall, S. Loo, G.M. Joynt, and B.L. Taylor. Pandemic preparedness. *Current opinion in critical care*, 13(6):742, 2007.

[32] M.L. Grayson, S. Melvani, J. Druce, I.G. Barr, S.A. Ballard, P.D.R. Johnson, T. Mastorakos, and C. Birch. Efficacy of soap and water and alcohol-based hand-rub preparations against live h1n1 influenza virus on the hands of human volunteers. *Clinical Infectious Diseases*, 48(3):285, 2009.

[33] L.V. Green and P.J. Kolesar. Improving emergency responsiveness with management science. *Management Science*, pages 1001–1014, 2004.

[34] M.S. Green, T. Swartz, E. Mayshar, B. Lev, A. Leventhal, P.E. Slater, and J. Shemer. When is an epidemic an epidemic? *Israel Medical Association Journal*, 4(1):3–6, 2002.

[35] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[36] M.E. Halloran, N.M. Ferguson, S. Eubank, I.M. Longini, D.A.T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T.C. Germann, et al. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639, 2008.

[37] J.L. Hick, M.D. Christian, and C.L. Sprung. Chapter 2. surge capacity and infrastructure considerations for mass critical care. *Intensive care medicine*, 36:11–20, 2010.

[38] A.N. Hill, I.M. Longini, et al. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical biosciences*, 181(1):85–106, 2003.

[39] J. Hoffbuhr. Utilities prepare for potential pandemic. *American Water Works Association Journal*, 98(6):48–60, 2006.

[40] Implementation plan for the national strategy for pandemic influenza. Last accessed June 2012. Available at www.flu.gov/planning-preparedness/federal/pandemic-influenza.pdf, May 2006.

[41] Edward Kaplan. Operations research and public health. University Lecture.

[42] M.J. Keeling and P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.

[43] W.O. Kermack and A.G. McKendrick. Contributions to the mathematical theory of epidemics. ii. the problem of endemicity. *Proceedings of the Royal society of London. Series A*, 138(834):55–83, 1932.

[44] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics. iii. further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A*, 141(843):94–122, 1933.

[45] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemicsŮi. *Bulletin of mathematical biology*, 53(1):33–55, 1991.

[46] H.J. Kushner and D.S. Clark. *Stochastic approximation methods for constrained and unconstrained systems*. Springer, 1978.

[47] R.C. Larson. Simple models of influenza progression within a heterogeneous population. *Operations research*, 55(3):399–412, 2007.

[48] H.L. Lee, V. Padmanabhan, and S. Whang. The bullwhip effect in supply chains1. *Sloan management review*, 38(3):93–102, 1997.

[49] V.J. Lee and M.I. Chen. Effectiveness of neuraminidase inhibitors for preventing staff absenteeism during pandemic influenza. *Emerging Infectious Diseases*, 13(3):449, 2007.

[50] I.M. Longini Jr, M.E. Halloran, A. Nizam, and Y. Yang. Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology*, 159(7):623, 2004.

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

[52] A.C. Lowen, S. Mubareka, J. Steel, and P. Palese. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS pathogens*, 3(10):e151, 2007.

[53] C.E. Mills, J.M. Robins, and M. Lipsitch. Transmissibility of 1918 pandemic influenza. *Nature*, 432(7019):904–906, 2004.

[54] N.A.M. Molinari, I.R. Ortega-Sanchez, M.L. Messonnier, W.W. Thompson, P.M. Wortley, E. Weintraub, and C.B. Bridges. The annual impact of seasonal influenza in the us: measuring disease burden and costs. *Vaccine*, 25(27):5086–5096, 2007.

[55] S.S. Morse. Building academic-practice partnerships: the Center for Public Health Preparedness at the Columbia University Mailman School of Public Health, before and after 9/11. *Journal of Public Health Management and Practice*, 9(5):427, 2003.

[56] Stephen Morse, February 2011. Professor of Clinical Epidemiology, Founding Director & Senior Res. Scientist, Center for Public Health Preparedness. Personal communication.

[57] Influenza pandemic plan guide for healthcare facilities. Available at http://www.nj.gov/health/flu/plan.shtml, August 2005.

[58] Pandemic preparedness planning template for federally qualified health centers. July 2007.

[59] Facts and statistics 2010. Retrieved October, 2011. http://nyp.org/about/facts-statistics.html, 2010.

[60] A.T. Newall, J.G. Wood, N. Oudin, and C.R. MacIntyre. Cost-effectiveness of Pharmaceutical-based Pandemic Influenza Mitigation Strategies. *Strategies*, 2:3.

[61] M. Nuno, G. Chowell, and AB Gumel. Assessing the role of basic control measures, antivirals and vaccine in curtailing pandemic influenza: scenarios for the us, uk and the netherlands. *Journal of the Royal Society Interface*, 4(14):505–521, 2007.

[62] Emergency management. Retrieved July, 2011. http://www.nyc.gov/html/doh/html/em/emergency-homepage.shtml, 2011.

[63] What employers can do to protect workers from pandemic influenza. Retrieved May, 2011. http://www.osha.gov/Publications/employers-protect-workers-flu-factsheet.html, 2009.

[64] National Institute of Allergy and Infectious Diseases. Flu (influenza). Retrieved May 1, 2012. http://www.niaid.nih.gov/topics/flu/understandingflu/pages/definitionsoverview.aspx, August 2011.

[65] Department of Transportation. Pandemic flu planning, March 2009. http://www.dot.gov/pandemicflu/.

[66] M.T. Osterholm. Preparing for the next pandemic. *New England Journal of Medicine*, 352(18):1839, 2005.

[67] Raúl Rabadán, November 2010. Assistant Professor of Biomedical Informatics, Columbia University College of Physicians and Surgeons. Personal communication.

[68] H. Robbins and S. Monro. On a stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[69] A. Shapiro and H. Homem de Mello. On rate of convergence of monte carlo approximations of stochastic programs. *SIAM J. on Optimization*, 11:70–86, 2001.

[70] D. Shmoys and C. Swamy. Sampling-based approximation algorithms for multi-stage stochastic optimization. *Proc. FOCS 2005*, pages 357–366, 2005.

[71] J. Taubenberger and D. Morens. 1918 Influenza: the mother of all pandemics. *Rev Biomed*, 17:69–79, 2006.

[72] B.L. Taylor, H.E. Montgomery, A. Rhodes, and C.L. Sprung. Chapter 6. protection of patients and staff during a pandemic. *Intensive care medicine*, 36:45–54, 2010.

[73] Swine flu (h1n1 virus). Retrieved May, 2011. http://nyti.ms/SiDWD, January 2010.

[74] The evolution of bird flu, and the race to keep up. Retrieved June, 2012. http://nyti.ms/N2g7mi, June 2012.

[75] S. Towers and Z. Feng. Pandemic h1n1 influenza: Predicting the course of vaccination programme in the united states. *Eurosurveillance*, 14, 2009.

[76] Lauren Trainor, Joseph; Barsky. Reporting for duty? a synthesis of research on role conflict, strain, and abandonment among emergency responders during disasters and catastrophes. Technical report, Disaster Research Center, 2011. Available at http://dspace.udel.edu:8080/dspace/handle/19716/9885.

[77] Business planning. Last accessed July 2011. http://www.pandemicflu.gov.

[78] Pandemic influenza plan, supplement 3, healthcare planning. http://www.hhs.gov/pandemicflu/plan/sup3.html.

[79] J. Wallinga, P. Teunis, and M. Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164(10):936, 2006.

[80] J. Wallinga, M. Van Boven, and M. Lipsitch. Optimizing infectious disease interventions during an emerging epidemic. *Proceedings of the National Academy of Sciences*, 107(2):923, 2010.

[81] What is public health? Retrieved June, 2012. http://www.whatispublichealth.org/.

[82] What is a pandemic? Reviewed on May, 2012. http://www.who.int/csr/disease/swineflu/en/.

[83] J.T. Wu, S. Riley, C. Fraser, and G.M. Leung. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Med*, 3(9):e361, 2006.