Managing Two-Dose COVID-19 Vaccine Rollouts with Limited Supply: Operations Strategies for Distributing Time-Sensitive Resources*

Ho-Yin Mak[†] Tinglong Dai[*] Christopher S. Tang[‡]

[†]McDonough School of Business, Georgetown University, Washington, DC 20057, HoYin.Mak@georgetown.edu

[*]Johns Hopkins University, Carey Business School, Baltimore, Maryland 21202, dai@jhu.edu

[‡]UCLA Anderson School of Management, Los Angeles, California 90095, chris.tang@anderson.ucla.edu

* Corresponding author: Tinglong Dai

Abstract

Distributing scarce resources such as COVID-19 vaccines is often a highly time-sensitive and mission-critical operation. Our research was prompted by a significant obstacle that the U.S. and other nations encountered during the early months of the COVID-19 vaccination campaign: most COVID-19 vaccines require two doses given three or four weeks apart. Given the severely limited supply and mounting pressure on many countries to reduce hospitalizations and mortality, how to effectively roll out two-dose vaccines was a critical policy decision. In this paper, we first model and analyze inventory dynamics of the rollout process under three rollout strategies: (1) holding back second doses, (2) releasing second doses, and (3) stretching the lead time between doses. Then we develop an SEIR model that incorporates COVID-19 asymptomatic and symptomatic infections to evaluate these strategies in terms of infections, hospitalizations, and mortality.

Among our findings, we show releasing second doses reduces infections but creates uneven vaccination patterns. In addition, to ensure second doses are given on time without holding back inventory, strictly less than half of the supply can be allocated to first-dose appointments. Stretching the between-dose lead time flattens the infection curve and reduces both hospitalizations and mortality compared with the strategy of releasing second doses. We also consider an alternative single-dose vaccine with lower efficacy and show the vaccine can be more effective than its two-dose counterparts in reducing infections and mortality. We conduct extensive sensitivity analyses related to age composition, risk-based prioritization, supply disruptions, and disease transmissibility. Our paper provides important implications for policymakers to develop effective vaccine rollout strategies in developed and developing countries alike. More broadly, our paper sheds light on how to develop effective operations strategies for distributing time-sensitive resources in times of crisis.

Keywords: COVID-19 vaccine rollout, vaccine inventory, public health operations, SEIR model, healthcare operations management

History: Received: January 2022; accepted: March 2022 for the 30th Anniversary of Production and Operations Management after one revision.

1. Introduction

This research was motivated by a major challenge that COVID-19 vaccination campaigns faced during the first half of 2021. One year after the onset of the COVID-19 pandemic, global efforts aimed at ending the pandemic had evolved from developing vaccines at a record pace to rapidly putting doses into people's arms. Experts predicted that manufacturing or deployment delays will cause the benefit of vaccination to "decline substantially" (Paltiel et al. 2021, p. 42). Adding to the urgency of distributing vaccines, SARS-CoV-2, the virus underlying the disease, continued to rapidly spread and mutate. Racing against time, the mass vaccination effort was arguably "the biggest logistics challenge in history" (Arnold 2020, p. 36), requiring vaccine manufacturers to produce billions of doses of newly developed vaccines in a matter of months.

Yet, expanding vaccine supply was challenging for two reasons. First, the vaccine production rate is limited by physical infrastructure, requisite raw materials (e.g., enzymes for converting DNA to mRNA for both the Pfizer and Moderna vaccines),[1] and packaging materials (e.g., glass vials and rubber stoppers). Second, the production yield can be uncertain. For example, uncertainty in AstraZeneca's vaccine production process, which requires the growth of cells inside bioreactors, caused major delays in its delivery to the E.U. in January 2021. In addition, developed countries have locked up much of the vaccine supply through 2023, leaving developing countries struggling with stringent supply constraints (Hopkins 2021). Specifically, the vaccine shortage was particularly salient in Africa, which accounts for 18% of the world's population but has only received 2% of global COVID-19 vaccine supply by June 2021 (Muhumuza and Mutsaka 2021).

In addition to limited supply, the authorization of three major COVID-19 vaccines (AstraZeneca, Moderna, and Pfizer) was for two-dose regimens, creating unprecedented *timing and inventory planning* issues. For example, the Pfizer vaccine required two doses with a recommended three-week interval between them.[2] To ensure adherence to the two-dose regimen, the U.S. initially (from December 2020) followed a "hold-back policy" for each dose offered to a first-time recipient; that is, one extra dose is held in reserve until the recipient returns for the second dose (Gottlieb 2021).

The hold-back policy can ensure that two doses are reserved for each recipient. However, when facing severely limited vaccine supplies and rapidly rising COVID-19 case counts, the hold-back policy can cause a delay for other individuals to receive their first doses. To accelerate the vaccination process in order to reach herd immunity sooner, the U.S. and the U.K. had proposed

[1] Throughout the paper, for brevity, we refer to the Pfizer/BioNTech vaccine as the Pfizer vaccine. Similarly, we refer to the AstraZeneca/Oxford vaccine as the AstraZeneca vaccine.

[2] The U.S. Food and Drug Administration (FDA) has only authorized the use of two specified doses of both Pfizer and Moderna vaccines at specified intervals—a three-week interval for the Pfizer vaccine and a four-week interval for the Moderna vaccine. See https://www.fda.gov/news-events/press-announcements/fda-statement-following-authorized-dosing-schedules-covid-19-vaccines for details.

unconventional solutions intended to address the logistical requirement of offering two doses of the COVID-19 vaccine for each recipient.

In the U.S., then-President-elect Joe Biden proposed on January 8, 2021, that he would order a switch from the hold-back policy to a stock "release policy," which means that (1) no vaccine inventory would be held for *future* vaccine recipients who return to receive their second doses and (2) all vaccine inventory is used as either first doses for new recipients or second doses for returning recipients (O'Donnell and Spalding 2021). Unlike the hold-back policy that can guarantee that the two-dose regimen is administered according to the recommended time interval, the release policy can increase access of the first doses to more recipients. However, without dedicated inventory for the second doses, shortages may arise when vaccine supply remains limited.

Across the Atlantic, the U.K. government announced a related and somewhat controversial policy on December 30, 2020. Instead of administering two vaccine doses three weeks apart as recommended (in the case of the Pfizer vaccine), the lead time would be stretched to 12 weeks. Logistically, this "stretching policy" defers the required demand from second-dose recipients and frees up more stock to cover first-dose recipients. Hence, the stretching policy generates an effect similar to that of the release policy: It can expand the access to the first dose for more recipients, but it will delay the administration of the second dose, especially when the vaccine supply is limited. Delaying the time for recipients to receive their required second doses can increase the chance of infection for these recipients during the delay.

Both the U.S. and U.K. policies aimed to maximize the number of people developing partial immunity. As *The Guardian* reported, prioritizing the first dose "would mean each vulnerable person who has received the vaccine would be afforded less protection, [while] the number of people given at least some protection would double" (Rawlinson 2020). Because clinical data suggest the marginal gain in protection of the first dose is significantly larger than that of the second dose, such a trade-off appeared to increase overall protection to the population.

On the other hand, both policies were controversial for myriad reasons (Mancini and Gross 2021). In particular, the issue of whether to hold back second doses garnered significant public attention in the U.S.: Some experts are in favor of holding back second doses, citing the risk of insufficient protection of single-dose inoculations in times of potential disruptions (Mak and Tang 2021); others believe the supply rate of COVID-19 vaccines is expected to grow over time, at least in the U.S., so holding back second doses is unnecessary (Dai and Yadav 2021). These diverging views motivate us to model and compare the release and stretching policies against the benchmark hold-back policy. Essentially, both the release and stretching policies are intended to address the challenge of limited vaccine supply through an operational approach: both policies call for *shifts in inventory-control policies* for vaccine stocks. Therefore, debates over these policies must involve

the operations management perspective, which has been largely absent in the public discourse, in addition to the clinical, public health, and political perspectives. To facilitate such discussions, we develop a modeling framework to examine and compare these policies.

Due to limited vaccine supply (especially in many developing countries), allocating vaccines has become a time-sensitive and mission-critical issue. In our context, these rollout policies of a vaccine with a two-dose requirement involve the following trade-offs: (i) The hold-back policy can secure two doses for fewer recipients with high efficacy, yet it delays other recipients' access to the first doses (which can give them some immunity); (ii) the release policy can increase access to the first doses, but it may cause shortages or delays for administering the required second doses for returning recipients; and (iii) the stretching policy can also vaccinate more recipients, but it can cause even longer delays for the recipients receiving their required second doses.

Critically, through modeling the rollout process under the alternative policies, we point out that the popular belief that "the number of people vaccinated would double" (Rawlinson 2020) is *incorrect*. In particular, the clinical requirement to administer the second dose imposes an *inter-temporal* constraint on the rollout problem: As more first doses are given early on, supplies in later periods become more tightly constrained.

In light of these trade-offs, the limited supply of vaccines, and how the required second dose can boost immunity against COVID-19, the implications of these policies are not immediately clear, especially when the timing for individuals to receive the first and second doses would vary for each policy. These observations motivate us to develop a dynamic model to examine the following questions arising from the rollout process of a COVID-19 vaccine with a two-dose requirement:

(i) Given the production rate of vaccines (constant or linearly increasing), what are the cumulative inoculation rates for the first and second doses under the hold-back policy, the release policy, and the stretching policy?

(ii) Relative to the hold-back policy, will the release and stretching policies generate higher cumulative inoculation rates?

(iii) In light of an age-stratified SEIR (susceptible, exposed, infectious, recovered) model that predicts the infection process under different vaccination policies, which policy is the most effective in reducing infections, hospitalizations, and mortality?

(iv) How would a single-dose vaccine with a lower efficacy (e.g., the Johnson & Johnson COVID-19 vaccine) perform relative to its two-dose counterparts?

As an initial attempt to examine the above questions arising from the two-dose regimen of all approved vaccines and the limited supply of vaccines, we develop a dynamic model by examining the first- and second-dose inoculation rates of a single vaccine under all three policies (hold-back,

release, and stretching). We also extend our model to capture the case of rolling out a new single-dose vaccine with a lower efficacy.

Through analytical comparisons, we find that under a constant or linearly increasing vaccine production rate, both release and stretching policies generate a higher cumulative vaccination population than the hold-back policy. Somewhat counter-intuitively, we show that even if one intends to release all second doses, no more than half of the available doses can be allocated to first-dose appointments, in order to avoid any delays in administering those required second doses for returning recipients, no matter how quickly supply grows over time.

One can view the single-dose regimen as an extreme case of a stretching policy in which the timing of the required second dose is extended to the infinite future. Using this logic, we show that the single-dose regimen can generate a higher cumulative vaccinated population than the hold-back policy and the release policy.

We also develop a compartmental SEIR model that incorporates COVID-19 asymptomatic and symptomatic infections in a population that consists of high- and low-risk groups. Our SIER model captures the evolution of the COVID-19 pandemic among these two groups who receive zero, one or two doses of vaccine under different rollout strategies over time. We also calibrate our parameter values using the latest vaccine effectiveness studies (e.g., Haas et al. 2021). Our simulation study of our SEIR model reveals that both release and stretching policies can generate lower disease transmissions than the hold-back policy; however, these two policies create uneven vaccination patterns. Stretching the between-dose lead time flattens the infection curve, and reduces hospitalization and mortality (more than under the strategy of releasing second doses). We also consider an alternative single-dose vaccine (e.g., J&J vaccine) with a lower overall efficacy and show that it can be more effective than its two-dose counterparts in reducing infections, hospitalizations, and deaths. We conduct extensive sensitivity analyses related to age composition, risk-based prioritization, supply disruptions, and the basic reproduction number R_0.

Although our work was motivated by the initial months of COVID-19 vaccination efforts in the E.U., U.K., and U.S., it has broader policy implications. Developing countries are set to face limited vaccine supply as developed countries have locked up most of the vaccine supply up to 2023 (Hopkins 2021). This is the latest episode of the longstanding ineffectiveness of global health supply chains that calls for better operations management practices (Kraiselburd and Yadav 2013, Gallien et al. 2017). Our results can help policymakers in various developing countries understand the pros and cons of different rollout policies. Furthermore, our modeling framework and our SEIR model can be applied to a variety of vaccination environments (when relevant data become available), including different variants and different vaccines with varying administration regimens and levels of efficacy.

The remainder of the paper is organized as follows. We review the relevant literature in Section 2. Section 3 describes our model and the three rollout policies. Using the hold-back policy as a benchmark, we examine the implications of the release policy in Section 4 and the stretching policy in Section 5. In Section 6, we examine an alternative single-dose vaccine. In Section 7, we develop an SEIR model to evaluate the three policies and the single-dose regimen. We conclude our paper in Section 8. All proofs are provided in Section OA1 of the online appendix.

2. Literature Review

Vaccine distribution is a vibrant line of research that has attracted substantial interest from the healthcare operations management community. Most of this stream of literature focuses on the influenza vaccine (for exceptions, see, e.g., Cummings et al. 2020; Hall et al. 2008). Whereas several papers in this literature (e.g., Arifoğlu et al. 2012; Arifoğlu and Tang 2021; Chick et al. 2008, 2016; Yamin and Gavious 2013) study governments' vaccine contracting, procurement, and subsidies decisions, others (e.g., Cho and Tang 2013; Dai et al. 2016) examine optimal contract design between vaccine manufacturers and healthcare providers who administer vaccines. Duijzer et al. (2018) use a susceptible-infectious-recovered (SIR) model to study the critical vaccine coverage leading to herd effect for a pandemic and consider the optimal allocation of vaccine doses across different populations. Unlike this literature, which focuses primarily on distributing *single*-dose vaccines, we study a unique problem of distributing a two-dose vaccine during a pandemic in which whether to hold back second doses becomes a consequential national policy. In doing so, our paper speaks to a number of related public health papers on various rollout strategies, including, for example, Paltiel et al. (2021a, 2021b), Tuite et al. (2021), and Voysey et al. (2021).

Another related stream of literature studies inventory management—a classical theme of the operations research and management science community. This literature emerges from the seminal work by Arrow et al. (1951), Dvoretzky et al. (1953), and Scarf (1960) and has become the foundation of much of today's work in manufacturing and supply chain operations (Porteus 1990; Song et al. 2020). Despite the immense breadth and depth of the extant inventory literature, our paper represents a first attempt to address a novel time-sensitive resource allocation problem in which each customer requires exactly *two units* of the product at a *specified timing*. In addition, unlike the literature that focuses on a single organization's decision, we examine a national policy with public health implications. The only other work in the inventory literature that shares similar characteristics, to the best of our knowledge, is the paper by Natarajan and Swaminathan (2017), which considers the dynamic allocation of inventory of treatment drugs among patients with different (and evolving) health states. Compared with this work, which considers patients' health conditions in each state to evolve exogenously over time, our vaccine rollout problem involves the evolution of a pandemic that is governed by an epidemiology model.

Our paper also contributes to the rapidly growing operations management literature that addresses the COVID-19 pandemic. For example, Kaplan (2020) documents a number of basic mathematical models to support local-level COVID-19 decisions ranging from event crowd control to hospital capacity planning. As another example, Chen et al. (2020) use mobility data around nursing homes to establish the relationship between cross-facility staff movement and COVID-19 infections. Shumsky et al. (2021) study a healthcare provider's COVID-19 vaccine inventory management problem. We enrich this literature by studying a novel problem relevant to how health policymakers roll out COVID-19 vaccines—the endgame of the COVID-19 pandemic—and our findings have important implications for vaccination policies in future pandemics.

3. Modeling a Two-Dose Vaccine Rollout Process

In this section, we present a parsimonious model of the vaccine rollout process in the spirit of scratch models as articulated by Kaplan (2020). To examine the implications of the three aforementioned policies—including (1) holding back second doses, (2) releasing second doses, and (3) stretching the between-dose lead time—our model focuses on the dynamic first- and second-dose inoculation rates over time by taking into consideration the limited availability of vaccine supply. Specifically, our model involves the following elements.

Vaccine supply rate $y(t)$. Consider a time horizon with a length of T (e.g., a year) during which the vaccine is rolled out. For any $t \in [0,T]$, let $y(t)$ be the deterministic supply rate of vaccine doses available at t;[3] $y(t)$ is exogenously determined by the rate of manufacturing, quality and safety checks, and transportation. We assume $y(t)$ is (weakly) increasing in t; that is, the supply ramps up over time. Let $Y(t)$ be the cumulative supply up to time t, where $Y(t) = \int_0^t y(\tau)d\tau$.

Vaccination rate $v_i(t)$ and cumulative vaccination $V_i(t)$. We denote by $v_1(t)$ and $v_2(t)$ the instantaneous rates of the first- and second-dose inoculations at time t, respectively. Thus, the cumulative vaccinations completed by time t, denoted by $V_i(t)$, satisfies

$$V_i(t) = \int_0^t v_i(\tau)d\tau \text{ for } i \in \{1,2\}.$$

The cumulative vaccinations $V_1(t)$ (and $V_2(t)$) can be interpreted as the partially (and fully) protected populations who have received the first dose (and the second dose) by time $t \in [0,T]$.

Lead time L. A parameter of key concern is the lead time L; i.e., the recommended time interval between the two doses of the same vaccine. The lead time L depends on the specific vaccine and the vaccination policy. In the U.S., Pfizer recommends $L = 3$ weeks and Moderna recommends

[3] Besides tractability, the deterministic production rate is a reasonable assumption for the following reason: The production process for the Pfizer and the Moderna vaccines is relatively stable, because the process involves converting DNA into mRNA rather than growing cells in a bioreactor (Boseley 2021).

$L = 4$. For simplicity of our analysis, we assume the second dose must be received exactly L periods after the first dose.[4] However, in the U.K., $L = 12$ weeks for the AstraZeneca, Moderna, and Pfizer vaccines under the dose-stretching policy.

Vaccination rollout policy $(v_1(.), v_2(.), L)$**.** As we will explain later, the hold-back, release, and stretching policies correspond to a particular choice of functions $v_1(\cdot)$, $v_2(\cdot)$ and the lead time L. In general, a feasible policy $(v_1(.), v_2(.), L)$ must satisfy the following constraints at any time t:[5]

$$V_1(t) + V_2(t) \le Y(t) \qquad \text{for } t \in [0, T] \tag{1}$$

$$V_2(t) = 0 \qquad \text{for } t \in [0, L) \tag{2}$$

$$V_2(t) \le V_1(t - L) \qquad \text{for } t \in [L, T]. \tag{3}$$

Due to limited vaccine supply, constraint (1) suggests cumulative first- and second-dose vaccinations offered are bounded by the cumulative supply at every instant t. In addition, to comply with the two-dose regimen requirement, constraints (2) and (3) indicate second doses cannot be offered until L time units have elapsed after first-dose inoculations. (If there is sufficient supply of vaccine at all times, then (3) will hold as equality; i.e., $V_2(t) = V_1(t - L)$.)

Cumulative protection $P_i(t)$**.** Building on the (partially and fully) protected population sizes ($V_1(\cdot)$ and $V_2(\cdot)$), we now define the cumulative partial protection rate (due to the first dose) $P_1(t)$ and full protection rate (due to both doses) $P_2(t)$ up to time t (in person-weeks):

$$P_i(t) = \int_0^t V_i(\tau) d\tau \text{ for } i = 1, 2.$$

The cumulative protection levels keep track of the number of people protected by vaccination *and* the duration over which protection has been developed. Hence, this measure reflects the amount of protection offered to the population as a whole. To elaborate, consider the following example.

Example 1. Consider two hypothetical scenarios: (i) One million people are given the first dose instantaneously at time $t = 0$ and none thereafter; and (ii) 50,000 people are given the first dose per day at a continuous rate over 20 days and none thereafter. In both cases, the cumulative first-dose

[4] The recommended lead time $L = 21$ days for the Pfizer vaccine, and CDC guidelines in 2021 allow for a grace period of 4 days earlier than the recommended date, see https://www.cdc.gov/vaccines/covid-19/info-by-product/clinical-considerations.html for details. In February 2022, the CDC modified its recommended lead time to 8 weeks for 12- to 64-year-olds, particularly males aged 12 to 39 (https://khn.org/morning-breakout/cdc-advising-some-to-wait-8-weeks-between-doses-of-mrna-vaccines/). Despite these new guidelines, our modeling framework and approach are applicable to any lead time L and any number of doses.

[5] To isolate the effect of different vaccine rollout policies, we do not model the issue of setting up a physical infrastructure (i.e., location and the size of vaccination sites) or mobilizing personnel (controlled by FEMA, state/county governments, or local healthcare providers) to administer vaccines. Instead, we abstract away those issues and assume all policies will be operated under the same environment.

vaccinated population sizes by Day 20 are the same at $V_1(t) = 1$ million when $t = 20$. However, the cumulative protection levels are different: $P_1(t)$ is equal to 20 million person-days in scenario (i), and is equal to 10 million person-days in scenario (ii). Thus, the amount of protection in scenario (i) is double that under scenario (ii).

The functions $y(t), Y(t), v_i(t)$, and $V_i(t)$ are meaningfully defined only for $t \in [0, T]$; for notational brevity, we adopt the convention that their values equal zero when $t < 0$. In addition to these measures, we further consider a broader class of performance metrics, including symptomatic infections, hospitalizations, and mortality, based on an SEIR model in Section 7 and Section OA2. This involves a system of differential equations that utilizes $v_i(t)$ and $V_i(t)$ for $i = 1, 2$.

3.1. Stock Hold-Back Policy $(v_1^h(t), v_2^h(t))$

As explained in Section 1, the U.S. government followed the manufacturer's recommended lead time of L weeks by adopting the hold-back policy as follows. Specifically, each facility only administers half of the vaccine stock received for first doses, and holds back the other half for L weeks to ensure recipients will have guaranteed stock when they return for the second dose. By considering the supply rate $y(t)$, the hold-back policy $(v_1^h(t), v_2^h(t))$ stipulates the following[6]:

$$v_1^h(t) = \frac{1}{2} y(t) \quad \text{and} \quad v_2^h(t) = v_1^h(t - L).$$

This policy ensures constraints (1)–(3) are satisfied at all time.

We now represent the cumulative protection functions $P_i^h(t)$ using the external supply function $Y(t)$. First, by the definition of $V_i^h(t)$, we have

$$V_1^h(t) = \frac{1}{2} Y(t) \quad \text{and} \quad V_2^h(t) = \frac{1}{2} Y(t - L).$$

Then, the amount of protection by time t can be given by

$$P_1^h(t) = \frac{1}{2} \int_0^t Y(\tau) d\tau \quad \text{and} \quad P_2^h(t) = \frac{1}{2} \int_0^{t-L} Y(\tau) d\tau.$$

3.2. Stock-Release Policy $(v_1^r(t), v_2^r(t))$

Under the stock-release policy, vaccine stock is no longer put on reserve for second doses.[7] The intention of the release policy is to maximize the rate of first-dose inoculations to accelerate the build-up of partial protection. Using our notation, the release policy stipulates that $v_1(t) + v_2(t) = y(t)$, such that all available stock is offered (as first or second doses) without delay. However, if the supply rate does not increase fast enough over time, then increasing the number of first-dose recipients earlier would constrain the number of first-dose recipients later.

[6] We use the superscripts h, r, and s to denote the hold-back, release, and stretching policies, respectively.

[7] Though small amounts of smoothing stock will still be kept in reality, we ignore this fact in our model for simplicity.

To illustrate different rollout dynamics under the hold-back and the release policies with limited vaccine supply, let us consider a simple example.

EXAMPLE 2. Consider a constant supply rate of 100 doses per day (i.e., $y(t) = 100$) and a lead time of $L = 21$ days (i.e., 3 weeks between dose time interval for the Pfizer vaccine).

(i) *Hold-back policy.* Under the hold-back policy, 50 first doses will be offered per day, whereas the remaining 50 doses of supply will be reserved for the same 50 recipients who return 21 days later for their second doses. This pattern repeats every day.

(ii) *Release policy*. Under the release policy, from Day 1 to Day 21, 100 doses per day (i.e., the entire supply for each day) are administered to 100 people as their first doses. However, on Day 22, those 100 recipients who received the first doses on Day 1 are now due for their second doses. To adhere to the required dosage schedule, it is necessary to allocate all 100 doses (i.e., the entire supply) on Day 22 for those 100 returning second-dose recipients. In doing so, the provider has no vaccine to offer to any first-dose recipient on Day 22. This situation will continue until Day 43, when there are no more recipients who are due for the second dose (because no first-dose recipient received a first dose from Day 22 to Day 42). Therefore, on Day 43, all 100 doses will be released to 100 first-dose recipients, and the pattern repeats as on Day 1. Overall, this "alternating pattern" repeats in 42-day (i.e., twice of the lead time $L = 21$ days) intervals.[8]

We now define the stock-release policy $(v_1^r(t), v_2^r(t))$ more formally. Initially, all available supply is released as first-dose vaccination over $t \in [0, L)$ so that $(v_1^r(t), v_2^r(t))$ satisfy

$$v_1^r(t) = y(t) \text{ and } v_2^r(t) = 0.$$

When $t \geq L$, the available supply will be split between first- and second-dose appointments. Let $B^r(t) = V_1^r(t-L) - V_2^r(t)$ denote the amount of backlog by time t; that is, the difference between the number of people who have received their first doses L weeks ago and the number of people who have received their second doses. For clinical reasons, the backlog $B^r(t)$ should be minimized. Therefore, the rate of second doses offered should match as closely as possible the rate of first doses completed L periods ago. To capture this preference, the release policy $(v_1^r(t), v_2^r(t))$ for $t \in [L, T]$ must satisfy:

$$\begin{aligned} v_1^r(t) &= y(t) - v_2^r(t) \\ v_2^r(t) &= \begin{cases} \min\{v_1^r(t-L), y(t)\} & \text{if } B^r(t) = 0 \\ y(t) & \text{otherwise.} \end{cases} \end{aligned}$$

[8] This alternating pattern has been observed in various parts of the world. For example, in early February 2021, COVID-19 vaccine sites in Los Angeles County postponed first-dose appointments and only administered second doses to returning recipients; see http://cbsloc.al/3bo5uiS for details.

That is, when the backlog of second-dose appointments exists (i.e., $B^r(t) > 0$, or, equivalently, $V_2^r(t) < V_1^r(t-L)$), the entire supply will be allocated for the second dose. In the absence of a backlog, the second-dose inoculation rate follows the first-dose rate lagged by lead time L, up to the available capacity.

The backlog of second-dose appointments $B^r(t)$ will build up if the rate of recipients due for second doses, $v_1(t-L)$, exceeds the available supply, $y(t)$. When the supply rate ramps up slowly, most (or all) of the available supply might be allocated for second doses in order to avoid backlog. This scenario is not desirable for the purpose of steadily increasing the protected population.

The next result evaluates the fraction of supply that can be allocated for first-dose appointments. In preparation, let us define β as the minimum fraction of available supply that is allocated for first-dose appointments at all times, i.e., $\beta = \min_{t \geq L} v_1(t)/y(t)$.

LEMMA 1. *No backlogs exist if and only if β satisfies*

$$(1-\beta)\cdot y(L+t) \geq \begin{cases} y(t) & \text{for all } t \in [0, L) \\ y(t) - y(t-L) & \text{for all } t \in [L, 2L) \\ y(t) - \beta \cdot y(t-L) & \text{for all } t \in [2L, T]. \end{cases}$$

Lemma 1 suggests that, to ensure a minimum fraction (β) of supply to be allocated for first doses (in order to continuously build up the protected population) at all times, the supply rate $y(t)$ must grow sufficiently quickly. The first two conditions of Lemma 1 hold when $(1-\beta)$, the proportion of supplies made available for second doses, is high and/or supply accelerates quickly in the initial weeks (especially during the first L periods). To illustrate the third condition, consider the case in which $\beta = \frac{1}{2}$; that is, when at least half of the available supply is to be used on first doses. In that case, the third condition holds if $y(t)$ is convex.

If vaccine supply does not accelerate sufficiently quickly, the rate of first-dose inoculations cannot be sustained. To avoid backlogs for second-dose inoculations, first-dose offerings must be temporarily reduced. As we show in the next section, this reduction gives rise to an alternating pattern between offering first and second doses (see Example 2).

3.3. Dose-Stretching Policy $(v_1^s(t), v_2^s(t))$

The dose-stretching policy is intended to optimize the use of vaccine supply and maximize first-dose inoculations. We define the dose-stretching policy $(v_1^s(t), v_2^s(t))$ by focusing on the effect of stretching the lead time from the recommended L to $L^s = mL$, where $m \geq 1$ is an integer.[9] Using this definition, we may view the stretching policy as a generalization of the stock-release policy with a longer lead time $L^s = mL$, where $m > 1$. Thus, we omit the analysis of the policy for brevity.

[9] In the case in which $m = 1$, the stretching policy reduces to the release policy. For practicality of implementation, we restrict m to being integer-valued. For example, the U.K.'s dose-stretching policy increases the lead time from $L = 3$ weeks to $L^s = 12$ weeks, that is, $m = 4$.

4. Is Releasing Second Doses the Right Move?

We now compare the performance between the hold-back policy $(v_1^h(t), v_2^h(t))$ and release policy $(v_1^r(t), v_2^r(t))$. To do so, we first derive the cumulative vaccinated population $V_i(t)$ and the cumulative protection $P_i(t)$ associated with these policies, by considering the case in which the supply rate grows linearly over time.[10] In particular, we assume, without loss of generality, that $y(t) = 1 + \alpha t$, where $\alpha \geq 0$ and the initial supply rate is normalized to one (i.e., $y(0) = 1$).

4.1. Benchmark: Hold-Back Policy

By using the supply rate $y(t) = 1 + \alpha t$ and the hold-back policy $(v_1^h(t), v_2^h(t))$ as defined in Section 3.1, we derive the corresponding performance measures $V_i^h(t)$ and $P_i^h(t)$ for $i = 1, 2$ below.

LEMMA 2. *Under the hold-back policy $(v_1^h(t), v_2^h(t))$ and a linear supply growth, the cumulative vaccinations and protection evolve over time as follows:*

$$V_1^h(t) = \frac{t}{2} + \frac{\alpha}{4}t^2 \text{ for } t \in [0, T]$$

$$V_2^h(t) = \begin{cases} 0 & \text{for } t \in [0, L) \\ \frac{1}{2}(t-L) + \frac{\alpha}{4}(t-L)^2 & \text{for } t \in [L, T] \end{cases}$$

$$P_1^h(t) = \frac{1}{4}t^2 + \frac{\alpha}{12}t^3 \text{ for } t \in [0, T]$$

$$P_2^h(t) = \begin{cases} 0 & \text{for } t \in [0, L) \\ \frac{1}{4}(t-L)^2 + \frac{\alpha}{12}(t-L)^3 & \text{for } t \in [L, T]. \end{cases}$$

The measures characterized in Lemma 2 later serve as the benchmark for ensuing comparisons.

4.2. Implications of the Release Policy

Using the same approach as in Lemma 2, we can derive the expressions of the performance measures $V_i^r(t)$ and $P_i^r(t)$ associated with the stock-release policy $(v_1^r(t), v_2^r(t))$ as defined in Section 3.2. We omit the details for conciseness.

By considering the cumulative vaccinated population $V_i^r(t)$ associated with the release policy, we can determine the backlog $B^r(t)$ as defined in Section 3.2, where $B^r(t) = V_1^r(t-L) - V_2^r(t)$. The next result follows from Lemma 1 and highlights the conditions under which the backlog $B^r(t) = 0$ so that a smooth rollout of first-dose appointments (without causing delays in administering the second dose) can be guaranteed under the linear-supply-growth scenario $y(t) = 1 + \alpha t$.

PROPOSITION 1. *Under a linearly growing supply, the stock-release policy $(v_1^r(t), v_2^r(t))$ can allocate β fraction of supply to first-dose appointments without incurring backlogs for second-dose appointments if and only if $\beta \leq \frac{1}{2} - \frac{1}{2(1+2\alpha L)}$.*

[10] For tractability, we consider the linear case. However, one can use the same approach to examine the case of a non-linear production rate numerically.

To avoid backlogs at any time t (i.e., $B^r(t) = V_1^r(t-L) - V_2^r(t) = 0$ for any t) so that enough second doses are available in stock for those who have received the first dose L weeks ago, Proposition 1 suggests the stock-release policy can *never* guarantee half of available supplies are allocated to first-dose appointments (i.e., $\beta < 0.5$), no matter how quickly the supply rate grows (i.e., for any value of α). The minimum guaranteed fraction available for first-dose appointments is, intuitively, increasing in α. (Furthermore, as part of the proof of Proposition 1 in the online appendix, (OA2)-(OA3) suggest that when α approaches zero, that is, when supply grows slowly, first- and second-dose appointments will have to be almost shut down in alternating periods of lead time (L), in order to avoid backlogs.)

Next, through direct comparison between the performance measure associated with the hold-back (i.e., $V_i^h(t)$ and $P_i^h(t)$ as stated in Lemma 2) and the release policies, we have the following proposition.

PROPOSITION 2. *Under linear supply growth, the stock-release policy $(v_1^r(t), v_2^r(t))$ can generate higher cumulative vaccinated populations than the stock hold-back policy $(v_1^h(t), v_2^h(t))$ by:*

$$\Delta V_1^r(t) \triangleq V_1^r(t) - V_1^h(t) = \begin{cases} \frac{\alpha}{2}nL^2 + \frac{1}{2}(t-2nL) + \frac{\alpha}{4}(t-2nL)^2 & \text{for } t \in [2nL, (2n+1)L) \\ \frac{\alpha}{2}nL^2 + \frac{1}{2}L + \frac{\alpha}{2}L^2 - \frac{1}{2}[t-(2n+1)L] & \\ -\frac{\alpha}{4}[2(n+1)L - t]^2 & \text{for } t \in [(2n+1)L, (2n+2)L) \end{cases}$$

$$\Delta V_2^r(t) \triangleq V_2^r(t) - V_2^h(t) = \Delta V_1^r(t-L).$$

Furthermore, over a full cycle $[0, 2nL)$, where $n \geq 1$, the stock-release policy achieves better cumulative protection than the stock hold-back policy by

$$\begin{aligned} \Delta P_1^r(2nL) &\triangleq P_1^r(2nL) - P_1^h(2nL) = \frac{n}{2}L^2 + \frac{\alpha}{2}n(n+1)L^3 \\ \Delta P_2^r(2nL) &\triangleq P_2^r(2nL) - P_2^h(2nL) = \Delta P_1^r((2n-1)L) \\ &= \left(\frac{n-1}{2} + \frac{1}{4}\right)L^2 + \frac{\alpha}{2}L^3\left(n^2 - \frac{5}{6}\right). \end{aligned}$$

From Proposition 2, we can verify that $\Delta V_i^r(t) > 0$ and $\Delta P_i^r(2nL) > 0$ for $i = 1, 2$. Hence, the net changes in cumulative inoculations as well as cumulative protection due to switching from the stock hold-back policy $(v_1^h(t), v_2^h(t))$ to the stock-release policy $(v_1^r(t), v_2^r(t))$ are all strictly positive. Therefore, releasing second doses accelerates the buildup of protection, as long as the supply rate is (weakly) growing.

This improvement from releasing second doses, however, comes at the expense of potentially uneven offering patterns that alternate between first and second doses (Proposition 1) and the associated scheduling complexity, especially when the supply rate grows sluggishly. In practice, without assurance that second doses are reserved, many healthcare providers struggle with the

uncertainty of scheduling second-dose appointments. In the state of Maryland, as recently as in February 2021, despite the Biden administration's decision not to hold back second doses at the federal level, Dennis Schrader, Acting Maryland Health Secretary, urged healthcare providers to hold back second doses (Miller and Cohn 2021). In other words, the impact of the stock-release policy on infection control is dampened by its complexity in implementation.

Proposition 2 reveals that the release policy dominates the hold-back policy. Will the stretching policy create additional value over the release policy by extending the lead time? We analytically examine this issue in the section that follows.

5. Will Stretching the Dose Time Interval Help?

We now examine the implication of the dose-stretching policy $(v_1^s(t), v_2^s(t))$. Recall from Section 3.3 that the stretching policy is essentially a stock-release policy with a longer lead time that has $L^s = mL$, where $m > 1$. Hence, we can extend our analysis presented in Section 4.2 to determine the corresponding performance measures $V_i^s(t)$ and $P_i^s(t)$ when $L^s = mL$ with $m > 1$. We omit the details to avoid repetition. In addition, through direct comparison between the performance measure associated with the release (i.e., $V_i^r(t)$ and $P_i^r(t)$ as examined in Section 4.2) and the stretching policies, we obtain the following result:

PROPOSITION 3. *Over a cycle* $[0, 2mL)$*, the dose-stretching policy* $L^s = mL$ *with* $m > 1$ *can increase the cumulative first-dose inoculations and reduce the second-dose inoculations over the stock-release policy by*

$$\Delta V_1^s(2mL) \triangleq V_1^s(2mL) - V_1^r(2mL) = \frac{\alpha}{2} L^2 m(m-1) > 0$$
$$\Delta V_2^s(2mL) \triangleq V_2^s(2mL) - V_2^r(2mL) = -\frac{3\alpha}{4} L^2 (m-1)^2 < 0.$$

In addition, the dose-stretching policy offers higher cumulative partial and lower full protection than the stock-release policy by

$$\Delta P_1^s(2mL) \triangleq P_1^s(2mL) - P_1^r(2mL) = \frac{m(m-1)L^2}{2} + \alpha[2m^3 - m(m+1)]L^3 > 0$$
$$\Delta P_2^s(2mL) \triangleq P_2^s(2mL) - P_2^r(2mL) = -\frac{3(m-1)^2 L^2}{4} - \frac{\alpha(m-1)(m^2+1)L^3}{2} < 0.$$

Proposition 3 highlights the trade-off between prioritizing the first and second doses. Through stretching the lead time (i.e., by increasing m), more supplies are made available for the first dose; thus, the cumulative population having completed the first dose and the cumulative partial protection both increase. However, deferring the second dose leads to lower levels of cumulative second-dose completions and the level of full protection.

To better explore the trade-off between first and second doses, let us consider the case when the first and second doses lead to efficacy levels of $1-\theta_1$ and $1-\theta_2$, respectively, that is, the rate that an otherwise unimmunized person being infected by COVID-19 drops from λ without vaccination to $\theta_1\lambda$ after the first dose, and $\theta_2\gamma$ after the second.[11] For notational brevity, let $\zeta=(\theta_2-\theta_1)/(1-\theta_1)$. If the first dose provides most of the protection, we have $\zeta<1$ (e.g., for the case of Pfizer, where the efficacy rates are $1-\theta_1=0.423$ and $1-\theta_2=0.047$ after the first and second doses, we have $\zeta=0.376/0.577=0.652$). Then, the efficacy-weighted protection level $P(t)$ measures the amount of protection provided by the vaccination program weighted by the strength of protection:

$$P(t)=(1-\theta_1)\lambda P_1(t)+(\theta_1-\theta_2)\lambda P_2(t)=\lambda(1-\theta_1)(P_1(t)+\zeta P_2(t)). \tag{4}$$

PROPOSITION 4. *Switching from the stock-release policy* $(v_1^r(t), v_2^r(t))$ *to the dose-stretching policy* $(L^s=mL)$ *can increase the efficacy-weighted protection* $P(t)$ *over the cycle* $[0,2mL]$ *if* $\zeta\le\frac{2m}{3(m-1)}$.

Proposition 4 suggests dose stretching (i.e., by increasing m) increases efficacy-weighted protection $P(t)$ as long as ζ is below the threshold value $\frac{2m}{3(m-1)}$. The threshold value drops as m increases, indicating dose stretching becomes less favorable as m increases. Recall from Section 3.3 that $m=4$ under the stretching policy for the Pfizer vaccine when the U.K. government stretched the lead time from three weeks to 12. The threshold value in Proposition 4 is $\frac{2m}{3(m-1)}=0.89$ when $m=4$. Combine this observation with the fact that $\zeta=0.652$ for the Pfizer vaccine, the condition $\zeta\le\frac{2m}{3(m-1)}$ is satisfied so that the stretching policy adopted by the U.K. government can generate a higher efficacy-weighted protection $P(t)$ over the release policy.

6. Implications of an Alternative Single-Dose Vaccine

Thus far, our model and analysis have been motivated by three approved vaccines developed by Pfizer, Moderna, and AstraZeneca that require two doses. Another COVID-19 vaccine, developed by Johnson & Johnson, was approved for emergency use on February 27, 2021. Unlike Pfizer, Moderna, and AstraZeneca, Johnson & Johnson ran its Phase III trials based on a single-dose regimen, which achieved an efficacy rate of 66%.[12] Without the need to deal with the second-dose timing issue, how would the single-dose vaccine perform relative to those two-dose vaccines? Clearly, if the single-dose vaccine has a higher efficacy, it will dominate the two-dose vaccine. However, the

[11] For example, the Pfizer vaccine offers 57.7% and 95.3% efficacy levels after the first and second doses; thus, $\theta_1=0.423$ and $\theta_2=0.048$ (Haas et al. 2021).

[12] In late January 2021, Johnson & Johnson announced the Phase III trial results of its single-dose COVID-19 vaccine with a relatively lower efficacy (66% worldwide and 72% in the U.S. trial). However, this vaccine has major operational advantages. In particular, it has a less stringent temperature requirement than the Pfizer vaccine and only requires a single dose. Thus, it circumvents the trade-off between offering first and second doses.

relative performance becomes unclear when the single-dose vaccine (e.g., the Johnson & Johnson vaccine) has a lower efficacy than the two-dose vaccine.

This observation motivates us to compare the rollout dynamics $V_i(t)$ and $P_i(t)$ of the single-dose regimen against the hold-back and stock-release policies in this section. In preparation, observe that when offered as a single-dose vaccine, all stock can be immediately released (because the second dose is no longer needed). Hence, this single-dose regimen can be viewed as an extreme case of the stretching policy with lead time $L^s = mL$, where $m = \infty$. Hence, we can apply the analysis of the stretching policy (see Section 5) to examine the performance of a single-dose vaccine.

Under linear supply growth, the vaccination rate associated with a single-dose regimen satisfies $v^d(t) = y(t) = 1 + \alpha t$. Using the same approach as presented in Sections 4 and 5 to compare the vaccinated population and protection levels for the single-dose regimen with the hold-back and stock-release policies under the two-dose regimen, we obtain the following proposition.

PROPOSITION 5. *Compared with the stock hold-back policy under a two-dose regimen, the single-dose regimen increases the cumulative first-dose inoculations and cumulative first-dose protection over a full cycle* $[0, 2nL)$ *by:*

$$V_1^d(2nL) - V_1^h(2nL) = nL + \alpha n^2 L^2 > 0$$
$$P_1^d(2nL) - P_1^h(2nL) = n^2 L^2 + \frac{2\alpha}{3} n^3 L^3 > 0.$$

Compared with the stock-release policy, the single-dose regimen increases the cumulative first-dose inoculations and cumulative first-dose protection over a full cycle $[0, 2nL)$ *by:*

$$V_1^d(2nL) - V_1^r(2nL) = nL + \frac{\alpha}{2}(2n^2 - n) > 0$$
$$P_1^d(2nL) - P_1^r(2nL) = \frac{n}{2} L^2 + \frac{\alpha}{2}\left[\frac{4}{3} n^3 - n(n+1)\right] L^3 > 0.$$

From Proposition 5, we can see that the single-dose regimen speeds up the buildup of protection. (Because only one dose is administered under the single-dose regimen, our results focus only on pertinent measures V_1^d and P_1^d.)

However, as articulated above, when the single-dose vaccine has a lower efficacy rate than its two-dose counterparts, the performance of the single-dose vaccine over the two-dose vaccine is unclear. Specifically, even though the two-dose regimen builds up protection more slowly, it can reach a higher level of protection due to its higher efficacy. To evaluate this trade-off, we use the "efficacy-weighted" protection level $P(t)$ for the two-dose vaccine as defined in (4) in §5 to define the corresponding $P^d(t)$ for the single-dose vaccine.

Recall from Section 5 that for the two-dose regimen, one and two doses reduce the risk of infection by factors of $1 - \theta_1$ and $1 - \theta_2$, respectively, and $\zeta = \frac{1-\theta_1}{\theta_1 - \theta_2}$ reflects the relative contribution of the

first dose. Let $1-\theta^d$ be the efficacy rate of the single-dose vaccine. To make a fair comparison with a two-dose vaccine, we define $\zeta^d = \frac{1-\theta^d}{1-\theta_1}$. (For example, the Pfizer vaccine has $\theta_1 = 0.423$ and the Johnson & Johnson vaccine has $\theta^d = 0.34$ so that $\zeta^d = 0.66/0.577 = 1.14$.) In addition, by using the results presented in Proposition 5, we can define the corresponding $P^d(t) = P_1^d(t)$ for the single-dose vaccine. The next proposition follows from direct comparisons.

PROPOSITION 6. *Switching from the two-dose regimen under the stock-release policy to the single-dose regimen can increase the efficacy-weighted protection over the cycle* $[0, 2nL)$ *by*

$$\Delta P^d(2nL) \triangleq P^d(2nL) - P^r(2nL) = \lambda(1-\theta_1)L^2\left[2n^2(\zeta^d-1) + \frac{n-1}{2} + \left(\frac{1}{2} - \frac{3\zeta}{4}\right)(2n^2-2n+1)\right]$$
$$+ \lambda(1-\theta_1)\frac{\alpha L^3}{12}\left[(16\zeta^d - 16)n^3 - 12n + 6 + (1-\zeta)(8n^3 - 6n^2 + 6n - 6)\right].$$

A sufficient condition for this improvement to be non-negative is that $\zeta^d \geq \frac{11}{8}$ *and* $\zeta \leq \frac{2}{3}$.

Recall from above that for the two-dose regimen of the Pfizer vaccine, we have $\zeta = 0.652$, and for the single-dose regimen of the Johnson & Johnson vaccine, $\zeta^d = 1.14$. Even though these values do not satisfy the sufficient condition in Proposition 6, we can compute:

$$\Delta P^d(2nL) = \lambda(1-\theta_1)L^2(0.5n^2 + 0.58n - 0.54) + \lambda(1-\theta_1)\frac{\alpha L^3}{12}(6.56n^3 - 1.68n^2 - 10.32 + 4.32).$$

We can check that the first term is positive for all $n \geq 1$ and that the second term is negative when $n = 1$ and positive for $n \geq 2$. When $n = 1$, the overall expression is positive when $\alpha L \leq 5.36$. Because the supply rate is unlikely to grow more than 5.36 times over lead time L, we are more likely to observe that the Johnson & Johnson single-dose regimen offers better cumulative efficacy-weighted protection than the Pfizer vaccine under the stock-release policy for any n. Overall, the above analysis suggests that the single-dose regimen can underperform in the short run, but will eventually outperform the higher-efficacy two-dose regimen.

So far, we have obtained three key analytical results. First, we show that the release policy can improve the cumulative vaccination population and the cumulative protection in Proposition 2. Second, we prove that the stretching policy can generate a higher efficacy-weighted protection $P(t)$ than the release policy in Proposition 4 when $\zeta \leq \frac{2m}{3(m-1)}$. Third, we show that a single-dose vaccine with a lower efficacy can outperform a two-dose vaccine in terms of a higher efficacy-weighted protection when $\zeta^d \geq \frac{11}{8}$ and $\zeta \leq \frac{2}{3}$ in Proposition 6. Besides the performance measures $V(t)$ and $P(t)$, one should evaluate and compare the symptomatic infection rate, hospitalization rate, and the mortality rate of different groups of people across all policies over time. To do so, we present a variant of the SEIR epidemic model to examine these issues numerically in the next section.

7. Evaluating Rollout Policies via an SEIR Model

In this section, we simulate the pandemic trajectories under the four rollout policies: (1) the stock hold-back policy $(v_1^h(t), v_2^h(t))$, (2) the stock-release policy $(v_1^r(t), v_2^r(t))$, (3) the dose-stretch policy ($L^s = mL$ with $m > 1$), and (4) the alternative single-dose vaccine. We begin by presenting our SEIR epidemic model in Section 7.1. In Section 7.2, we simulate the rollout policies for the "baseline case" in which the parameter values are listed in Table OA1 of the online appendix. To analyze the robustness of our simulation results, we conduct sensitivity analyses with respect to age composition in Section 7.3, risk-based vaccine prioritization in Section 7.4, vaccine supply disruption in Section 7.5, and the basic reproductive number (R_0) in Section 7.6. We caution that, as our simulation parameters are calibrated to data obtained on the original strain of the vaccines and the earlier (e.g., Alpha) variants, the results do not necessarily reflect the effects of future variants. Yet, the same methodology developed in our paper supports new analysis using actual or estimated parameters associated with future variants.

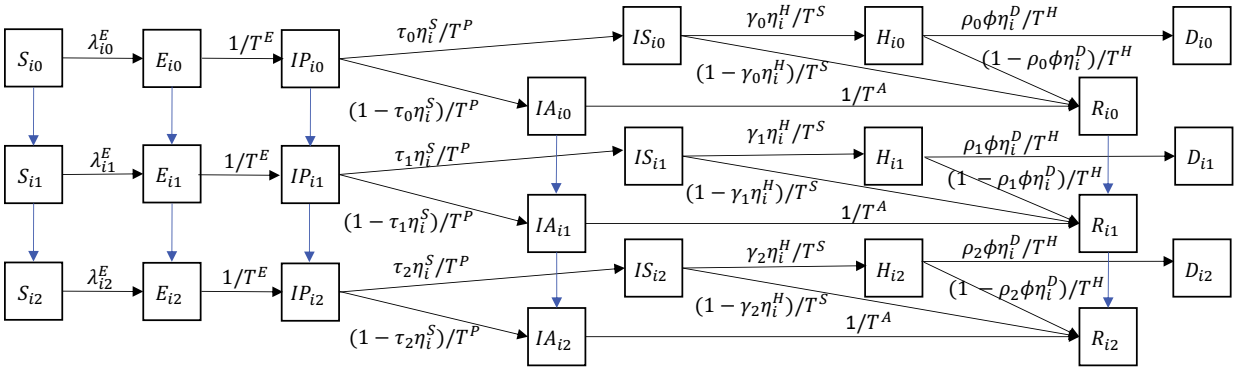
7.1. SEIR Model

We develop a compartmental SEIR epidemic model to capture the evolution of the COVID-19 pandemic and the vaccine rollout dynamics. Using this SEIR model, we evaluate different rollout policies by comparing three performance metrics, including symptomatic infections, hospitalizations, and mortality. These metrics provide an epidemiological perspective and complement the operational perspective presented in Sections 4 to 6.

7.1.1. Risk Groups and Vaccination States. To capture various age-dependent factors such as hospitalization rate, mortality rate due to COVID, and frequencies of social contact, we consider a population of size N (e.g., the U.S. adult population) that consists of a high-risk group H (e.g., the population aged 65 and above) of size N_H and a low-risk group L (e.g., the population under the age of 65) of size N_L; this setting is consistent with the literature on modeling compartments with age groups (e.g., Longini et al. 1978). Based on the U.S. demographics, we normalize the adult population to 100 and set N_H and N_L to 13 and 87 (according to the profile of the adult population in the U.S.), respectively. To incorporate the vaccination status of the two-dose regimen, we denote by (i,k) each individual's type, where $i \in \{H, L\}$ is the risk group and $k \in \{0, 1, 2\}$ is the number of vaccine doses that the individual has received.

7.1.2. Compartments. In line with Keskinocak et al. (2020), an individual of type (i,k) is in one of the following eight compartments at any time t: susceptible $(S_{ik}(t))$, exposed $(E_{ik}(t))$, infectious and pre-symptomatic $(IP_{ik}(t))$, infectious and asymptomatic $(IA_{ik}(t))$, infectious and symptomatic $(IS_{ik}(t))$, hospitalized $(H_{ik}(t))$, recovered $(R_{ik}(t))$, or deceased $(D_{ik}(t))$. By considering this expanded set of states instead of the standard infectious and recovered compartments,

Figure 1 **An expanded SEIR model for group $i \in \{H, L\}$. The downward arrows (in blue) indicate the change in vaccination status. $\phi = 1$ if $\sum_{i,k} H_{ik} \le K$ and $\phi = \kappa$ otherwise.**



our model captures key features of the COVID-19 pandemic such as *asymptomatic transmission* and the demand on the healthcare system due to *hospitalization*. With two risk groups (i), three vaccination states (k), and eight compartments for each individual type (i,k), our SEIR model comprises a total of 48 compartments. Figure 1 illustrates our SEIR model for group $i \in \{H, L\}$.[13]

7.1.3. Transition between compartments. For each individual of type (i,k), we explain in Section OA2 of the online appendix about how we determine the transition rate from one compartment to a different compartment as depicted in Figure 1. In particular, COVID transmissions depend on the numbers of infectious and susceptible individuals, their transmissibility and susceptibility (which depend on their vaccination status), and the frequencies of social contact between the two age groups as governed by a "contact matrix" (Bubar et al. 2020).

After contracting the virus, individuals will go through the exposed and presymptomatic stages, at the end of which they will either develop symptoms or remain asymptomatic, governed by the age- and vaccination-dependent probabilities. Those who are symptomatically infected will be hospitalized (and if so, recover or die) according to the age- and vaccination-dependent probabilities. We calibrate different age- and vaccination-dependent parameter values using the CDC data and the recent literature and summarize the values in Table OA1 of the online appendix.

In addition to the transition rates among different compartments, our SEIR model depicted in Figure 1 captures the impact of the healthcare system utilization on mortality rate. Specifically, as more patients are hospitalized, critical facilities (e.g., intensive care units) become overloaded and mortality rates increase substantially (Wilde et al. 2021). Details of our mortality rate formulation are provided in Section OA2.

[13] We assume the time horizon to be short enough so that individuals do not move from the low- to high-risk group and vice versa (their underlying health status does not change). Therefore, the two groups interact only through sharing the same vaccine supply and hospitalization capacity.

7.1.4. The Role of Vaccination. To evaluate different vaccine rollout policies via the SEIR model for an individual type (i,k) as depicted in Figure 1, we can use the vaccination rates $v_k(t)$ $(k=1,2)$ under each rollout policy presented in Sections 4 to 6 as the key inputs for our SEIR model. In doing so, we can determine the corresponding transition rates between compartments associated with different vaccination status (e.g., from $S_{i0}(t)$ to $S_{i1}(t)$ after receiving the first vaccine dose). Based on the system dynamics captured by our SEIR model, we simulate the following three performance metrics for each risk group $i=H,L$, under all four aforementioned rollout policies: (1) The number of active symptomatic infections: $IS_i(t)=\sum_{k=0}^{2} IS_{ik}(t)$, (2) the number of active hospitalizations: $H_i(t)=\sum_{k=0}^{2} H_{ik}(t)$, and (3) the cumulative number of deaths: $D_i(t)=\sum_{k=0}^{2} D_{ik}(t)$. In the following subsections, we compare the simulated performance metrics $IS_i(t)$, $H_i(t)$, and $D_i(t)$ via our SEIR model across all four rollout policies.

We simulate our SEIR model over a one-year horizon. The parameter values and the settings of the simulation are summarized in Table OA1. Consistent with our theoretical analysis in the previous sections, we consider a supply function that is growing linearly unless otherwise specified.

7.2. Baseline Case

We begin our numerical examples by simulating the performance of the rollout policies under the baseline case; all the parameter values are listed in Table OA1 of the online appendix.

Recall from Sections 4 to 6 that $v_1(t)$ and $v_2(t)$ denote the rate of vaccine doses to be given as first and second doses at time t, under the hold-back, release, stretch or single-dose policies. To determine who receives the vaccines, we consider a simple proportional allocation rule between the high- and low-risk groups, e.g., $\pi\cdot v_k(t)$ and $(1-\pi)\cdot v_k(t)$ of dose-k supplies $(k=1,2)$ are allocated to the high- and low-risk groups at any time t, respectively. In the baseline case, we have $\pi=1$ (i.e., 100% of supply is given to eligible high-risk individuals until they have all been vaccinated). This is consistent with the strict age-based rollout practice in countries such as the U.S. and U.K.. In §7.3, we shall consider an alternative case with a lower value of π for comparison.

Figure 2 shows the SEIR simulation results for the baseline case when we scale the population size $N=100$. The panels plot the number of active symptomatic cases (left), hospitalizations (middle), and cumulative deaths (right), for the high-risk group (top), low-risk group (middle) and total population (bottom). Because the size of the population is normalized to 100, the vertical axes represent the percentage of the total population in each status of interest. In each panel, the four rollout policies are compared against the no vaccination benchmark (solid line in black).

From the top panels of Figure 2, we find that vaccination is highly effective in reducing symptomatic cases, hospitalizations and deaths for the high-risk group. Under all four rollout policies, each of these key metrics are reduced by at least 75%, compared with the no-vaccination benchmark.

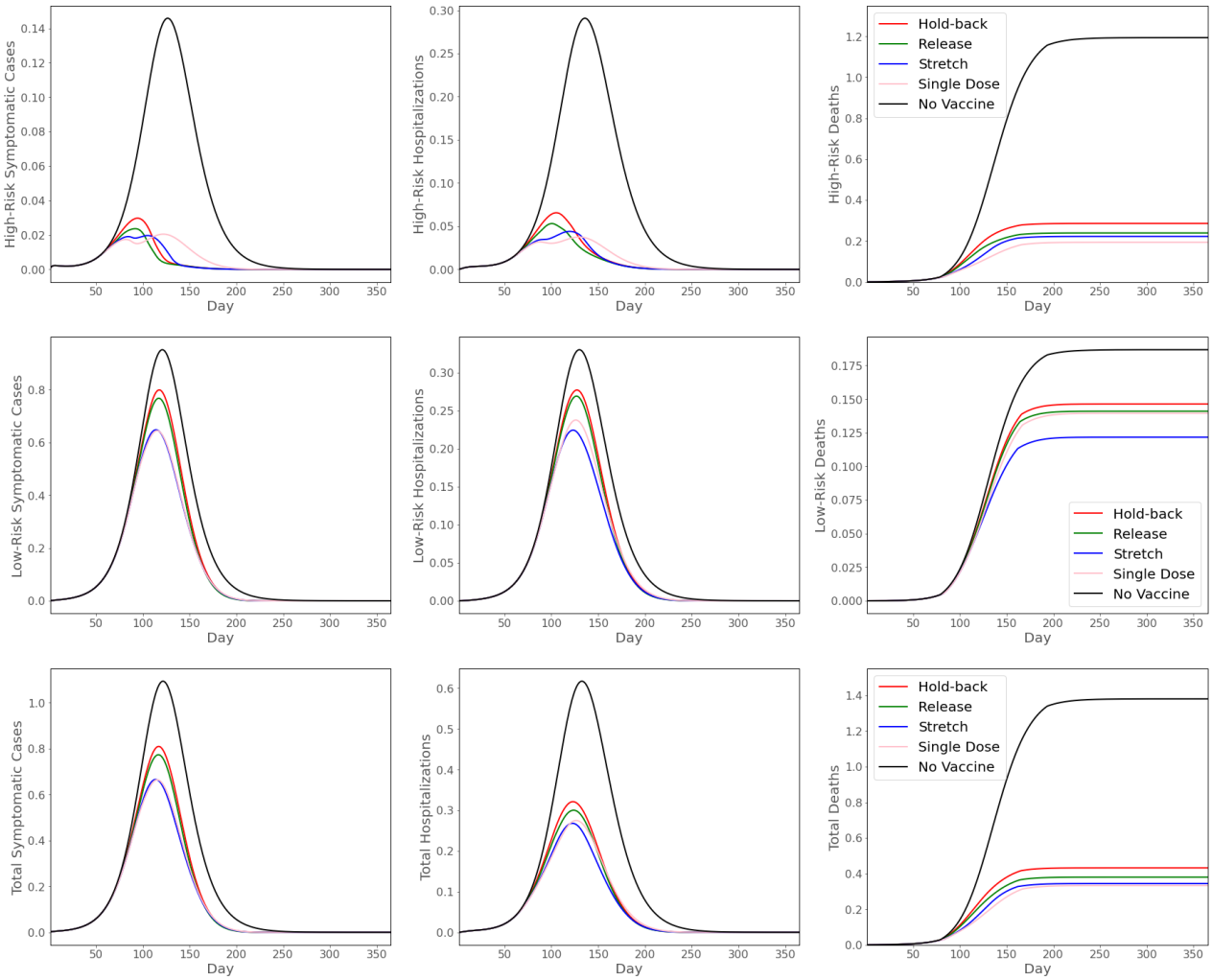


Figure 2 **SEIR Simulation Results Under Baseline Case**

In comparison, the corresponding reductions are smaller for the low-risk group, due to prioritizing the high-risk group and the overall larger population of the low-risk group (thus taking longer to get vaccinated). For the whole population, due to strict priority of vaccinating the high-risk group (i.e., $\pi = 1$), the reduction in deaths (i.e., 69%, 72%, 75% and 76% for the hold-back, release, stretch and single-dose policies, respectively) are more substantial than the reductions in hospitalizations (i.e., 48%, 51%, 57% and 51%, respectively) and in symptomatic infections (i.e., 26%, 29%, 39% and 39%, respectively).

Consistent with the implications from our analytical findings, Figure 2 shows that both the stretching policy and the single-dose regimen are effective in accelerating vaccine impact and reducing symptomatic infections, hospitalizations, and mortality. The release policy also outperforms the hold-back policy in all three metrics, although the difference is modest in the baseline case.

7.3. Age Composition

We now evaluate how the demographic composition may influence the choice of rollout policies. Since hospitalization and mortality risks of COVID are highly dependent on age, the pandemic impact on relatively younger (e.g., in India and South Africa, with a median ages of about 28) and older (e.g., in the U.S., with median age of about 38) populations can differ significantly. All else being the same, we now change N_H to half as in the baseline case (so that 6.5% of the total population is above 65, compared with 93.5% below 65) so that the resulting population is younger. Figure 3 depicts our results.

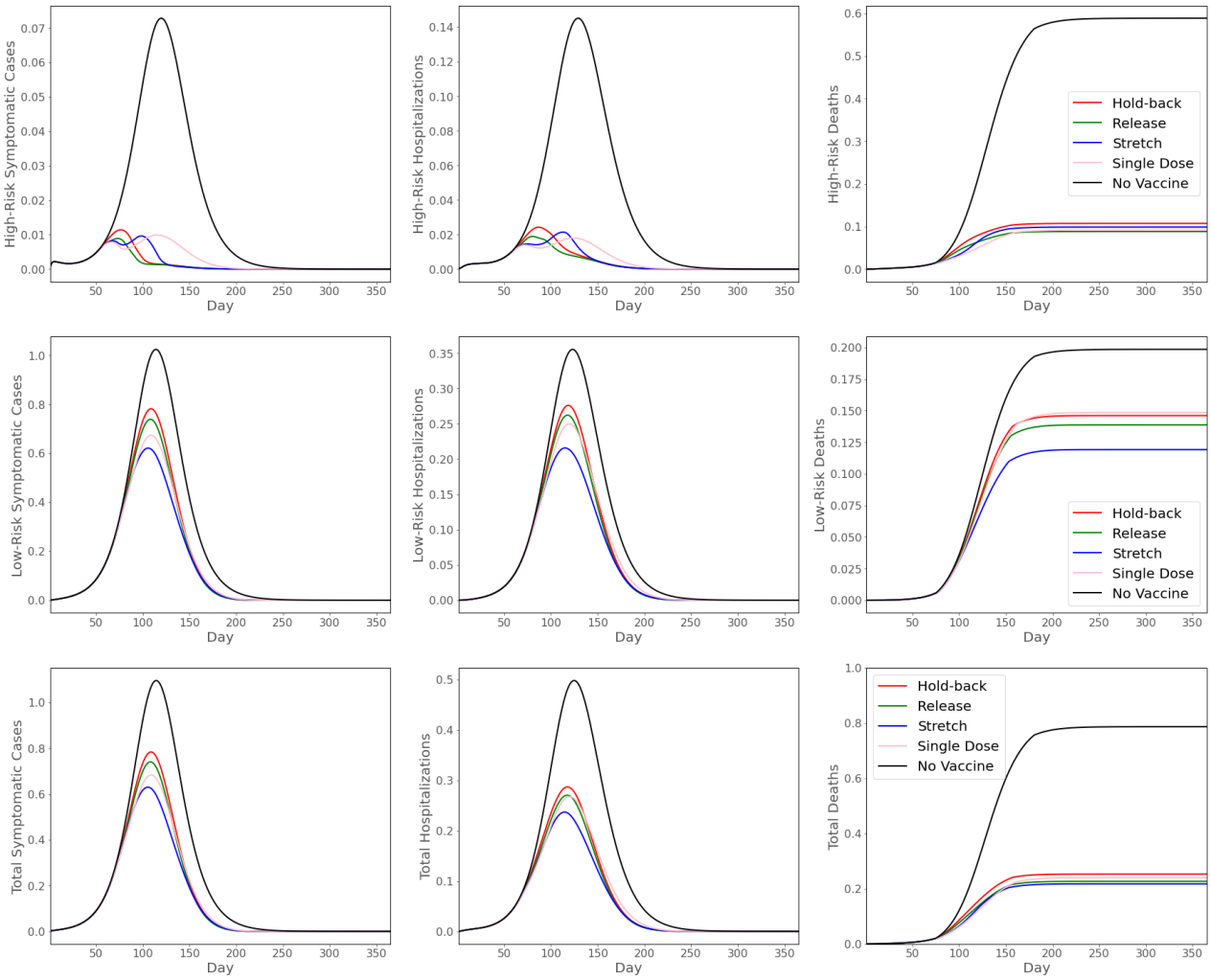


Figure 3 SEIR Simulation Results with Younger Population

Figure 3 resembles the baseline case as reported in Figure 2 in terms of, for example, the effectiveness of the stretching policy and the single-dose regimen. Observe from Figure 3 that even in a

younger population with a 6.5% high-risk population, vaccination reduces mortality significantly—the four policies achieve a mortality reduction from 68% (hold-back policy) to 72% (stretching policy). Despite similar percentage reductions in mortality compared to the baseline case (Figure 2), the absolute number of deaths is considerably lower because we consider a younger population. However, this is not the case for the other two key metrics, as the younger population records similar absolute levels of symptomatic cases and hospitalizations as the older population. This is due to the fact that most symptomatic cases and about half of hospitalizations are associated with the low-risk group due to its higher transmission rate (albeit lower mortality rate).

Interestingly, Figure 3 shows that, relative to the stretching policy, the single-dose regimen does not perform as well when the population is younger. Specifically, the single-dose regimen reduces peak symptomatic cases, peak hospitalizations, and cumulative mortality by 38%, 46%, and 69%, respectively. However, it is outperformed by the stretching policy, which reduces the three key metrics by 43%, 52% and 72%, respectively. In terms of preventing mortality, the single dose regimen performs worse than the release policy (71% reduction) and is similar to the hold-back policy (68% reduction). The reason is that, when the high-risk group is relatively small, the main advantage of the single-dose regimen (i.e., rollout speed) becomes less appealing, whereas its weakness (i.e., lower efficacy, especially in preventing infections) becomes more pronounced.

This observation has important practical implications. Prior to the halt of the Johnson & Johnson vaccine in the U.S. due to blood clot concerns, its "one-and-done" logistical convenience was considered highly promising for younger age groups (Dai et al. 2021). Various countries have since refocused the rollout of Johnson & Johnson to older age groups on the basis of lower blood clot risks (Strauss 2021). Our findings suggest that doing so is sensible logistically and epidemiologically as well, since the single-dose regimen would perform better in an older population.

7.4. Risk-based Prioritization

In the baseline case (Figure 2), we have set $\pi = 1$ (i.e., 100% of supply is given to eligible high-risk individuals until they have all been vaccinated). In this subsection, we compare this with an alternative case that has $\pi = 0.5$ (i.e., half of the supply is allocated to the high-risk group of age above 65 and the remaining half is allocated to the low-risk group of age below 65). Because the low-risk group has higher frequencies of social contacts than the high-risk group, the alternative case is intended to examine the impact of allocating some vaccine to the low-risk group earlier as a mechanism to curb the infection rates.

Figure 4 plots the simulation results for the alternative case with $\pi = 50\%$ priority for the high-risk group. Despite a different vaccine priority rule, Figure 4 resembles Figure 2 in that the single-dose regimen and stretch policies are the most effective in reducing symptomatic infections,

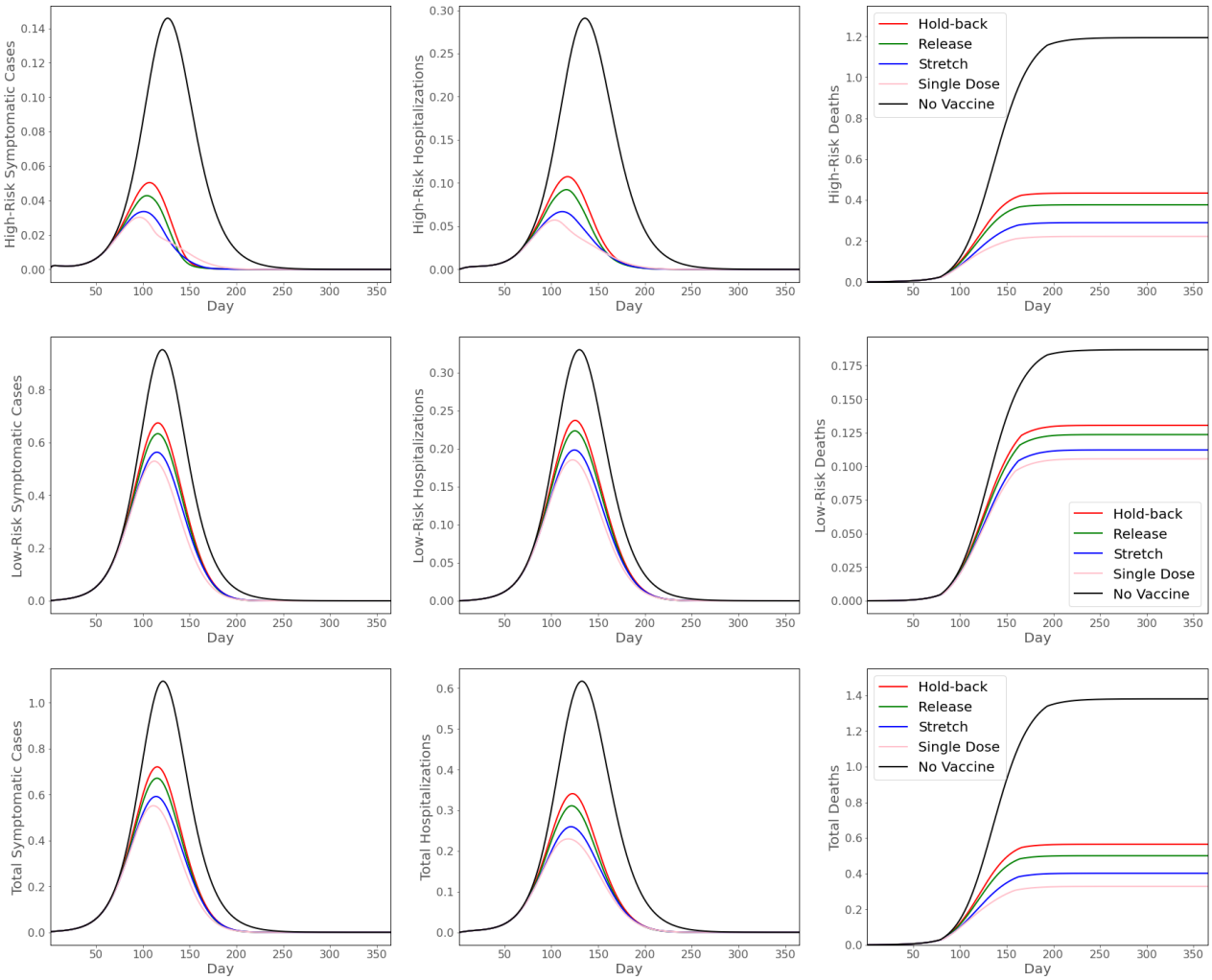


Figure 4 **SEIR Simulation Results with $\pi = 0.5$**

hospitalizations and mortality. In addition, we observe from Figure 4 that the effects of vaccination are abated for the high-risk group and improved for the low-risk group under 50% priority, compared with the baseline case (Figure 2) which strictly prioritizes the high-risk group. Because most deaths arise from the high-risk group, this leads to higher overall mortality (bottom-right panel) under the release and hold-back policies. On the other hand, because most symptomatic cases arise from the low-risk group that has a larger population size and more social contacts, a less stringent risk-group priority leads to fewer symptomatic cases, and a similar or lower number of hospitalizations.

Our findings reflect an important trade-off. Notwithstanding the importance of protecting the most vulnerable people by prioritizing the high-risk group, a strict priority (e.g., the appointment system in the U.K., where younger age groups cannot make bookings until the majority of older age groups receive their vaccines) can be inefficient in containing the spread of the pandemic due to the more frequent social contacts of younger age groups. This is because the vaccines (e.g.,

Pfizer and AstraZeneca) have significant effects on not only preventing infections, but also reducing infectiousness of those who get infected despite vaccinations, as found by Public Health England (Harris et al. 2021) in the U.K. rollout. Many countries and regions, including various states of the U.S., also prioritize the vaccination of younger individuals in high-contact essential professions, such as public transit drivers, delivery drivers, and postal service workers, together with the elderly; vaccinating them helps prevent secondary infections.

Furthermore, we notice that the differences (in all three key metrics) between the rollout policies become more pronounced under less stringent (50%) priority. In particular, the single-dose regimen and the stretching policy, which are effective in accelerating vaccine coverage, become relatively more beneficial when supply is split between the risk groups. This is intuitive, because splitting supply leads to fewer doses available for the high-risk group, which in turn heightens the importance of rollout efficiency. On the whole, we find that both the stretching policy and single-dose regimen offer good trade-offs under the 50% priority case: these policies lead to lower symptomatic cases (46% and 50% reduction from the no-vaccination benchmark, respectively, compared with 39% and 39% under strict priority), and a similar number of hospitalizations (58% and 63% reduction from the no-vaccination benchmark, respectively, compared with 57% and 55% under strict priority) through preventing transmissions, and record similar mortality figures (71% and 76% reduction from the no-vaccination case, compared with 75% and 76% under strict priority) thanks to accelerated build-up of protection for the high-risk group.

7.5. Supply Disruptions

So far, our analytical model and simulation analysis are based on the assumption that the supply process is smooth. However, the supply of the COVID-19 vaccines is subject to disruptions to the production[14] or distribution process.[15] We conduct additional simulations to examine the performance of all four rollout policies in the presence of potential supply disruptions.

We simulate a scenario with production hiccups occurring at the initial stage of vaccine rollouts. In particular, the supply rate is assumed to be growing linearly, as explained earlier. However, the supply rate $y(t)$ suffers from two disruptions occurring Days 81–110 and Days 141–170, during which the supply rate is decreased to half of the initial rate. Figure 5 depicts our simulated results.

Figure 5 shows that under supply disruptions, the results obtained from the baseline case continue to hold: the single-dose regimen and stretch policies are most effective. Because rollouts prioritize the high-risk group, it is natural that the improvements for the low-risk group are limited with

[14] For example, the AstraZeneca vaccine faces an uncertain yield due to uneven cell growth in bioreactors.

[15] Disruption can occur due to weather conditions as experienced in the U.S. during the snow storm in February.

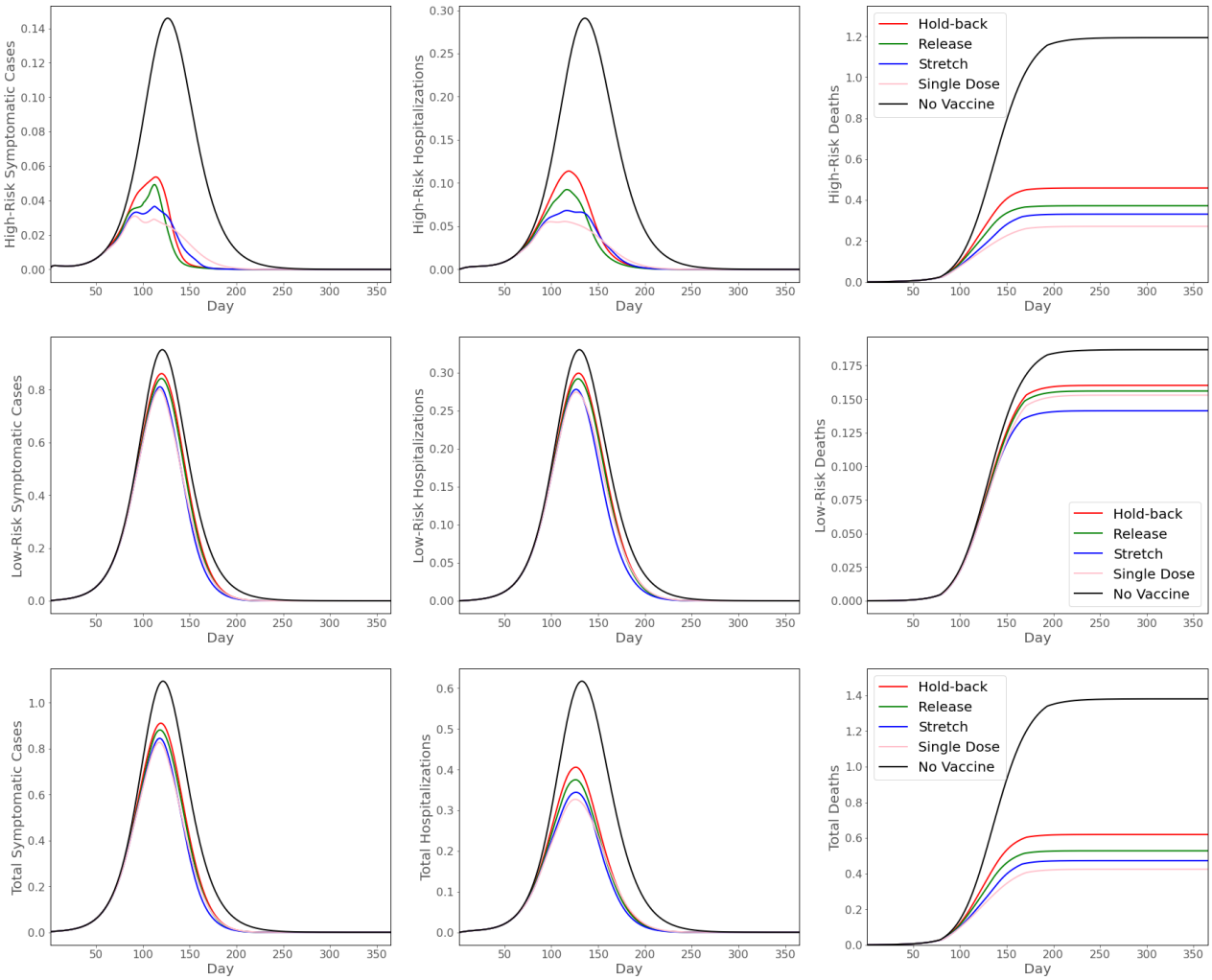


Figure 5 **SEIR Simulation Results with Supply Disruptions**

lumpy supply. Overall, substantial reduction in mortality can still be achieved, especially under the stretching policy or the single-dose regimen policy.

We also observe that the qualitative insights from the baseline case with smooth supply (Figure 2) remain intact; that is, the single-dose regimen and stretch policies help prevent mortality better (by 47% and 44% compared with the no-vaccination case) than do the release and hold-back policies (39% and 34% reductions, respectively). These figures suggest that the performance gaps between the policies become more pronounced during supply disruptions, because making full use of the limited (and unreliable) supply to build up protection becomes more essential.

7.6. Basic Reproduction Number (R_0)

Finally, we evaluate the impact of the speed at which the pandemic propagates, as measured by the basic reproduction number R_0. In the COVID-19 pandemic, the R_0 number could change due to the rise of more transmissible variants or the relaxation of social distancing measures. In Figure 6,

we conduct our simulation based on the case when $R_0 = 1.8$, which is higher than the baseline case that has $R_0 = 1.5$ (Figure 2).

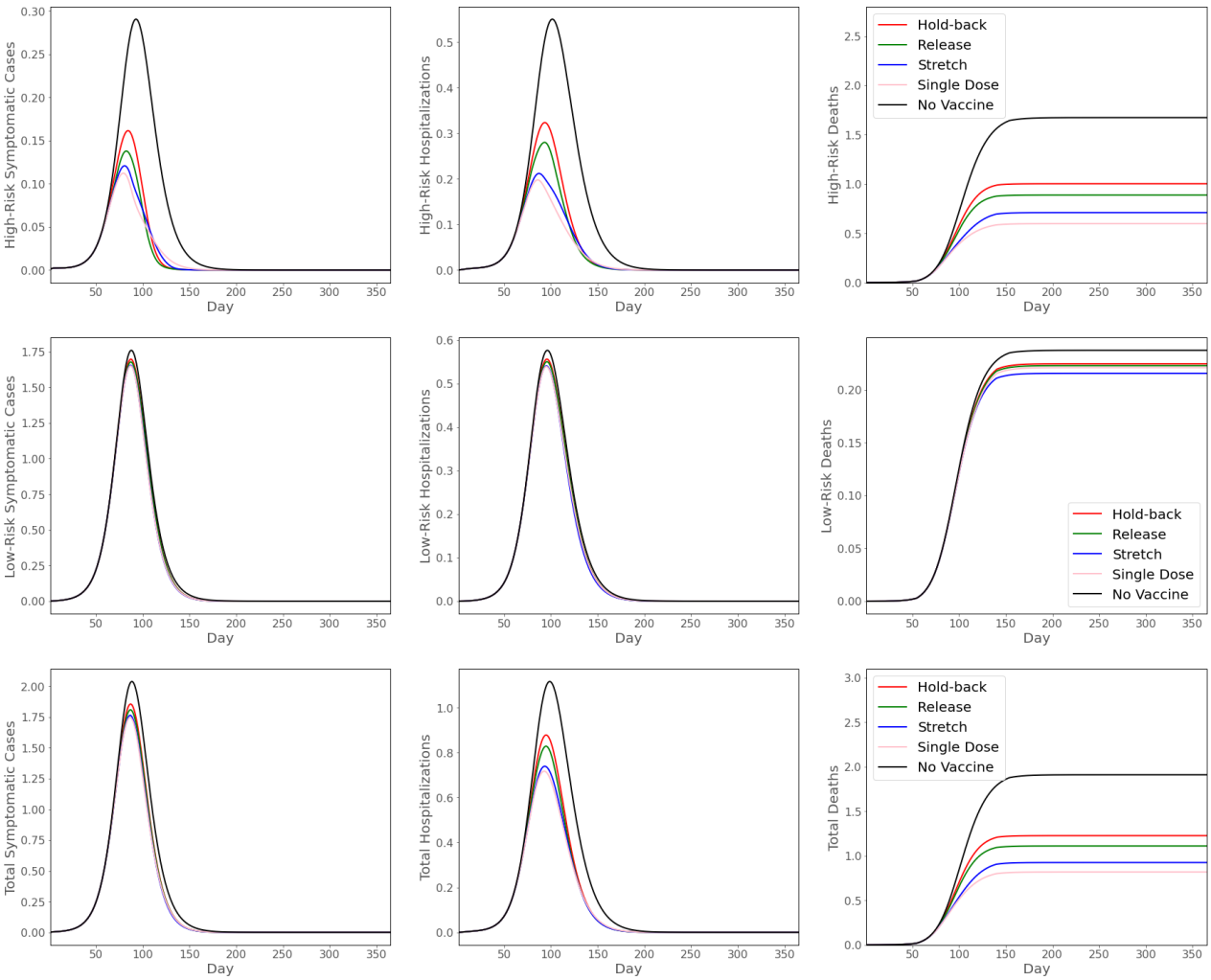


Figure 6 **SEIR Simulation Results with** $R_0 = 1.8$

Figure 6 reveals that whereas all key metrics have worsened under faster transmissions, the results obtained from the baseline case continue to hold qualitatively. Figure 6 also offers insights similar to those from the case with supply disruptions (Figure 5). In particular, vaccine rollouts mostly improve the key metrics for the high-risk group and show limited benefits for the low-risk group. Overall, we observe significant improvements in mortality reduction—by 57%, 52%, 42% and 36% under the single-dose, stretch, release and hold-back policies, respectively. Comparing these figures with those in the baseline case (which range from 69% to 76% reduction across the policies), we observe that under a higher R_0, choosing the right rollout policy has a greater impact on reducing mortality.

The reason that the high R_0 and the supply disruption cases are qualitatively similar is that vaccination outcomes depend on the *relative* speeds of the pandemic and of vaccine rollout. When supply is lumpy or when R_0 is high, vaccination rollout lags behind the pandemic spread, and thus the rollout policies that accelerate protection (i.e., single dose and stretching policies) perform relatively better in preventing mortality of the high-risk group and thus the total population.

8. Conclusions

Rolling out a two-dose COVID-19 vaccine such as that developed by Pfizer offers a rare and compelling research context for examining the impact of different inventory control policies (i.e., hold-back, releasing, and stretching) on public health. We develop a modeling framework of a vaccine rollout process and complement it with an SEIR model to capture the disease transmission process. We compare—both analytically and numerically—various vaccine rollout policies, including (1) holding back all the second doses, (2) releasing all the doses, and (3) stretching the lead time between two doses. We also examine the rollout of a single-dose vaccine with a lower efficacy level than its two-dose counterparts.

Our key findings are three-fold. First, one might expect releasing all the second doses would lead to an oversized effect in slowing down the pandemic. Our analytical results show that compared with the hold-back policy, releasing doses does help slow down infections and end the pandemic sooner; we numerically show that such an impact can be modest. This result is in part due to the fact that releasing more first doses also necessitates more second doses in the future; we show that even if one intends to release all second doses, *fewer than half* of the available doses can be used for first-dose appointments to avoid delays in administering the required second doses, regardless of how fast supply grows over time. Second, stretching the between-dose lead time helps reduce mortality as well as infections and hospitalizations (more than under the strategy of releasing second doses). Third, a single-dose vaccine, even with an overall efficacy significantly lower than its counterparts, can be more effective in infection control.

It should be acknowledged that the logistics parameters of vaccine regimens (i.e., one or two doses, recommended lead time) depend on the design of trials and the data presented for regulatory approval. For instance, the fact that Pfizer requires two doses but Johnson & Johnson requires only one is primarily due to how the respective trials were designed and thus how the specific regimens were approved, but is not necessarily reflective of the differences in the *science* of the vaccines.[16] In this sense, our results underline the importance of taking the *operational* aspect of the rollout process into consideration when designing clinical trials.

[16] The authors acknowledge Andrew Pollard of the University of Oxford for this important insight.

Our model has several limitations that can serve as subjects for future research. First, COVID-19 vaccine rollouts inspire future research into a wide range of operational and logistics issues specific to vaccine administration, including cold chain management, inventory management, and manufacturing capacity planning (Dai and Song 2021). One particularly interesting area of research entails incorporating and managing the uncertainty associated with the number of doses that can be extracted from each vial. Second, as Mak and Tang (2021) note, no-shows for second-dose appointments may worsen when the time interval between two doses is stretched from three weeks to 12 weeks for a variety of reasons, including, for example, the perception that the second dose provides limited incremental benefit due to the stretching. To incorporate potential no-shows, one may extend our SEIR model by adding a new parameter capturing the reduction in the second-dose vaccination rate. Third, we consider in this paper a time horizon long enough that the pandemic subsides. One related matter that we do not consider is the possibility of waning vaccine effects, which could affect infections beyond the time horizon. Lastly, several highly contagious new variants of SARS-CoV-2, the virus causing COVID-19, have been found in various countries. Examining other rollout policies that can expedite herd immunity even more quickly to reduce the spread and mutation of the virus would be of interest.

Our modeling framework represents an initial attempt to examine different vaccine rollout policies under limited supply. In addition to helping inform vaccine rollout policymaking, our modeling framework can be applied to analyze similar issues facing local vaccination planners and health systems leadership in both the current and future epidemics. At the global level, our work draws from the experience of the E.U., U.K., and U.S. Yet our results have important implications for vaccine planning in various countries, especially because developing countries are expected to face extremely limited vaccine supply up to 2023 (Hopkins 2021). More broadly, our paper sheds light on how to develop effective operations strategies for distributing time-sensitive resources in times of crisis.

Acknowledgments

The authors are grateful to Stephen Chick, Eric Denardo, Edward Kaplan, Steven Lippman, Elisa Long, John Mamer, David Paltiel, Andrew Pollard, Prashant Yadav, seminar and session participants at INFORMS 2021 Annual Meeting, INSEAD, The 3rd Johns Hopkins Symposium on Healthcare Operations, McGill University, and Rutgers University for their helpful comments and suggestions.

References

Amit S, Regev-Yochay G, Afek A, Kreiss Y, Leshem E (2021) Early rate reductions of SARS-CoV-2 infection and COVID-19 in BNT162b2 vaccine recipients. *Lancet* (February 18, 2021), ePub ahead of print.

Arifoğlu K, Deo S, Iravani S (2012) Consumption externality and yield uncertainty in the influenza vaccine supply chain: Interventions in demand and supply sides. *Management Sci.* 58(6):1072–1091.

Arifoğlu K, Tang CS (2021) A two-sided incentive program for coordinating the influenza vaccine supply chain. *Manufacturing Service Oper. Management*, forthcoming.

Arnold C (2020) The biggest logistics challenge in history. *New Scientist* 248(3309):36–40.

Arrow KJ, Harris T, Marschak J (1951) Optimal inventory policy. *Econometrica* 19(3):250–272.

Bar-On Y, Baron T, Cornfeld O, Milo R, Yashiv E (2020) COVID19: Erroneous modelling and its policy implications. *Weizmann Institute of Science Working Paper.*

Boseley S (2021) Why has AstraZeneca reduced promised vaccine supply to EU and is UK affected? *Guardian* (Jan 26). http://bit.ly/azredeuuk Accessed February 18, 2021.

Bubar KM, Reinholt K, Kissler SM, Lipsitch M, Cobey S, Grad YH, Larremore DB (2021) Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Sci.* 371(6532) 916–921.

Chen MK, Chevalier JA, Long EF (2020) Nursing home staff networks and COVID-19. *Proc. National Acad. Sci.* 118(1):e2015455118.

Chick SE, Mamani H, Simchi-Levi D (2008) Supply chain coordination and influenza vaccination. *Oper. Res.* 56(6):1493–1506.

Chick SE, Hasija S, Nasiry J (2017) Information elicitation and influenza vaccine production. *Oper. Res.* 65(1):75–96.

Cho S-H, Tang CS (2013) Advance selling in a supply chain under uncertain supply and demand. *Manufacturing Service Oper. Management* 15(2):305–319.

Cummings K, Behzad B, Martonosi S (2020) Centers for Disease Control and Prevention as a strategic agent in the pediatric vaccine market: An analytical approach. *Manufacturing Service Oper. Management*, forthcoming.

Dai T, Cho S-H, Zhang F (2016) Contracting for on-time delivery in the US influenza vaccine supply chain. *Manufacturing Service Oper. Management* 18(3):332–346.

Dai T, Mak HY, Tang CS (2021) The great promise of a one-dose vaccine. *Barron's* (February 26, 2021). http://bit.ly/barronsjnj Accessed June 11, 2021.

Dai T, Song JS (2021) Transforming COVID-19 vaccines into vaccination: Challenges and opportunities for management scientists. *Health Care Management Sci.* ePub ahead of print.

Dai T, Yadav P (2021) Release second doses to speed vaccinations. *USA Today* (January 13) 7A.

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

Dvoretzky A, Kiefer J, Wolfowitz J (1953) On the optimal character of the (s, S) policy in inventory theory. *Econometrica* 21(4):586—596.

Gallien, J, Rashkova, I, Atun, R, Yadav, P (2017) National drug stockout risks and the global fund disbursement process for procurement. *Production Oper. Management*, 26(6):997–1014.

Gottlieb S (2021) Pharmacies can get shots in arms. *Wall Street Journal* (January 3). http://on.wsj.com/3aim4Be Accessed February 18, 2021.

Hall SN, Jacobson SH, Sewell EC (2008) An analysis of pediatric vaccine formulary selection problems. *Oper. Res.* 56(6):1348–1365.

Haas EJ, Angulo FJ, McLaughlin JM, Anis E, Singer SR, Khan F, Brooks N, et al. (2021) Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* 397(10287):1819–1829.

Hopkins JS (2021). Developed countries lock up Covid-19 vaccines through 2023. *Wall Street Journal* (June 2). https://on.wsj.com/3pL3qbh Accessed June 11, 2021.

Kaplan EH (2020) OM Forum—COVID-19 scratch models to support local decisions. *Manufacturing Service Oper. Management* 22(4):645–655.

Keskinocak P, Oruc BE, Baxter A, Asplund J, Serban N (2020) The impact of social distancing on COVID19 spread: State of Georgia case study. *PLoS ONE* 15(10):e0239798.

Kraiselburd, S, Yadav, P (2013) Supply chains and global health: an imperative for bringing operations management scholarship into action. *Production Oper. Management*, 22(2):377–381.

Longini IM, Ackerman E, Elveback LR (1978) An optimization model for Influenza A epidemics. *Math. Biosciences* 38(1–2):141–157.

Mak HY, Tang CS (2021) More vaccine supplies are being released. It's a gamble. *Barron's* (Jan 12, 2021). http://bit.ly/2ndvaxgamble Accessed February 18, 2021.

Mancini DP, Gross A (2021) UK decision to space vaccination doses sparks fierce debate over risk. *Financial Times* (January 8). http://on.ft.com/3pgBRVr Accessed February 18, 2021.

Miller H, Cohn M (2021) Maryland facing shortage of second COVID vaccine doses, health secretary says. *Baltimore Sun* (February 1). http://bit.ly/md2dosehb Accessed February 18, 2021.

Muhumuza R, Mutsaka F (2021) In poorest countries, surges worsen shortages of vaccines. Associated Press (June 17). http://bit.ly/africavaxap Accessed June 20, 2021.

Natarajan, KV, Swaminathan, JM (2017) Multi-treatment inventory allocation in humanitarian health settings under funding constraints. *Production Oper. Management* 26(6): 1015–1034.

O'Donnell C, Spalding R (2021) U.S. releases millions of COVID-19 doses, urges states to broaden vaccine access. *Reuters News* (Jan 12, 2021). http://reut.rs/3rBPJv1

Paltiel D, Schwartz J, Zheng A, Walensky RP. (2021a) Clinical outcomes of a COVID-19 vaccine: Implementation over efficacy. *Health Affairs* 40(1): 42-52.

Paltiel AD, Zheng A, Schwartz JL (2021b) Speed versus efficacy: Quantifying potential tradeoffs in COVID-19 vaccine deployment. *Ann. Intern Medicine* 174(4):568–570.

Porteus EL (1990) Chapter 12 Stochastic inventory theory. D.P. Heyman, M.J. Sobel, eds. *Handbooks in Operations Research and Management Science*. (Elsevier, Amsterdam, The Netherlands), 605–652.

Rawlinson K (2020) More data needed before giving just one vaccine dose, says Covid adviser. *The Guardian*. https://bit.ly/3aEz6tj Accessed February 18, 2021.

Scarf H (1960) The optimality of (S, s) policies in the dynamic inventory problem. K. J. Arrow, S. Karlin, P. Suppes, eds. *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford, CA, 196–202.

Shumsky RA, Smith J, Hoen A, Gilbert M (2021) Allocating COVID-19 vaccines: Save one for the second dose? *Dartmouth College Working Paper*.

Song JS, van Houtum GJ, Van Mieghem JA (2020) Capacity and inventory management: Review, trends, and projections. *Manufacturing Service Oper. Management* 22(1):36–46.

Strauss M (2021) Belgium halts J&J COVID vaccine for under 41s after first EU death. Reuters (May 26). https://reut.rs/3gsfKJl Accessed June 11, 2021.

Tuite AR, Zhu L, Fisman DN, Salomon JA (2021) Alternative dose allocation strategies to increase benefits from constrained COVID-19 vaccine supply. *Ann. Intern Medicine* 174(4):570–572.

Voysey M, Costa Clemens SA, Madhi SA, Weckx LY, Folegatti PM, Aley PK, Angus B, et al. (2021) Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. *Lancet* 397(10277):881–891.

Wilde H, Mellan T, Hawryluk I, Dennis JM, Denaxas S, Pagel C, Duncan A, et al. (2021) The association between mechanical ventilator availability and mortality risk in intensive care patients with COVID-19: A national retrospective cohort study. medRxiv 2021.01.11.21249461.

Yamin D, Gavious A (2013) Incentives' effect in influenza vaccination policy. *Management Sci.* 59(12):2667–2686.

Online Appendix to "Managing Two-Dose COVID-19 Vaccine Rollouts with Limited Supply"

OA1. Technical Proofs

Proof of Lemma 1

No backlogs exist any time at or before t if and only if

$$(1-\beta)y(u) \geq v_1^r(u-L) \text{ for all } u \leq t.$$

The above statement can be rewritten as

$$(1-\beta)y(nL+\tau) \geq v_1^r((n-1)L+\tau) \text{ for } n=1,2,\dots \text{ and all } \tau \in [0,L]. \tag{OA1}$$

We prove (OA1) by induction. First, the case in which $n=1$ and $n=2$ obviously holds for all $\tau \in [0,L]$ given the first two conditions of the lemma. Next, suppose the statement holds for $n=1,\cdots,k$. Then, for $n=k+1$, it holds that

$$\begin{aligned} v_1^r(kL+\tau) &= y(kL+\tau) - v_2^r(kL+\tau) \\ &= y(kL+\tau) - v_1^r((k-1)L+\tau) \qquad \text{(because } B(t)=0 \text{ for all } t \leq kL+\tau) \\ &= y(kL+\tau) - y((k-1)L+\tau) + v_1^r((k-2)L+\tau) \\ &\leq y(kL+\tau) - \beta y((k-1)L+\tau) \qquad \text{(by induction assumption).} \end{aligned}$$

Because $(1-\beta)y((k+1)L+\tau) \geq y(kL+\tau) - \beta y((k-1)L+\tau)$, the statement also holds for $n=k+1$. Thus, it holds for any positive integer n by induction, which completes the proof. *Q.E.D.*

Proof of Lemma 2

The rates of vaccination are given by

$$\begin{aligned} v_1^h(t) &= \frac{1}{2} + \frac{\alpha}{2}t \\ v_2^h(t) &= \begin{cases} 0 & \text{for } t \in [0,L) \\ \frac{1}{2} + \frac{\alpha}{2}(t-L) & \text{otherwise.} \end{cases} \end{aligned}$$

The result follows from integrating the vaccination rates with respect to t. *Q.E.D.*

Proof of Proposition 1

First, in $t \in [0,L)$, all available supply is released for first-dose appointments:

$$\begin{aligned} v_1^r(t) &= 1 + \alpha t \\ V_1^r(t) &= t + \frac{\alpha}{2}t^2 \end{aligned}$$

$$P_1^r(t) = \frac{1}{2}t^2 + \frac{\alpha}{6}t^3$$
$$v_2^r(t) = V_2^r(t) = P_2^r(t) = 0.$$

For $t \in [L, 2L)$, the demand for the second dose is $v_1^r(t-L) = 1+\alpha(t-L)$ and the total supply available is $1+\alpha t$. For no backlogs to exist, all second-dose demand will be fulfilled, that is, $v_2^r(t) = 1+\alpha(t-L)$, and the supply available for first doses will be $y(t) - v_2^r(t) = \alpha L$. To ensure this amount exceeds the fraction β, we require $\beta \le \frac{\alpha L}{1+\alpha t}$ for all $t \in [L, 2L)$, that is, $\beta \le \frac{\alpha L}{1+2\alpha L} = \frac{1}{2} - \frac{1}{2(1+2\alpha L)}$.

For $t \in [2L, 3L)$, the demand for the second dose is $v_1^r(t-L) = \alpha L$ and the supply is $y(t) = 1+\alpha t$. Fulfilling all second-dose demand, the capacity available for first doses is given by $1+\alpha(t-L)$. For this amount to exceed a β fraction of available supply, we require $\beta \le \frac{1+\alpha(t-L)}{1+\alpha t}$ for all $t \in [2L, 3L)$, that is, $\beta \le \frac{1+\alpha L}{1+2\alpha L}$, which is implied by the previous necessary condition, $\beta \le \frac{\alpha L}{1+2\alpha L} = \frac{1}{2} - \frac{1}{2(1+2\alpha L)}$.

Continuing this argument, the vaccination rate follows an oscillating pattern:

$$v_1^r(t) = \begin{cases} 1+\alpha(t-nL) & \text{for } t \in [2nL, (2n+1)L) \\ \alpha(n+1)L & \text{for } t \in [(2n+1)L, (2n+2)L) \end{cases}. \tag{OA2}$$

$$v_2^r(t) = \begin{cases} \alpha nL & \text{for } t \in [2nL, (2n+1)L) \\ 1+\alpha(t-nL) & \text{for } t \in [(2n+1)L, (2n+2)L). \end{cases} \tag{OA3}$$

The upper bound on β increases over time. Thus, $\beta \le \frac{1}{2} - \frac{1}{2(1+2\alpha L)}$ is sufficient. *Q.E.D.*

Proof of Proposition 2

The stock-release policy increases vaccination rates by:

$$\begin{aligned} \Delta v_1^r(t) &= \begin{cases} 1+\alpha(t-nL) - \frac{1}{2}(1+\alpha t) & \text{for } t \in [2nL, (2n+1)L) \\ \alpha(n+1)L - \frac{1}{2}(1+\alpha t) & \text{for } t \in [(2n+1)L, (2n+2)L) \end{cases} \\ &= \begin{cases} \frac{1}{2} + \frac{\alpha}{2}(t-2nL) > \frac{1}{2} & \text{for } t \in [2nL, (2n+1)L) \\ -\frac{1}{2} + \frac{\alpha}{2}[(2n+2)L - t] > -\frac{1}{2} & \text{for } t \in [(2n+1)L, (2n+2)L). \end{cases} \end{aligned}$$

In the interval $[2nL, t)$ where $t \in [2nL, (2n+1)L)$, the release policy increases the cumulative vaccinations by:

$$\begin{aligned} \int_{2nL}^{t} \Delta v_1^r(\tau)d\tau &= \frac{1}{2}(t-2nL) + \frac{\alpha}{2}\int_{2nL}^{t}(\tau - 2nL)d\tau \\ &= \frac{1}{2}(t-2nL) + \frac{\alpha}{4}(t-2nL)^2. \end{aligned}$$

Similarly, in the interval $[(2n+1)L, t)$, where $t \in [(2n+1)L, (2n+2)L)$, the release policy increases the cumulative vaccinations by

$$\begin{aligned} \int_{(2n+1)L}^{t} \Delta v_1^r(\tau)d\tau &= -\frac{1}{2}[t-(2n+1)L] + \frac{\alpha}{2}\int_{(2n+1)L}^{t}[(2n+2)L - \tau]d\tau \\ &= -\frac{1}{2}[t-(2n+1)L] + \frac{\alpha}{4}\left\{L^2 - [2(n+1)L - t]^2\right\}. \end{aligned}$$

The above implies that over $[2nL,(2n+2)L)$, the gain amounts to $\frac{\alpha}{2}L^2$. Thus, over the time period $t\in[0,t)$, the release policy increases the cumulative vaccinations by:

$$\Delta V_1^r(t)=\begin{cases}\frac{\alpha}{2}nL^2+\frac{1}{2}(t-2nL)+\frac{\alpha}{4}(t-2nL)^2 & \text{for } t\in[2nL,(2n+1)L)\\ \frac{\alpha}{2}nL^2+\frac{1}{2}L+\frac{\alpha}{2}L^2-\frac{1}{2}[t-(2n+1)L] & \\ -\frac{\alpha}{4}[2(n+1)L-t]^2 & \text{for } t\in[(2n+1)L,(2n+2)L).\end{cases}$$

Going through similar steps, in the interval $[2nL,t)$, where $t\in[2nL,(2n+1)L)$, the release policy increases the cumulative protection by:

$$\begin{aligned}\int_{2nL}^{t}\Delta V_1^r(\tau)d\tau&=\frac{\alpha}{2}nL^2(t-2nL)+\int_{2nL}^{t}\left[(\tau-2nL)+\frac{\alpha}{4}(\tau-2nL)^2\right]d\tau\\&=\frac{\alpha}{2}nL^2(t-2nL)+\frac{1}{4}(t-2nL)^2+\frac{\alpha}{12}(t-2nL)^3\\&=\frac{1}{4}L^2+\frac{\alpha}{2}L^3\left(n+\frac{1}{6}\right)\qquad(\text{when } t=(2n+1)L).\end{aligned}$$

In the interval $[(2n+1)L,t)$, where $t\in[(2n+1)L,(2n+2)L)$, the release policy increases the cumulative protection by:

$$\begin{aligned}\int_{(2n+1)L}^{t}\Delta V_1^r(\tau)d\tau&=\left(\frac{\alpha}{2}nL^2+\frac{1}{2}L+\frac{\alpha}{2}L^2\right)[t-(2n+1)L]-\int_{(2n+1)L}^{t}\left\{[\tau-(2n+1)L]-\frac{\alpha}{4}[(2n+2)L-\tau]^2\right\}d\tau\\&=\left(\frac{\alpha}{2}nL^2+\frac{1}{2}L+\frac{\alpha}{2}L^2\right)[t-(2n+1)L]-\frac{1}{4}[t-(2n+1)]^2-\frac{\alpha}{12}\left\{L^3-[(2n+2)L-t]^3\right\}\\&=\frac{1}{4}L^2+\frac{\alpha}{2}L^3\left(n+\frac{5}{6}\right)\qquad(\text{when } t=(2n+2)L).\end{aligned}$$

Thus, the gain in cumulative protection over the period $[0,t)$ is given by

$$\Delta P_1^r(t)=\begin{cases}\frac{n}{2}L^2+\frac{\alpha}{2}L^3\left(2\sum_{i=0}^{n}i\right)&\\ +\frac{\alpha}{2}nL^2(t-2nL)+\frac{1}{4}(t-2nL)^2+\frac{\alpha}{12}(t-2nL)^3 & \text{for } t\in[2nL,(2n+1)L)\\ \left(\frac{n}{2}+\frac{1}{4}\right)L^2+\frac{\alpha}{2}L^3\left(2\sum_{i=0}^{n}i+n+\frac{1}{6}\right)&\\ +\left(\frac{\alpha}{2}nL^2+\frac{1}{2}L+\frac{\alpha}{2}L^2\right)[t-(2n+1)L]&\\ -\frac{1}{4}[t-(2n+1)]^2-\frac{\alpha}{12}\left[L^3-[2(n+1)L-t]^3\right] & \text{for } t\in[(2n+1)L,(2n+2)L).\end{cases}$$

Over a full cycle $[0,2nL)$, the release policy improves the cumulative protection by

$$\begin{aligned}\Delta P_1^r(2nL)&=\frac{n}{2}L^2+\frac{\alpha}{2}L^3\left(2\sum_{i=0}^{n}i\right)\\&=\frac{n}{2}L^2+\frac{\alpha}{2}L^3n(n+1),\ \text{and}\\\Delta P_2^r(2nL)&=\Delta P_1^r((2n-1)L)\\&=\left(\frac{n-1}{2}+\frac{1}{4}\right)L^2+\frac{\alpha}{2}L^3\left(2\sum_{i=0}^{n-1}i+n-1+\frac{1}{6}\right)\\&=\left(\frac{n-1}{2}+\frac{1}{4}\right)L^2+\frac{\alpha}{2}L^3\left(n^2-\frac{5}{6}\right).\end{aligned}$$

Q.E.D.

Proof of Proposition 3

Following Proposition 2, the cumulative first- and second-dose inoculations under the stock-release policy are given by:

$$V_1^r(t) = \begin{cases} \frac{t}{2} + \frac{\alpha t^2}{4} + \frac{\alpha}{2} nL^2 + \frac{1}{2}(t-2nL) + \frac{\alpha}{4}(t-2nL)^2 & \text{for } t \in [2nL, (2n+1)L) \\ \frac{t}{2} + \frac{\alpha t^2}{4} + \frac{\alpha}{2} nL^2 + \frac{1}{2}L + \frac{\alpha}{2}L^2 - \frac{1}{2}[t-(2n+1)L] & \\ -\frac{\alpha}{4}[2(n+1)L - t]^2 & \text{for } t \in [(2n+1)L, (2n+2)L) \end{cases}$$
$$V_2^r(t) = V_1^r(t-L).$$

At $t = 2mL$, we have

$$\begin{aligned} V_1^r(2mL) &= mL + \alpha\left(m^2 + \frac{m}{2}\right)L^2 \\ V_2^r(2mL) &= \frac{(2m-1)L}{2} + \frac{\alpha}{4}(2m-1)^2L^2 + \frac{\alpha}{2}(m-1)L^2 + \frac{1}{2}L + \frac{\alpha}{4}L^2 \\ &= mL + \frac{3\alpha}{4}L^2(2m^2 - 2m + 1). \end{aligned}$$

For the stretching policy, the same derivation holds with the increased lead time:

$$\begin{aligned} V_1^s(2mL) &= L^s + \alpha\left(1 + \frac{1}{2}\right)(L^s)^2 \\ &= mL + \frac{3\alpha m^2 L^2}{2} \\ &= V_1^r(2mL) + \alpha \cdot \frac{m^2 - m}{2} \cdot L^2 \\ V_2^s(2mL) &= L^s + \frac{3\alpha}{4}(L^s)^2 \\ &= mL + \frac{3\alpha}{4}L^2 m^2 \\ &= V_2^r(2mL) - \frac{3\alpha}{4}L^2(m^2 - 2m + 1). \end{aligned}$$

Similarly, following Proposition 2, the cumulative protection under the stock-release policy is given by:

$$\begin{aligned} P_1^r(2mL) &= m^2L^2 + \frac{2\alpha}{3}m^3L^3 + \frac{m}{2}L^2 + \frac{\alpha}{2}L^3 m(m+1) \\ &= \frac{2m^2 + m}{2}L^2 + \alpha L^3\left[\frac{2}{3}m^3 + \frac{1}{2}m(m+1)\right] \\ P_2^r(2mL) &= \frac{1}{4}(2m-1)^2L^2 + \frac{\alpha}{12}(2m-1)^3L^3 + \left(\frac{m-1}{2} + \frac{1}{4}\right)L^2 + \frac{\alpha}{2}L^3\left(m^2 - \frac{5}{6}\right) \\ &= \frac{3}{4}(2m^2 - 2m + 1)L^2 + \frac{\alpha L^3}{12}(8m^3 - 6m^2 + 6m - 6). \end{aligned}$$

Similarly, under the dose-stretching policy, the cumulative protection can be written as:

$$P_1^s(2mL) = \frac{3}{2}(L^s)^2 + \frac{5}{3}\alpha(L^s)^3$$

$$
\begin{aligned}
&= P_1^r(2mL) + \frac{m^2-m}{2}L^2 + \alpha L^3[2m^3 - m(m+1)] \\
P_2^s(2mL) &= \frac{3}{4}m^2L^2 + \frac{\alpha L^3}{12}(2m^3) \\
&= P_2^r(2mL) - \frac{3}{4}(m^2-2m+1)L^2 - \frac{\alpha L^3}{2}(m^3-m^2+m-1) \\
&= P_2^r(2mL) - \frac{3}{4}(m-1)^2L^2 - \frac{\alpha L^3}{2}(m-1)(m^2+1),
\end{aligned}
$$

which completes the proof. *Q.E.D.*

Proof of Proposition 4

Following Proposition 3, dose stretching increases efficacy-weighted protection by:

$$
\begin{aligned}
\Delta P^s(2mL) &= \lambda(1-\theta_1)\left[\Delta P_1^s(2mL) + \zeta \Delta P_2^s(2mL)\right] \\
&= \frac{L^2}{4}\lambda(1-\theta_1)\left[2m(m-1) - 3(m-1)^2\zeta\right] + \frac{\alpha L^3}{2}\lambda(1-\theta_1)\left[4m^3 - 2m(m+1) - (m-1)(m^2+1)\zeta\right] \\
&= \frac{L^2}{4}\lambda(1-\theta_1)\left[2m(m-1) - 3(m-1)^2\zeta\right] + \frac{\alpha L^3}{2}\lambda(1-\theta_1)\left[2(m-1)(2m^2+m) - (m-1)(m^2+1)\zeta\right].
\end{aligned}
$$

If $m=1$ or $\zeta \le \frac{2m}{3(m-1)}$, the term $[2m(m-1) - 3(m-1)^2\zeta] \ge 0$. Also, for any $m \ge 1$, $2m^2+m) > m^2+1$. Thus the term $[2(m-1)(2m^2+m) - (m-1)(m^2+1)\zeta] > 0$ if $\zeta \le 2$. Because $\frac{2m}{3(m-1)} < 2$ for all $m > 1$, $\Delta P^s(2mL) > 0$ as long as $\zeta \le \frac{2m}{3(m-1)}$. *Q.E.D.*

Proof of Proposition 5

Because $v^d(t) = 1 + \alpha t$, we have $V^d(t) = t + \frac{\alpha}{2}t^2$ and $P^d(t) = \frac{t^2}{2} + \frac{\alpha}{6}t^3$. These expressions directly lead to the comparisons with the hold-back policy.

Comparing against the stock-release policy gives

$$
\begin{aligned}
V_1^d(2nL) - V_1^r(2nL) &= 2nL + \frac{\alpha}{2}(2nL)^2 - \frac{2nL}{2} - \frac{\alpha}{4}(2nL)^2 - \frac{\alpha}{2}nL^2 \\
&= nL + \frac{\alpha}{2}(2n^2 - n) \\
P_1^d(2nL) - P_1^r(2nL) &= \frac{1}{4}(2nL)^2 + \frac{\alpha}{12}(2nL)^3 - \frac{n}{2}L^2 - \frac{\alpha}{2}L^3 n(n+1) \\
&= \frac{n}{2}L^2 + \frac{\alpha}{2}L^3\left[\frac{4}{3}n^3 - n(n+1)\right],
\end{aligned}
$$

which completes the proof. *Q.E.D.*

Proof of Proposition 6

Recall from the proof of Proposition 4 that

$$
\begin{aligned}
P_1^r(2nL) &= \frac{2n^2+n}{2}L^2 + \alpha L^3\left[\frac{2}{3}n^3 + \frac{1}{2}n(n+1)\right] \\
P_2^r(2nL) &= \frac{3}{4}(2n^2-2n+1)L^2 + \frac{\alpha L^3}{12}(8n^3 - 6n^2 + 6n - 6).
\end{aligned}
$$

For the single-dose regimen,

$$P_1^d(2nL) = 2n^2L^2 + \frac{4\alpha}{3}n^3L^3.$$

Comparing the efficacy-weighted protection levels,

$$\begin{aligned}
\Delta P^d(2nL) &= \lambda(1-\theta_1)\zeta^d P_1^d(2nL) - \lambda(1-\theta_1)[P_1^r(2nL) + \zeta P_2^r(2nL)] \\
&= \lambda(1-\theta_1)\left\{\zeta^d\left(2n^2L^2 + \frac{4\alpha}{3}n^3L^3\right) - \frac{2n^2+n}{2}L^2 - \alpha L^3\left[\frac{2}{3}n^3 + \frac{1}{2}n(n+1)\right]\right. \\
&\quad \left. -\frac{3}{4}\zeta(2n^2-2n+1)L^2 - \zeta\frac{\alpha L^3}{12}(8n^3-6n^2+6n-6)\right\} \\
&= \lambda(1-\theta_1)L^2\left[2n^2\zeta^d - \frac{2n^2+n}{2} - \frac{3\zeta}{4}(2n^2-2n+1)\right] \\
&\quad + \lambda(1-\theta_1)\frac{\alpha L^3}{12}\left[16n^3\zeta^d - 8n^3 - 6n^2 - 6n - \zeta(8n^3-6n^2+6n-6)\right] \\
&= \lambda(1-\theta_1)L^2\left[2n^2(\zeta^d-1) + \frac{n-1}{2} + \left(\frac{1}{2} - \frac{3\zeta}{4}\right)(2n^2-2n+1)\right] \\
&\quad + \lambda(1-\theta_1)\frac{\alpha L^3}{12}\left[(16\zeta^d-16)n^3 - 12n + 6 + (1-\zeta)(8n^3-6n^2+6n-6)\right].
\end{aligned}$$

In the above, the L^2 term is non-negative if $\zeta^d \geq 1$ and $\zeta \leq \frac{2}{3}$. The L^3 term is non-negative if $\zeta \leq 1$ and $\zeta^d \geq \frac{11}{8}$, which ensures $(16\zeta^d - 16)n^3 \geq 6n^3$. Thus, a sufficient condition for $\Delta P^d(2nL) \geq 0$ is that $\zeta^d \geq \frac{11}{8}$ and $\zeta \leq \frac{2}{3}$.

Q.E.D.

OA2. SEIR Epidemic Model

In this section, we first provide technical details of our SEIR model in Section OA2.1. Next, we provide the rationale for the parameter values used in our baseline case in Section OA2.2.

OA2.1. SEIR Model Description

We consider a variant of the standard SEIR model to incorporate the effect of vaccination on the evolution of the pandemic.

- We consider the population to comprise of a high-risk group (age 65 and above) and a low-risk group (64 or below), indexed by $i \in \{H, L\}$. The initial population size of group i is denoted by N_i.[17]

- At any time t, following the vaccine rollout policy, the high- and low-risk groups can further be divided into three subgroups each: those who have received k doses of the vaccine, where $k = 0, 1, 2$. We refer to individuals in risk group i who have received k vaccine doses as type (i, k).

[17] We thank the Associate Editor for proposing to use the age as the basis for determining the cutoff between two risk groups. We note that under the cutoff of 65, the death rate of the elderly is more than 1,000% higher than the younger group. In addition, the data regarding vaccine efficacy rate, hospitalization, transmission rate, etc. are available for these two groups.

- The structure of the compartmental model follows that of Keskinocak et al. (2020). At any time, a each type (i,k) patient is in one of the following states: susceptible (S), exposed (E), infectious-presymptomatic (IP), infectious-asymptomatic (IA), infectious-symptomatic (IS), hospitalized (H), recovered (R), or deceased (D). See Figure 1 in Section 7.1 for a graphical illustration.
- We assume that vaccination reduces susceptibility, infectiousness, probability of symptomatic disease, probability of hospitalization and probability of death.

Infection. Similar to Bubar et al. (2020) and consistent with the notion of contact discussed by Diekmann and Heesterbeek (2000), a susceptible person of type (i,k) faces a rate of exposure expressed by

$$\lambda_{ik}^{E} = u_i \theta_k \sum_{j\in\{H,L\}} c_{ij} \frac{\sum_{l=0}^{2} \phi_l (IS_{jl} + \delta^P IP_{jl} + \delta^A IA_{jl})}{N_j - \Omega_j},$$

where u_i is the probability of virus transmission given contact with an infectious individual and c_{ij} is the expected number of contacts between a group i and a group j individual (per unit time). The adjustment factors δ^P and δ^A reflects the lower infectiousness of presymptomatic and asymptomatic patients. The parameters θ_k and ϕ_k reflect susceptibility (of an uninfected individual) and infectiousness (of an infected individual) after receiving k doses of the vaccine ($k \in \{0,1,2\}$). Naturally, $0 \le \theta_2 \le \theta_1 \le \theta_0 = 1$ and $0 \le \phi_2 \le \phi_1 \le \phi_0 = 1$, i.e., vaccination reduces susceptibility and infectiousness. The variables IP_{jl}, IS_{jl}, and IA_{jl} denote the type (j,l) individuals who are in the infectious presymptomatic, symptomatic and asymptomatic states, respectively, and Ω_j denotes the cumulative number of deceased individuals in risk group j. We assume that hospitalized individuals are isolated and do not transmit the virus to others. The term $\frac{\sum_{l=0}^{2} \phi_l (IS_{jl} + \delta^P IP_{jl} + \delta^A IA_{jl})}{N_j - \Omega_j}$ can be interpreted as the probability that a randomly-encountered group-j person is infectious.

Disease Severity. Let T^E and T^P be the average time that an infected patient spends in the exposed and infected-presymptomatic states, respectively.

At the end of the presymptomatic state, a group i patient develops symptoms with probability η_i^S. In this case, the patient spends an average of T_i^S in the infected-symptomatic state, at the end of which the patient can either become hospitalized (with probability η_i^H) or recovered.

To reflect how the fatality rate depends on the load of the healthcare system (e.g., ICU utilization), we consider a two-step function of death rates. In particular, if the total number of hospitalizations remains lower than a threshold K (effective capacity of the healthcare system), a hospitalized group i patient transitions to the deceased state at the base probability of η_i^D; if the number of hospitalized individuals exceeds K, the probability becomes $\kappa\eta_i^D$, where $\kappa > 1$.

Similar to Keskinocak et al. (2020), we assume that all fatalities go through the hospitalized (H) state, i.e., a patient cannot transition directly from infected-symptomatic (IS) to deceased

(D). Thus, one can interpret the H state in our model as cases of severe disease that *require* hospitalization, although not all patients in this state may be admitted to hospital in reality due to factors such as access or capacity. Nevertheless, all such cases constitute the potential load on the healthcare system and are thus counted toward the capacity threshold for higher death rates.

On the other hand, if a patient becomes asymptomatic (milder disease) following the presymptomatic stage, the patient recovers after an average time of T^A.

To reflect the vaccine's effect on disease severity, we consider that group i patients who have received the k-th dose will see their probabilities of symptomatic disease, hospitalization and death (uncapacitated) drop to $\tau_k\eta_i^S$, $\gamma_k\eta_i^H$ and $\rho_k\eta_i^D$, respectively. Similarly, once the capacity threshold for hospitalization is reached, the death probability rises to $\rho_k\kappa\eta_i^D$.

Vaccination Rollout. The external vaccine supply process is specified as follows. The supply rate $y(t)=0$ for $t\in[0,50)$; thereafter, we have $y(t)=y(0)+\alpha t$ for $t\geq 50$, where $y(0)=0.001$ (supply at day 50 equals 0.1% of the population) and $\alpha=0.02y(0)$ (2% increase in supply rate per day).

At time t, following the vaccination rates determined from the chosen rollout policy (as discussed in Sections 4 to 6), $v_1(t)$ and $v_2(t)$ doses are made available for first and second doses. We assume that these are proportionally allocated to the H and L groups; that is, the rate at which type (i,k) patients are vaccinated is denoted by

$$v_{ik}(t)=\begin{cases}\pi\cdot v_k(t) & \text{if } i=H\\ (1-\pi)\cdot v_k(t) & \text{if } i=L\end{cases}.$$

With a finite population and the probability of mortality, it is possible that the allocated vaccination rates exceed the number of remaining eligible individuals (particularly in group H). In such case, the excess supplies are reallocated to group L.

Consistent with practice, we assume that all individuals without symptoms (i.e., the S, E, IP, IA, and R compartments) have uniform probability of receiving the vaccine allocated to the risk group i. As an approximation, we do not consider a minimum time gap between recovery and eligibility to receive the vaccine.

Differential Equation Formulation. Based on the assumptions discussed above, the SEIR model can be formulated as follows (suppressing (t) for brevity):

$$\begin{aligned}S'_{ik}=&-S_{ik}\lambda^E_{ik}+\frac{v_{i,k-1}S_{i,k-1}}{S_{i,k-1}+E_{i,k-1}+IP_{i,k-1}+IA_{i,k-1}+R_{i,k-1}}\\&-\frac{v_{ik}S_{ik}}{S_{ik}+E_{ik}+IP_{i,k}+IA_{i,k}+R_{ik}}\end{aligned}\tag{OA4}$$

$$\begin{aligned}E'_{ik}=&\,S_{ik}\lambda^E_{ik}-E_{ik}\frac{1}{T^E}+\frac{v_{i,k-1}E_{i,k-1}}{S_{i,k-1}+E_{i,k-1}+IP_{i,k-1}+IA_{i,k-1}+R_{i,k-1}}\\&-\frac{v_{ik}E_{ik}}{S_{i,k}+E_{ik}+IP_{i,k}+IA_{i,k}+R_{i,k}}\end{aligned}\tag{OA5}$$

$$IP'_{ik} = E_{ik}\frac{1}{T^E} - IP_{ik}\frac{1}{T^P} + \frac{v_{i,k-1}IP_{i,k-1}}{S_{i,k-1}+E_{i,k-1}+IP_{i,k-1}+IA_{i,k-1}+R_{i,k-1}} - \frac{v_{ik}IP_{ik}}{S_{i,k}+E_{i,k}+IP_{i,k}+IA_{i,k}+R_{i,k}} \tag{OA6}$$

$$IA'_{ik} = IP_{ik}\frac{1-\tau_k\eta_i^S}{T^P} - IA_{ik}\frac{1}{T^A} + \frac{v_{i,k-1}IA_{i,k-1}}{S_{i,k-1}+E_{i,k-1}+IP_{i,k-1}+IA_{i,k-1}+R_{i,k-1}} - \frac{v_{ik}IA_{ik}}{S_{i,k}+E_{i,k}+IP_{i,k}+IA_{i,k}+R_{i,k}} \tag{OA7}$$

$$IS'_{ik} = IP_{ik}\frac{\tau_k\eta_i^S}{T^P} - IS_{ik}\frac{1}{T^S} \tag{OA8}$$

$$H'_{ik} = IS_{ik}\frac{\gamma_k\eta_i^H}{T^S} - H_{ik}\frac{1}{T^H} \tag{OA9}$$

$$D'_{ik} = \begin{cases} H_{ik}\frac{\rho_k\eta_i^D}{T^H} & \text{if } \sum_{i,k} H_{ik} \le K \\ H_{ik}\frac{\rho_k\kappa\eta_i^D}{T^H} & \text{if } \sum_{i,k} H_{ik} > K \end{cases} \tag{OA10}$$

$$R_{ik} = \begin{cases} IA_{ik}\frac{1}{T^A} + IS_{ik}\frac{1-\gamma_k\eta_i^H}{T^S} + H_{ik}\frac{1-\rho_k\eta_i^D}{T^H} \\ \quad + \frac{v_{i,k-1}R_{i,k-1}}{S_{i,k-1}+E_{i,k-1}+IP_{i,k-1}+IA_{i,k-1}+R_{i,k-1}} - \frac{v_{ik}R_{ik}}{S_{i,k}+E_{i,k}+IP_{i,k}+IA_{i,k}+R_{i,k}} & \text{if } \sum_{i,k} H_{ik} \le K \\ IA_{ik}\frac{1}{T^A} + IS_{ik}\frac{1-\gamma_k\eta_i^H}{T^S} + H_{ik}\frac{1-\rho_k\kappa\eta_i^D}{T^H} \\ \quad + \frac{v_{i,k-1}R_{i,k-1}}{S_{i,k-1}+E_{i,k-1}+IP_{i,k-1}+IA_{i,k-1}+R_{i,k-1}} - \frac{v_{ik}R_{ik}}{S_{i,k}+E_{i,k}+IP_{i,k}+IA_{i,k}+R_{i,k}} & \text{if } \sum_{i,k} H_{ik} > K. \end{cases} \tag{OA11}$$

OA2.2. Parameter Calibration for the SEIR model

In Table OA1, we provide the rationale for the choice of parameter values used in the simulation of our baseline case (Section 7.2).

References

Bubar KM, Reinholt K, Kissler SM, Lipsitch M, Cobey S, Grad YH, Larremore DB (2021) Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Sci.* 371(6532) 916–921.

Centers for Disease Control and Prevention (CDC) (2021) COVID-19 pandemic planning scenarios. Accessed May 10, 2021, https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html.

Diekmann, O., J. A. P. Heesterbeek. 2000. *Mathematical Epidemiology of Infectious Diseases* (John Wiley & Sons, West Sussex, England).

Haas EJ, Angulo FJ, McLaughlin JM, Anis E, Singer SR, Khan F, Brooks N, et al. (2021) Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* 397(10287):1819–1829.

Harris RJ, Hall JA, Zaidi A, Andrews NJ, Dunbar JK, Dabrera G (2021) Impact of vaccination on household transmission of SARS-COV-2 in England. Accessed May 16, 2021, https://khub.net/documents/135939561/390853656/Impact+of+vaccination+on+household+transmission+of+SARS-COV-2+in+England.pdf/35bf4bb1-6ade-d3eb-a39e-9c9b25a8122a?t=1619551571214.

Table OA1 Parameter Values Used in the Baseline Case (Section 7.2)

Parameter	Value	Rationale
$[N_L, N_H]$	[87, 13]	In 2019, 13% of the U.S. population (over 20) was age 65 or above (U.S. Census Bureau 2021). We normalize $N_L + N_H = 100$.
$E_{ik}(0)$	0.01 for $i = H, L$ and $k = 0$; 0 otherwise	$E_{ik}(t)$ is the size of infectious group for type (i,k). We assume the pandemic begins with 0.01% (1 in 10,000) of the population becoming exposed in each of the two groups.
$[u_L, u_H]$	[0.84, 0.84]	u_i is the base probability of virus transmission given contact. We take the mid-point of values calibrated for above-20 age groups in Bubar et al. (2021).
$\begin{bmatrix} c_{LL} & c_{HL} \\ c_{LH} & c_{HH} \end{bmatrix}$	$\begin{bmatrix} 0.841 & 0.071 \\ 0.258 & 0.318 \end{bmatrix}$	The number of group-j individuals contacted by an group-i individual per day. We first regroup the U.S. contact matrix from Prem et al. (2017) into the 20-64 ($i = L$) and 65+ ($i = H$) bins. Then, the matrix is rescaled such that the reproductive coefficient is 1.5.
$[\delta^P, \delta^A]$	[0.48, 0.48]	Transmissibility adjustment factors in pre/asymptomatic states (Keskinocak et al. 2020).
T^E	4.6 days	The average duration of exposed state (Keskinocak et al. 2020).
T^P	0.5 days	The average duration of presymptomatic state (Keskinocak et al. 2020).
T^S	2.9 days	The average duration of symptomatic state (Keskinocak et al. 2020).
T^H	10.4 days	The average duration of hospitalization (Keskinocak et al. 2020).
T^A	1.93 days	The average duration of asymptomatic state, computed using the symptomatic-asymptomatic duration ratio of 1.5 (Keskinocak et al. 2020).
$[\eta_L^S, \eta_H^S]$	[0.431, 0.653]	The probability of symptomatic given infection (from Israel's nationwide rollout, Haas et al. 2021).
$[\eta_L^H, \eta_H^H]$	[0.104, 0.593]	The probability of hospitalization given symptomatic (from Israel's nationwide rollout, Haas et al. 2021).
$[\eta_L^D, \eta_H^D]$	[0.043, 0.303]	The probability of death given hospitalization (from Israel's nationwide rollout, Haas et al. 2021).
K	0.108	On average, about 25% of hospitalized patients require ICU admission (CDC 2021). The U.S. ICU capacity is 0.027 beds per 100 population (Kaiser Family Foundation 2021), which translates into $0.027/0.25 = 0.108$ effective hospitalizations.
κ	1.92	When ICU capacity is reached, fatality rates increase by 92% (Wilde et al. 2021).
$[\theta_1, \theta_2]$	[0.423, 0.047] for Pfizer, [0.34, 0.34] for Johnson & Johnson	One minus the first- and second-dose efficacy rates of the Pfizer vaccine (from Israel's nationwide rollout, Haas et al. 2021)
$[\phi_1, \phi_2]$	[0.57, 0.57]	The Pfizer vaccine reduces transmissibility by 43% (from study by Public Health England, Harris et al. 2021).
$[\tau_1, \tau_2]$	[0.929, 0.884]	Reduction factors for the conditional probability of symptomatic disease given infection (from Israel's nationwide rollout, Haas et al. 2021) with vaccination.
$[\gamma_1, \gamma_2]$	[0.901 2.477]	Reduction factors for the conditional probability of hospitalization given symptomatic disease (from Israel's nationwide rollout, Haas et al. 2021) with vaccination. The second dose increases the *conditional* probability, despite reducing the overall probability of hospitalization.
$[\rho_1, \rho_2]$	[1.077, 1.799]	Reduction factors for the conditional probability of death given hospitalization (from Israel's nationwide rollout, Haas et al. 2021) with vaccination. Vaccine increases the *conditional* probabilities, despite reducing the overall probabilities of death.
π	1 (except in Section 7.4, where its value is 0.5)	The fraction of vaccine supply allocated to H group is 100%, following age-based prioritization (e.g., in the U.S. and U.K.). We consider the effect of prioritization by considering allocating only half of supplies to H group in in Section 7.4.

Kaiser Family Foundation (2021) State health facts – ICU beds. Accessed May 10, 2021, https://www.kff.org/other/state-indicator/icu-beds/.

Keskinocak P, Oruc BE, Baxter A, Asplund J, Serban N (2020) The impact of social distancing on COVID19 spread: State of Georgia case study. *PLoS ONE* 15(10):e0239798.

Prem K, Cook AR, Jit M (2017) Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* 13(9):e1005697.

U.S. Census Bureau (2021) U.S. and world population clock. Accessed May 10, 2021, https://www.census.gov/popclock/.

Wilde H, Mellan T, Hawryluk I, Dennis JM, Denaxas S, Pagel C, Duncan A, et al. (2021) The association between mechanical ventilator availability and mortality risk in intensive care patients with COVID-19: A national retrospective cohort study. *medRxiv* 2021.01.11.21249461.