Managing medical equipment capacity with early spread of infection in a region

Apurva Jain  |  Swapnil Rayal

Foster School of Business, University of Washington, Seattle, Washington, USA

Correspondence
Apurva Jain, Foster School of Business, University of Washington, Seattle, WA 98195, USA.
Email: apurva@uw.edu

Handling Editor: Special Issue Pandemic Editors

Abstract
We develop a model for a regional decision-maker to analyze the requirement of medical equipment capacity in the early stages of a spread of infections. We use the model to propose and evaluate ways to manage limited equipment capacity. Early-stage infection growth is captured by a stochastic differential equation (SDE) and is part of a two-period community spread and shutdown model. We use the running-maximum process of a geometric Brownian motion to develop a performance metric, probability of breach, for a given capacity level. Decision-maker estimates costs of economy versus health and the time till the availability of a cure; we develop a heuristic rule and an optimal formulation that use these estimates to determine the required medical equipment capacity. We connect the level of capacity to a menu of actions, including the level and timing of shutdown, shutdown effectiveness, and enforcement. Our results show how these actions can compensate for the limited medical equipment capacity in a region. We next address the sharing of medical equipment capacity across regions and its impact on the breach probability. In addition to traditional risk-pooling, we identify a peak-timing effect depending on when infections peak in different regions. We show that equipment sharing may not benefit the regions when capacity is tight. A coupled SDE model captures the messaging coordination and movement across regional borders. Numerical experiments on this model show that under certain conditions, such movement and coordination can synchronize the infection trajectories and bring the peaks closer, reducing the benefit of sharing capacity.

KEYWORDS
capacity planning, COVID-19 pandemic, optimal stopping, stochastic differential equations

1 | INTRODUCTION

During the early spread of COVID-19, there has been widespread reporting about shortages of medical equipment like ventilators, vital sign monitors, and intensive-care-unit beds (Halpern & Tan, 2020). Institutions like the FDA and WHO have provided information about monitoring and tracking such shortages (US FDA, 2020). The demand for such equipment is driven by the evolution of infections in different regions. The supply is severely constrained. The regional administrators and healthcare authorities must determine the level of capacity they need in their regions. Given the limited availability, they must also identify ways to use this capacity to improve health outcomes in their regions. Our goal in this paper is to develop an analytical model to guide such decisions.

The determination of the right capacity and its management to improve performance is a core concern of the production and operations management (POM) research. The literature, however, has paid limited attention to managing demands that are driven by the spread of a contagious infection. The growth pattern of infections, an early exponential rise followed by a decline or a plateau, determines how much capacity will be needed and when. The literature is also relatively silent about situations where decision-makers, facing considerable uncertainty about infection-growth patterns and other model inputs, exhibit a wide range in the decisions they make (KFF, 2021). We aim to develop a model that can capture the essential new features of this context: uncertainty in how infections grow and drive demand, a wide range of decision-making choices in different regions, and the need to compensate for the limited capacity.
We take the perspective of a regional decision-maker (she) to address the following questions. Based on her region’s infection growth and her estimates of its impact on the region’s health and economy, how should she determine the medical equipment capacity? Given that such capacity may not be fully available, what is the impact of compensating actions—from early shutdowns to consistent enforcement messaging and regional capacity sharing—on her region’s health outcomes? Our approach to answering these questions is analytical, but the model provides an understanding of trade-offs and insights into the value of collaboration that is practically useful. On the theoretical side, our contributions include developing a model that (i) derives medical equipment demand from the stochastic spread of infections, (ii) accounts for the variety in decision-makers’ actions, and (iii) proposes and evaluates ideas for improving health outcomes with constrained supply. Our focus is on the demand side; we do not model the acquisition of medical equipment supply.

Specifically, we analyze a stochastic differential equation (SDE) of infection growth in the early stages. Based on this growth, we develop a measure of the chance that the need for medical equipment over a period will exceed a given capacity, the breach probability. We then describe a formulation of a regional decision maker’s beliefs about costs related to economy and health and the expectation of a cure. We show how her preferred estimates interact with the spread of infections to determine the infection level at which a region should shut down to reduce its infection rate. We also analyze the capacity needed at the shutdown time. We present a heuristic rule for this decision and an optimal stopping formulation. We show that while a lack of capacity can drive up the breach probability, actions like reusability, early or high-intensity shutdowns, and consistent messaging can reduce it. We then develop a coupled SDE model to analyze the impact of capacity coordination across two regions. The model allows for travel between the regions and for the coordination of messaging and shutdown policy. We analyze the impact of such coordination on the regions’ breach probability.

We develop the following results and insights:

- Modeling uncertainty in infection rates in an SDE formulation induces the needed realism in the model while still maintaining tractability for analyzing outcomes like the breach probability.
- A region can shut down to reduce infection rate; should it do so and when? The model clarifies the trade-off between incurring a cost to shutting down the economy and reducing the probability of a breach. A heuristic decision rule determines the level of infections at which to shut down.
- At the shutdown level of infections, capacity-determination is a function of the decision maker’s two estimates: the relative cost impacts of an economic shutdown versus a breach of health capacity and the time until the availability of a cure or a vaccine. A simple visual guide lets her determine the capacity she needs corresponding to her estimates. The guide maps the variety of decision-makers’ beliefs and estimates to the range of preparatory actions we observe across various US states.
- An optimal stopping formulation of the problem leads to observations similar to those based on the heuristic rule.
- If capacity is limited, what actions can a region take to reduce its breach probability? The toolbox includes reusability, early shut down or high-intensity shut down, and consistent enforcement messaging. The model can determine the reduction in breach probability that these actions can achieve. Consistency in messaging reduces noise by limiting the variation in people’s behavior; its impact on the breach probability is not always beneficial.
- Looking beyond a region’s boundaries, what is the impact of sharing capacity with a neighboring region? We model the dynamics as a coupled SDE and prove a technical result for the solution’s existence and positivity. The model captures travel between regions and the correlation between the random processes underlying the two regions’ infections.
- Without any travel and correlation, capacity-sharing between the two regions will reduce the breach probability. Interestingly, this holds only if the capacity is not too tight. We identify two separate drivers for this effect: a pooling effect and a peak-timing effect.
- Movement across the border of the two regions can increase the breach probability with shared capacity.
- Consistency in pandemic-related messaging and joint efforts to maintain data accuracy correlates with the random processes underlying the two regions’ infections. An increase in correlation can increase the breach probability with shared capacity.

In addition to these practical insights, our theoretical contribution to the POM literature includes offering a tractable model to determine the equipment capacity as a function of the running maximum of infections evolving as a geometric Brownian motion. We believe that the paper contributes to the POM literature as an early effort to introduce uncertainty in the pandemic-driven demand model while keeping it tractable for capacity questions. The coupled SDE formulation for regional coordination is new to the literature as well; the proof of the condition for its solution’s existence and positivity is a technical contribution to the literature. The insights into the drivers of capacity-sharing benefits introduce the idea of peak-timing, which is new to the literature.

Section 2 describes how our model connects to three different streams of literature. Section 3 presents the model, and Section 4 offers an analysis of the shutdown level and the capacity. Section 5 presents the impact of compensatory actions. Section 6 presents the regional collaboration coupled SDE model. Section 7 discusses possible extensions and concludes the paper.
2 | LITERATURE REVIEW

We draw on several different streams of literature: epidemiology literature, stochastic finance literature, and capacity models with diffusion-driven demand in operations management. This section highlights how our model connects to these streams and how we differentiate ourselves from the existing models. Some of the recent and rapidly growing literature on pandemic modeling is also discussed.

Our starting point is a model of how infections spread in a pandemic. There is a vast body of literature in epidemiology that develops and analyzes such models. The best-known class of such models is compartment models, with a well-known example being the susceptible, infectious, recovered (SIR) model. These types of models divide the population into separate compartments and estimate transition rates between these compartments. If we drop the recovered compartment, it is called an susceptible, infectious, susceptible (SIS) model. If we add an exposed compartment, it is a susceptible, exposed, infectious, recovered model. If post-infection immunity fades away, the model is called susceptible, infectious, recovered, susceptible. Demographic and geographic details can be added to population transition rate inputs. The resulting ordinary differential equations can be solved numerically or simulated to project the number of infections over time. For a good introduction, see Martcheva (2015).

Our goal is to analyze capacity questions in a setting that captures the uncertainty inherent in the spread of infections. Most compartment models, however, do not offer analytical solutions even without any uncertainty. Our approach is to present a simplified, early-stage version of the SIR model before we add uncertainty to it. The early-stage approach, discussed in Britton (2009) among others, relies on the observation that at the beginning of the pandemic, the number of susceptible (almost the whole population) is extremely large, compared to the number of infections. The next section discusses the early-stage model in detail and offers other supporting references for this approach.

We add uncertainty by explicitly modeling the probabilistic spread using an SDE. Some recent reporting (see, e.g., Roberts, 2020) has argued that classic epidemiological models tend to use constant reproduction rate numbers even though, in reality, the spread depends on uncertain factors. There is, however, plenty of discussion of stochastic modeling in epidemiology. For basic computational methods for analyzing probabilistic spread, see Vynnycky and White (2010). The direct approach to introduce stochasticity in a compartment model is to use continuous-time Markov chains (CTMC). For the CTMC approach, now known as the stochastic general epidemic model, see Bailey (1975). In general, these models yield only to computational analysis. The use of SDEs to model the spread is relatively new. The literature justifies the use of SDE models for epidemic modeling in several ways. Allen (2015) presents discrete-state CTMC models and then shows how they can be approximated by SDEs. In some cases, SDEs are introduced as a reasonable way to model growth without any need for further justification; see, for example, Gard (1988). Another approach is to perturb the infection rate parameter by adding noise to it. This is the approach we take; the next section describes it in detail with supporting references. Our contribution is to combine the SDE method with the early-stage modeling discussed above to present an approach that is tractable enough to analyze questions about capacity, the area of interest in the production and operations management literature.

In contrast to population-based compartment models, an alternative approach is to treat each individual as an agent and develop the epidemic dynamics based on interactions between agents. These individual-based models allow for the incorporation of agents’ choices and reactions to testing and shutdown policies. See, for example, Acemoglu et al. (2020) and Drakopoulos and Randhawa (2021). Given our focus on the aggregated need for capacity and on how it evolves over time, we do not take this individual agent-based approach. Our paper fits a rapidly growing literature on modeling the impact of economic, social, and operational policies in the context of the recent pandemic. Some examples of papers that focus on operations issues include Mehrotra et al. (2020), Alvarez et al. (2020), and Yang et al. (2022).

The second stream of literature that we draw upon is stochastic finance. While we do not interpret our model using the finance terminology, the analytical techniques we use are similar to those used in finance. Analysis and solution of SDEs is a foundation of stochastic finance. See, for example, Shreve (2007) and Oksendal (2003), but they are also used in finance to make investment decisions, see, for example, Dixit and Pindyck (1994).

We extend this literature by presenting a new application of SDEs to model the spread of infections in two neighboring regions. The regions influence each other due to the possibility of people traveling across the border. They are also related through the correlation of Wiener processes that underlie the evolution of infections. This leads to the creation of a set of two coupled SDEs. Our contribution to this literature is a new result to show the existence and positivity of a solution to these equations.

Finally, the third stream of literature we relate to is the analysis of capacity or inventory decisions in the presence of demands that follow the increasing-then-decreasing pattern discussed above. While we are not aware of any work that directly addresses demands influenced by the spread of infections, the first-increasing-then-decreasing pattern is similar to the new product diffusion. There has been some work in the operations management literature that relates to diffusion demand. Ho et al. (2002) and Kumar and Swaminathan (2003) compare the option of delaying the launch with the option to reject customers when capacity is limited. Shen et al. (2011) include pricing as a decision variable in addition to capacity and sales policy. These papers follow a new
product demand model based on Bass (1969) diffusion, a deterministic approach. The use of SDEs and optimal stopping formulation for other POM problems is comparatively limited; see Kwon (2010) and Angelus (2020) for two recent examples. We contribute to this literature by offering a stochastic diffusion model to consider capacity questions in a new context, the spread of infections in a pandemic.

3 | THE MODEL

We model a region with a decision-maker. The region is observing the beginning of the spread of infections. A continuous variable $I_t$ represents the number of infections in the region at time $t$. We will first present the dynamics of the spread and then discuss its underlying assumptions based on epidemiology literature.

3.1 | Dynamics of the infection spread

We model the growth of infections by focusing on the new infections generated by an infected individual. We define the transmission rate $b$ to be equal to the new infections as the instantaneous fraction of the infected population. To consider the recovery of infected individuals, we model the recovery rate $d$ as the instantaneous fraction of the infected population that leaves the infected compartment due to recovery or death. For notational brevity, we define the effective infection spread rate as $r = b - d$, a constant estimated based on actual observations. The rate of change in the number of infections is as follows:

$$\frac{dI_t}{dt} = (b - d)I_t = rI_t. \quad (1)$$

3.2 | Early-stage modeling

Our approach is a simplification of the SIR model as applied to the early spread stage. In the standard SIR model, the infection dynamics are presented as $-\frac{dI_t}{dt} = \left(\frac{\beta S_t I_t}{N}\right) - \gamma I_t$, where $S_t$ represents the number of susceptible at time $t$, $N$ is the fixed total population, and $\beta, \gamma$ are rate parameters. For detailed explanations and parameter estimation, please see the references available in the previous section. The first term captures the increase in the number of infections as a fraction of interactions between the susceptible and infected populations. In the early stage spread of infections, the whole population $N$ of the region, an extremely large number, is susceptible to infection. That is, there is an early stage in which $S_t / N \approx 1$ is a reasonable assumption. Substituting this in the above SIR equation leads to our model. In other words, when the number of susceptible (almost the whole population) is extremely large, compared to the number of infections, it is reasonable to assume that each infection generates a certain number of new infections irrespective of how many susceptible remain. Therefore, it is sufficient to track only the number of infections to model early-stage spread. This early-stage approach is well supported in the literature. Employ this approach to model the spread of COVID-19 in stages. For a unit-size population, they assume $S \approx 1$ in the first stage of the epidemic. Britton (2009) suggests that “The key reason for having an approximation during the early stages of an outbreak when there are many initial susceptible is that it is then very unlikely that any of the first numbers of infectious contacts happen to be with the same susceptible individual. Conversely, it is very likely that all of the first set of infectious contacts happen with distinct individuals.” This formulation is equivalent to our model. Sazonov et al. (2011) model the spread in two stages where the first (early) stage is analyzed without the recovered compartment.

The modeling advantage of this early-stage approach is that it allows us to introduce uncertainty in the model while keeping it tractable for capacity considerations. The disadvantage is that in the absence of an intervention, the number of infections in the model will continue to grow. We address this later by adding a second stage where a shutdown decreases the infection rate.

3.3 | Stochastic spread

The above dynamics are deterministic, but there is a great deal of variation in how infected individuals interact with others and spread the infection. At the beginning of the pandemic, the drivers of this variation, such as people’s behavior, the accuracy of information, penetration of testing, and so forth, are poorly understood. The decision-maker perceives them as unpredictable variability in the spread. We model the infection spread rate as a stochastic process $\tilde{r}$ where $d\tilde{r} = rd\tau + s \, dW_t$, $s$ is the volatility, and $dW_t$ is the increment of the standard Brownian motion (Wiener process) $\{W_t\}_{t \geq 0}$. We can now rewrite the dynamics of the infection spread as follows:

$$dI_t = (rd\tau + s \, dW_t)I_t = rI_t \, d\tau + sI_t \, dW_t. \quad (2)$$

The stochastic process $I_t$, which satisfies the above SDE is defined on an underlying probability space $(\Omega, F, P)$, which is equipped with a filtration $F = \{F_t\}_{t \geq 0}$ and the stochastic process is adapted to $F$. Parameters $r$ and $s$ are constants.

We have briefly reviewed the literature on stochastic epidemic models in the previous section. The parameter perturbation approach (addition of a time-dependent noise in the infection rate) that we have used to create the SDE model is directly supported by Gray et al. (2011), Cai et al. (2019), Tornatore et al. (2005), and references therein.

It is useful to consider the interpretation of the volatility parameter $s$ and the underlying Wiener process $W_t$ because these notations are not usually part of infection-spread models. The SDE is interpreted to mean that in successive small intervals $[t, t + dt), [t + dt, t + 2dt), \ldots, [t + (n - 1)dt, t + ndt),$
the numbers of new infections created by an infected individual are independent and identically distributed random variables that follow a normal distribution with mean $rt$ and variance $s^2 dt$. Thus, the Wiener process captures the time-dependent variability of the number of contacts an infected individual will make in a small period, and $s$ can be adjusted to scale its impact on the spread. This variability is driven by the inconsistency in the behavior of the infected individual and the behavior of others, a natural part of human interaction. A lack of standard messaging will add to this variability. Incomplete and non-uniform testing leads to noise in the observed number of infections and that too will add to this variability. A possible way to estimate $s$ is to collect estimates of infection rates from various sources like hospitals, neighborhood health centers and use this sample to estimate it. For the interpretation and estimation of the infection rate, we refer to Fisher (2020). In general, for the entire timeline defined below, we assume $r - s^2/2 \geq 0$ in order to avoid scenarios in which $I_t \to 0$ a.s. when $t \to \infty$.

### 3.4 The timeline

We frame our model in terms of two periods, beginning at two discrete points in time. At time $t_0$, when a very small number of initial infections $I_{t_0}$ are observed, the regional decision-maker starts tracking the number of infections. The infection rate is $r_0$ and the volatility is $s_0$, both given constants. At a time $t_1$ of her choosing, she observes the number of infections $I_{t_1}$ and decides if she will take action (e.g., mask mandates, shutdowns) to decrease the infections’ spread. If she chooses to shut down, we change the infection rate to $r_1 < r_0$ and the volatility to $s_1$, assumed to be equal to $s_0$, $s_1 = s_0 = s$, unless otherwise mentioned. If she chooses not to shut down, the rate and volatility parameters remain unchanged. Even though it is a shutdown-or-not decision, for convenience, we refer to it as a shutdown decision.

$t_0$: A given constant, representing the start of the decision-maker’s observations.

$I_{t_0}$: A given constant, representing the very small number of infections observed at the time $t_0$.

$t_1$: Decision time at which the decision-maker observes the number of infections $I_{t_1}$ and determines whether to shut down. For now, it is assumed to be a given constant $t_1 > t_0$. Later, we develop a decision rule to determine $t_1$.

$I_t = I_{t_0}(r, s):$ Stochastic process $I_t$ at time $t > u$, starting at time $u$ with a known constant value $I_u$, and following the early-stage SDE described above with known constant parameters $r$ and $s$. For $t_0 \leq t < t_1$, we have $I_u = I_{t_0}, r = r_0, s = s_0$. For $t > t_1$, the process evolves conditional on the filtration up to the time $t_1$ and the decision taken at that time: If the decision at $t_1$ is to shut down, we have $I_u = I_{t_1}, r = r_1 < r_0, s = s_1$; and if the decision is to not shut down, we have $I_u = I_{t_1}$.

$T$: The decision-horizon, a constant; the decision-maker who is considering a shutdown decision at a time $t_1 > t_0$ will consider the impact of her decisions in $[t_1, t_1 + T]$.

If a shutdown is implemented at a time $t_1$, we refer to $[t_1, t_1 + T]$ as the shutdown period, but we emphasize that $t_1 + T$ does not model the actual reopening time. It only represents the decision maker’s estimate of how far she prefers to look ahead. Given our focus on the capacity question during the early spread, we do not specifically model the reopening decision. In the extensions section, we provide a brief discussion of the reopening decision. Figure 1 shows two sample trajectories with and without shutdown.

We believe that this two-period framing reflects how the recent pandemic unfolded in various regions, U.S. states, and other countries. At the same time, it is flexible enough to
capture the variety of choices made by the regional decision-makers about the timing and intensity of shutdowns. Our focus in this paper is on the decision time $t_1$ and the shutdown period. This is the period in which the region experiences a peak in infections and in which the questions about medical equipment capacity have the most relevance.

### 3.5 Cost and performance considerations

As the decision-maker watches the infections increase in the early community spread period, it becomes clear that the disease will continue to grow and impose adverse health consequences. The degree of this adverse impact will depend on the region’s level of preparation (medical equipment capacity). In the absence of therapeutics or vaccines, the decision-maker must decide if she should take any preventive measures and, if so, when. These measures can range from a ban on large gatherings to a wholesale shutdown of businesses. Different measures may bring different levels of improvements in the infection rates, but they will all impose an adverse economic consequence on the region. In our basic model, we capture these adverse health and economic consequences by the following parameters:

- $m$: Preparation level or medical equipment capacity, a known constant, expressed in terms of the maximum number of infections that the regional health system can simultaneously manage. Note that there are two types of equipment, those that cannot be reused and those that can be. The former’s required capacity depends on the cumulative number of infections and the latter on the maximum number of infections. The required capacity for equipment like ventilators is of the latter type; this is how we define $m$ here.
- $PrB(I_t, m; r_i, s_i, T)$: Breach probability; the probability that starting with $I_1$ number of infections at the time $t_1$ and evolving with the rate $r_i$ and volatility $s_i$, $i = 0, 1$, the maximum number of infections in $[t_1, t_1 + T]$ will exceed the capacity $m$.
- $C_h$: A fixed health cost incurred if the maximum number of infections exceeds the capacity $m$.
- $C_e$: A fixed economy cost that the decision-maker believes will be incurred if she chooses to implement a regional shutdown.

The primary performance measure in our model is the probability of a breach. If a breach occurs, the region incurs a fixed health cost. The practical relevance of these measures is evident from reporting such as Sanger-Katz et al. (2020) that discusses the risk of shortage and efforts to prepare for a breach. The region can influence this probability by incurring a fixed economy cost. Once again, the extreme uncertainty surrounding these events makes it unlikely that regional decision-makers have precise estimates for such costs. We believe that a model built on the decision-makers’ preferred estimates or beliefs regarding such costs better suits the reality. See, for example, Leachman and Sullivan (2020) to support our modeling of differences in decision-makers’ estimates of these costs. Our performance measures capture this situation’s essential trade-off while directly connecting it to the paper’s main new feature, infection rate uncertainty. In the extensions section, we briefly discuss other, more detailed metrics to capture the impact on health and economic performance.

Decision-makers in different regions have different preferences about how they view the health versus economy debate and how optimistic they are about the availability of a cure. These preferences guide their estimates of $C_h, C_e$, and $T$. Given these preferences, a decision-maker can compare the costs of shutdown versus not shutting down as we analyze below.

### 4 THE ANALYSIS

In this section, we analyze the expression for the number of infections, develop a formula for the breach probability, and then analyze the decision to shut down and how it relates to the capacity level.

#### 4.1 Expressions for $I_t$

The first step in the analysis is to obtain an expression for the number of infections at time $t$, $I_t$. We use standard methods (Shreve, 2004) from the analysis of SDEs to analyze the infection dynamics to obtain the following solution.

**Proposition 1.**

$$I_t = \begin{cases} 
I_0 e^{\left(\frac{r_0 - \frac{s_0^2}{2}}{2}\right)(t-t_0) + s_0 \left(W_t - W_{t_0}\right)} & \text{for } t_0 \leq t < t_1 \\
I_1 e^{\left(\frac{r_1 - \frac{s_1^2}{2}}{2}\right)(t-t_1) + s_1 \left(W_t - W_{t_1}\right)} & \text{for } t \geq t_1 \text{ if a shutdown is implemented} \\
I_1 e^{\left(\frac{r_0 - \frac{s_0^2}{2}}{2}\right)(t-t_1) + s_0 \left(W_t - W_{t_1}\right)} & \text{for } t \geq t_1 \text{ if a shutdown is not implemented} 
\end{cases} \quad (3)$$
It is straightforward to recognize the dynamics of infections as a geometric Brownian motion, leading to the above result. Proofs are available in the Appendix in Supporting Information. The equations describe how the number of infections evolves, first with community spread in the period \( t_0 \leq t < t_1 \) and then with shutdown interventions in the period \( t \geq t_1 \).

### 4.2 The breach probability

We next develop an expression for the probability that the number of infections exceeds the capacity, a measure we have defined as the breach probability. The explicit expression of \( I_t \) in the above result allows us to determine the probability that the maximum number of simultaneous infections will exceed a given level of capacity over a time horizon. We evaluate this probability at the decision time \( t_1 \), conditional on the filtration up to the time \( t_1 \). Let \( \hat{I}_{t_1} = \max_{t_1 \leq t \leq t_1 + T} I_t \). For \( I_{t_1} \geq m \), the breach probability is trivially 1. For \( 0 < I_{t_1} < m \), we have the following result:

**Proposition 2.** \( PrB(I_{t_1}, m; r_i, s_i, T) = Pr[\hat{I}_{t_1} > m] \)

\[
= 1 - \Theta \left( \frac{\frac{1}{s_i} \log \left[ \frac{m}{I_{t_1}} \right] - \frac{1}{s_i} \left( r_i - \frac{\sigma_i^2}{2} \right) T}{\sqrt{T}} \right)
\]

\[
+ \left[ \frac{m}{I_{t_1}} \right] ^{\left( \frac{2s_i^2}{\sigma_i^2} - 1 \right)} \Theta \left( \frac{\frac{1}{s_i} \log \left[ \frac{m}{I_{t_1}} \right] - \frac{1}{s_i} \left( r_i - \frac{\sigma_i^2}{2} \right) T}{\sqrt{T}} \right)
\]

for \( i = 0, 1 \),

where \( \Theta(\cdot) \) is the standard normal cumulative distribution function.

We briefly discuss the plan for the rest of this section. As the time progresses forward from \( t_0 \), we arrive at the given decision time \( t_1 > t_0 \) and observe \( I_{t_1} \). We first propose a heuristic shutdown rule to determine if the decision-maker should shut down at time \( t_1 \). Next, rather than using a given time \( t_1 \), we use the rule to determine \( t_1 \) when the region should shut down and, at this time, what should be the region’s shutdown level capacity. We then extend the basic model and present an optimal stopping formulation of the problem.

### 4.3 The shutdown rule

The decision-maker has her preferred estimates \( C_h, \sigma_i, \) and \( T \), knows the current rate and volatility parameters \( r_0, s_0 \), and estimates that a shutdown will change them to \( r_1 < r_0, s_1 \). At time \( t_1 \) with \( I_{t_1} \) infections, should she shut down or not? We note that she is making this decision in a context characterized by a great deal of uncertainty about the pandemic’s evolution and difficulty estimating precise cost parameters. Therefore, our first approach is to present a simple heuristic rule that can capture the major drivers of this decision and analyze their relationships and trade-offs. The rule is as follows.

At a given time \( t_1 \), with \( I_{t_1} \) infections, a decision-maker will choose to remain open if for all \( i \leq I_{t_1} \)

\[
PrB(i, m; r_i, s_i, T) * C_h + \frac{C_e}{T} \geq PrB(i, m; r_0, s_0, T) * C_h
\]

that is if, \( \Delta PrB(i, m; r_0, s_0, r_1, s_1, T) = PrB(i, m; r_0, s_0, T) - PrB(i, m; r_1, s_1, T) \leq C_e/C_h \) and will choose to shut down otherwise.

The heuristic is directly motivated by the underlying trade-off. If a regional decision-maker chooses not to shut down, the higher infection rate \( r_0 \) will continue, and the region will incur the health cost \( C_h \) with a probability that, given her decision horizon, she will estimate as \( PrB(I_{t_1}, m; r_0, s_0, T) \). If she chooses to shut down, the infection rate will decrease to \( r_1 \). The region will incur the economic cost \( C_e \) and will also incur the health cost \( C_h \) with probability \( PrB(I_{t_1}, m; r_1, s_1, T) \). The heuristic decision rule suggests that the decision-maker will remain open as long as the shutdown has a higher cost. The rule relies on the assumption that the region starts its observation at a very small number of infections, \( I_{t_0} \). As it observes the number of infections grow from a small number, the rule recommends a shutdown at the first time when the estimated economic cost of shutdown exceeds the health cost of remaining open. We will highlight the limitations of this rule in the next section, but its advantage is clear. It is a simple connection between the decision-maker’s economy and health beliefs, and it roots the shutdown decision in the dynamics of how infections evolve.

Next, to make sure that the rule is easily implementable, it will be useful to show that the decision can simply be based on the observed number of infections at the decision time. The following result shows that there exist thresholds on different input parameters (costs, decision horizon, capacity level) such that if they are crossed, the region will not shut down. Otherwise, the rule is well-formed in the sense that there is a unique threshold on the number of infections that would determine the decision.

**Proposition 3.** At the decision time \( t_1 \), there exist thresholds \( \frac{C_e}{C_h}(m, T, I_{t_1}) \), and \( T(C_e/C_h, m, I_{t_1}) \) such that if either \( C_e/C_h > \frac{C_e}{C_h}(m, T, I_{t_1}) \) or \( T < T(C_e/C_h, m, I_{t_1}) \) holds, the decision rule will not recommend a shutdown; otherwise, there exists a unique shutdown level \( I^* \) such that the decision rule recommends a shutdown if \( I_{t_1} \geq I^*(C_e/C_h, m, T) \).

The benefit we gain by shutting down, represented by the \( \Delta PrB \) function, is the difference in breach probabilities achievable due to a shutdown-induced reduction in the infection rate. As capacity increases, individual breach probabilities may decrease, but the difference first increases then
decreases, achieving its maximum when the capacity is neither too low nor too high. Therefore, at some high level of cost ratio \( C_e / C_h \), even the peak benefit may be less than the cost. That is why the result states that if the preferred estimate of \( C_e / C_h \) is too high, the rule will not recommend a shutdown. An overly optimistic outlook for the availability of a cure or vaccine will cause the decision-maker to estimate a very short horizon \( T \) and that too will lead to the rule to not recommend a shutdown. Otherwise, at a given decision time \( t_1 \), if the number of infections is more than a threshold \( I^h \) (where \( h \) refers to the heuristic rule), then the rule recommends a shutdown.

The above result also shows the impact of some parameters on the threshold \( I^h \). The more effective the shutdown is, the less the shutdown-induced infection rate \( r_1 \) is. This increases the benefit of the shutdown leading the rule to recommend shutdown at a lower threshold. The proof of Proposition 2 allows us to understand that it is actually the ratio of capacity to the number of infections \( m/I_1 \) that determines the threshold behavior. A higher capacity allows for a higher number of infections before shutting down. The shutdown level capacity, \( m^h \) as a function of the regional decision-makers preferences \( (T, C_e/C_h) \) [Color figure can be viewed at wileyonlinelibrary.com]

Figure 2 visualizes the decision rule and the relationship between \( m \) and \( I_1 \). The heuristic decision rule trades off the shutdown-induced reduction in the breach probability (due to a lower infection spread rate) against the fixed economy cost. As the number of infections increases, the breach probability with different infection rates increases at different rates. The difference, \( \Delta PrB \) (the yellow surface in the figure), can be shown to increase first and then decrease. If the capacity or the economy-to-health cost ratio \( (C_e/C_h) \), the blue plane in the figure) is too high, \( \Delta PrB \) never reaches the threshold, and no shutdown is recommended. Otherwise, the rule recommends a shutdown when the difference in breach probabilities hits the cost ratio. As can be seen in the figure, an increase in the number of infections decreases the benefit, which can be compensated by an increase in capacity.

4.4 The shutdown level capacity

At this stage, a natural question for the decision-maker is to ask how much capacity to acquire? There was a great shortage of medical equipment like ventilators at the beginning of the pandemic; prices were reported to be temporarily high and variable. In this context, we focus only on the demand side and ask the following question: what should be the regional decision-maker’s target capacity at the time she chooses to shut down? In the severely capacity-constrained environment, this can guide the decision-maker’s efforts to acquire capacity.

We note that the capacity question is intertwined with the shutdown decision. As the number of infections grows, the decision-maker is working to acquire capacity and also considering a shutdown. Rather than taking the shutdown time \( t_1 \) as given, we consider a shutdown time when starting with a low \( I_{th} \), the number of infections reaches the threshold \( I^h \) from below, and the heuristic rule recommends a shut down for the first time. At this time, the equation \( \Delta PrB = C_e/C_h \) is satisfied (we note that a rigorous formulation of stopping time is delayed till the next section.) We are interested in the capacity at the shutdown time, referred to as the shutdown level capacity, \( m^h = m(C_e/C_h, T, I_1) \) satisfying \( \Delta PrB = C_e/C_h \). This measure of capacity level has practical significance because it represents the capacity a decision-maker will aim to have at the time she decides to shut down. This level, of course, depends on the decision-maker’s preferred estimates of \( C_e/C_h \) and \( T \). Therefore, we can now directly analyze the influence of the decision-maker’s preferences on the capacity level. Figure 3 displays this relationship.

Figure 3 is instructive as it allows us to map the variety of preferences and beliefs decision-makers have on their shutdown and capacity choices. As discussed earlier, different regional decision-makers bring different preferences to this decision. We organize these preferences along two dimensions: (i) estimate of \( C_e/C_h \), and (ii) estimate of \( T \). The former is an expression of the decision-maker’s desire not to mandate rules for closing business activity; the stronger is this desire, the higher is her estimate of \( C_e/C_h \). The latter is an expression of her hopefulness towards discovering a quick solution, a new drug, a vaccine, or just a gradual disappearance of the virus. It is not difficult to find quotes from various state governors expressing their differences of the more optimistic she is, the smaller is her estimate of \( T \) opinions.
regarding the importance they give to keeping the economy open. Nor is it difficult to find differences in governors’ opinions regarding how soon the pandemic will go away. Various state officials are on record expressing their differences regarding the importance they give to keeping the economy open and their opinions about how soon the pandemic will end.

The figure shows the shutdown level capacity as a function of these two preferences. For a given $T$, there is a threshold $C_e/C_h$ value above which the decision-maker will choose not to shut down; this gives the black dashed line in the figure. Regions whose preferences place them at or above this line will not shut down. For example, North Dakota, where the decision-maker put a high value on keeping the economy open, is one of the few states where no stay-at-home order was ever issued. Regions whose preferences intersect below this line can use this figure as a decision aid to determine capacity by noting the solid curve the intersection lies on; each solid curve indicates a specific $m^h$ value (curves for only some selected $m^h$ values are shown in the figure). The purpose of drawing these constant $m^h$ curves is to observe that different preferences can lead to the same capacity level.

Following any one of the constant $m^h$ curve shows that an increase in decision horizon requires higher capacity but can be compensated by a decrease in $C_e/C_h$ which would reduce the required capacity.

4.5 Per-infection cost: The optimal stopping formulation

In the previous section, we proposed a decision rule heuristic for the regional decision-maker based on the trade-offs she observes. We defended the decision rule approach based on the simplicity it offers in an uncertain environment with hard-to-estimate parameters. A main limitation of the decision rule is that it is limited to considering only fixed costs related to health and economy; there is no per-infection cost charged. Additionally, it takes a myopic view of the shutdown decision in the sense that a decision is made the first time when the cost of shutting down becomes less than the cost of keeping it open; the option value of waiting under uncertainty is not considered. It also does not offer a rigorous formulation and definition of the stopping (shutdown) time of the underlying stochastic process. To include the per-infection cost and address other issues, we present an optimal stopping formulation of the problem.

The optimal stopping formulation offers a more general form of the problem than the decision rule formulation. Another way to consider the progression from the decision rule to the optimal stopping formulation is to think of the decision rule as capturing early decision-making with less clarity about costs and more heterogeneity in inputs and outcomes and to think of the optimal stopping as capturing later-period, more mature decision-making with additional cost information and more homogeneity in outcomes. Given the complexity of the objective function in the optimal stopping formulation presented below, our goal in this section is relatively limited. We primarily focus on rigorously proving the existence of a stopping time. We believe that the analysis below can serve as a first step towards considering extensions mentioned in a later section.

For the optimal stopping formulation, we assume that $t_0 = 0$, with $t_0 = x$. The total cost function is given by $J^e(x)$ where the decision-maker implements a shutdown at an admissible stopping time, $\tau$, which is adapted to the underlying filtration. The number of infections at stopping time $\tau$, $I_\tau$, is a random variable. In the period before the shutdown, a per-infection cost rate $0 < c < C_e$ is incurred. If a shutdown is implemented at time $\tau$, we incur a cost equal to $C_e - C_h \Delta PrB$, which includes a fixed economy cost of the shutdown, less the saving in health cost that would be achieved due to shutdown-induced lower probability of a breach. The objective is to determine the optimal stopping time (shutdown time) that will minimize the total cost function. The optimal stopping time, $\tau^*$, is the time $\tau$ at which the expected total cost $J^e(x)$ achieves its minimum. The function $C_e - C_h \Delta PrB$ is referred to as terminal cost and $V(x)$ is the optimal cost function. The formulation is as follows:

$$J^e(x) = E_x \left[ \int_0^\tau c I_t(x, r_0, s_0) dt + C_e \right.$$

$$- C_h \Delta PrB(I_\tau, r_0, s_0, r_1, s_1, T) \left. \right]$$

$$\tau^* := \arg \inf_{\tau \geq 0} J^e(x), \quad (6)$$

$$V(x) = J^{e*}(x) = \inf_{\tau \geq 0} E_x \left[ \int_0^\tau c I_t(x, r_0, s_0) dt + C_e \right.$$

$$- C_h \Delta PrB(I_\tau, r_0, s_0, r_1, s_1, T) \left. \right]. \quad (7)$$

With some modifications to address the integration term in the optimal cost function, the above formulation has the form of a classical optimal stopping problem (see, Oksendal (2003, Chap. 10)). The solution approach involves using the characteristic function of the underlying stochastic process to set up a differential equation. This equation, along with the value-matching and smooth-pasting conditions, leads to a candidate solution $V(x)$ for the optimal cost function and the definition of a continuation region $D$ that specifies the range of $x$ in which the solution proposes not shutting down. We then use the verification theorem (see, e.g., Van Handel (2007)) to show that this candidate function satisfies the requirements to be the optimal cost function. The proof of Proposition 4 shows the above steps and is supported by three lemmas to prove that the verification conditions are satisfied. Finally, this leads us to the following existence result.

**Proposition 4.** If $c > \frac{r_0}{m^{2/n}/\rho}$ then there exists a threshold $I^0 < m$ such that it is optimal to shut down if $I_t > I^0$. 
The result shows that under the optimal policy, the continuation region can be written as \( D := \{ x | x \leq I^0 \} \) and the shutdown time is the optimal stopping time defined as \( t^* := \inf \{ t > 0 : I_t \notin D \} \). The proof of Proposition 4 holds for any initial infection level and defines the threshold \( I^0 \) as the supremum of values where smooth pasting condition is valid. If the infections exceed \( I^0 \), the cost of infections will exceed the cost of imposing a shutdown. The role of the assumption \( c > \frac{-\tau_0}{m^{\frac{2\sigma^2}{\sigma^2}} < 0} \) is to ensure \( I^0 < m \). In the case of small \( c \), \( 0 < c < \frac{-\tau_0}{m^{\frac{2\sigma^2}{\sigma^2}}} \), we still have a finite stopping time and a threshold \( I^0 \) such that if \( I_t > I^0 \), it would always be optimal to shut down. This is so because the continuing infection cost increases in the number of infections, but in contrast, the terminal economy and health costs are bounded and also the underlying geometric Brownian motion has a positive drift. Though there would be no guarantee that \( I^0 < m \), making the result less insightful in a practical setting.

It is useful to consider the difference between the decision rule and the optimal stopping approaches. Without a per-infection cost, the decision rule looks for the first time that the trade-off between fixed economy and health costs justifies shutdown. With the per-infection cost, the optimal stopping formulation shifts focus to how that cost determines a level at which the influence of the per-infection cost justifies a shutdown. The qualitative behaviors of the optimal stopping problem and the decision rule strategy are fundamentally different. For the decision rule strategy, without the per-infection cost, it is possible that it is never optimal to shut down. But with the optimal stopping problem, with a non-zero per-infection cost, the shutdown should eventually occur with Probability 1. This is because the total per-unit infection cost will continue to increase and exceed the combined impact of health and economy costs as long as the latter costs are bounded.

## 5 COMPENSATING FOR LIMITED CAPACITY

Starting with a decision maker’s preferred estimates of costs of health and economy (\( C_h, C_e \)), and the level of her optimism about the availability of a cure or vaccine (\( T \)), the previous section analyzed the shutdown level of capacity. In a capacity-constrained capacity environment, the desired level of capacity may not be available. The objective would then be to reduce the probability of breach. Therefore, in this section, we analyze other actions a decision-maker can take to reduce the breach probability if the medical equipment capacity is constrained to a given value.

### 5.1 Reusability

We start with a simple observation regarding the impact of reusability of the capacity on the breach probability. See, for example, Srinivasan et al. (2020) for ventilator sharing. Recall that the capacity \( m \) was introduced in terms of the maximum number of infections that can be simultaneously treated. For example, let \( V \) represent the ventilator capacity and let the fraction \( \alpha \) (Gelles & Petras, 2020) represent the fraction of infected persons who will need the equipment. Then, \( m = V / \alpha \). Any reuse decreases \( \alpha \) and, for a fixed \( V \), increases \( m \). As a result, the breach probability decreases.

### 5.2 Shutdown intensity \( r_1 \)

Next, we discuss the impact of \( r_1 \), the infection spread rate in the shutdown period. The decision-maker may have a range of mandated behaviors as possible options, from stay-at-home orders to mask orders, which translate into different degrees of reduction in the infection rate post shutdown. For example, Flaxman (2020) lists recommended strategies ranging from self-isolation, ban on public events, social distancing, school closures, and lockdown with the corresponding percentage decrease in transmissibility from 5% to 45%. Our model can predict the impact of such reduction on the breach probability. Part (a) of the following result formalizes this observation.

**Proposition 5.**

\[
\frac{\partial \Pr(B | I_t, m, r_1, s_1, T)}{\partial r_1} \geq 0.
\]

\[
\frac{\partial \Pr(B | I_t, m, r_1, s_1, T)}{\partial h_1} \geq 0.
\]

### 5.3 Shutdown level \( I_t \) or, equivalently, timing \( t_1 \)

The previous section proposed a heuristic rule and an optimal stopping formulation to determine the shutdown level, but if the region does not have the capacity it needs, it can choose to deviate from the prescribed shutdown level. To formalize this argument, we first look at the impact of the shutdown level on the breach probability. Part (b) of the result shows that the breach probability increases as the shutdown level infections increase.

These observations allow us to point out substitutability between the capacity and the shutdown infection rate and shutdown level. As Figure 4 shows, a lack of capacity increases the breach probability, but it can be compensated by taking action to increase shutdown intensity. The figure also shows that any decrease in the breach probability due to lack of capacity can be compensated by starting the shutdown early at a smaller number of infections. These strategies can be observed in real contexts as well. For example, by some measures of available bed capacity (adjusted for size), states like Oregon and Washington are the ones with the tightest capacity, while Texas and Minnesota come out at the top (Jones, 2020). Both Washington and Oregon were the earliest states to issue stay-at-home orders, and Texas was one of the last ones (see Mervosh et al., 2020).
MANAGING MEDICAL EQUIPMENT CAPACITY

Production and Operations Management

FIGURE 4  Variation in breach probability with variation in preparation level \( m \), rate of spread during shutdown \( r_1 \) and shutdown level \( I_{t_1} \) [(a) \( T = 5, s_1 = 1, I_{t_1} = 10 \); (b) \( T = 5, r_1 = 1.2, s_1 = 1 \)] [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 5  Variation in breach probability with volatility for short time horizon [\( m = 50, I_{t_1} = 10, r_1 = 1.5, T = \{0.5, 1.0, 1.5\} \)] [Color figure can be viewed at wileyonlinelibrary.com]

5.4  Shutdown enforcement \( s_1 \)

Recalling our discussion for the introduction of volatility parameter in an earlier section, we note that volatility exists due to variations in how people behave in their interactions with each other. Before the shutdown, people follow their normal behavior patterns; some interact more and others less, with each interaction a possible source of infection. After the shutdown, the variation in people’s behavior may be influenced by the strength of shutdown messaging and enforcement of its rules. Improvements in testing capacity and accuracy may also reduce noise in the system. We interpret these as reductions in volatility. Its impact on the breach probability can be complex. The intuition suggests that reduction in variability should benefit the performance. As Figure 5 shows, this is indeed the case. But, in cases with long time horizons, it can also increase the breach probability.

This section considered strategies that a region can implement on its own. In the next section, we consider strategies that require coordination across regions.

6  COORDINATION ACROSS REGIONS

A major theme in reporting about the early spread of the disease in the United States is the lack of coordination between different states and regions. In the absence of coordination, many states and regions had closed down their borders, charted their own communication strategies, and relied on their own medical equipment capacity (Jancowicz et al., 2020). This description fits closely with the model we have developed in the previous section. We will refer to the previous model as the individual-region model.

There are different ways we can conceive coordination between individual regions in order to improve outcomes. In the United States, a lack of coordination on acquiring and managing the ventilator capacity came in for early criticism. Later, several states changed their policies to make their beds and ventilators available to both neighboring and non-adjacent states (Mulvihill, 2020; Ronayne, 2020). A natural question to ask is about the improvement in outcomes that can be achieved by combining two individual regions’ capacity. When two regions agree to make their total medical equipment capacity available to treat patients from either region, we refer to it as medical equipment capacity-sharing. It appears intuitively clear that such a sharing of capacity should be better for the regions. Our model offers a direct way to check this intuition; does this always hold true? We address this question as the first step in this section.

Our interest, however, is not just in showing the existence of such capacity-sharing improvement. We are also interested in identifying and investigating other features of regional coordination that can influence the benefit of capacity sharing. As we better understand the drivers of this value, we aim to show that other features of regional coordination, such as information and movement coordination, can have either a positive or negative influence on the benefit of capacity sharing. Later sub-sections address such features.

6.1  Medical equipment capacity sharing in two independent regions

Let us identify two regions by assigning a superscript index \( k \) to each relevant parameter where \( k = a, b \), with each index value representing one region. The regions are assumed to be independent in the sense that the infections evolve independently in each region. The next section considers the
case of interdependent regions. We will focus on the dynamics of infections in the shutdown periods $t \geq t_1$ in each region, assuming a shutdown is implemented at $t_1$. Recall that in the individual region model discussed earlier, $\hat{I}_1$ represents the maximum number of infections in $[t_1, t_1 + T]$ and the capacity $m$ represents the number of simultaneous infections that the existing medical equipment capacity can treat. The health cost is incurred if the maximum number of infections exceeds the capacity, an event we refer to as a breach. To keep the focus on the operational performance, we consider the impact of capacity sharing on the breach probability represented by $\Pr[\hat{I}_1 > m + m^\prime]$. The proof of the above result is based on the consideration of an intermediate system. This system focuses on the sum of maximum infections in the two regions, $\hat{I}_1 + \hat{I}_1$, and corresponding breach probability $\Pr[\hat{I}_1 + \hat{I}_1 > m + m^\prime]$. We note that this sum, $\hat{I}_1 + \hat{I}_1$, does not necessarily represent the true requirement, $\hat{I}_1 + \hat{I}_1$, of medical equipment at any time because the chance that two independent regions will experience the maximum need simultaneously is vanishingly small. As we will discuss in the next section, interdependence between regions will bring $\hat{I}_1 + \hat{I}_1$ closer to $\hat{I}_1 + \hat{I}_1$. The consideration of the sum of the maximums and the breach probability in the intermediate system provides a bridge to understand the breach probability we are really interested in, the maximum of the sum versus total capacity in the shared-capacity case. The proof first shows that the probability of $\hat{I}_1 + \hat{I}_1$ breaching the total capacity is less than the probability of each region breaching its own capacity if and only if the capacity is above a threshold. The proof then shows that the probability of $\hat{I}_1 + \hat{I}_1$ breaching the total capacity will be even less.

Our construction of this intermediate-system measure, $\hat{I}_1 + \hat{I}_1$, not just helps with the above proof, it also helps build an intuitive explanation of the drivers of the benefits of capacity sharing, an explanation that will lay a foundation for the next section. We find it helpful to differentiate between two drivers of the capacity-sharing benefit. First, there is the traditional risk-pooling benefit. At any given instant, the sum of two random variables representing independent regional infections is less variable than individual regions’ infections. One would expect it to drive some part of the sharing benefits. That is represented by the difference between $\hat{I}_1$ or $\hat{I}_1$ and, $\hat{I}_1 + \hat{I}_1$. Second, there is the peak-timing benefit. Given that each region’s infections evolve independently, the time difference between the two regions’ peaks should benefit the deployment and use of the limited total equipment capacity. That is represented by the difference between $\hat{I}_1 + \hat{I}_1$ and $\hat{I}_1 + \hat{I}_1$. We will use these two drivers, risk pooling and peak timing, to discuss our next section’s observations. The discussion of peak-timing as a source of capacity-sharing benefits adds new insight to the literature.

However, why may capacity sharing not be beneficial if each region’s capacity is below a threshold? If an individual region’s capacity is so small that $\hat{I}_1$ (or $\hat{I}_1$) is very likely to cross it, then $\hat{I}_1 + \hat{I}_1$ is also very likely to cross the sum of the two capacities. If so, there should not be much value in sharing capacity. With a larger capacity, it becomes possible that one region crosses its capacity, but the other does not. Such cases open up the possibility that total infections may
still be below the combined capacity. Both risk-pooling and peak-timing play a role in this phenomenon.

6.2 Evolution of infections with information and movement coordination

We are now ready to consider other aspects of coordination across regions that can promote interdependence between the two regions’ infection evolutions. Specifically, we analyze two different features of regional coordination. First, inter-regional coordination may allow movement across the physical boundaries. We label such physical coordination leakage and seek to understand its impact on capacity sharing. Second, coordination may allow regions to exercise uniformity in communication strategies. We label this practice information coordination. We propose a way to incorporate it into our model and ask how this may influence capacity sharing benefits. Later, we briefly discuss another feature, shutdown coordination, where regions may synchronize their shutdown policies. This section models the evolution of infections with these new features, and the next section considers the impact of such interdependent evolution on the benefit of capacity sharing.

Unlike our discussion of capacity sharing in the previous section, where each region’s infections evolved independently, the coordination features we describe in this section require us to model the interdependence between how regional infections evolve. To that end, we first propose a new dynamics of infection evolution that links the two regions. We introduce two new parameters: a correlation index \( \rho \), and a leakage parameter \( l \). They are explained in detail right after the dynamics presented below. Assuming a shutdown at \( t_1 \), for \( t \geq t_1 \):

\[
dI^a_t = r^a_t \left( (1 - l) I^a_t + l I^b_t \right) dt + s^a_t I^a_t dW^a_t, \tag{8}
\]

\[
dI^b_t = r^b_t \left( (1 - l) I^b_t + l I^a_t \right) dt + s^b_t I^b_t dW^b_t, \tag{9}
\]

where \( W^a_t \) and \( W^b_t \) are correlated Brownian motions. That is, \( [W^a_t, W^b_t] = \rho t, \rho \in (-1, 1) \).

To model physical coordination between different regions such that the travel between the regions is permitted but the travel from outside to either region is not, we use the leakage parameter \( l \). Physical coordination across the two regions allows some movement across the borders, thus enlarging the area that can be insulated from outside influences. Such a policy has its advantages in providing the residents with a larger area for movement, and thus lessening the pressure brought on by the shutdown orders. For example, such physical coordination existed between states like New York and New Jersey in the early days of the pandemic. As seen in the equations above, a fraction \( l \) of the current infected population in one region may leak to the other region. Once there, this infected fraction will contribute to the other regions’ spread.

To model information coordination, we focus on the sources of noise in the infection rate. Each region has its idiosyncratic character that drives the variation in the range of people’s social behavior. The lack of testing capacity and inconsistent data reporting also introduce uncertainty. The Wiener process captures this variation. With information coordination, regions implement coordinated messaging and uniform standards of data capture and reporting. Initially, the lack of such coordination was heavily criticized. Later, multi-state collaborations were launched to coordinate data accuracy; see, for example, Battaglia (2020). The evidence of heterogeneity in health-related social behavior is available in Allcott (2020). The connection between consistent messaging and uniformity in social behavior is discussed in Gadian et al. (2021). In this framing, coordination of information and messaging across two regions will induce consistency in people’s social behavior. We model such coordination as an increase in the correlation between the two regions’ Wiener processes.

Since we are proposing a new system of SDEs to model the spread of infections due to interactions between two regions, it is important to derive the conditions under which there exists a unique positive solution to the coupled equations. We proceed to do so in the following result.

**Theorem 1.** The coupled system of SDEs presented above has a unique positive strong solution if

\[
\max \left\{ r^a_t (1 - l) - r^b_t (1 - l) + \frac{(s^a_t)^2}{2}, -r^a_t (1 - l) + r^b_t (1 - l) \right\} \leq 0.
\]

We show the existence of a unique solution using the almost linear growth and Lipschitz continuity condition. The proof of the positivity of the solution is more involved. We create a sequence of stopping times, which are the first hitting times of a sequence of boundaries very close to 0. We then show that as the boundaries tend toward 0, the stopping time tends to infinity. The detailed proof is presented in the Appendix in Supporting Information. The literature on SDEs to model the spread of infections is limited. We believe that with good parameter estimation, the above equations can be used to model the early spread of infections.

Because the set of equations above does not yield an analytical solution, we present further insights based on the above system’s simulation analysis. The analysis is carried out on a testbed of parameter ranges. The observations reported in the next section hold true for this testbed. For clarity, we present graphs for specific parameter combinations, which are mentioned with each graph.
6.3 Capacity sharing with information and movement coordination

Before analyzing the impact of regional interdependence, driven by information and movement coordination, on capacity sharing, we first use the simulation to set a baseline by revisiting Section 6.1 where there was no movement coordination (zero leakage) and no information coordination (zero correlation). Figure 6a shows that as capacity increases, the breach probability for both the individual region case, \( \Pr[\hat{I}^*_a > m] \), and the shared capacity case, \( \Pr[\hat{I}^*_a + \hat{I}^*_b > 2m] \), decreases as we expect them to. There is, however, a unique threshold capacity level below in which sharing capacity does not reduce breach probability. Benefits of capacity-sharing occur only if the capacity is above a threshold; this is in line with Proposition 7. Figure 6b digs deeper into this by adding the breach probability for the intermediate system, \( \Pr[\hat{I}^*_a + \hat{I}^*_b > 2m] \). When capacity is above the threshold, the graph can differentiate between the risk-pooling effect and the peak timing effect as discussed below Proposition 7.

6.4 Impact of physical coordination

Physical coordination across the two regions allows movement across the borders increasing the leakage \( l \) and inducing interdependence between two regions’ evolutions. Our focus is on its impact on the breach probability. We note that unlike Figure 6 and Section 6.1, the following discussion computes, \( \hat{I}^*_a, \hat{I}^*_b, \) and \( \hat{I}^*_a + \hat{I}^*_b \), based on simulation of the coupled SDEs presented in Section 6.2.

While the existence of a capacity threshold mentioned in Proposition 7 was proved for the independent region case, it is notable that a similar notion holds in the case of interdependent regions. Figure 7 shows that capacity sharing improves breach probability for the high-capacity case but not for the low-capacity case. When capacity is high, the figure shows that the breach probability in both individual-region and capacity-sharing cases is relatively indifferent to increasing leakage, but when capacity is tight, increasing leakage increases breach probability in the capacity-sharing case. At any given time, the SDEs with leakage replace the individual regions’ number of infections with a combination of both regions’ infections. This synchronizes the two trajectories bringing their peaks closer. This peak-timing effect then leads to an increase in breach probability when capacity is shared. The sample trajectories in Figure 8 show a visual example of how an increase in leakage can bring the two regions’ trajectories together. If we assume increased travel to be a proxy for increased leakage then this suggests inverse outcomes for an increase in travel, an observation similar to that made in some empirical work (see, e.g., Knittel & Ozaltun, 2020).

6.5 Impact of information coordination

Information coordination across the two regions means coordination on messaging and communications transmitted to populations in two regions. We interpret this as an increase in correlation between the two Wiener processes underlying the spread of infections in the regions. We consider the impact of the increasing correlation on the breach probability.

Once again, Figure 9 supports the existence of a capacity threshold even with interdependent regions; capacity-sharing reduces breach probability only with high capacity. The figure shows that increasing correlation does not have a measurable influence on the individual region’s breach probability. As the SDEs show, combining two Wiener processes still
FIGURE 7 Variation in breach probability with leakage under capacity coordination and no capacity coordination. The graphs are generated under the scenarios of high and low individual capacity. \[ r_1^a = r_1^b = 1.2, s_1^a = s_1^b = 0.5, I_{1a}^t = I_{1b}^t = 2, \rho = 0, \text{high } m = 1500, \text{low } m = 200 \] [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 8 Difference in the spread of infection in region B due to highly infected region A. Left: Leakage \( l = 0 \) signifying closed borders; right: Leakage \( l = 0.1 \) signifying some interactions [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 9 Variation in breach probability with correlation under capacity coordination and no capacity coordination. The graphs are generated under the scenarios of high and low individual capacity. \[ r_1^a = r_1^b = 1.2, s_1^a = s_1^b = 0.5, I_{1a}^t = I_{1b}^t = 2, l = 0, \text{high } m = 1500, \text{low } m = 200 \] [Color figure can be viewed at wileyonlinelibrary.com]
means the same degree of noise for the individual regions. The figure further shows that with high capacity, increasing correlation brings the individual-region and capacity-sharing breach probabilities closer, thus reducing the benefit derived by capacity-sharing. This too can be explained by the peak timing effect. The larger the correlation, the more likely it is that peaks occur close to each other, driving up the breach probability for capacity-sharing.

### 6.6 Shutdown coordination

We close this section with an observation related to the coordination of shutdown timing. In our analysis till now, we have assumed that the two regions’ shutdown levels are the same and the shutdowns at the same time. However, given the random evolution of infection in the early spread period, the two shutdown times will not be the same. This would be a further reason for the peaks to occur at different times. Therefore, the peak timing effect will drive the breach probability down even further. Conversely, we expect coordination on shutdown timing to decrease the breach probability under capacity sharing.

To summarize, this section shows that medical equipment capacity sharing across two regions can be beneficial but not when capacity is very tight. Allowing movement across the border, interestingly, can smooth the infection trajectories and offer further benefits. Any coordination that increases the correlation between the underlying random processes can bring peaks in sync and reduce the benefits. We note that these conclusions say nothing about other benefits regional coordination can bring. However, our results suggest that regions should assess the benefit of capacity sharing in light of other elements, such as leakage and correlation, included in the coordination agreement.

### 7 DISCUSSION OF EXTENSIONS AND CONCLUSION

We believe that a significant part of the paper’s contribution stems from being the first to apply SDE-based modeling to pandemic-related capacity. The model can capture the uncertainty in infection spread and yet remain tractable enough to address capacity questions. Being early in this research stream, we have mainly focused on building a model that captures this situation’s essential features and asks fundamental questions. However, other questions remain, raising the possibility of a variety of extensions. We discuss them in this section.

#### 7.1 Additional details in the infection-spread dynamics

Our focus on the early-stage dynamics led to the SDE model we used here. The main modeling goal is to produce a trajectory that can capture an early exponential increase and then follow it up with a variety of possibilities leading to a plateau, long-term decrease, or resurgence. One extension would be to use time-dependent rate and volatility parameters in the geometric Brownian motion. Depending on how these parameters change, it is possible to produce a variety of trajectories. Such a model can better fit the wide range of long-term infection-spread patterns we have observed across US states and other regions.

Another extension would be to bring the infection-spread SDE closer to the epidemiological literature by modeling additional compartments. Specifically, we can include the susceptible compartment in the model where the sum of the numbers of susceptible people and infected people equals a fixed population size.

#### 7.2 Additional details in performance measures

We have focused on the breach probability as our primary performance measure. There are, however, other performance measures that directly depend on medical equipment capacity. The total number of infections over the horizon or the cumulative number of infections for which capacity is unavailable are important performance measures to compute. On the economy side, the region’s size and the shutdown length will be important determinants of the total cost. An extension can include these considerations in the optimal stopping problem.

Combining the previous two extensions, modeling long-term trajectories and additional performance metrics, would also allow us to include other decisions in the optimal stopping problem. The most important example is the reopening decision. The reopening decision will depend on how the infections are likely to spread after reopening and its implications for health and the economy.

#### 7.3 Supply-side model

We limited the scope of our model to modeling the demand-side impact of capacity. In our judgment, that is the most relevant question in a constrained capacity environment. However, this leaves open the question of how this capacity is acquired. The most straightforward extension would be to include a per-unit capacity cost in the model. The reality, however, is more complex. Given the severe shortages, each region is competing against others to acquire limited capacity. Thus, an extension with a competitive supply model would be more appropriate.

#### 7.4 Additional details in the coordination model

The previous point also offers a reason to extend our analysis of regional coordination. While we have focused on the
coordination of existing capacity, the coordination on acquiring new capacity can be a worthy extension. This extension would be akin to the development of cooperative game models for distributing coordination benefits in newsvendor-model setting; such models followed the earlier literature showing the benefits of pooling coordination. Other extensions for the coordination model would include consideration of non-identical regions and shutdown policies.

7.5 | Testable propositions

Our theoretical model delivers conclusions that relate differences in decision-makers’ preferences (about the relative costs of economy and health and expectation of a cure) to differences in their shutdown policies. Another conclusion we draw relates the value of capacity sharing to the tightness of capacity and the extent of movement and messaging coordination across neighboring regions. An extension that tests these theoretical conclusions against actual observations in the US states and other countries would be a useful exercise.

We conclude by reciting some of the main themes in the paper. This paper is a first attempt to explicitly include infection rate uncertainty in the pandemic modeling while maintaining tractability to analyze the impact of limited capacity on the region’s health performance. We develop an SDE model of the spread of infection in the early stages of a pandemic, analyze its running maximum process, and use it to compute the required medical equipment capacity. The performance measure of our choice is the breach probability.

Our results relate regional decision-makers’ preferred estimates of economy and health costs and time-till-cure to the shutdown policy, the level of infections at which the region should shut down, and the corresponding medical equipment capacity. Given the shortage of capacity, we analyze a list of options to compensate for it. These options include reusability of the equipment, shutdown intensity, earlier shut down, and reduction in the messaging noise. The model analyzes the impact of these actions on the breach probability. We then consider sharing capacity across regions as a way to influence the breach probability. We show that shared capacity may not benefit individual regions if the capacity is too tight. Movement across borders decreases the breach probability, and an increasing correlation between the random processes underlying two regions’ infections increases the breach probability. As discussed above, we offer several ways to extend the model.

REFERENCES

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2020). Testing, voluntary social distancing and the spread of an infection. NBER Working Paper Series Working Paper 27483.

Allcott, H. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. http://web.stanford.edu/~gentzkow/research/social_distancing.pdf

Allen, L. (2015). Stochastic population and epidemic models: Persistence and extinction. Springer.

Alvarez, F. E., Argente, D., & Lippi, F. (2020). A simple planning problem for COVID-19 lockdown. [Technical Report]. National Bureau of Economic Research. Working paper 2698.

Angelus, A. (2020). Distributed renewable power generation and implications for capacity investment and electricity prices. Production and Operations Management, 30(12), 4614–4634. https://doi.org/10.1111/poms.13241

Bailey, N. T. J. (1975). The mathematical theory of infectious diseases and its applications. Griffin.

Bass, F. M. (1969). A new product growth model for consumer durables. Management Science, 15(5), 215–227. https://doi.org/10.1287/mnsc.15.5.215

Battaglia, R. (2020, July 28). State auditor unveils new multi-state collaboration to improve COVID-19 data accuracy. Delaware Public Media. https://www.delawarepublic.org/news/2020-07-28/state-auditor-unveils-new-multi-state-collaboration-to-improve-covid-19-data-accuracy

Breuninger, K. (2020, April 13). N.Y. Gov. Cuomo says the ‘ultimate resolution’ of coronavirus will come with vaccine in 18 months. CNBC.COM. https://www.cnbc.com/2020/04/13/coronavirus-cuomo-says-ultimate-resolution-will-be-vaccine-in-18-months.html

Britton, T. (2009). Stochastic epidemic models: A survey. Math. Biosciences, 225(1), 24–35. https://doi.org/10.1016/j.mbs.2010.01.006

Buchen, P. (2012). An introduction to exotic option pricing. CRC Press.

Cai, S., Cai, Y., & Mao, S. (2019). A stochastic differential equation SIS epidemic model with two independent Brownian motions. Journal of Mathematical Analysis and Applications, 474(2), 1536–1550. https://doi.org/10.1016/j.jmaa.2019.02.039

Dixit, R. K., & Pindyck, R. S. (1994). Investment under uncertainty. Princeton University Press.

Drakopoulos, K., & Randhawa, R. S. (2021). Demand control of information products: Why perfect tests may not be worth waiting for. Management Science, 67(11), 6678–6693. https://doi.org/10.1287/mnsc.2021.4029

Fisher, M. (2020, April 23). R0, the messy metric that may soon shape our lives, explained. New York Times. https://www.nytimes.com/2020/04/23/world/europe/coronavirus-R0-explainer.html

Flaxman, S. (2020). Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Imperial College COVID-19 Response Team. https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-30-COVID19-Report-13.pdf

Gadarian, K., Goodman, S., Wallace, S., & Pepinsky, T. B. (2021). Partisan health, behavior, and policy attitudes in the early stages of the COVID-19 pandemic. PloS One, 16(4), e0249596. https://doi.org/10.1371/journal.pone.0249596

Gard, T. C. (1988). Introduction to stochastic differential equations. Marcel Dekker.

Gelles, K., & Petras, G. (2020, April 10). How ventilators work and why COVID-19 patients need them to survive coronavirus. USA Today. https://www.usatoday.com/in-depth/news/2020/04/10/coronavirus-ventilator-how-works-why-covid-19-patients-need/2942996001/

Goldberg, D. (2020, March 17). ‘It’s going to disappear’: Trump’s changing tone on coronavirus. POLITICO.COM. https://www.politico.com/news/2020/03/17/how-trump-shifted-his-tone-on-coronavirus-134246

Gray, A., Greenhalgh, D., Hu, L., Mao, X., & Pan, J. (2011). A stochastic differential equation SIS epidemic model. SIAM Journal on Applied Mathematics, 71(3), 876–902. https://doi.org/10.1137/10081856X

Halpern, N. A., & Tan, K. S. (2020). United States resource availability for COVID-19. Society of Critical Care Medicine. https://www.sccm.org/getattachment/Blog/March-2020/United-States-Resource-Availability-for-COVID-19/United-States-Resource-Availability-for-COVID-19.pdf?lang=en-US

Harrison, M. (2013). Brownian models of performance and control. Cambridge University Press.

Ho, T. -H., Savin, S., & Terwiesch, C. (2002). Managing demand and sales dynamics in new product diffusion under supply constraint. Management Science, 48(2), 187–206. https://doi.org/10.1287/mnsc.48.2.187.257

Jancowicz, D. J., Handfield, R., & Guinto, P. (2020, September 17). Why the U.S. still has a severe shortage of medical supplies.
JAIN AND RAYAL

Production and Operations Management

Harvard Business Review. https://hbr.org/2020/09/why-the-u-s-still-has-a-severe-shortage-of-medical-supplies

Jones, R. P. (2020). Would the United States have had too few beds for universal emergency care in the event of a more widespread COVID-19 epidemic? International Journal of Environmental Research and Public Health, 17(14), 5210. https://doi.org/10.3390/ijerph17145210

KFF. (2021). State COVID-19 data and policy actions. Kaiser Family Foundation Reports. https://www.kff.org/report-section/state-covid-19-data-and-policy-actions-policy-actions/#socialdistancing

Knittel, C. R., & Ozaltun, B. (2020). What does and does not correlate with COVID-19 death rates (Working Paper 27391), NBER.

Kumar, S., & Swaminathan, J. M. (2003). Diffusion of innovations under supply constraints. Operations Research, 51(6), 866–879. https://doi.org/10.1287/opre.51.6.866.24918

Kwon, D. H. (2010). Invest or exit? Optimal decisions in the face of a declining profit stream. Operations Research, 58(3), 638–649 https://doi.org/10.1287/opre.1009.0740

Leachman, M., & Sullivan, J. (2020, March 20). Some states much better prepared than others for recession. Center on Budget and Policy Priorities. https://www.cbpp.org/research/state-budget-and-tax/some-states-much-better-prepared-than-others-for-recession

Martcheva, M. (2015). Introduction to mathematical epidemiology. Springer.

Mehrotra, S., Rahimian, H., Barah, M., Luo, F., & Schantz, K. (2020). A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19. Naval Res Logistics, 67, 303–320 https://doi.org/10.1002/nav.21905

Mervosh, S., Lu, D., & Swales, V. (2020, April 20). See which states and cities have told residents to stay at home. New York Times. https://www.nytimes.com/interactive/2020/03/17/upshot/hospital-bed-shortages-coronavirus.html

Mulvihill, G. (2020, April 10). U.S. states share, get creative in hunt for medical supplies. A.P. News. https://apnews.com/article/nd-state-wire-wa-state-wire-ar-state-wire-ia-state-wire-virus-outbreak-bba8c890271ed9517682e81eece2c5e4

Oksendal, B. (2003). Stochastic differential equations: An Introduction with applications. Springer.

Roberts, S. (2020, August 4). How to think like an epidemiologist, New York Times. https://www.nytimes.com/2020/08/04/science/coronavirus-bayes-statistics-math.html

Ronayne, K. (2020, April 7). California lends 500 ventilators to 4 states, 2 territories. U.S. News. https://apnews.com/article/nv-state-wire-district-of-columbia-wa-state-wire-or-state-wire-virus-outbreak-b38e3388dc5742a87ec652602854a1b1

Sanger-Katz, M., Kliff, S., & Parlapiano, A. (2020, March 17). These places could run out of hospital beds as coronavirus spreads. New York Times. https://www.nytimes.com/interactive/2020/03/17/upshot/hospital-bed-shortages-coronavirus.html

Sazonov, I., Kelbert, M., & Gravenor, M. B. (2011). A two-stage model for the SIR outbreak: Accounting for the discrete and stochastic nature of the epidemic at the initial contamination stage. Mathematical Biosciences, 234(2), 108–117. https://doi.org/10.1016/j.mbs.2011.09.002

Shaked, M., & Shanbhikumar, J. G. (2007). Stochastic orders. Springer Science & Business Media.

Shen, W., Duenyas, I., & Kapuscinski, R. (2011). New product diffusion decisions under supply constraints. Management Science, 57(10), 1802–1810. https://doi.org/10.1287/mnsc.1100.1277

Shiryaev, A. N. (2007). Optimal stopping rules. Springer Science & Business Media.

Shreve, S. E. (2004). Stochastic calculus for finance II. Springer Finance.

Srinivasan, S., Ramadi, K. B., Vicario, F., Gwynne, D., Hayward, A., Lagier, D., Langer, R., Frassica, J. J., Baron, R. M., & Traverso, G., (2020). A rapidly deployable individualized system for augmenting ventilator capacity. Science Translational Medicine, 12(549), eabb9401. https://doi.org/10.1126/scitranslmed.abb9401

Tornatore, E., Baccellato, S. M., & Vetro, P., (2005). Stability of a stochastic SIR system. Journal of Physics A, 354, 111–126. https://doi.org/10.1016/j.physa.2005.02.057

US FDA. (2020). Medical device shortages during the COVID-19 public health emergency. https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/medical-device-shortages-during-covid-19-public-health-emergency

Van Handel, R. (2007). Stochastic calculus, filtering, and stochastic control lecture notes. https://web.math.princeton.edu/~rvan/acm217/ACM217.pdf

Vynnycky, E., & White, R. (2010). An Introduction to infectious disease modelling. Oxford University Press.

Yang, L., Cui, S., & Wang, Z. (2022). Design of Covid-19 testing queues. Production and Operations Management, https://doi.org/10.1111/poms.13673

Wikipedia. (2020). U.S. state and local government responses to the COVID-19 pandemic. https://en.wikipedia.org/wiki/U.S._state_and_local_government_responses_to_the_COVID-19_pandemic

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: Jain, A., & Rayal, S. (2023). Managing medical equipment capacity with early spread of infection in a region. Production and Operations Management, 32, 1415–1432. https://doi.org/10.1111/poms.13684