OPTIMAL VACCINE SUBSIDIES FOR ENDEMIC AND EPIDEMIC DISEASES

Matthew Goodkin-Gold
Michael Kremer
Christopher M. Snyder
Heidi L. Williams

Optimal Vaccine Subsidies for Endemic and Epidemic Diseases
Matthew Goodkin-Gold, Michael Kremer, Christopher M. Snyder, and Heidi L. Williams
NBER Working Paper No. 28085
November 2020
JEL No. D4,I18,L11,L65,O31

ABSTRACT

Vaccines exert a positive externality, reducing spread of disease from the consumer to others, providing a rationale for subsidies. We study how optimal subsidies vary with disease characteristics by integrating a standard epidemiological model into a vaccine market with rational economic agents. In the steady-state equilibrium for an endemic disease, across market structures ranging from competition to monopoly, the marginal externality and optimal subsidy are non-monotonic in disease infectiousness, peaking for diseases that spread quickly but not so quickly as to drive all consumers to become vaccinated.

Motivated by the Covid-19 pandemic, we adapt the analysis to study a vaccine campaign introduced at a point in time against an emerging epidemic. While the nonmonotonic pattern of the optimal subsidy persists, new findings emerge. Universal vaccination with a perfectly effective vaccine becomes a viable firm strategy: the marginal consumer is still willing to pay since those infected before vaccine rollout remain a source of transmission. We derive a simple condition under which vaccination exhibits increasing social returns, providing an argument for concentrating a capacity-constrained campaign in few regions. We discuss a variety of extensions and calibrations of the results to vaccines and other mitigation measures targeting existing diseases.

Matthew Goodkin-Gold
Department of Economics
Harvard University
Littauer Center
1805 Cambridge Street
Cambridge, MA 02138
United States
mgoodkingold@g.harvard.edu

Michael Kremer
University of Chicago
Kenneth C. Griffin Department of Economics
1126 E. 59th St.
Chicago, IL 60637
and NBER
kremermr@uchicago.edu

Christopher M. Snyder
Department of Economics
Dartmouth College
301 Rockefeller Hall
Hanover, NH 03755
and NBER
chris.snyder@dartmouth.edu

Heidi L. Williams
Department of Economics
Stanford University
579 Serra Mall
Office 323
Stanford, CA 94305
and NBER
hlwill@stanford.edu

1. Introduction

Technologies that help prevent infectious diseases such as vaccines, condoms, and mosquito nets can generate positive health externalities. While standard economic models provide a justification for public subsidies of such preventive technologies, existing models provide little guidance on the appropriate magnitude of these subsidies. This gap in understanding makes it difficult for economists to provide guidance—even at a conceptual level—on both the optimal level of government subsidies for infectious disease control and the ways in which the level of such subsidies should vary across diseases. Furthermore, most existing work by economists has been oriented toward endemic diseases and policies which play out over years or decades such as vaccination campaigns to eradicate polio or circumcision to reduce the spread of HIV. However, the Covid-19 pandemic has underscored the urgency of understanding short-run policy responses to the epidemics that can result from the emergence of novel diseases or from localized outbreaks of otherwise dormant diseases such as Ebola or MERS.

To address these questions, we construct a tractable model integrating epidemiological and economic considerations. For concreteness, the discussion is focused on the market for a vaccine, but the analysis applies to any technology (circumcision, bed nets, social distancing) preventing consumers from contracting a disease. Consumers and producers base their economic decisions on rational expectations of disease dynamics based on a susceptible-infected-recovered (SIR) model standard in the epidemiology literature. Section 2 adopts a long-run perspective suited to an endemic disease. Consistent with this perspective, we incorporate population turnover into the SIR model and analyze the steady-state equilibrium of the vaccine market. Section 3 shifts to a short-run perspective of a vaccine campaign introduced at a single point in time into an SIR epidemic without population turnover. Absent preventive measures, such an epidemic can rapidly peak before ultimately burning out, leaving some untouched. Consumers in the short-run analysis make vaccination decisions based on their probability of being infected before the disease burns out, while consumers in the long-run analysis base their decisions on the probability of infection in the steady state.

For a continuum of market structures—including perfect competition, monopoly, and Cournot competition nesting these extremes—the long-run analysis generates a closed-form solution for the marginal vaccine externality in steady-state equilibrium as a function of estimable parameters including disease infectiveness, recovery rate, vaccine cost, and efficacy. A key finding is that this marginal externality is nonmonotonic in the disease's basic reproductive ratio $\mathcal{R}_0$, a widely used measure of infectiveness. For sufficiently low values of $\mathcal{R}_0$ (technically, $\mathcal{R}_0 \leq 1$), the disease dies

out in the steady state even without a vaccine, so there is no vaccine market and no marginal externality. For sufficiently high values of $\mathcal{R}_0$, vaccinating a given consumer does not provide much protection to others since they are almost certain to contract the disease from another source anyway. To be sure, a consumer's vaccination provides a substantial social benefit when $\mathcal{R}_0$ is extremely high, but most of that benefit is internalized by the consumer themselves. The marginal externality on others is greatest not for extreme but for intermediate values of $\mathcal{R}_0$. The optimal vaccine subsidy (minimum subsidy needed to obtain the first-best quantity) is likewise nonmonotonic in $\mathcal{R}_0$. Holding constant disease burden (the product of prevalence and harm), our results suggest that the best candidates for vaccine subsides are rarer but more serious diseases.

Our crude calibrations suggest that optimal subsidies may be very large relative to current levels in the case of competitively supplied products. For example, the external benefit from using circumcision to prevent HIV in developing countries could justify subsidies of more than $1,000, while programs studied to date paid participants no more than $15 beyond procedure costs. For products produced by a monopoly selling directly to consumers, a per-dose subsidy may not always be a viable policy option since the minimum subsidy achieving the first best can be enormous: in our measles calibration, for example, fifteen times the harm from actually having the disease. Given plausible values for the social cost of public funds, such subsidies would be prohibitively expensive. Bulk purchase of the optimal quantity coupled with subsidized distribution may be a cheaper route to the first best.

Moving from the long-run to the short-run analysis, the latter is technically more difficult because equilibrium variables no longer have closed-form expressions, requiring the Lambert W function to express analytically. Still, we are able to recover the result from the long-run analysis that the marginal externality and optimal subsidy are nonmonotonic in $\mathcal{R}_0$ for market structures ranging from perfect competition, to Cournot, to monopoly. The featured calibration of the short-run analysis is to the Covid-19 pandemic.

A major point of departure between the long- and short-run analyses regards the viability of universal vaccination as a business strategy. In the long-run analysis, a perfectly effective vaccine would never be universally purchased if sold at a positive price because, with all other consumers protected, the marginal consumer obtains no private benefit. A number of previous game-theoretic analyses of vaccine uptake make exactly this point (Geoffard and Phillipson 1997, May 2000, Bauch and Earn 2004). In our short-run analysis, however, universal vaccination with a perfectly effective vaccine may be a profitable business strategy. The risk of contracting the disease from those infected

2

before the arrival of the vaccine but not yet recovered preserves a positive willingness to pay for the marginal consumer even if all other susceptibles are protected. Not only is universal vaccination with a perfectly effective vaccine a possible equilibrium for some parameters for some market structure, this outcome is guaranteed in equilibrium even under monopoly for sufficiently low cost and sufficiently high infectiousness.

Another point of departure regards capacity constraints. The long run leaves little room for capacity constraints, but they may be relevant in the short run. We can thus apply the short-run analysis to understand how to allocate a limited capacity across states, countries, or other regions with separate epidemiologies, owing, say, to travel restrictions. We show that if the product of $\mathcal{R}_0$ and $\hat{S}_0$ (the susceptible proportion of the population upon vaccine rollout) exceeds 2, there are initially increasing social returns to vaccination, meaning that a small capacity would be more efficiently concentrated in a single region than spread evenly across them. The condition $\mathcal{R}_0\hat{S}_0 > 2$ has intuitive appeal, which can be understood by analogy to forest fires. Infectiousness $\mathcal{R}_0$ promotes disease spread much as prime weather conditions (low moisture, hot temperature, high winds) promote fires' spread; the stock of susceptibles $\hat{S}_0$ acts like the stock of flammable material providing fuel for the fire. For an epidemic to grow initially rather than diminishing from the outset requires a certain level of both factors in the absence of a vaccine, technically $\mathcal{R}_0\hat{S}_0 > 1$. When $\mathcal{R}_0\hat{S}_0$ exceeds the higher threshold of 2, the epidemic is so explosive that a small amount of vaccine does little to slow it. To make a measurable dent in the epidemic requires concentrating supplies in one region. If $\mathcal{R}_0\hat{S}_0$ exceeds a yet higher threshold than 2 that we specify, vaccination exhibits increasing social returns for all capacity levels, meaning that would be efficient to serve all susceptibles in a region before moving to the next region.

Section 4 extends the model to allow for consumer heterogeneity, conducting several exercises. We show that the nonmonotonicity in the marginal externality, observed for both long- and short-run analyses and for all market structures is robust to the addition of consumer heterogeneity (in, say, disease harm). We illustrate how even imprecise targeting can help an otherwise expensive program subsidizing consumers' heterogeneous adoption costs to pass a cost-benefit test. We provide conditions under which a naïve researcher who estimates the marginal externality using aggregate data ignoring heterogeneity in infectiousness across groups can arrive at an unbiased estimate of the average marginal externality across consumers; in other cases, a simple adjustment is all that is needed to remove the bias.

To the various ways in which the vaccine's epidemiological externality impacts the market, Sec-

tion 5 compares the results to a market for a drug is similar in all ways to the vaccine except that it treats symptoms but does nothing to reduce disease spread from treated individuals. We show that a monopolist would always prefer to develop the drug but parameters exist in both the long- and short-run analyses for which social welfare is higher with the vaccine. Consistent with nonmonotonicities found elsewhere, the monopolist's bias toward a drug is greatest for intermediate values of $\mathcal{R}_0$.

While the SIR model we use is standard in the epidemiology literature by design, our novel contribution to that literature is to incorporate an epidemiological model into a welfare economics framework, facilitating the analysis of market equilibrium and optimal policy. Much of the epidemiology literature focuses on characterizing when disease eradication is feasible, whereas our framework admits nontrivial analysis even in settings in which eradication is impossible (long-run equilibrium with a perfectly effective vaccine) or certain (the disease always dies out in our short-run analysis). An important exception is Althouse, Bergstrom, and Bergstrom (2010), who also consider welfare analysis of vaccination, calibrating a simple model for four prominent diseases to estimate optimal subsidies under perfect competition and perfectly effective vaccination. We view our paper as building on their work, allowing for imperfect vaccines, including a supply-side model of firm behavior, and generating comparative statics which allow theoretical insights into how epidemiological and economic parameters impact market outcomes and optimal policy.

The epidemiology literature previously recognized the possibility that the nonlinear nature of epidemics may dictate optimal policy concentrating a scarce stockpile in one population rather spreading across them. Keeling and Shattock (2012) provided an early contribution, subsequently refined by work including Keeling and Ross (2015), Nguyen and Carlson (2016), and Enayati and ʿOzaltin (2020). This literature has the advantage of studying increasingly rich epidemiological models, the results are simulated in numerical examples. We contribute a formal conceptualization of initial and eventual increasing social returns and aid understanding by providing a necessary and sufficient condition for these outcomes in analytical form.

Economists have long observed that vaccines may provide positive externalities that could affect consumers' and firms' decisions (see, among others, Brito, Sheshinski, and Intrilligator 1991; Chen and Toxvaerd 2014; Francis 1997; Geoffard and Philipson 1997; Gersovitz 2003; Gersovitz and Hammer 2004, 2005).[1] Boulier, Datta, and Goldfarb (2007) use a standard epidemiological model alone (i.e., neither interacted with consumer decisions nor a supply-side model of firm behavior)

---

[1]Recent work in behavioral epidemiology implicitly incorporates externalities, considering, for example, game-theoretic analyses of decisions around whether to vaccinate or to free ride on herd immunity. (Funk *et al.* 2010; Manfredi and D'Onofrio 2013)

to examine properties of vaccination externalities that arise solely due to epidemiological concerns. Geoffard and Philipson (1997) use an epidemiological model similar to ours to show that a vaccine producer with market power will not choose to eradicate the disease in the steady-state. Galeotti and Rogers (2013) model vaccination choices in a heterogenous population, and consider the effect of network structures in determining optimal vaccine allocation.[2] Economists have attempted to provide well-identified empirical estimates of vaccine externalities (Cook *et al.* 2009, Ward 2014, Bethune and Korinek 2020). Given the difficulty of using randomized controlled trials to estimate externalities, Manski (2017) provides theoretical guidance on how to optimize the decision to mandate a vaccine when the extent of externality is unknown. We make a number of contributions to this economics literature. First, we incorporate a model of firm behavior and explicitly characterize equilibrium solutions for both positive and normative outcomes—specifically, externalities and optimal subsidies—in terms of estimable parameters. Second, our short-run analysis provides what to our knowledge is the first treatment of vaccination externalities and subsidies oriented to the initial stages of a novel disease or localized outbreak and the first to characterize increasing returns to vaccination in this setting.

2. Long-Run Analysis of Endemic Disease

This section analyzes the market for vaccine against an endemic disease such as HIV that persists in the population over the long run. We begin by outlining a standard epidemiological model into which we introduce an economic model of the vaccine market. Reflecting our long-run perspective, we focus on the steady-state equilibrium. We solve for the equilibrium under various market structures ranging from perfect competition to monopoly, analyzing comparative statics, welfare implications, and optimal subsidies.

2.1. Epidemiological Model

The foundation of our analysis is the standard susceptible-infected-removed (SIR) epidemiological model due to Kermanck and McKendrick (1927). See Martcheva (2015) for a textbook treatment of the basic model as well as the enhanced version with vaccination and demographic flows (births and deaths) that we specifically employ. Time is continuous, indexed by t. Let $\mu \in (0, 1)$ denote the

---

[2]Mecholan (2007) provides some analysis of treatments (conditional on infection) for communicable diseases in the context of a monopoly manufacturer, but provides no analytical results, instead focusing on numerical simulations, primarily related to issues of drug resistance.

mortality rate from causes other than the disease, referred to as "natural causes." For simplicity, we will study the case of a non-fatal disease and set the birth rate to equal to μ, leaving the population size constant over time. Anticipating their role in the vaccine market, we call members of this population "consumers." Normalize the mass of consumers to 1. Consumers are partitioned into four compartments: susceptible to infection S_t, currently infected I_t, recovered from an infection R_t, or immunized V_t.

Since the compartments partition the mass of consumers (i.e., the compartments are mutually exclusive and exhaustive),

$$S_t + I_t + R_t + V_t = 1. \tag{1}$$

With the total mass on the right-hand of (1) normalized to 1, the compartments on the left-hand side can be interpreted as either masses or proportions.

Turn next to the equations determining the evolution of compartments over continuous time, starting with V_t. Let $Q \geq 0$ be the quantity of vaccine purchased each instant. For now take Q as given; later, we will solve for its equilibrium value using the economic model and substitute this value back into the epidemiological model. In principle, Q could vary over time, but we omit a time subscript anticipating that we will solve for its equilibrium value in the steady state. Let $\theta \in (0, 1)$ denote vaccine efficacy. Assume that vaccinations are adminstered to newborns[3] and that if the initial dose is not effective for a person, further doses will not be effective for that person either. The mass of newborns that are immunized (i.e., successfully vaccinated) each instant is θQ. The rate of change of the immunized population, $\partial V_t / \partial t \equiv \dot{V}_t$, is

$$\dot{V}_t = \theta Q - \mu V_t, \tag{2}$$

the number of newborns who are immunized minus the number of immunized individuals who die (of causes other than the disease).

The rate of change of the infected population is

$$\dot{I}_t = \beta I_t S_t - (\alpha + \mu) I_t. \tag{3}$$

The first term captures the flow of new infections. A susceptible consumer contracts the disease

---

[3]Given the Poisson structure of the model, and hence the stationarity of consumers' life cycles, assuming vaccines are adminstered only to newborns is without loss of generality; we could equivalently have assumed that the vaccine is administered to any subset of susceptible consumers who have not yet been vaccinated.

from an infected consumer at rate $\beta > 0$, embodying the rate of contact between people and the rate at which a contact leads to infection. Assuming the infection rate is linear in the number of infected consumers, a single susceptible consumer is infected with probability βI_t, and the mass of susceptibles generates $\beta I_t S_t$ new infections. The mass of infected consumers is reduced by the mass αI_t of them who recover, where $\alpha \in (0,1)$ denotes the recovery rate, and by the number of infected consumers who die. Consistent with our assumption that the disease is non-fatal, the mortality rate is the same μ for infected consumers as for other compartments.

Assume that recovered people cannot be reinfected. Then the rate of change of the recovered population is

$$\dot{R}_t = \alpha I_t - \mu R_t, \tag{4}$$

the number of infected people who recover that instant minus the number of recovered vaccinated individuals who die.

Newborns enter the population at rate as susceptibles. To maintain a stationary population size, the birth rate must equal the mortality rate, μ. A fraction θQ of newborns are immunized at birth. Mass $\beta I_t S_t$ of susceptibles become infected each instant, and a further μS_t die. Thus, the rate of change of the susceptible population is

$$\dot{S}_t = \mu - \theta Q - \beta I_t S_t - \mu S_t. \tag{5}$$

Because the compartment masses must sum to 1, one of the equations (2)–(5) is redundant given the others. For instance, (5) can be derived by substituting from (1)–(4) into the derivative of (1) with respect to t.

That completes the specification of the SIR model. In lieu of the transmission parameter β, epidemiologists often work with a related parameter called the basic reproductive ratio $\mathcal{R}_0$, traditionally defined as the expected number of secondary infections generated by adding an infected individual to a fully susceptible population.[4] One can see that the disease eventually dies out in an unvaccinated population if $\mathcal{R}_0 < 1$ and remains endemic if $\mathcal{R}_0 > 1$.

In our model,

$$\mathcal{R}_0 = \frac{\beta}{\alpha + \mu}. \tag{6}$$

To understand this expression, each instant the individual remains infected, he or she infects a num-

---

[4]The modern definition of $\mathcal{R}_0$ due to Dikemann, Heersterbeek, and Metz (1990) is the dominant eigenvalue of the next-generation operator in the epidemiological system. Martcheva (2015, p. 51) shows that equation (6) provides the value of $\mathcal{R}_0$ implied by this definition.

ber of others equal to β times the size of the susceptible population, which is approximately 1 over the period since the infected individual is introduced into a fully susceptible population. The individual remains infected for an expected duration of $1/(\alpha+\mu)$. To see this, note that there are two competing risks for exiting the infected state: recovery, which has hazard $\lambda_R(t) = \alpha$, and mortality, which has hazard $\lambda_M(t) = \mu$. The combined hazard of exiting the infected state is $\lambda_{EI}(t) = \lambda_R(t) + \lambda_M(t)$. As is well known for Poisson duration models, the duration of a spell equals the reciprocal of the hazard, here $1/\lambda_{EI}(t) = 1/(\alpha+\mu)$.

The subsequent analysis takes $\mathcal{R}_0$ as the key exogenous parameter, capturing the disease's infectiveness. Estimates of $\mathcal{R}_0$ vary considerably across disease—from 1.1 for SARS at the low end to 16–18 for measles and pertussis at the high end—as well as across time and region. A systematic review of $\mathcal{R}_0$ estimates is deferred to Section 2.9, where we calibrate our results to the parameters of a variety of actual diseases.

Let variables without time t subscripts (S, I, R, V, Q, etc.) denote steady-state values. For example, $S = \lim_{t\uparrow\infty} S_t$. For a given value of Q, the steady-state values of the compartments can be found by solving the system of equations formed by setting $\dot{V}_t = \dot{I}_t = \dot{R}_t = \dot{S}_t = 0$ in equations (2)–(5). The unique stable solution is[5]

$$S(Q) = \min\left(1 - \frac{\theta}{\mu}Q, \frac{1}{\mathcal{R}_0}\right) \tag{7}$$

$$I(Q) = \max\left[0, \frac{\mu}{\alpha+\mu}\left(1 - \frac{1}{\mathcal{R}_0} - \frac{\theta}{\mu}Q\right)\right] \tag{8}$$

$$R(Q) = \frac{\alpha}{\mu}I(Q) \tag{9}$$

$$V(Q) = \frac{\theta}{\mu}Q. \tag{10}$$

By equation (8), if $\mathcal{R}_0 \leq 1$, then $I = 0$ for all $Q \geq 0$. The disease dies out in the steady state with or without a vaccine, confirming a claim from the previous subsection. For a non-trivial vaccine market to exist, $\mathcal{R}_0 > 1$. In that case, whether the disease dies out in the steady state depends on Q: $I(Q) = 0$ if and only if $Q \geq (\mu/\theta)(1 - 1/\mathcal{R}_0)$. Combining the various cases in a way that will be convenient for the subsequent analysis, we can rewrite

$$I(Q) = \begin{cases} 0 & Q \geq \mathcal{Q}_0 \\ \frac{\mu}{\alpha+\mu}\left(1 - \frac{1}{\mathcal{R}_0} - \frac{\theta}{\mu}Q\right) & Q < \mathcal{Q}_0, \end{cases} \tag{11}$$

---

[5]A trivial solution involving $I = 0$ always exists, but it is unstable when $\mathcal{R}_0 > \mu/(\mu - \theta Q)$.

where

$$\mathcal{Q}_0 = \max \left[0, \frac{\mu}{\theta} \left(1 - \frac{1}{\mathcal{R}_0} \right) \right] \tag{12}$$

is the threshold vaccine quantity above which the disease dies out and below which the disease remains endemic.

2.2. Consumer Demand

Having specified the epidemiological features of the model, we next turn to economic features of the vaccine market, starting with demand. Assume consumers are risk neutral. For now, assume in addition that consumers are homogeneous; later sections and appendices will extend the results to different dimensions of consumer heterogeneity. To avoid computing transition paths, focus on the steady-state in the limiting case without discounting.

A consumer demands the vaccine if his or her marginal private benefit $MPB(Q)$ from the vaccine exceeds its price. For a risk-neutral consumer, $MPB(Q) = \theta H \Phi_I(Q)$, where efficacy parameter θ is the probability that the vaccine allows the consumer to avoid harm, H denotes the expected harm over the infection's spell conditional on being infected, and $\Phi_I(Q)$ denotes the probability that a susceptible person becomes infected.

To compute H, harm measured as a stock over the infection spell, let h be flow harm to a consumer each instant he or she is infected. The consumer experiences that flow harm from the initial date of infection until he or she exits the infected state. Section 2.1 showed that the expected duration of the infected state equals $1/(\alpha + \mu)$. Hence,

$$H = \frac{h}{\alpha + \mu}. \tag{13}$$

To compute $\Phi_I(Q)$, consider the competing risks that can lead a susceptible to exit that state. A susceptible consumer who has chosen not to be vaccinated faces two remaining competing risks: infection, which has hazard $\lambda_I(t, Q) = \beta I(Q)$ each instant, and mortality from natural causes, which has hazard $\lambda_M(t) = \mu$. The combined hazard of exiting the susceptible state is $\lambda_{ES}(t, Q) = \lambda_I(t, Q) + \lambda_M(t)$. By standard results for competing Poisson risks, the consumer's cumulative risk of exiting the susceptible state by age t is $\Lambda_{ES}(t, Q) = \int_0^t \lambda_{ES}(\tau, Q) d\tau$, probability of surviving as susceptible to age t is $e^{-\Lambda_{ES}(t, Q)}$, and likelihood of exit due to infection is $\phi_I(t, Q) = \lambda_I(t, Q) e^{-\Lambda_{ES}(t, Q)}$. To compute the probability that the consumer experiences an infection at some point over his or her lifetime, we

integrate this cause-specific likelihood:

$$\Phi_I(Q) = \int_0^\infty \phi_I(t,Q)dt = \frac{\beta I(Q)}{\beta I(Q) + \mu} = \frac{(\alpha+\mu)\mathcal{R}_0 I(Q)}{(\alpha+\mu)\mathcal{R}_0 I(Q) + \mu}. \tag{14}$$

Substituting for $I(Q)$ from (11) yields

$$\Phi_I(Q) = \begin{cases} 0 & Q \geq \mathcal{Q}_0 \\ 1 - \dfrac{1}{(1-\theta Q/\mu)\mathcal{R}_0} & Q < \mathcal{Q}_0. \end{cases} \tag{15}$$

The marginal private benefit inherits this branched structure from $\Phi(Q)$:

$$MPB(Q) = \begin{cases} 0 & Q \geq \mathcal{Q}_0 \\ \theta H \left[1 - \dfrac{1}{(1-\theta Q/\mu)\mathcal{R}_0} \right] & Q < \mathcal{Q}_0. \end{cases} \tag{16}$$

It can be verified that the second branch of $MPB(Q)$ is strictly decreasing in Q, ensuring the demand curve derived next is well defined and weakly downward sloping.

Let P be the vaccine's price. Consumers strictly prefer to purchase the vaccine if $P < MPB(Q)$. Substituting the quantity $Q = \mu$ that results when all consumers who are eligible to purchase—all newborns—end up purchasing into the preceding inequality, we have $D(P) = \mu$ if $P < MPB(\mu)$. Consumers strictly prefer not to purchase if $P > MPB(Q)$. Substituting $Q = 0$ into the preceding inequality, we have $D(P) = 0$ if $P > MPB(0)$. For P strictly between $MPB(\mu)$ and $MPB(0)$, some but not all consumers purchase. Given they are homogeneous, consumers must be indifferent between purchasing and not, implying $P = MPB(Q)$. Given they are indifferent, any fraction of them purchasing is an equilibrium; demand is pinned down by the value of Q satisfying (16) when the right-hand side is set equal to P. Demand in that case is given by

$$d(P) = \frac{\mu}{\theta} \left[1 - \frac{1}{(1-P/\theta H)\mathcal{R}_0} \right]. \tag{17}$$

Combining these facts yields the demand curve

$$D(P) = \begin{cases} 0 & P > MPB(0) \\ d(P) & P \in [MPB(\mu), MPB(0)] \\ \mu & P < MPB(\mu). \end{cases} \tag{18}$$

Equivalently, the demand curve is given by $d(P)$ unless this violates a boundary condition $d(P) \in [0, \mu]$, in which case demand is given by the violated boundary.

Demand considerations alone allow us to establish that an important result due to Geoffard and Philipson (1997)—that nontrivial equilibria cannot entail the disease's eradication in the steady state—holds quite generally in our model as well. The result holds independent of market structure and holds whether or not the government subsidizes vaccines.

Proposition 1. *Suppose that the disease is eradicated in steady-state equilibrium of a market with or without government subsidies, i.e., $I^* = 0$. Then either the vaccine is free to consumers ($P^* = 0$) or no vaccine is purchased ($Q^* = 0$).*

The proof is provided in Appendix A. Intuitively, if the disease is eradicated in steady-state equilibrium, then $\Phi_I(Q) = 0$, implying $MPB(Q) = 0$. But consumers with no marginal private benefit will not purchase the vaccine at a positive price.

2.3. Firm Supply

We analyze two different market structures in the text: perfect competition and monopoly. Appendix B provides results from a more general model of Cournot competition among N firms that nests these extremes.

Assume firms produce at constant marginal and average cost $c > 0$ per vaccine course (where a course involves multiple doses when needed to provide immunity). Under perfect competition, vaccine supply is perfectly elastic at price c. Under monopoly, the firm sets a price maximizing industry profit Π, measured as a flow each instant in the steady state.

2.4. Normative Measures

Define the total social benefit from the vaccine in the steady state as $SB(Q) = h[1 - I(Q)]$, the product of flow health benefit h to an individual and the number of healthy individuals $1 - I(Q)$ at any instant in steady-state equilibrium.[6] Substituting from (8) yields

$$
SB(Q) = \begin{cases} (\alpha + \mu)H & Q \geq \mathcal{Q}_0 \\ (\alpha + \theta Q + \mu/\mathcal{R}_0)H & Q \leq \mathcal{Q}_0. \end{cases} \tag{19}
$$

---

[6]A natural alternative measure of social benefit would be $SB(Q) = h[I(0) - I(Q)]$. We use the maximum conceivable infection rate 1 rather than the rate in the absence of a vaccine $I(0)$ as the benchmark against which the equilibrium outcome is compared to generate more natural comparative statics. The alternative measure could lead to the odd conclusion that an increase in $\mathcal{R}_0$ is socially beneficial if the increase in $\mathcal{R}_0$ increases $I(0)$ sufficiently to increase $SB(Q)$.

Let $W(Q)$ denote social welfare on the vaccine market, i.e., the difference between total social benefit and total vaccine production costs:

$$W(Q) = SB(Q) - cQ. \tag{20}$$

Let $MSB(Q) = \partial SB(Q)/\partial Q$ denote marginal social benefit from a vaccine course. Differentiating (19),[7]

$$MSB(Q) = \begin{cases} 0 & Q \geq \mathcal{Q}_0 \\ \theta H & Q < \mathcal{Q}_0. \end{cases} \tag{21}$$

Let $MEX(Q) = MSB(Q) - MPB(Q)$ denote the marginal externality from a vaccine course. Subtracting the formulas provided yields

$$MEX(Q) = \begin{cases} 0 & Q \geq \mathcal{Q}_0 \\ \dfrac{\theta H}{(1 - \theta Q/\mu)\mathcal{R}_0} & Q < \mathcal{Q}_0. \end{cases} \tag{22}$$

A comprehensive analysis of these normative measures is postponed until after we have solved for the equilibrium value of Q, on which the measures depend. Some observations can be made even at this preliminary stage. For $Q \geq \mathcal{Q}_0$, enough vaccine is available to eradicate the disease in the steady state, eliminating the marginal benefit of vaccine. Hence, $MPB(Q) = MSB(Q) = MEX(Q) = 0$ for $Q \geq \mathcal{Q}_0$.

For $Q < \mathcal{Q}_0$, $MSB(Q) = \theta H$, which is the expected benefit of treating the an infected person with a drug that does not prevent disease transmission but has efficacy θ in preventing harmful symptoms. It may be at first puzzling that vaccinating an individual not certain to contract the disease amounts to the same social benefit as treating a certainly infected individual. The puzzle is resolved upon recognizing the positive externality associated with the vaccine's preventing a whole chain of possible infections in the unvaccinated population. Conditional on the vaccine's being effective for individual, an event with probability θ, his or her direct benefit plus the chain of expected external benefits sum exactly to H in the model.

Also remarkable is the fact that, for $Q < \mathcal{Q}_0$, $MSB(Q)$ is a constant that does not vary with the infectiveness of the disease as measured by $\mathcal{R}_0$. When $\mathcal{R}_0$ increases, the vaccinated individual's direct benefit $MPB(Q)$ increases; but, perhaps counterintuitively, the marginal externality $MEX(Q)$

---

[7]The derivative $\partial SB(\mathcal{Q}_0)\partial Q$ does not exist, but the left and right derivatives, respectively $\partial SB(\mathcal{Q}_0)/\partial Q^-$ and $\partial SB(\mathcal{Q}_0)/\partial Q^+$, do. We set $MSB(\mathcal{Q}_0) = \partial SB(\mathcal{Q}_0)/\partial Q^+$.

decreases. An individual's vaccination provides less external benefit to contacts because they are increasingly likely to pick up the disease from someone else. The increase in *MPB*(Q) and decrease in *MEX*(Q) exactly offset so that their sum *MSB*(Q) is independent of $\mathcal{R}_0$.

As a benchmark, we will characterize the first-best quantity Q^{**}, which maximizes social welfare. To accommodate cases in which a possibly open set of quantities maximize social welfare, we take Q^{**} to be the infimum of the set. Comparing the marginal social benefit of a vaccine in (21)—which we see is either 0 or a positive constant—with its marginal social cost—given simply by c—it is immediate that $Q^{**} = 0$ if $c \geq \theta H$. To rule out this trivial case, we assume throughout the remainder of the paper that the disease is harmful enough that the vaccine's social benefit exceeds its cost: $\theta H > c$. Rescaling cost as $\tilde{c} = c/\theta H$, the assumption can be equivalently written

$$\tilde{c} < 1. \tag{23}$$

When (23) holds, we can see from (21) that $Q^{**} = \mathcal{Q}_0$ unless this entails vaccinating more than the flow of newborns each instant, in which case $Q^{**} = \mu$. Thus,

$$Q^{**} = \min(\mathcal{Q}_0, \mu). \tag{24}$$

2.5. Perfectly Competitive Equilibrium

Equilibrium values of steady-state variables will be distinguished with asterisks, with an added subscript indicating the relevant market structure. Thus, for example, P_c^* denotes the perfectly competitive equilibrium price.

Under perfect competition $P_c^* = c$. The remaining equilibrium variables can be computed using straightforward algebra applied to the supplied formulas taking due care to ensure the appropriate branch of each formula is used in the relevant case. The relevant cases turn out to depend on the interval in which $\mathcal{R}_0$ falls.

First consider $\mathcal{R}_0 \leq 1$. Infectiveness is so low in this case—labeled (LR1)—that the disease disappears in the steady state even without a vaccine. The vaccine has no marginal social or private value, and no vaccine is sold.

For values of $\mathcal{R}_0$ just above 1, the disease is infective enough not to disappear in the steady state but not infective enough to justify its purchase at the competitive price. Although the vaccine has a positive marginal social and private benefit, the marginal private benefit is below c even if no other

consumer purchases. The upper bound on $\mathcal{R}_0$ for this case—labeled (LR2)—is determined by the condition $MPB(0) \leq c$, which using (16) yields $\mathcal{R}_0 \leq 1/(1-\tilde{c})$.

For $\mathcal{R}_0$ in an intermediate range—labeled case (LR3)—some but not all newborns purchase. The interior solution for equilibrium quantity Q_c^* can be found by substituting $P_c^* = c$ into demand expression (17). The values of other equilibrium variables can be found by substituting Q_c^* into the relevant equations—(8), (21), etc.

For extreme values of $\mathcal{R}_0$, the disease can conceivably be so infective that all μ newborns purchase at the competitive price. The existence of this case requires an imperfectly effective vaccine, $\theta < 1$. Vaccinating all consumers with a perfectly effective vaccine would eliminate the infection and demand, which would be inconsistent with all newborns purchasing. Enough unsuccessfully vaccinated consumers must remain to generate an infection rate that justifies purchase at the competitive price. The relevant condition determining this case—labeled (LR5)—is $MPB(\mu) \geq c$, implying $\mathcal{R}_0 \geq 1/(1-\theta)(1-\tilde{c})$. Since all consumers purchase in this case, there is no underconsumption distortion; the first best is obtained.

Table 1 reports the steady-state equilibrium values of selected variables under perfect competition as a function of $\mathcal{R}_0$. See Appendix A for formal derivations of table entries. To visualize how the variables in Table 1 vary with $\mathcal{R}_0$, Figure 1 graphs a selection of them, one per panel, as functions of $\mathcal{R}_0$. Focus for now on the dotted curves representing equilibrium under perfect competition. Vaccine quantity Q_c^*, graphed in the first panel, equals 0 in case (LR1) because there is no infection in equilibrium. It continues to be 0 in case (LR2) despite positive disease prevalence because the prevalence is too low for any consumer to justify spending c on it. In case (LR3), we start to see a positive vaccine quantity that increases in $\mathcal{R}_0$ until case (LR4) is reached and all μ purchase.

It is unsurprising that equilibrium quantity is weakly increasing in the infectiveness of the disease measured by $\mathcal{R}_0$. Other equilibrium variables also display expected comparative statics in $\mathcal{R}_0$. MPB_c^* is weakly increasing and W_c^* is weakly decreasing in $\mathcal{R}_0$. It is noteworthy that MPB_c^* levels off at c in case (LR3). Given that some but not all consumers purchase in this case, consumers must be indifferent between purchasing and not, implying that the equilibrium price $P_c^* = c$ must extract the entire marginal private benefit, implying $MPB_c^* = c$ over the whole interval.

Other variables display interesting non-monotonicities. Initially at zero in case (LR1), disease prevalence I_c^* begins to increase in case (LR2) due to the epidemiological effects of the higher $\mathcal{R}_0$ with no vaccine purchases in that case to offset it. In case (LR3), when consumers begin purchasing vaccine, I_c^* reverses course and begins to slope downward in $\mathcal{R}_0$. The direct, epidemiological effect

14

of an increase in $\mathcal{R}_0$ continues to be to increase I_c^*. Working in the oppositive direction is an indirect, economic effect of an increase in $\mathcal{R}_0$, inducing consumers to purchase more vaccine. In (LR3), the indirect effect is strong enough to dominate the direct effect, so that I_c^* decreases in $\mathcal{R}_0$. Mathematically, to maintain the constant marginal private benefit ($MPB_c^* = c$) observed throughout (LR3), the increase in infectiveness $\mathcal{R}_0$ must be offset by the exact reduction in prevalence I_c^* to maintain a constant probability of contracting the disease. In cases (LR4) and (LR5), I_c^* again rises with $\mathcal{R}_0$ because the direct effect of an increase in infectiveness cannot be offset by an increase in Q_c^* given that all newborns are already purchasing in these cases.

The marginal externality MEX_c^* also exhibits an interesting nonmonotonicity. Although MEX_c^* is monotonic over the subinterval ($\mathcal{R}_0 > 1$) for which it is positive, it is nonmonotonic over the entire range of $\mathcal{R}_0$, starting out at 0 for $\mathcal{R}_0 < 1$ and jumping to its highest level for $\mathcal{R}_0$ just above 1. Lower levels of $\mathcal{R}_0$ eliminate the disease in the steady state. Higher levels of $\mathcal{R}_0$ than 1 increase the infectiveness of the disease, but this leads consumers to internalize an increasing share of the social benefit of their own vaccination.

The next proposition, proved in Appendix A, summarizes the comparative-static effects of an increase in $\mathcal{R}_0$ on the steady-state equilibrium under perfect competition observed from Figure 1.

Proposition 2. *Consider the comparative-static effect of $\mathcal{R}_0$ on steady-state equilibrium under perfect competition.*

- *Price and industry profit are constant for $\mathcal{R}_0 > 0$, with $P_c^* = c$ and $\Pi_c^* = 0$.*

- *Q_c^* and MPB_c^* are weakly increasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$.*

- *I_c^* is nonmonotonic, reaching a local maximum $I_c^* = \mu\tilde{c}/(\alpha+\mu)$ at $\mathcal{R}_0 = 1/(1-\tilde{c})$, which is a global maximum if $\tilde{c} > 1-\theta$.*

- *Marginal social benefit is constant for all $\mathcal{R}_0 > 1$, with $MSB_c^* = \theta H$.*

- *The marginal externality is nonmonotonic over $\mathcal{R}_0 > 0$, approaching the supremum $MEX_c^* = \theta H$ as $\mathcal{R}_0 \downarrow 1$.*

- *W_c^* is weakly decreasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$.*

The weak changes in Q_c^*, MPB_c^*, and W_c^* are strict for a nonempty interval of $\mathcal{R}_0$ for each variable.

2.6. Monopoly Equilibrium

Since a monopolist charges a markup above cost, $P_m^* \geq c = P_c^*$, implying $Q_m^* \leq Q_c^*$. Thus, in cases (LR1) and (LR2) in which $Q_c^* = 0$, we have $Q_m^* = 0$. Cases (LR1) and (LR2) are thus trivially

identical across perfect competition and monopoly. In the remaining cases, competitive firms are able to make positive sales at price c. By continuity, the monopolist can make positive sales at some small markup above c, implying $Q_m^* > 0$ for $\mathcal{R}_0$ in case (LR3) and above.

To solve for Q_m^* in these other cases, it is convenient to express the monopoly's maximization problem so that the choice variable is quantity rather than price. Letting $P(Q)$ denote inverse demand, monopoly profit can be written $\Pi = [P(Q) - c]Q = [MPB(Q) - c]Q$, where the second equality follows since the monopolist optimally charges a price extracting the private benefit of the marginal consumer. The monopoly maximizes profit subject to the constraint that no more than the population of newborn consumers can be served: $Q \leq \mu$. Applying the Kuhn-Tucker method yields the following solution. If $\mathcal{R}_0 < 1/(1-\theta)^2(1-\tilde{c})$, then the constraint $Q \leq \mu$ does not bind, yielding solution

$$Q_m^* = \frac{\mu}{\theta}\left[1 - \sqrt{\frac{1}{(1-\tilde{c})\mathcal{R}_0}}\right]. \tag{25}$$

Otherwise, the constraint holds with equality, implying $Q_m^* = \mu$. For the constraint to bind requires the vaccine to be imperfectly effective ($\theta < 1$). The monopolist would never sell a perfectly effective vaccine to all consumers because this would eradicate the disease, leaving the monopolist with zero steady-state profit according to Proposition 1.

Substituting Q_m^* into the formulas supplied for the other variables yields the entries in Table 2. The analysis is straigthforward, so we omit formal verification.

To visualize how the variables in Table 2 vary with $\mathcal{R}_0$, focus on the solid curves in Figure 1. The two market structures overlap in cases (LR1) and (LR2), neither generating any vaccine output. The two market structures overlap again in (LR5), both generating the first-best quantity $Q^{**} = \mu$. In between—in (LR3) and (LR4)—the two market structures diverge, with monopoly generating strictly lower output. The strictly lower output under monopoly in these cases leads to a weakly higher infection rate and marginal private benefit and weakly lower welfare under monopoly. Perhaps counterintuitively, the marginal externality is also lower under monpoly. This follows from the fact that $MPB_m^* \geq MPB_c^*$, implying $MEX_m^* = MSB_m^* - MPB_m^* \leq MSB_c^* - MPB_c^* = MEX_c^*$ since $MSB_m^* = MSB_c^* = \theta H$.

While it is not surprising that welfare is lower under monopoly than competition as this is true in typical markets, our model shuts down the typical channel for monopoly deadweight loss by taking consumers to be homogeneous. In our model, the epidemiological externality confers market power: starting from a price that extracts purchasers' entire marginal private benefit, a positive albeit reduced fraction of consumers will continue to purchase at a higher price since the reduction

16

in vaccine quantity increases their marginal private benefit through an increase in disease prevalence. The monopolist's exercise of this market power generates deadweight loss. The large gap between W_c^* and W_m^* for an intermediate range of $\mathcal{R}_0$ in the bottom panel of Figure 1 suggests that monopoly distortions may be worst for moderate levels of infectiveness. Market power generates little welfare loss for the lowest values of R_0 and—with an imperfect vaccine—for the highest values of $\mathcal{R}_0$.

Figure 1 illustrates the consequences of Proposition 1 for the market structures we study. Under both perfect competition and monopoly, the disease has zero prevalence if and only if $\mathcal{R}_0$ falls in case (LR1). Consistent with Proposition 1, no vaccine is purchased in (LR1) under both market structures.

The comparative statics for Q^*, MPB^*, and MEX^* are similar across the two market structures, but the comparative statics for I^* are considerably more complex for monopoly than perfect competition. The direct, epidemiological effect leading an increase in infectiveness $\mathcal{R}_0$ to increase prevalence I_m^* is still present as is the indirect, economic effect of increasing vaccine purchases. However, the monopolist can leverage its market power to absorb some of the indirect effect in the form of price increases. The balance between the two forces need not tip in the same direction as under perfect competition. As the next proposition indicates, states, for some parameters, I_m^* is decreasing throughout cases (LR3) and (LR4); for other parameters, I_m^* is increasing throughout (LR3) and (LR4); and for yet other parameters—including those chosen for Figure 1, I_m^* is nonmonotonic, attaining a local maximum at a point such as A in the figure.

The graph of W^* under monopoly illustrates the remarkable possibility that increasing $\mathcal{R}_0$ can increase welfare. One would think that society would always be harmed by an increase in infectiveness. While the direct, epidemiological effect of an increase in $\mathcal{R}_0$ harms society, the indirect effect of increasing vaccinations can counteract the direct effect, increasing welfare over some parameter ranges. In the bottom panel of Figure 1, we see this possibility emerging for the interval of $\mathcal{R}_0$ between point C and the boundary of (LR4). Under monopoly, not only do consumers fail to consider the external benefit their vaccination provides other consumers, but the monopolist compounds this by placing negative value on consumption to the extent it reduces others' willingness to pay for a vaccine. Mitigating this compounded underconsumption problem via an increase in $\mathcal{R}_0$ can provide such a large indirect benefit that it swamps the direct harm from an increase in $\mathcal{R}_0$, leading to an increase in social welfare.

The next proposition, proved in Appendix A, summarizes the comparative-static effects of an increase in $\mathcal{R}_0$ on the steady-state equilibrium under monopoly observed from Figure 1.

17

Proposition 3. *Consider the comparative-static effect of $\mathcal{R}_0$ on steady-state equilibrium under monopoly.*

- *Q_m^*, MPB_m^*, and Π_m^* are weakly increasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$.*

- *A sequence of equilibrium prices exists for which P_m^* is weakly increasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$.*

- *If $\tilde{c} < 1/2$ and $\tilde{c} \leq (1-2\theta)/2(1-\theta)$, then I_m^* is weakly increasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$. Otherwise, I_m^* is nonmonotonic, reaching a local maximum at $\mathcal{R}_0 = 1/(1-\tilde{c})$ if $\tilde{c} \geq 1/2$, which is a global maximum if $\tilde{c} \geq 1-\theta$, and reaching a local maximum at $\mathcal{R}_0 = 4(1-\theta)$ if $\tilde{c} < 1/2$, which is a global maximum if $\tilde{c} \geq (3-4\theta)/4(1-\theta)$.*

- *Marginal social benefit is constant for all $\mathcal{R}_0 > 1$, with $MSB_m^* = \theta H$.*

- *The marginal externality is nonmonotonic for $\mathcal{R}_0 > 0$, approaching the supremum $MEX_m^* = \theta H$ as $\mathcal{R}_0 \downarrow 1$.*

- *If $\theta \leq 1/4$, then W_m^* is weakly decreasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$. Otherwise, W_m^* is nonmonotonic in $\mathcal{R}_0$, reaching a local minimum at $\mathcal{R}_0 = 4/(1-\tilde{c})$, which is a global minimum if $\theta \geq 3/4$.*

The weak changes in Q_c^*, MPB_c^*, Π_m^*, and P_m^* are strict for a nonempty interval of $\mathcal{R}_0$ for each variable.

To help parse the complicated conditions behind the nonmonotonicity of I_m^* and W_m^*, the next proposition provides simple sufficient conditions for those variables to achieve local and global interior optima. Appendix A derives these sufficient conditions as a corollary of Proposition 3.

Proposition 4. *If $\theta > 1/2$, I_m^* reaches a local maximum and W_m^* a local minimum over $\mathcal{R}_0 > 0$. These local optima are global optima if $\theta > 3/4$.*

According to the proposition, if the vaccine is at least 75% effective, then I_m^* and W_m^* attain global optima for interior values of $\mathcal{R}_0$. In practice, agencies such as the U.S. Food and Drug Administration (FDA) do not typically license vaccines with efficacy below 80% (Brennan 2009). For realistic parameter values we therefore have the surprising result that if a vaccine is sold by a monopolist directly to consumers, disease prevalence will be greatest and social welfare lowest not for diseases with the most extreme infectiveness as indexed by $\mathcal{R}_0$ but for moderate infectiveness.

The next proposition, proved in Appendix A, measures how substantial the nonmonotonicity in both variables can be by comparing the interior optimum to the limiting value of each.

Proposition 5. *There exist values of parameters besides $\mathcal{R}_0$ such that the interior maximum for I_m^* over $\mathcal{R}_0$ can be made arbitrarily higher in percentage terms than $\lim_{\mathcal{R}_0 \uparrow \infty} I_m^*$. The interior minimum for W_m^* over $\mathcal{R}_0$ can be up to 25% larger than $\lim_{\mathcal{R}_0 \uparrow \infty} W_m^*$.*

2.7. Government Subsidies

We have that the positive externality associated with vaccine consumption can lead to underconsumption relative to the first best for both perfectly competitive and monopoly market structures. This naturally raises the question of whether the government can intervene to correct the market failure. In this section, we characterize the optimal government subsidy and determine its comparative-static properties.

Assume a benevolent government with the objective of maximizing social welfare commits to a per-dose subsidy $GS \geq 0$ at the outset of the game. Adopting the accounting convention that the subsidy is paid to firms, the subsidy is equivalent to a reduction in firms' marginal cost from c to $c - GS$. Since social welfare is maximized by the first-best quantity Q^{**}, the first-best subsidy GS^{**} is that implementing Q^{**}. To accommodate cases in which the government is indifferent among a possibly open set of subsidies maximizing social welfare, we take GS^{**} to be the infimum of the set. Focusing on the infimum could be justified by assuming that the government has lexicographic preferences with economizing on the subsidy being its next priority after social welfare.

Before analyzing particular market structures, we provide some general principles behind the optimal subsidy that apply to any market structure, presented in the next proposition.

Proposition 6. *Consider any market structure in which a subsidy increase weakly reduces the equilibrium price induced. If $\mathcal{R}_0 \leq 1$, then $GS^{**} = 0$; the disease is eradicated in the steady state without a vaccine or subsidy. If $\mathcal{R}_0 \in (1, 1/(1-\theta))$, then GS^{**} induces equilibrium quantity $Q^{**} = \mathcal{Q}_0$ and price $P^{**} = 0$, resulting in eradication of the disease in the steady state. If $\mathcal{R}_0 > 1/(1-\theta)$, then GS^{**} induces equilibrium quantity $Q^{**} = \mu$ and price*

$$P^{**} = \theta H \left[1 - \frac{1}{(1-\theta)\mathcal{R}_0} \right], \tag{26}$$

resulting in universal vaccination but not eradication.

Appendix A provides a proof. Intuitively, the government would like to eradicate the disease in all circumstances if this were possible. When $\mathcal{R}_0 \geq 1/(1-\theta)$, however, the disease is so infective relative to vaccine efficacy that eradication cannot be achieved even if all newborns are vaccinated. The government settles for the goal it can achieve, universal vaccination. Equation (26) characterizes the highest price at which all consumers are still willing to purchase, which is associated with the lowest subsidy required for universal vaccination under the maintained assumption that equilibrium price is weakly decreasing in the subsidy.

With these general principles in hand, we turn to an analysis of specific market structures. Under perfect competition, the firms pass the subsidy directly to consumers, implying $P^{**} = c - GS_c^{**}$ and

thus $GS_c^{**} = c - P^{**}$. Substituting the relevant values of P^{**} from Proposition 6 yields $GS_c^{**} = 0$ if $\mathcal{R}_0 \leq 1$, $GS_c^{**} = c$ if $\mathcal{R}_0 \in (1, 1/(1-\theta))$, and

$$GS_c^{**} = c - \theta H \left[1 - \frac{1}{(1-\theta)\mathcal{R}_0} \right] \tag{27}$$

if $\mathcal{R}_0 \geq 1/(1-\theta)$.

For an imperfectly effective vaccine, (27) becomes negative for sufficiently large $\mathcal{R}_0$. Had we not ruled out negative subsidies by assumption, these negative values of (27) would indeed constitute GS_c^{**}. Given the nonnegativity constraint on subsidies, we have $GS_c^{**} = 0$ for $\mathcal{R}_0 \geq 1/(1-\theta)(1-\tilde{c})$. It is no coincidence that this is same threshold for the perfectly competitive equilibrium to obtain the first best in the absence of a subsidy. No subsidy is needed if equilibrium generates the first best without one. For $\mathcal{R}_0$ strictly above this threshold, the government would like to tax vaccines since some revenue can be raised without impairing universal vaccination. Table 3 lists these results for reference.

Next turn to computing the optimal subsidy under monopoly. By Proposition 6, $GS_m^{**} = 0$ if $\mathcal{R}_0 \leq 1$. If $\mathcal{R}_0 \in (1, 1/(1-\theta))$, Proposition 6 implies $Q^{**} = \mathcal{Q}_0$. Setting the quantity in (25) derived from the monopoly's first-order condition equal to $\mathcal{Q}_0$, substituting the effective marginal cost $c - GS$ under a subsidy for c in the formula, and rearranging yields $GS_m^{**} = c + \theta H(\mathcal{R}_0 - 1)$. If $\mathcal{R}_0 \geq 1/(1-\theta)$, calculations that are similar except that the monopoly quantity in (25) needs to equated with the relevant first-best quantity in this case, $Q^{**} = \mu$, yielding

$$GS_m^{**} = \max \left\{ 0, c + \theta H \left[\frac{1}{(1-\theta)^2 \mathcal{R}_0} - 1 \right] \right\}, \tag{28}$$

where the max operator has been added to reflect the nonnegativity constraint on subsidies. The nonnegativity constraint binds in case (LR5)—not coincidentally the case in which the first-best would be obtained in monopoly equilibrium without a subsidy. Table 3 reports these results for reference, suitably rescaled.

Figure 2 graphs GS^{**} as a function of R_0 for the two market structures. As the figure documents, the optimal subsidy is weakly lower under competition than monopoly. The optimal subsidy is not monotonically increasing in $\mathcal{R}_0$ as might be inferred based solely on epidemiological considerations but is hump shaped.[8] The difficulty in addressing a disease depends not only on its infectiousness

---

[8]While GS^{**} has the same general hump shape in $\mathcal{R}_0$ as MEX^*, the two are not identical for several reasons. Since consumers are homogeneous, the first best is a corner solution involving eradication or universal vaccination, which may be achieved with a subsidy less than MEX^*. This point can be seen most clearly in case (LR5), in which the first best

but also on consumers' response to this infectiousness. Consumers respond to extremely infectious diseases by getting vaccinated even if many others also do. Moderately infectious diseases provide consumers more leeway to free ride on the vaccination of others.

The next proposition catalogs relevant observations concerning GS^{**}. The proof in Appendix A fills in the few details that cannot be directly gleaned from Table 3.

Proposition 7. *The following results characterize optimal subsidies, attaining the first best at minimum government expenditure.*

- $GS_m^{**} \geq GS_c^{**}$, *with strict inequality for all* $\mathcal{R}_0 \in (1, 1/(1-\theta)^2(1-\tilde{c}))$.

- GS_c^{**} *and* GS_m^{**} *are weakly increasing in c and* θ.

- GS_c^{**} *and* GS_m^{**} *are nonmonotonic in* $\mathcal{R}_0 > 0$. GS_c^{**} *reaches its global maximum, c, for all* $\mathcal{R}_0$ *in the interval* $(1, 1/(1-\theta)]$; GS_m^{**} *reaches its global maximum,* $c + H\theta^2/(1-\theta)$, *for* $\mathcal{R}_0 = 1/(1-\theta)$.

2.8. Relevant Limiting Results

While Tables 1 and 2 help organize the results, they still have numerous entries, some of which contain complex formulas. Further clarity can be obtained by taking limits of certain parameters justified by pedagogical or practical considerations.

For vaccine efficacy, we take the limiting case of a perfectly effective vaccine, $\theta \uparrow 1$. This limit is interesting on pedagogical grounds because it allows disease eradication to be technologically feasible. This limit is also interesting on practical grounds, reflecting the high efficacy of many existing vaccines.[9] For rescaled cost, $\tilde{c} = c/\theta H$, we take the limiting case of a costless vaccine, $\tilde{c} \downarrow 0$, reflecting the low cost c for existing vaccines, especially in comparison to the potential disease harm H.[10] For the recovery rate, we take the limiting case of no recovery, $\alpha \downarrow 0$, meaning that the

---

is obtained without a subsidy under both market structures. Even though all consumers are vaccinated in equilibrium, eliminating the familiar positive externality of vaccination on the unvaccinated, *MEX** is still positive, reflecting the externality on those who purchased vaccine but for whom it was ineffective. Other factors can point in the opposite direction, leading lead *MEX** to exceed *GS***. The benevolent government regards transfers as neutral, so it is willing to pay any amount—more than *MEX** if need be—to increase consumption. There is no danger that exorbitant subsidies would induce overconsumption because the first best is a corner solution. An additional wedge between the optimal subsidy and marginal externality under monopoly is that firms with market power may only incompletely pass through the subsidy.

[9]The US Centers for Disease Control and Prevention (2020b) report efficacies from the recommended vaccine courses of 95% for hepatitis B and tetanus, 97% for measles and shingles, 98% for pertussis, and 99% for polio.

[10]Taking the example of a hepatitis B vaccine, the US Centers for Disease Control and Prevention (2002a) lists a price of $25 per dose, totalling $75 for the recommended three-dose course. In the study of the cost effectiveness of a vaccine program against the disease, Hoerger *et al.* (2013) estimate that 4,527 infections would be avoided, generating

person continues to experience harm and can transmit the disease over his or her remaining lifespan, characteristic of diseases such as HIV, syphillis, and malaria.[11]

Imposing these limits and restricting attention to $\mathcal{R}_0 > 1$ considerably simplifies the analysis. Only case (LR3) remains from Tables 1 and 2 and case (LR2$'$) from Table 3. The following proposition is then immediate from inspection of the tables.

Proposition 8. *Suppose $\mathcal{R}_0 > 1$ and consider the limits $\theta \uparrow 1$, $\tilde{c} \downarrow 0$, and $\alpha \downarrow 0$.*

- *Under perfect competition, $P_c^* = MPB_c^* = \Pi_c^* = I_c^* = GS_c^{**} = 0$, $Q_c^* = \mu(1 - 1/\mathcal{R}_0)$, and $MEX_c^* = H$.*

- *Under monopoly, $P_m^* = MPB_m^* = H(1 - 1/\sqrt{\mathcal{R}_0})$, $Q_m^* = \mu(1 - 1/\sqrt{\mathcal{R}_0})$, $\Pi_m^* = \mu H(1 - 1/\sqrt{\mathcal{R}_0})^2$, $I_m^* = (1/\sqrt{\mathcal{R}_0}) - (1/\mathcal{R}_0)$, $MEX_m^* = H/\sqrt{\mathcal{R}_0}$, and $GS_m^{**} = H(\mathcal{R}_0 - 1)$.*

According to the proposition, under perfect competition, for all $\mathcal{R}_0 > 1$, enough consumers are vaccinated to eradicate the disease, attaining the first best. Price, profit, and the infection rate are all 0. Under monopoly, price, profit, and the share of consumers vaccinated approach 0 in the limit $\mathcal{R}_0 \downarrow 1$, while price approaches 100% of the harm from contracting the disease and quantity approaches 100% share of newborns in the limit $\mathcal{R}_0 \uparrow \infty$. Equilibrium prevalence is nonmonotonic in $\mathcal{R}_0$ under monopoly. Maximizing the formula given for I_m^*, one can show that equilibrium prevalence is greatest for a disease with $\mathcal{R}_0 = 4$.

2.9. Calibrations

This section provides a series of calibrations to help understand the implications of the analysis for existing vaccines. The calibrations are meant more as illustrations than forecasts. Our present model is too stylized on many fronts to provide accurate forecasts. We abstract from heterogeneity in infectiousness, heterogeneity in costs of prevention among consumers, and mortality effects of disease. Certain parameters are set to convenient limiting values rather than being estimated from data. A host of political-economy considerations lead real-world vaccine markets to depart from our theoretical construct of firms selling directly to individual consumers without third-party funding.

1,218 quality adjusted life years (QALYs), implying a loss per infection of 0.27 QALYs. Scaling by the \$200,000 per QALY standard for cost effectiveness advocated for U.S. health programs by Ubel, Hirth, and Chernew (2003) yields estimate $H = \$54,000$ of the health benefit from avoiding the certainty of a hepatits B infection, implying $\tilde{c} = c/\theta H = \$75/(0.95)(\$54,000) = 0.0045$.

[11]This normalization is not crucial but slightly simplifies one formula.

2.9.1. $\mathcal{R}_0$ Estimates

Table 4 presents estimates of $\mathcal{R}_0$ from the epidemiology literature that can be substituted into the formulas in Proposition 8 to calibrate equilibrium outcomes for various diseases. As previously noted, $\mathcal{R}_0$ varies considerably across diseases, from 1.1 for SARS at the low end to 16–18 for measles and pertussis at the high end. Since estimates of $\mathcal{R}_0$ can also vary widely across time and region, the table lists the relevant time period and region from which each estimate comes.

2.9.2. HIV Calibration

Our first calibration considers a disease, HIV, with a moderate $\mathcal{R}_0$. Table 4 provides a range of estimates for $\mathcal{R}_0$ for HIV. For convenience, we take the round number $\mathcal{R}_0 = 4$ from this range. Incidentally, recall that this is the value of $\mathcal{R}_0$ for which disease prevalence is greatest in equilibrium under monopoly.

Substituting $\mathcal{R}_0 = 4$ into the formulas provided by Proposition 8, the model suggest that a monopolist selling an HIV vaccine to consumers would price it at half the harm from contracting the disease. At this price, half of consumers purchase the vaccine and the other half free ride. Half of the free riders become infected, resulting in an overall HIV prevalence rate of 1/4. This is much lower than the 3/4 prevalence rate that would emerge in the absence of a vaccine for a disease with $\mathcal{R}_0 = 4$. The monopolist captures one third of the potential social surplus from a vaccine. Consumers capture one third, and the remaining third is lost to monopoly distortion.

As noted in the previous subsection, for all $\mathcal{R}_0 > 1$, including $\mathcal{R}_0 = 4$ assumed here, the first best for this essentially costless vaccine involves vaccinating enough people to eradicate the disease in the steady state. The first best is realized under perfect competition. The minimum subsidy that would have to be paid to the monopolist to attain the first best is $GS_m^{**} = 3H$. In other words, the monopolist would have to receive a per-course subsidy of at least three times the lifetime harm experienced from certainly contracting HIV, easily running into many thousands of dollars per course. The model abstracts from any distortion involved in raising government funds. With any deadweight loss of taxation, such enormous subsidies would be prohibitively expensive, forcing governments to use other instruments, such as bulk purchases, to attain the first-best level of vaccination or give up on reaching the first best.

The situation is brighter if the HIV vaccine were competitively supplied since, as mentioned, the first best is obtained in the perfectly competitive equilibrium without a subsidy. Unfortunately, as discussed in Section 2.3, it may be unrealistic to suppose that competition would emerge even in the

long run for high-tech product like a potential HIV vaccine that are extremely difficult for a generic competitor to reverse engineer.

Certain other health interventions apart from vaccines may be appropriately modeled as being competitively supplied. Consider the use of adult male circumcision as an HIV preventive. Far from universal adoption predicted by the model for a costless, perfectly effective intervention, in the meta-analysis by Kennedy *et al.* (2020), adult circumcision rates in control samples were negligible. Thus the assumptions of zero cost and perfect efficacy are far-fetched for circumcision and need to be relaxed.

Consider the introduction of a small program subsidizing costly circumcision in a country with little current uptake.[12] The relevant theoretical case is (LR2), the only case with a positive infection rate but negligible adoption. The subsidy required to induce the initial participants in the small program to become circumcized in case (LR2) equals $c - MPB_c^* = c - \theta H(1 - 1/\mathcal{R}_0)$. A planner would be willing to provide such a subsidy up to the level of $MEX_c^* = \theta H/\mathcal{R}_0$. Substituting $\mathcal{R}_0 = 4$ assumed in this subsection for HIV, $\theta = 0.6$, the estimate of efficacy in randomized controlled trials in several studies including Bailey *et al.* (2007), and $H = 7,000$, the estimated lifetime cost of first-line drug treatments for HIV (UNAIDS/WHO/SACEMA Expert Group on Modelling the Impact and Cost of Male Circumcision for HIV Prevention 2009), a lower bound on the health benefit from avoiding an HIV infection, yields $MEX_c^* = 1,050$, implying that the required subsidy is somewhere between \$0 and \$1,050. Of the eight studies surveyed by Kennedy *et al.* (2020), none paid a subsidy beyond the cost of the procedure of more than \$15. While most of these studies obtained statistically significant estimates for the effect of subsidies on circumcision, Kennedy *et al.* (2020, p. 11) note that "the overall uptake ... in these studies was low, and the absolute differences between groups were small," suggesting that the subsidies were far below the optimum.

2.9.3. Measles Calibration

As a contrast to the HIV calibration, we next calibrate the market for a vaccine for measles, a disease with a much higher value of $\mathcal{R}_0$ than HIV. For convenience, take the round number $\mathcal{R}_0 = 16$ from the range offered by Table 4. The model suggests that a monopolist selling a measles vaccine to consumers would price it at fully 75% of the harm from contracting the disease. Of consumers, 75% purchase the vaccine, and 25% free ride. A substantial majority, 75%, of free riders contract the

---

[12]Since male circumcision is only relevant for half the population, analyzing a program targeting universal adoption would require modifying the model to allow for gendered subpopulations. We avoid this complication by supposing the program is small. Thus, instead of computing the optimal subsidy GS_c^{**} achieving universal adoption, this calibration analyzes a subsidy provided to the initial adopter.

disease, resulting in an overall measles prevalence rate of about 19%, less than the 25% prevalence calibrated for HIV—as expected given the prevalence rate is maximized at the value of $\mathcal{R}_0$ used in the HIV calibration.

The minimum subsidy required to attain the first best under monopoly is $GS_m^{**} = 15H$. In other words, the monopolist would have to receive a per-course subsidy of at least 15 times the lifetime harm experienced from certainly contracting measles. This enormous expense highlights even more strongly than the HIV calibration that the optimal subsidy, while providing a useful theoretical benchmark, would not be a realistic policy alternative in practice for diseases with estimated values of $\mathcal{R}_0$ toward the higher end.

2.9.4. SARS Calibration

Just as the measles calibration provided a useful contrast to HIV, the next calibration for SARS provides a useful contrast but in the other direction. The value $\mathcal{R}_0 = 1.1$ for SARS in Table 4 is considerably lower than for the other diseases, almost lying on the boundary ($\mathcal{R}_0 = 1$) at which the disease disappears in the steady state without a vaccine. The model suggests that a monopolist selling a SARS vaccine to consumers would price it at less than 5% of the harm from contracting the disease. Less than 5% of consumers purchase the vaccine. Almost all consumers free ride. They have the luxury to do so since, despite the low immunization rate, the low infectiousness of SARS means equilibrium prevalence remains low at about 4%.

3. Short-Run Analysis of Disease Outbreak

This section analyzes the market for a vaccine against a disease like Covid that rises and falls as an epidemic in a much shorter spell than a human generation. We seek to model a campaign that rolls out the vaccine to substantial fraction of the population quickly to mitigate the harm experienced by the current generation rather than continued sales to each new cohort of newborns as with vaccines against typical childhood diseases. Future generations may be spared much damange by the epidemic's natural decline.

Rather than modeling a continuous flow vaccine rate each instant, in this section we take the polar opposite approach that a stock of vaccine is administered to the population in a short span. For pedagogical purposes, we adopt the extreme assumption that all doses of vaccine that will ever be administered are administered in a single instant. If feasible, this would be the profit-maximizing approach for firms selling to the single, relevant generation of consumers. It would also be the

welfare-maximizing approach for a social planner. Logistical constraints would prevent such rapid vaccine rollout in practice, but the model may still be a reasonable approximation to an intensive vaccine campaign against Covid or other epidemic disease.

3.1. Epidemiological Model

We seek to maintain as much of the machinery from the earlier analysis as possible, including the SIR framework, adapting it to incorporate a discrete vaccination event. Continuous time indexed by t is divided into the ex ante period, before the vaccine's arrival, and the ex post period, after the vaccine's arrival. Fixing the arrival date at $t = 0$, negative values of t index time in the ex ante period, the earlier the date, the greater the absolute value of the index. Positive values of t index time in the ex post period.

During the ex post period, the disease evolves according to SIR conditions similar to those laid out in Section 2.1:

$$S_t + I_t + R_t + V_t = 1 \tag{29}$$

$$\dot{V}_t = 0 \tag{30}$$

$$\dot{I}_t = \beta I_t S_t - \alpha I_t \tag{31}$$

$$\dot{R}_t = \alpha I_t \tag{32}$$

$$\dot{S}_t = -\beta I_t S_t. \tag{33}$$

There are only two differences between these equations and the system (1)–(5) underlying the long-run analysis. First, to reflect the short span of the epidemic, occurring within a generation, we assume there are no births and deaths by setting $\mu = 0$. Second, no further vaccine is administered after the initial tranche at date T. Hence, the vaccine compartment does not change, reflected by equation (30).[13]

Let hats distinguish the compartment values during the ex ante period $\hat{S}_t$, $\hat{I}_t$, and $\hat{R}_t$ (there is no need for the compartment $\hat{V}_t$ since there is no vaccine ex ante). Compartments follow the same SIR dynamics ex ante as they do ex post. It remains to specify how compartments transition from the ex ante period to determine the initial conditions for the ex post period. Letting Q denote the vaccine

---

[13]Epidemiology texts label the vaccination process involved in our short-run analysis "vaccination at recruitment" (Martcheva 2015, Section 9.2.1), contrasting the process of "continuous vaccination" involved in our long-run analysis (Martcheva 2015, Section 9.2.2).

quantity administered at date 0, the initial conditions for the ex post period are

$$V_0 = \theta Q \tag{34}$$

$$S_0 = \hat{S}_0 - \theta Q \tag{35}$$

$$I_0 = \hat{I}_0 \tag{36}$$

$$R_0 = \hat{R}_0 = 1 - \hat{S}_0 - \hat{I}_0. \tag{37}$$

Equation (34) implies that the mass of successfully immunized at the start of the ex post period equals vaccine efficacy $\theta \in (0,1)$ times the number of courses Q purchased. Equation (35) implies that the ex post period starts with discretely fewer susceptibles than the ex ante period ended with, the difference given by the mass $V_0 = \theta Q$ who are immunized at date 0. Equations (36) and (37) imply that there are no discrete change to either the mass of infected or the mass of recovered individuals from the ex ante to ex post period. To ensure $S_0 \geq 0$, we must have

$$Q \leq \hat{S}_0. \tag{38}$$

We treat $\hat{I}_0$ and $\hat{S}_0$ as exogenous parameters, allowing them to take on any admissible values $\hat{I}_0 \in (0,1)$, $\hat{S}_0 \in (0, 1-\hat{I}_0]$.[14] As before, we will work with the basic reproductive ratio $\mathcal{R}_0$ instead of the transmission rate β. Substituting the short-run value $\mu = 0$ into (6) yields the following expression for the basic reproductive ratio in the short-run model:

$$\mathcal{R}_0 = \frac{\beta}{\alpha}. \tag{39}$$

The short-run analysis will require some different methods than the long-run analysis, warranting explanation given the analyses are based on similar SIR frameworks. In the long-run analysis, social benefit was measured as $SB(Q) = h[1 - I(Q)]$. Recall the notational convention that dropping the subscript t on a time-varying variables indicates its steady-state value: e.g., $I(Q) = \lim_{t \uparrow \infty} I_t(Q)$. Also notice that we have started adding argument Q to indicate variables' dependence on that key variable, a convention that we will continue from now on.

---

[14]An alternative approach would treat them as endogenous and derive them from ex ante initial conditions $\hat{S}_{-T}$ and $\hat{I}_{-T}$ and the duration T of the ex ante period. Our approach has two advantages. It is more parsimonious, involving the two parameters $\hat{I}_0$ and $\hat{S}_0$ rather than the three $\hat{I}_{-T}$, $\hat{S}_{-T}$, and T. The parameters involved in our approach are also easier to measure. Pinning down the exact start of an epidemic is notoriously difficult to say nothing of estimating the proportions of susceptibles and infecteds before surveillance protocols have been instituted.

A constant, positive infection could be maintained in the long-run model since it involved a continuous flows of births and deaths. The flow harm at any instant over the long run thus constituted a natural measure of social benefit. In the short-run analysis, with a single generation experiencing no demographic flows, the previous measure of social benefit is not useful since, as will be seen in Lemma 4, the disease eventually dies out in fixed population for all $Q \geq 0$, i.e., $I(Q) = 0$. In a short-run model, a better measure of illness is not the infection rate to which the population converges, which is always 0 by Lemma 4, but the mass of people who contract the disease at some point during the epidemic, captured by the limiting value of the recovered population, $R(Q)$. The measure of social benefit we will use in the short-run analysis is harm avoided in the population who never contract the disease:

$$SB(Q) = H[1 - R(Q)] = H[S(Q) + \theta Q], \tag{40}$$

where the second equality follows from equation (29) and Lemma 4. Besides using $R(Q)$ in place of $I(Q)$, (40) differs from the previous expression for $SB(Q)$ in that it involves the stock of harm parameter H rather than flow h. Assuming, as we do for simplicity, that there is no discounting and noting, by (32), that recovery follows a Poisson process, (40) properly measures the total potential harm avoided in the population over the entire epidemic.

The following series of lemmas characterize $S_t(Q)$ and $I_t(Q)$ both for finite and limiting values of t. Appendix A provides proofs.

Lemma 1. $I_t(Q) > 0$ and $S_t(Q) > 0$.

Lemma 2. $S_t(Q)$ is strictly decreasing in t.

Lemma 3. If $\mathcal{R}_0 S_0(Q) \leq 1$, then $I_t(Q)$ is strictly decreasing in t for all $t > 0$. Otherwise, $I_t(Q)$ is hump-shaped, peaking at time $T > 0$ satisfying $S_T(Q) = 1/\mathcal{R}_0$, strictly increasing for $t < T$, and strictly decreasing for $t > T$.

Lemma 4. The limits $I(Q)$ and $S(Q)$ exist. In particular, $I(Q) = 0$ and $S(Q) \in (0, S_0(Q))$.

Lemma 5. $S(Q) < 1/\mathcal{R}_0$.

To explain the lemmas in intuitive terms, they tell us that the infection rate is always positive in finite time because, if not increasing, infections are at worst declining at a proportional rate less than 100% each instant, which can never force the infection rate to 0. The infection rate does asymptote to 0 as the stock of susceptibles is depleted and recovery takes over as the dominating force, reducing the stock of infecteds. Turning to results for the population of susceptibles, with an imperfectly effective

28

vaccine ($\theta < 1$), even a universal vaccination campaign cannot eliminate the stock of susceptibles at date 0. The stock of susceptibles is never forced to 0 after because the proportional decline is less than 100% each instant. The stock of susceptibles strictly decreases over time since it is subject to outflows but not inflows.

According to Lemma 3, there are two possible shapes for infection rate over the ex post period. One possibility is that the infection rate continuously diminishes from its initial level. Another possibility is that infections have a hump-shaped path, expanding up to a peak and declining thereafter.[15] For infections to initially expand rather than contract requires both infectiveness (i.e., high $\mathcal{R}_0$) and substantial fuel (i.e., high $S_0(Q)$). We will call $\mathcal{R}_0 S_0(Q) > 1$ the condition for epidemic expansion.

While no closed-form solution exists for the steady-state number of susceptibles $S(Q)$, the next lemma, proved in Appendix A, expresses $S(Q)$ as an implicit function of other model parameters (though the notation continues to emphasize only the variable to be endogenized later, Q, as an argument of S). The lemma also provides an expression for $S(Q)$ in terms of the principal branch of the Lambert W function, here denoted $\bar{L}$.[16]

Lemma 6. *$S(Q)$ satisfies*

$$\ln S(Q) - \mathcal{R}_0 S(Q) = \ln(\hat{S}_0 - \theta Q) - \mathcal{R}_0(\hat{I}_0 + \hat{S}_0 - \theta Q) \tag{41}$$

and can be written

$$S(Q) = \frac{1}{\mathcal{R}_0} \left| \bar{L} \left(-\mathcal{R}_0(\hat{S}_0 - \theta Q) e^{-\mathcal{R}_0(\hat{I}_0 + \hat{S}_0 - \theta Q)} \right) \right|. \tag{42}$$

Equations (41) and (42) can be used to derive the steady-state population of susceptibles in the extremes of uninfective or infinitely infective diseases. If a disease cannot be transmitted, initial susceptibles never contract the disease, so remain as susceptibles in the steady state, implying $\lim_{\mathcal{R}_0 \downarrow 0} S(Q) = S_0(Q) = \hat{S}_Q - \theta Q$. This can be proved by substituting $\mathcal{R}_0 = 0$ into (41).

Lemma 7. $\lim_{\mathcal{R}_0 \downarrow 0} S(Q) = \hat{S}_0 - \theta Q$ *and* $\lim_{\mathcal{R}_0 \downarrow 0}[\mathcal{R}_0 S(Q)] = 0$.

The second limit, recorded for reference, follows from $\lim_{\mathcal{R}_0 \downarrow 0}[\mathcal{R}_0 S(Q)] = (\hat{S}_0 - \theta Q) \lim_{\mathcal{R}_0 \downarrow 0} \mathcal{R}_0 = 0$.

In the opposite extreme of an infinitely infective disease, all susceptibles eventually become infected, implying that the steady-state susceptible population vanishes in the limit. This result can be

---

[15]Some epidemiologists take the hump-shaped path as the defining feature of an epidemic.

[16]The Lambert W function L frequently arises in epidemiological applications. By definition $L(x)$ is the implicit solution to the exponential equation $L(x)e^{L(x)} = x$. The principal branch $\bar{L}$ is the sole solution to the implicit equation or, if two solutions exist, the higher of the two. The lower branch $\underline{L}$ is defined when two solutions exist as the lower of the two. Though $\bar{L}$ and $\underline{L}$ do not have a closed-form solutions, they can be computed with built-in functions included in Matlab, R, and other standard software packages.

derived formally from Lemma 5: $0 \leq \lim_{\mathcal{R}_0 \uparrow \infty} S(Q) \leq \lim_{\mathcal{R}_0 \uparrow \infty} (1/\mathcal{R}_0) = 0$, implying $\lim_{\mathcal{R}_0 \uparrow \infty} S(Q) = 0$. A more subtle question regards the rate at which the steady-state susceptible population vanishes. The next lemma, proved in Appendix A, states that $S(Q)$ vanishes faster than $\mathcal{R}_0$ increases.

Lemma 8. $\lim_{\mathcal{R}_0 \uparrow \infty} S(Q) = \lim_{\mathcal{R}_0 \uparrow \infty} [\mathcal{R}_0 S(Q)] = 0$.

Comparative-static results can be obtained by applying the Implicit Function Theorem to (41). For instance,

$$\frac{\partial S(Q)}{\partial Q} = \frac{\theta S(Q)}{S_0(Q)} \left[\frac{\mathcal{R}_0 S_0(Q) - 1}{1 - \mathcal{R}_0 S(Q)} \right]. \tag{43}$$

Since its denominator is positive by Lemma 5, the sign of (43) is determined by whether the condition for epidemic expansion holds. The immediate effect of an increase in Q is to move an individual from the currently susceptible to the vaccinated compartment. If infections are waning, this immediate effect persists as a reduction in susceptibles that remain in the steady state. If the disease meets the condition for an epidemic, however, the reduction in current susceptibles has such a strong feedback effect in the form of reduced "fuel" for infections that a greater proportion of susceptibles remain in the steady state despite the immediate reduction in susceptibles.

3.2. Consumer Demand

We maintain the assumptions about consumers made in Section 2.2 for the long-run analysis. Consumers are homogeneous and risk neutral. Consistent with the present short-run perspective, assume agents do not discount the future. Consumers suffer harm H from contracting the disease over the expected spell before recovery.

Individuals in the susceptible compartment when the vaccine is introduced, numbering $\hat{S}_0$, are potential consumers. They make their demand decisions by comparing the vaccine's price P to their marginal private benefit. As in the long-run analysis, $MPB(Q) = \theta H \Phi_I(Q)$, recalling $\Phi_I(Q)$ is the probability a susceptible contracts the disease over his or her lifespan.

The expression for $\Phi_I(Q)$ here differs from the long-run analysis. Since there is no mortality here, a person's lifespan is the duration of the ex post period. The probability an unvaccinated individual does not contract the disease equals $S(Q)/S_0(Q)$, the number of people who remain susceptible over the model's horizon divided by the number of people who are susceptible at the start of the ex post period. The probability of infection is the complementary probability

$$\Phi_I(Q) = 1 - \frac{S(Q)}{S_0(Q)} = 1 - \frac{S(Q)}{\hat{S}_0 - \theta Q}, \tag{44}$$

which Lemma 4 guarantees is positive. Thus,

$$MPB(Q) = \theta H \Phi_I(Q) = \theta H \left[1 - \frac{S(Q)}{\hat{S}_0 - \theta Q}\right]. \tag{45}$$

Differentiating, substituting from (43) and (44), and rearranging yields

$$\frac{\partial MPB(Q)}{\partial Q} = \frac{-\theta \mathcal{R}_0 S(Q) MPB(Q)}{S_0(Q)[1 - \mathcal{R}_0 S(Q)]}, \tag{46}$$

which is negative by Lemma 5, confirming the intuition that vaccinating more consumers lowers their marginal private benefit.

Proceeding to derive the demand curve, all $\hat{S}_0$ consumers purchase the vaccine if $P < MPB(\hat{S}_0)$, and none purchase if $P > MPB(0)$. For P strictly between $MPB(\hat{S}_0)$ and $MPB(0)$, some but not all consumers purchase. Given they are homogeneous, consumers must be indifferent between purchasing and not, implying $P = MPB(Q)$. Given they are indifferent, any fraction of them purchasing is an equilibrium; demand is pinned down by the value of Q satisfying (45) when the right-hand side is set equal to P. Rearranging the resulting equation yields $S(Q) = (1 - P/\theta H)(\hat{S}_0 - \theta Q)$. Substituting this into (41) and solving for Q gives the following expression for demand when a subset purchase:

$$d(P) = \frac{1}{\theta} \left\{ \hat{S}_0 + \frac{\theta H}{P} \left[\frac{1}{\mathcal{R}_0} \ln\left(1 - \frac{P}{\theta H}\right) + \hat{I}_0\right] \right\}. \tag{47}$$

Combining these facts yields the demand curve

$$D(P) = \begin{cases} 0 & P > MPB(0) \\ d(P) & P \in [MPB(\hat{S}_0), MPB(0)] \\ \hat{S}_0 & P < MPB(\hat{S}_0). \end{cases} \tag{48}$$

Equivalently, the demand curve is given by $d(P)$ unless this violates a boundary condition $d(P) \in [0, \hat{S}_0]$, in which case demand is given by the violated boundary.

3.3. Firm Supply

Firm behavior is the same as that assumed in the long-run analysis. Firms produce at constant marginal and average cost $c \in (0, \theta H)$ per vaccine course. Perfectly competitive firms supply the vaccine at price c. A monopoly sets a price maximizing industry profit, Π, now earned from sales made at date 0.

3.4. Normative Measures

Equation (40) provided the expression for social benefit relevant for the present short-run analysis. The expression for welfare in the long-run analysis given in (20) continues to be relevant for the short-run analysis: $W(Q) = SB(Q) - cQ$. Marginal social benefit is the derivative $MSB(Q) = \partial SB(Q)/\partial Q$. Differentiating (40), substituting from (43) and (44), and rearranging yields

$$MSB(Q) = \frac{\theta H \Phi_I(Q)}{1 - \mathcal{R}_0 S(Q)}. \tag{49}$$

Substituting (45) and (49) into the definition $MEX(Q) = MSB(Q) - MPB(Q)$ yields

$$MEX(Q) = \frac{\theta H \Phi_I(Q) \mathcal{R}_0 S(Q)}{1 - \mathcal{R}_0 S(Q)}. \tag{50}$$

Combining (44), (45), and (50) yields an equivalent expression for the marginal externality providing some useful intuition:

$$MEX(Q) = \mathcal{R}_0 S(Q) MSB(Q) = \mathcal{R}_0 S_0(Q) \left[\frac{S(Q)}{S_0(Q)} \right] MSB(Q). \tag{51}$$

The external benefit from vaccinating individual A equals the social benefit of vaccinating everyone who would not be infected but for their interaction with A. By definition of the basic reproductive ratio, A causes $\mathcal{R}_0 S_0(Q)$ direct infections in the susceptible population. However, some of those would have been infected by someone else later; only the fraction $S(Q)/S_0(Q)$ would have survived as susceptibles to the end but for their interaction with A. The marginal externality equals the marginal social benefit cumulated over these "but for" infections.

As before, let Q^{**} denote the first-best quantity, maximizing $W(Q)$. In the long-run analysis, $MSB(Q)$ was a constant, leading Q^{**} to be a corner solution, involving universal vaccination (unless a lower quantity eradicates the disease). Here, $MSB(Q)$ is not a constant, so Q^{**} can be an interior value, in which case it solves the social planner's first-order condition $MSB(Q^{**}) = c$.

3.5. Perfectly Competitive Equilibrium

Under perfect competition, the equilibrium price is $P_c^* = c$ and profit is $\Pi_c^* = 0$. The remaining equilibrium variables can be computed using straightforward algebra applied to the supplied equations. Table 5 reports the equilibrium values of selected variables as a function of $\mathcal{R}_0$. The expressions have been simplified and their mathematical structure clarified via the change of variables indicated

with tildes as described in the table notes. We again use lower-case roman numerals to identify cases, but the cases do not correspond to the same ones as in the long-run analysis.

Although similar in many respects to its analogue, Table 1, that lists equilibrium variables under perfect competition for the long-run analysis, Table 5 has some notable differences. One difference is that R_c^* rather than I_c^* is reported. In the short-run analysis, the disease always dies out without a population inflow to sustain it within the single generation considered, implying $I_c^* = 0$. We track the total number of infections that ever happen, embodied in R_c^*. Another difference is that $\mathcal{R}_0 = 1$ is no longer an important threshold determining whether the disease dies out in the steady state without a vaccine. In the short-run analysis, as just explained, the disease always dies out in the steady state, leaving us to track the sum of infections over the epidemic, which is positive for positive $\mathcal{R}_0$ no matter how low.

The fact that $\mathcal{R}_0 = 1$ is no longer an important threshold leaves three relevant intervals for $\mathcal{R}_0$ to report in Table 5: case (SR1) in which $\mathcal{R}_0$ is so low that no consumer finds it worthwhile to purchase the vaccine, case (SR2) in which moderate values of $\mathcal{R}_0$ lead some but not all of the $\hat{S}_0$ susceptibles to purchase, and case (SR3) in which $\mathcal{R}_0$ is so high that all susceptibles find purchasing the vaccine worthwhile. The first best is obtained in this case: $Q_c^* = Q^{**}$.

The threshold dividing cases (SR1) and (SR2) is the value of $\mathcal{R}_0$ for which the interior solution for demand in (47) equals 0 (since no consumers purchase) at the equilibrium price $P_c^* = c$. Solving $d(c) = 0$ for $\mathcal{R}_0$ yields $\mathcal{R}_0 = |\ln(1-\tilde{c})|/(\hat{I}_0 + \tilde{c}\hat{S}_0)$, where $\tilde{c} = c/\theta H$. Similarly, the threshold dividing cases (SR2) and (SR3) is the value of $\mathcal{R}_0$ for which (47) equals $\hat{S}_0$ (since all susceptibles purchase) at $P_c^* = c$. Solving $d(c) = \hat{S}_0$ for $\mathcal{R}_0$ yields $\mathcal{R}_0 = |\ln(1-\tilde{c})|/[\hat{I}_0 + (1-\theta)\tilde{c}\hat{S}_0]$. Appendix A provides derivations of the table entries.

The next proposition summarizes the comparative-static effects of an increase in $\mathcal{R}_0$ on the equilibrium under perfect competition. Among other notable results, we have that the nonmonotonic behavior of infections and the marginal externality uncovered in the long-run analysis persists in the short-run analysis.

Proposition 9. *Consider the comparative-static effect of $\mathcal{R}_0$ on equilibrium under perfect competition in the short-run analysis.*

- *Price and industry profit are constant, with $P_c^* = c$ and $\Pi_c^* = 0$.*

- *Q_c^* and MPB_c^* are weakly increasing in $\mathcal{R}_0$.*

- *R_c^*, MSB_c^*, and MEX_c^* are nonmonotonic in $\mathcal{R}_0$. R_c^* attains a single interior local maximum, which is a global maximum if and only if $\tilde{c} \geq 1 - \theta$. Each of MSB_c^* and MEX_c^* attain no more than two interior local maxima, one of which is a global maximum.*

33

- W_c^* *is weakly decreasing in* $\mathcal{R}_0$.

For each Q_c^*, MPB_c^*, *and* W_c^*, *there exists a nonempty interval of* $\mathcal{R}_0$ *such that the weak change is strict.*

Appendix A provides proofs for results not obvious from Table 5. The proofs for the nonmonotonic variables are particularly intricate since they involve characterizing complex shapes of functions without closed-form solutions. Our approach uses the concavity of the sign of the function's derivative to count the derivative's roots, coupled with an examination of the limiting value of the derivative. Graphs of equilibrium variables closely resemble Figure 1, so for space considerations have been relegated to an online appendix (Appendix D).

3.6. Monopoly Equilibrium

Similar insights that guided the analysis of monopoly in the long-run model guide it here in the short-run model. It continues to be true that monopoly output is zero in any case it would be zero under perfect competition. Case (SR1) is thus trivially identical across perfect competition and monopoly. In the remaining cases, competitive firms are able to make positive sales at price c. It continues to be true that, by continuity, the monopolist can make positive sales at some small markup above c, implying $Q_m^* > 0$ in cases (SR2)–(SR4).

To solve for Q_m^* in these other cases, we again convert the monopolist's maximization problem so that the choice variable is quantity rather than price. The monopolist optimally sets a price to extract the entire private benefit of the marginal consumer, leading to inverse demand $P(Q) = MPB(Q)$. The monopolist chooses Q to maximize $[P(Q) - c]Q = [MPB(Q) - c]Q$ subject to $Q \leq \hat{S}_0$, a constrained maximization problem which can be solved using the Kuhn-Tucker method.

Differentiating the profit function with respect to Q yields

$$\theta H \left[1 - \frac{S(Q)}{S_0(Q)} \right] \left\{ 1 - \frac{\theta Q \mathcal{R}_0 S(Q)}{S_0(Q)[1 - \mathcal{R}_0 S(Q)]} \right\} - c. \tag{52}$$

If the constraint $Q \leq \hat{S}_0$ does not bind, the monopoly quantity can be found by setting (52) equal to zero, which after manipulation yields

$$MPB(Q_m^*) = P(Q_m^*) = \left\{ 1 - \frac{\theta Q_m^* \mathcal{R}_0 S(Q_m^*)}{S_0(Q_m^*)[1 - \mathcal{R}_0 S(Q_m^*)]} \right\}^{-1} c. \tag{53}$$

Several useful insights are immediate consequences of (53). Since the expression in braces is less than 1, its reciprocal is greater than 1, implying $P_m^* = P(Q_m^*) > c = P_c^*$. If Q_c^* is an interior solution—

i.e., $Q_c^* \in (0, \hat{S}_0)$, which is true in case (SR2) of Table 5—then the price relationship has a direct implication for quantities, namely, $Q_m^* < Q_c^*$.

Equation (53) can also be used to establish the existence of parameters for which the first best is attained under perfect competition but monopoly falls short. To see this, consider $\mathcal{R}_0$ on the boundary between cases (SR2) and (SR3) in Table 1, i.e., $\mathcal{R}_0 = |\ln(1 - \tilde{c})| / [\hat{I}_0 + (1 - \theta)\hat{S}_0]$. The entries for MPB_c^* in Table 1 imply $P(\hat{S}_0) = c$ at this value of $\mathcal{R}_0$, but (53) implies $P(Q_m^*) > c$. Hence, $P(Q_m^*) > P(\hat{S}_0)$, implying $Q_m^* < \hat{S}_0$ at this boundary value of $\mathcal{R}_0$. By continuity, there is a neighborhood of $\mathcal{R}_0$ above this boundary value for which the first best is not obtained under monopoly but is produced under perfect competition.

If the monopoly quantity is not an interior solution, then the constraint holds with equality, implying $Q_m^* = \hat{S}_0$. According the the Kuhn-Tucker conditions, the constraint holds with equality when the parameters are such that the derivative (52) is nonnegative when evaluated at the constraint quantity. Substituting $Q = \hat{S}_0$ into (52) and rearranging, this condition is

$$\Phi_I(\hat{S}_0) \left[1 - \left(\frac{\theta}{1 - \theta} \right) \frac{\mathcal{R}_0 S(\hat{S}_0)}{1 - \mathcal{R}_0 S(\hat{S}_0)} \right] \geq \tilde{c}. \tag{54}$$

The left-hand side equals 1 in the limit $\mathcal{R}_0 \uparrow \infty$ by Lemma 8, exceeding the right-hand side since $\tilde{c} < 1$ by assumption (23). Therefore, for sufficiently high $\mathcal{R}_0$, a monopoly produces the first-best quantity: $Q_m^* = \hat{S}_0 = Q^{**}$.

The values of equilibrium variables resulting from the preceding analysis are recorded in Table 6. The results reported in the column for cases (SR2) and (SR3) are not provided in analytic form, let alone in closed form. This need not preclude definitive comparative-statics results; one could apply the Implicit Function Theorem to the first-order condition (52) to determine how Q_m^* changes with $\mathcal{R}_0$ in (SR2) and (SR3). However, this approach does not deliver a definitive sign. We can be sure that Q_m^* increases in $\mathcal{R}_0$ for some $\mathcal{R}_0$ in (SR2) and (SR3)—since Q_m^* must rise from 0 to the first-best quantity $\hat{S}_0$ somewhere in that set by continuity—but we cannot rule out the possibility that the monopolist responds to an increase in $\mathcal{R}_0$ in some subintervals by reducing output in order to extract an even larger price increase than otherwise.

Despite these challenges, we are able to derive definitive comparative-statics results for Π_m^* by the envelope theorem. In addition, we are able to show that the nonmonotonic behavior of R^*, MSB^*, and MEX^* extends from perfect competition to monopoly. The key insight behind the proof is that perfect competition and monopoly generate the same output at both extremes of the lowest and

the highest values of $\mathcal{R}_0$. Certain nonmonotonicies can be inferred from the behavior of variables in those extremes, already characterized in the analysis of perfect competition. We report these comparative-statics results in the next proposition, proved in Appendix A.

Proposition 10. *Consider the comparative-static effect of $\mathcal{R}_0$ on equilibrium under monopoly in the short-run analysis.*

- *$\Pi_m^* = 0$ for $\mathcal{R}_0 \leq |\ln(1-\tilde{c})|/(\hat{I}_0 + \tilde{c}\hat{S}_0)$; Π_m^* is positive and strictly increasing for higher $\mathcal{R}_0$.*

- *If $\tilde{c} \geq 1-\theta$, R_c^* is nonmonotonic in $\mathcal{R}_0$ and attains an interior global maximum.*

- *MSB_c^* and MEX_c^* are nonmonotonic, attaining interior global maxima.*

3.7. Government Subsidies

As in Section 2.7, define GS^{**} as the infimum of the set of government subsidies implementing the first best quantity Q^{**} in equilibrium. We continue to constrain GS^{**} to be nonnegative.

It is easy to establish some broad results for any market structure. Since $MEX(Q) \geq 0$, equilibrium output cannot exceed the first best: $Q^* \leq Q^{**}$. If, in addition, $Q^{**} = 0$, then $0 \leq Q^* \leq Q^{**} = 0$, implying $Q^* = Q^{**}$, in turn implying $GS^{**} = 0$ since the first best can be achieved without a subsidy. The proof of the next proposition shows that $Q^{**} = 0$ for all $\mathcal{R}_0$ in a neighborhood above 0. For sufficiently small $\mathcal{R}_0$, then, $GS^{**} = 0$ for any market structure.

We can also draw broad conclusions about the optimal subsidy for high values of $\mathcal{R}_0$. Suppose monopoly output is the corner solution $Q_m^* = \hat{S}_0$. Then $\hat{S}_0 = Q_m^* \leq Q^{**} \leq \hat{S}^0$, implying $Q_m^* = Q^{**}$, in turn implying $GS^{**} = 0$ since the first best can be achieved without a subsidy under monopoly. By Table 6, monopoly attains the first best for all $\mathcal{R}_0$ in case (SR4), which the text argued includes an interval of sufficiently high values of $\mathcal{R}_0$. We conclude that for sufficiently high $\mathcal{R}_0$, $GS_m^{**} = 0$. This result also immediately extends to perfect competition or any market structure involving weakly higher output than monopoly.

Having established that $GS^{**} = 0$ for intervals of low and high values of $\mathcal{R}_0$ for general market structures, if it can be shown that $GS^{**} > 0$ for some intermediate value of $\mathcal{R}_0$, it is immediate that GS^{**} is nonmonotonic, attaining a maximum for some interior $\mathcal{R}_0 \in (0, \infty)$ as the next proposition states. The proof provided in Appendix A fills in this and other omitted details.

Proposition 11. *For monopoly—or any market structure involving weakly lower output including perfect competition—GS^{**} is nonmonotonic in $\mathcal{R}_0$, equaling 0 for sufficiently low and sufficiently high $\mathcal{R}_0$, and attaining a positive maximum for some $\mathcal{R}_0 \in (0, \infty)$.*

Turn to a more precise characterization of the optimal subsidy under perfect competition. Suppose the first best is given by an interior solution, i.e., $Q^{**} \in (0, \hat{S}_0)$. Then Q^{**} must satisfy the first-order condition for welfare maximization $MSB(Q^{**}) = c$, implying $MPB(Q^{**}) + MEX(Q^{**}) = c$, in turn implying $P_c^{**} = MPB(Q^{**}) = c - MEX(Q^{**})$. Since competitive firms pass the subsidy through to consumers, $P_c^{**} = c - GS_c^{**}$. Combining the preceding equations yields $GS_c^{**} = MEX(Q^{**})$, the familiar result that setting the subsidy equal to the marginal externality is optimal. If Q^{**} is a corner but Q_c^* is not, i.e., $Q_c^* < \hat{S}_0 = Q^{**}$, then the highest price at which output $\hat{S}_0$ is purchased satisfies $P_c^{**} = MPB(\hat{S}_0)$. Combined with competitive pass through, $P_c^{**} = c - GS_c^{**}$, we have $GS_c^{**} = c - MPB(\hat{S}_0)$. If $Q_c^* = Q^{**} = \hat{S}_0$, then the preceding proposition implies $GS_c^{**} = 0$. The various results for $Q^{**} = \hat{S}_0$ can be nested as $GS_c^{**} = \max[0, c - MPB(\hat{S}_0)]$.

Next, turn to a more precise characterization of the optimal subsidy under monopoly. The monopolist regards the subsidy as a reduction in marginal cost, maximizing $[P(Q) - c + GS]Q = [MPB(Q) - c + GS]Q$. To generate the first best, the optimal subsidy GS_m^{**} must force the monopoly's first-order condition to be satisfied by Q^{**}:

$$MPB(Q^{**}) - c + GS_m^{**} + Q^{**} \frac{\partial MPB(Q^{**})}{\partial Q} = 0. \tag{55}$$

If $Q^{**} \in (0, \hat{S}_0)$, then the analysis of perfect competition showed $MPB(Q^{**}) = P_c^{**} = c - GS_c^{**} = c - MEX(Q^{**})$. Substituting into (55) yields, after rearranging,

$$GS_m^{**} = MEX(Q^{**}) \left(\frac{\hat{S}_0}{\hat{S}_0 - \theta Q^{**}} \right). \tag{56}$$

Equation (56) also uses the fact that $\partial MPB(Q)/\partial Q = -\theta MEX(Q)/S_0(Q)$ for all Q, as can be seen by combining (46) and (50). Equation (56) shows that the monopoly subsidy is proportional to the marginal externality, scaled up by the factor in parentheses, which adjusts for the monopoly markup. This scale factor grows without bound as the first best approaches successful vaccination of all initial susceptibles. Equation (56) also implies that $GS_m^{**} > GS_c^{**}$ since $GS_c^{**} = MEX(Q^{**})$.

If $Q^{**} = \hat{S}_0$, then GS_m^{**} must force (55) to hold at that corner value of quantity. The next proposition summarizes the analysis for the two market structures.

Proposition 12. *If $Q^{**} = 0$, then $GS_c^{**} = GS_m^{**} = 0$. If $Q^{**} \in (0, \hat{S}_0)$, then $GS_c^{**} = MEX(Q^{**})$ and $GS_m^{**} = MEX(Q^{**})\hat{S}_0/(\hat{S}^0 - \theta Q^{**})$. If $Q^{**} = \hat{S}_0$, then $GS_c^{**} = \max[0, c - MPB(\hat{S}_0)]$ and*

$$GS_m^{**} = \max \left[0, c - MPB(\hat{S}_0) + \left(\frac{\theta}{1 - \theta} \right) MEX(\hat{S}_0) \right]. \tag{57}$$

Across all cases, $GS_m^{**} \geq GS_c^{**}$ with strict inequality if $Q^{**} > 0$ and $Q_m^* < \hat{S}_0$.

3.8. Increasing Social Returns

Typical products exhibit concave social benefits. The underlying logic is that initial units provide higher marginal social benefits than subsequent units since highest-value uses are served first, with subsequent units allocated to lower-value uses. Epidemiological externalities may lead this logic to fail with vaccines. Vaccinating a few individuals may do little to slow the spread of an epidemic if susceptibles are likely contract the disease from the many remaining unvaccinated people in any event. Doubling coverage may more than double the social benefit if the additional coverage is needed to make a dent in the infection rate.

In this section, we analyze conditions under which vaccines exhibit increasing rather than diminishing social returns. This issue is particularly relevant to the short-run analysis. For simplicity, the model assumes that any amount of vaccine can be produced at the constant marginal cost c, while in reality, capacity constraints may prevent production up to the point that marginal social benefit equals production cost. Rationing may be required. With the population divided into regional subunits experiencing relatively independent epidemiological processes because of restricted travel flows, it is natural to ask whether vaccine should be spread across regions in proportion to their populations (as considerations of fairness or heterogeneity in value within each region might dictate) or whether the benefits would be larger if vaccine were concentrated in fewer regions (chosen by lottery if urgency of need in certain regions does not provide sufficient reason for concentrating vaccine there). Capacity constraints and rationing were not as relevant to the long-run analysis because the long run presumably provides sufficient time to overcome capacity constraints.

Formally, a vaccine exhibits increasing social returns if $MSB(Q)$ is increasing in Q. Differentiating (49), substituting from (43), and rearranging yields

$$\frac{\partial MSB(Q)}{\partial Q} = \frac{\theta^2 H \mathcal{R}_0 S(Q) S_0(Q) \Phi_I(Q)}{[1 - \mathcal{R}_0 S(Q)]^3} \left\{ \mathcal{R}_0 [S_0(Q) + S(Q)] - 2 \right\}. \tag{58}$$

By Lemma 5, all the factors on the right-hand side are definitively positive except for the last. Thus, the sign of the last factor in braces determines whether the vaccine exhibits increasing social returns. Rearranging gives the following proposition.

Proposition 13. *The Qth unit of vaccine exhibits increasing social returns if and only if*

$$\mathcal{R}_0 \left[\frac{S_0(Q) + S(Q)}{2} \right] > 1. \tag{59}$$

Earlier, we identified the inequality $\mathcal{R}_0 S_0(Q) > 1$ as the condition for epidemic expansion, requiring a combination of sufficient infectiveness, $\mathcal{R}_0$, and sufficient fuel, $S_0(Q)$, for the disease to spread. Condition (59) is more stringent, requiring that the product of $\mathcal{R}_0$—not with the initial number of susceptibles—but with the average of the initial and eventual number of susceptibles exceed 1. One can see this is more stringent since $S(Q) < S_0(Q)$ by Lemma 4. Proposition 13 can thus be interpreted as saying that unit Q of the vaccine exhibits increasing social returns if the potential for epidemic expansion is sufficiently high at the current output level.

The next proposition provides simpler sufficient conditions for the vaccine to exhibit increasing social returns at initial output levels and at all output levels. It is proved in Appendix A as a straightforward corollary of Proposition 13.

Proposition 14. *The vaccine exhibits initial increasing social returns (i.e., at an output level of $Q = 0$) if $\mathcal{R}_0 \hat{S}_0 \geq 2$. The vaccine exhibits everywhere increasing social returns (i.e., at all output levels $Q \in (0, \hat{S}_0)$) if $\mathcal{R}_0 \hat{S}_0 \geq 2/(1-\theta)$.*

According to Proposition 14, if a federal authority only has access to a small stockpile of a vaccine to allocate across several similar states with independent epidemiological processes, allocating the entire stockpile to one state would produce more social benefit than spreading it evenly across them if $\mathcal{R}_0 \hat{S}_0 > 2$. If, for example, $\hat{S}_0 = 0.8$ in each state, then concentrating the vaccine would be efficient for any $\mathcal{R}_0 > 2.5$. If the more stringent condition $\mathcal{R}_0 \hat{S}_0 > 2/(1-\theta)$ holds, then even a starker form of concentration is efficient: not just for very small stockpiles but for any size, the federal authorities should vaccinate all susceptibles in one state before moving to the next. The starkness of the policy hinges on the modeled consumer homogeneity: if each state has some vulnerable consumers with a high benefit from vaccinating, a higher bar on $\mathcal{R}_0$ would need to be cleared for concentrating vaccines in one state to be more efficient than serving high-value consumers everywhere first.

Proposition 14 hints that efficacy θ plays a role in determining whether social returns are everywhere increasing. The next proposition, also proved in Appendix A as a straightforward corollary of Proposition 13, draws a clearer connection.

Proposition 15. *A perfectly effective vaccine ($\theta = 1$) cannot exhibit everywhere increasing social returns.*

The result is intuitive. Protecting all but the last consumer with a perfectly effective vaccine eliminates the externality from vaccinating the marginal consumer, the force behind increasing social returns. Thus, social returns must start to diminish with a perfectly effective at some point before universal vaccination is reached.

Proposition 15 does not automatically imply that a vaccine with $\theta < 1$ should be spread across regions. Marginal social benefits may be diminishing in a region yet still be above that in other regions not yet receiving a vaccine.

3.9. Relevant Limiting Results

This section undertakes the same exercise as undertaken in Section 2.8 for the long-run analysis: examining limits of certain parameters to clarify the results. We take the same limits as before for the subset of parameters carried over from the long-run analysis: $\theta \uparrow 1$ and $c \downarrow 0$. Taking these limits simplifies the results considerably. Only cases (SR3) and (SR4) remain from Tables 5 and 6, and analytic expressions can be obtained for table entries only defined implicitly for general parameters.

It is immediate that the first best is always obtained under perfect competition for the limits considered here since the first best is obtained in (SR3) and (SR4) in Table 5, and these are the only cases remaining in the limits considered here. Under monopoly, the first best is obtained if and only if condition (54) holds. We show in the proof of the next proposition that (54) holds for a disconnected set of the highest and lowest values of $\mathcal{R}_0$:

$$\mathcal{R}_0 \in \left(0, \frac{|\bar{L}(-\hat{I}_0/\hat{S}_0)|}{\hat{I}_0} \right) \cup \left(\frac{|\underline{L}(-\hat{I}_0/\hat{S}_0)|}{\hat{I}_0}, \infty \right), \tag{60}$$

where $\bar{L}$ is the principal branch of the Lambert W function and $\underline{L}$ is the lower branch (see footnote 16 for definitions). It is no surprise that the first best is obtained under monopoly for sufficiently high $\mathcal{R}_0$; we argued this above for general parameters. The surprise in (60) is that for the limiting value of parameters considered here, the first best is also obtained for sufficiently low $\mathcal{R}_0$. Intuitively, for sufficiently low $\mathcal{R}_0$, the disease is not infective enough to break out into an epidemic no matter how many susceptibles are left unvaccinated to fuel the epidemic. The disease peters out with just a few susceptibles mostly catching the disease from the stock of infected individuals at the date of vaccine introduction. With no complicated epidemiological factors at play, the monopolist maximizes profit by vaccinating all the susceptibles. If susceptibles were heterogeneous, the monopolist would raise price by restricting output, but they are homogenous in the model. If the monopolist faced substantial

costs, consumers' willingness to pay for a vaccine against such a low-$\mathcal{R}_0$ disease would not cover costs; but here in the limiting case of costless production the monopolist earns a profit (however slight) by producing.

If (60) does not hold, then output is positive under monopoly but less than the first best. We are able to derive an analytical expression for output and other variables stated in the next proposition.

Proposition 16. *Suppose $\mathcal{R}_0 > 1$ and consider the limits $\theta \uparrow 1$ and $c \downarrow 0$.*

- *Under perfect competition, the first best is obtained. We have $P_c^* = \Pi_c^* = GS_c^{**} = 0$, $Q_c^* = \mu$, $MPB_c^* = MSB_c^* = H(1 - e^{-\mathcal{R}_0 \hat{I}_0})$, and $MEX_c^* = 0$.*

- *Under monopoly, if (60) holds, the first best is obtained, delivering the same outcomes as those listed for perfect compeition, except the monopoly charges a higher price $P_m^* = MPB_m^* = H(1 - e^{-\mathcal{R}_0 \hat{I}_0})$. Otherwise, monopoly falls short of the first best, with $P_m^* = H(\mathcal{R}_0 \hat{S}_0 - 1)/\mathcal{R}_0 \hat{S}_0$,*

$$Q_m^* = \frac{\mu[\mathcal{R}_0(\hat{I}_0 + \hat{S}_0) - 1 - \ln(\mathcal{R}_0 \hat{S}_0)]}{\mathcal{R}_0 \hat{S}_0 - 1}, \tag{61}$$

$GS_m^{**} \uparrow \infty$, and closed-form expressions for other equilibrium variables provided in Appendix A.

As before in the long-run analysis, perfect competition delivers the first best when the vaccine is costless here in the short-run analysis. In contrast, monopoly can now also deliver the first best, and as explained does so for both very low and very high $\mathcal{R}_0$. Even if all other susceptibles purchase a perfectly effective vaccine, the marginal susceptible is induced to purchase by the threat of contracting the disease from those who were infected before the vaccine was introduced. That group is not a source of infections in the long-run analysis because they exit the population by time steady-state sales are being made.

Extreme nonmonotonicity emerges for GS_m^{**}. For the lowest and highest values of $\mathcal{R}_0$, (60) holds, implying that monopoly obtains the first best and hence that $GS_m^{**} = 0$. For intermediate values of $\mathcal{R}_0$, GS^{**} approaches infinity. Thus, as $\mathcal{R}_0$ increases on $(0, \infty)$, GS_m^{**} gyrates from its lowest possible value to its highest possible value back to its lowest possible value.

3.10. COVID Calibration

This section provides an illustrative calibration using parameters drawn from the current COVID-19. Section 2.9 issues a number of caveats about using the calibrations as a forecast, and the same caveats apply here.

Taking limits $\theta \uparrow 1$ and $c \downarrow 0$ from the previous section leaves three important model parameters to be filled in: $\mathcal{R}_0$, $\hat{I}_0$, and $\hat{S}_0$. We will take current estimates of those parameters, calibrating the counterfactual effect of the arrival of a vaccine at the present moment. We use estimates from U.K. government agencies, which provide some of the best estimates for a developed country currently available. Based on U.K. Government Office for Science (2020), we set $\mathcal{R}_0 = 1.5$. Based on U.K. Office for National Statistics (2020), we take the proportion of currently infected to be $\hat{I}_0 = 0.19\%$ and the proportion of recovered to be $\hat{R}_0 = 6.2\%$, implying $\hat{S}_0 = 1 - \hat{I}_0 - \hat{R}_0 = 93.61\%$.

Substituting these parameters into the formulas in Proposition 16, we have that the first best is obtained under perfect competition. Universal vaccination results in all current susceptibles being protected through the end of the pandemic. The condition (60) ensuring that monopoly attains the first best reduces to $\mathcal{R}_0 < 1.1$ or $\mathcal{R}_0 > 4,378$, which for practical purposes can be further reduced to $\mathcal{R}_0 < 1.1$ since the upper interval is outlandishly high. Since the observed value $\mathcal{R}_0 = 1.5 > 1.1$, monopoly does not attain the first best in this calibration. Instead, the monopoly price is set to 41% of harm of contracting the disease. At this price, only 29% of susceptible consumers buy, generating welfare equal to 68% of the available health benefit. The calibrated parameters are in the range in which, according to Proposition 16, it is impossible to induce the first best under monopoly with a finite subsidy.

Since $\mathcal{R}_0 \hat{S}_0 = 1.9$, the simple necessary condition for initially increasing social returns given in Proposition 14, $\mathcal{R}_0 \hat{S}_0 > 2$, is not met. However, the more complicated but weaker condition (59) provided by Proposition 13 does indicate initially increasing returns, as $\mathcal{R}_0 [\hat{S}_0 + S(0)]/2 = 1.16 > 1$. Examing (59) for a range of quantities shows that increasing returns persists through output equal to 36% of the susceptible population. Supposing that a stockpile has to be allocated to two identical states with independent epidemiological processes, concentrating the entire stockpile in one state generates higher welfare than dividing equally until the stockpile exceeds 46% of the population of one state. Larger stockpiles than this are more efficiently divided equally between the states.

4. Consumer Heterogeneity

The models used in the long- and short-run analyses thus far have assumed that consumers are homogeneous. This section introduces consumer heterogeneity. The next subsection demonstrates that key results continue to hold when consumers experience heterogeneous benefits from a vaccine. Subsequent subsections discuss issues including the importance of targeting subsidies and biases in the naïve estimation of the average marginal externality when aggregate data masks underlying

consumer heterogeneity.

4.1. Heterogeneity in Benefits

For concreteness, assume consumers, indexed by i, differ in disease harm, H_i. Similar analysis applies if consumers experience different efficacies θ_i or have different lifespans.[17] Denote the probability density function (pdf) by $f(H_i)$, the cumulative distribution function (cdf) by $F(H_i)$, and the complementary cdf by $\bar{F}(H_i) = 1 - F(H_i)$, and the expected value by $E(H_i) = \int_0^\infty H_i f(H_i) dH_i$. Assume H_i has full support on $(0, \infty)$.

Assume further that the population distribution of H_i is common knowledge but the specific realization of H_i is consumer i's private information. The model requires consumers to be aware of their heterogeneity, for example, differences in income leading to different willingnesses to pay to avoid harm, or a family history of disease. Undiagnosed conditions that lead harm to vary but are unknown to the consumer are better accommodated in the homogeneous-harm model.

In both long- and short-run analyses with homogeneous consumers, we wrote the marginal private benefit as $MPB(Q) = \theta H \Phi_I(Q)$, the product of efficacy, harm, and probability of contracting the disease. With consumer heterogeneity, consumer i's marginal private benefit becomes $MPB_i(Q) = \theta H_i \Phi_I(Q)$.

Incorporating heterogeneity in some of the normative measures requires additional work to keep track of the high-value consumers who end up purchasing. We have

$$SB(Q) = C \left\{ \left[1 - \Phi_I(Q)\right] \int_0^{\hat{H}} H_i f(H_i) dH_i + \left[1 - \Phi_I(Q) - \theta \Phi_I(Q)\right] \int_{\hat{H}}^\infty H_i f(H_i) dH_i \right\}. \tag{62}$$

The first integral reflects the expected health experienced by those whose harm is below the threshold $\hat{H}$ for purchase. With no vaccine to protect them, consumer i in this group obtains H_i with probability $1 - \Phi_I(Q)$. The second integral reflects the expected health experienced by those who purchase. Consumer i in this group obtains H_i if either they would not have been infected anyway (probability $1 - \Phi_I(Q)$) or would have been infected without a vaccine but receive the vaccine protection (probability $\theta \Phi_I(Q)$). The leading factor C allows the per-consumer surplus given by the integrals to be scaled up to the population of potential consumers: $C = \mu$ in the long-run analysis

---

[17] We conjecture that the analysis is also similar if consumers contract the disease at different rates, but modeling heterogeneity in that dimension requires delicacy to avoid changing the epidemiological process.

and $C = \hat{S}_0$ in the short-run analysis. Differentiating (62) yields

$$MSB(Q) = C\left\{-\frac{\partial\Phi_I(Q)}{\partial Q}\left[E(H_i) - \theta\int_{\hat{H}}^{\infty} H_i f(H_i)dH_i\right] + \theta\Phi_I(Q)\hat{H}f(\hat{H})\frac{\partial\hat{H}}{\partial Q}\right\}. \qquad (63)$$

To compute $\partial\hat{H}/\partial Q$, note threshold consumer type $\hat{H}$ is given as an implicit function of Q by $Q = C\bar{F}(\hat{H})$. Totally differentiating this identity with respect to Q and rearranging yields $\partial\hat{H}/\partial Q = 1/Cf(\hat{H})$. After substituting this derivative into (63), we see that the last term equals $\theta\hat{H}\Phi_I(Q)$. This is the private benefit of the threshold consumer, equal to MPB^* when evaluated at the equilibrium Q^*. Subtracting to compute $MEX^* = MSB^* - MPB^*$ leaves just the first term of (63), as stated in the following lemma.

Lemma 9. *In the model with heterogeneity in consumer harm H_i, the marginal externality in both the long- and short-run analyses equals*

$$MEX^* = -\frac{\partial\Phi_I(Q^*)}{\partial Q}\left[E(H_i) - \theta\int_{\hat{H}(Q^*)}^{\infty} H_i f(H_i)dH_i\right]C. \qquad (64)$$

Intuitively, Lemma 9 says that the marginal externality is proportional to $-\partial\Phi_I(Q^*)/\partial Q$, the decline in the equilibrium probability of infection for an unvaccinated individual when one additional susceptible is vaccinated. The proof of the next proposition shows that that leading factor approaches 0 as $\mathcal{R}_0 \downarrow 0$ in both long- and short-run analyses since a noninfectious disease presents no danger of infection in either analysis. The factor also approaches 0 as $\mathcal{R}_0 \uparrow \infty$ in both analyses since the individual will almost certainly contract the infinitely infectious disease in any event— from someone who was vaccinated but for whom the vaccine was ineffective if no one else. The remaining factors are obviously positive and finite for all $\mathcal{R}_0$. Thus, MEX^* approaches 0 for extreme values of $\mathcal{R}_0$, implying it is nonmonotonic in $\mathcal{R}_0$, as the following proposition states. The proof in Appendix A fills in omitted details.

Proposition 17. *In the model with heterogeneity in consumer harm H_i, MEX^* is nonmonotonic in both the long- and short-run analyses, under both perfect competition and monopoly.*

4.2. Targeting Subsidies

While various forms of consumer heterogeneity raise common issues, they can have subtle differences. When consumers differ in benefits obtained from a purchased product, the support of heterogeneity is bounded between zero and the threshold benefit for the marginal purchaser. A modest

subsidy covering purchase price may be sufficient to induce universal consumption. When consumers differ in personal costs of mitigation measures, non-adopters may have unboundedly high personal costs, requiring an inordinate subsidy to induce consumption. Targeting and other measures may help restrain program expenditures.

Consider an example in which vaccines are supplied for free by perfectly competitive firms producing at zero marginal cost. The only remaining cost is a personal cost c_i of receiving a vaccination. Such heterogeneity is plausible. The average cost of receiving a polio vaccine, for instance, is probably low; but for those remaining unvaccinated, since they decline the vaccine despite severe disease consequences, revealed preference suggests c_i may be high, perhaps reflecting travel difficulties, low tolerance for pain, or "anti-vax" sentiment. Using a uniform subsidy to induce more of these high-cost consumers to become vaccinated could require paying a high subsidy to inframarginal consumers, amplifying program expense, perhaps explaining the reluctance of authorities to endorse this approach more widely. In our model, the subsidy should be provided regardless of program expense as long as it moves quantity in the direction of the first best. In practical settings, authorities may face unmodeled frictions in raising or dispersing public funds making program expenses a relevant consideration.

While heterogeneity in c_i combined with practical program frictions could weaken the case for subsidies, it seems unlikely to overturn it. Calibration suggests that subsidies would be warranted, if they could be targeted, even very imprecisely, to people or areas with low vaccine take-up. Consider a "lentils for vaccination" program along the lines studied in India by Banerjee *et al.* (2010), here calibrated with conservative, round numbers for parameters for ease of analysis. Suppose the program provides $10 worth of lentils per child immunized (much more than the amount actually provided in India). Suppose that for every marginal child immunized, nine inframarginal children receive the subsidy and that either this subsidy is a pure social cost or that an additional dollar of deadweight loss is incurred for every dollar of lentils distributed. Thus, $100 of subsidy must be paid out for each new immunization induced. Even in this case, the program would still be justified as long as the social benefit of vaccination exceeds $100 (plus several dollars of marginal manufacturing and delivery costs of the vaccine plus any disutility of vaccination). Vaccine campaigns can be so cost effective that they remain so even if their costs are magnified, as one can argue in the case of the rollout of pneumococcal conjugate vaccine in countries eligible for support from GAVI (formerly the Global Alliance for Vaccines and Immunizations).[18] Similar exercises could be con-

---

[18]Sinha *et al.* (2007) put the purchase and program administration cost of a three-dose course of a 13-valent pneumococcal conjugate vaccine (PCV13) at $18 per course. They estimate a total cost in the 72 GAVI-eligible countries, after

ducted to determine the acceptable "leakage" rate in programs subsidizing mosquito nets or chlorine for water purification.

4.3. Aggregation Estimation on Heterogeneous Groups

This section discusses conditions under which a naïve estimate that does not take into account population heterogeneity can recover the population means of the marginal externality and other equilibrium variables.

To derive formal results, consider a simple setting of a population consisting of epidemiologically distinct subgroups indexed by $g \in G$. Assume the subgroups sharing all parameters in common except for subgroup size, denoted w_g, with $\sum_{g \in G} w_g = 1$, and disease transmissibility β_g, generating different values of the basic reproductive ratio, $\mathcal{R}_{0g}$ according to the relevant formula in the long- or short-run analysis. A researcher, unaware of the subgroup heterogeneity, seeks to estimate $\mathcal{R}_0$ and *MEX* using aggregate data. Assume no vaccine is yet available—skirting a potentially awkward situation in which the expert knows less about the disease than ordinary consumers, who are required to by our economic model to have rational expectations about the epidemiology of the disease when making their consumption decisions. The researcher seeks an estimate of the externality associated with a first dose of possible vaccine to be developed.

In the long-run variant of the analysis, the researcher estimates $\mathcal{R}_0$—in what would be a model-consistent way absent consumer heterogeneity—by substituting an observation of aggregate, steady-state disease prevalence *I* and known values of parameters besides $\mathcal{R}_0$ into equation (8) and solving for $\mathcal{R}_0$. In the short-run variant of the analysis, the researcher derives their estimating equation from (31), which upon rearranging and substituting $t = Q = 0$ yields

$$\mathcal{R}_0 = \frac{\beta}{\alpha} = 1 + \frac{1}{\alpha \hat{S}_0} \left(\frac{\dot{I}_0}{I_0} \right). \tag{65}$$

In this case, the research estimates $\mathcal{R}_0$ by substituting an observation of the percentage change in prevalence shortly after date 0, along with known values of parameters α and $\hat{S}_0$, into (65). The estimates of $\mathcal{R}_0$ can be translated into an estimate of *MEX* using the formulas from relevant tables.

---

netting out estimated $1.22 billion savings in averted medical costs, of $9.16 billion. The program would save 119.6 million disability adjusted life years (DALYs) at a cost of $77 per DALY (see their Table 5). After adding $100 to the cost of each course from an imperfectly targeted vaccines-for-lentils program, one can show the cost per DALY is $558 per DALY, about a quarter of the World Health Organization (WHO) threshold for highly cost effective intervention (as outlined in Marseille *et al.* 2015, one times per capita GDP, currently about $2,000 in India) and a tenth of the WHO threshold for a cost effective intervention (three times per capita GDP).

We relegate the analysis to Appendix C. Among other details, to isolate the effect of consumer heterogeneity, the original epidemiological model needs to be normalized to avoid the outcome that separation itself—even of a homogeneous population—has real effects on $\mathcal{R}_0$. If the per-capita contact rate is not adjusted, a reduction in the size of the interacting population in an SIR model reduces per-capita contacts.

To summarize the results, the researcher obtains an estimate of the mean of *MEX* across groups in both long- and short-run analysis with some caveats. In the long-run analysis, the researcher obtains an unbiased estimate of the weighted arithmetic mean of *MEX* across groups but only if $\mathcal{R}_{0g} > 1$ for all $g \in G$. If $\mathcal{R}_{0g} < 1$ for some g, then the researcher will overestimate mean *MEX*. The discontinuity in the functional relationship between $\mathcal{R}_{0g}$ and MEX_g at $\mathcal{R}_{0g} = 1$ throws off the calculation of the mean. Appendix C relates the overestimate to the share of the population in subgroups g with $\mathcal{R}_{0g} \leq 1$, providing a simple bias correction if that share is known.

In the short-run analysis, the researcher obtains an unbiased estimate of the weighted mean of MEX_g with no restrictions on $\mathcal{R}_{0g}$. However, the relevant mean is not the arithmetic mean of MEX_g but the harmonic mean.

5. Vaccines Versus Drugs

Commentators on the pharmaceutical industry frequently suggest that firms are biased in favor of developing drugs rather than vaccines. Kremer and Snyder (2016) list a variety of reasons for this bias, ranging from vaccines' complexity relative to drug molecules, to the scale often needed for vaccine clinical trials, to the evaporation of consumers' private disease-risk information when making drug purchases (that paper's focus).

The epidemiological externality analyzed in this paper provides another rationale. By preventing individuals from becoming infected, vaccines curtail their transmission of the disease to others. The reduction in others' disease risk is a public good that reduces others' willingness to pay for a vaccine. This public-good feature distinguishes vaccines from some drugs that treat symptoms without curing the underlying disease or inhibiting transmission. Firms would have more of incentive to develop a drug that does not have this demand-reducing public-good feature than a similarly effective vaccine.[19]

---

[19]If one takes the focus on steady states in the long-run analysis as literally implying that the discount rate is zero, then there would not be any bias in development decisions. All products producing any positive flow profit would be developed regardless of how large is the up-front development cost; the flow always swamps the up-front cost at a zero discount rate. We are taking the steady-state profit and welfare differentials as approximations of the present discounted

To quantify a monopoly's bias toward a drug and against a vaccine, consider a drug that is similar in all ways to the vaccine analyzed to this point except that the drug does not reduce disease transmission. Finding the right normalization to make drug and vaccine costs equivalent is somewhat delicate since at equal marginal production costs c the total cost of serving a population with a drug is lower if it only needs to be administered to infected consumers rather than the whole population in advance as with a vaccine. We finesse this normalization issue by assuming both products are costless to produce and administer, i.e., $c = 0$. Assume the drug is effective with probability θ. Efficacy for the drug means it eliminates any harm from the symptoms experienced by infected individuals but does not prevent them from transmitting the disease to susceptible individuals. One course of the drug is sufficient to eliminate symptoms for the rest of the consumer's life. If this first course is ineffective for an individual, further courses will be ineffective for that individual as well.

We start by computing monopoly profit and welfare from a drug, respectively Π^*_{md} and W^*_{md}, in steady-state equilibrium in the long-run analysis. If $\mathcal{R}_0 \leq 1$, the disease naturally dies out in the steady-state, implying $Q^*_{md} = \Pi^*_{md} = 0$. If $\mathcal{R}_0 > 1$, the monopolist can charge $P^*_{md} = \theta H$ for the drug to all newly infected consumers each instant. According to equation (3), new infections, βIS, and removals from the infected population, $(\alpha + \mu)I$, must balance each instant to maintain $\dot{I} = 0$ in the steady state. Hence, we can compute new infections as $(\alpha + \mu)I(0)$, where the argument added to $I(0)$ indicates that the drug does nothing to curtail infections, as in the vaccine model with no vaccine sales. Equilibrium drug quantity is thus $Q^*_{md} = (\alpha + \mu)I(0) = \mu(1 - 1/\mathcal{R}_0)$ by (8). Given production is costless,

$$\Pi^*_{md} = P^*_{md} Q^*_{md} = \mu\theta H\left(1 - \frac{1}{\mathcal{R}_0}\right). \tag{66}$$

Welfare is

$$W^*_{md} = h[1 - I(0) + \theta I(0)] = H\left[\alpha + \mu - \mu(1-\theta)\left(1 - \frac{1}{\mathcal{R}_0}\right)\right]. \tag{67}$$

Comparing these expressions against the analogous entries in Table 2 for a vaccine leads to the next proposition. The proposition uses the notation $\Delta\Pi^*_m = \Pi^*_{md} - \Pi^*_{mv}$ and $\Delta W^*_m = W^*_{md} - W^*_{mv}$ for differences between equilibrium variables for the two products and $\Delta W^{**} = W^{**}_d - W^{**}_v$ for difference between first-best welfare.

Proposition 18. *Consider the long-run analysis with $c = 0$. For all $\mathcal{R}_0 > 0$, $\Delta W^{**} \leq 0$ and $\Delta\Pi^*_m \geq 0$, with strict inequality if and only if $\mathcal{R}_0 > 1$. $\Delta\Pi^*_m$ is quasiconcave in $\mathcal{R}_0$, with $\lim_{\mathcal{R}_0 \downarrow 0} \Delta\Pi^*_m = \lim_{\mathcal{R}_0 \uparrow \infty} \Delta\Pi^*_m = \inf_{\mathcal{R}_0 > 0} = 0$, reaching an maximum of $\theta^2 H/(1-\theta)$ at $\mathcal{R}_0 = (1+\theta)^2$. $W^*_m < 0$ if and only if $\mathcal{R}_0 > [\theta/(1-\theta)]^2$.*

value of streams with a positive discount rate.

The proposition states that the monopoly prefers to develop the drug over the vaccine as long as there is a nontrivial market for the products ($\mathcal{R}_0 > 1$). However, if $\mathcal{R}_0 > [\theta/(1-\theta)]^2$, social welfare is higher with a vaccine, implying that the monopoly is biased toward the "wrong" product for sufficiently high $\mathcal{R}_0$. While the drug has the advantage that the monopolist sells the first-best quantity in equilibrium, a drug dose is socially inferior to a vaccine dose because the drug offers no positive externality. Like other variables studied so far that capture the impact of the epidemiological externality on economic outcomes (including the marginal externality and government subsidy), here the magnitude of the monopoly's bias toward the "wrong" product, as quantified by $\Delta\Pi_m^*$, is nonmonotonic in $\mathcal{R}_0$, greatest for some interior value. The externality disappears if the disease is noninfective and is swallowed by consumers' private benefit if the disease is infinitely infective.

Moving to the short-run analysis, for all $\mathcal{R}_0 > 0$, the drug monopoly can charge $P_m^* = \theta H$ to the $\hat{I}_0$ individuals infected at the moment the drug is developed as well as the $\hat{S}_0 - S(0)$ individuals who become infected at some point afterwards, yielding drug profit

$$\Pi_{md}^* = \theta H \left[\hat{I}_0 + \hat{S}_0 - S(0) \right]. \tag{68}$$

To compute equilibrium welfare with a drug, the $\hat{I}_0$ individuals infected initially along with the $\hat{S}_0 - S(0)$ infected later obtain health benefit H with probability θ from the drug. The $S(0)$ remaining susceptibles are never infected and obtain health benefit H with certainty, yielding the following expression for equilibrium welfare after rearranging:

$$W_{md}^* = H \left[(1-\theta)S(0) + \theta(\hat{I}_0 + \hat{S}_0) \right]. \tag{69}$$

Comparing these expressions against the analogous entries in Table 6 for a vaccine leads to the next proposition. Details are provided in the proof of this and the previous proposition, provided in Appendix A.

Proposition 19. *Consider the short-run analysis with $c = 0$. For all $\mathcal{R}_0 > 0$, $\Delta\Pi_m^* > 0$. $\Delta\Pi_m^*$ is nonmonotonic in $\mathcal{R}_0$, with $\lim_{\mathcal{R}_0 \downarrow 0} \Delta\Pi_m^* = \lim_{\mathcal{R}_0 \uparrow \infty} \Delta\Pi_m^* = \inf_{\mathcal{R}_0 > 0} = \theta H \hat{I}_0$. For all parameters besides $\mathcal{R}_0$, $\lim_{\mathcal{R}_0 \downarrow 0} W_m^* > 0$ and $\lim_{\mathcal{R}_0 \uparrow \infty} W_m^* > 0$. There there exist parameters including $\mathcal{R}_0$ for which $W_m^* < 0$.*

According to the proposition, many of the points of comparison between a drug and vaccine from the long-run analysis carry over to the short-run analysis. The monopoly is biased toward the drug for all parameters, and this bias leads the firm to choose the socially inferior product for some

parameters.

A new feature of the short-run analysis is that equilibrium welfare never falls below $\theta H \hat{I}_0$ for a drug monopoly, even for extreme values of $\mathcal{R}_0$. Administering a drug to the $\hat{I}_0$ initially infected provides a social benefit even if $\mathcal{R}_0$ is so low that the infection does not spread to others. A vaccine cannot provide this social benefit because it is useless unless administered prior to infection in the model. Thus, equilibrium welfare is higher with a drug than vaccine in the limit $\mathcal{R}_0 \downarrow 0$. Equilibrium welfare is also higher with a drug than vaccine in the limit $\mathcal{R}_0 \uparrow \infty$. The externality associated with vaccine disappears because susceptibles are certain to contract the disease, if no one else, from an unsuccessfully vaccinated person with an infinitely infective disease. Hence, apart from the drug's remaining social benefit of treating the $\hat{I}_0$ initially infected, the drug and vaccine provide equal welfare in the limit $\mathcal{R}_0 \uparrow \infty$. The opposing welfare factors—the drug helps initially infected but the vaccine reduces subsequent spread to others—prevent many firm conclusions from being drawn about the sign of the equilibrium or first-best welfare differentials.

6. Conclusion

We analyzed the market for technologies preventing individuals from contracting a disease. Such products are interesting since, by preventing the consumer from contracting a disease, they exert a positive externality, reducing the spread to others. Though the analysis applies to a variety of technologies such as circumcision, bed nets, or social distancing, the discussion focused on vaccines for concreteness. Vaccines (and other aforementioned technologies) are not pure public goods since they are physical products that exhibit rivalry and excludability in consumption, yet they share with public goods the feature that one's consumption reduces others' demand for that product, a feature that can potentially lead to large distortions in firms' supply decisions.

Such distortions and policy correctives were the focus of this paper. To study them, we constructed and analyzed a theoretical model of the vaccine market involving economic agents that base their consumption and production decisions on rational expectations of the disease's evolution consistent with a standard SIR epidemiological model. We provided two separate analyses. Our long-run analysis studied of the steady-state equilibrium for an endemic disease such as HIV. Our short-run analysis studied a vaccine campaign launched at one point in time along the path of an epidemic disease such as Covid-19 that is expected to wane before population turnover becomes a relevant issue. We sought to provide a comprehensive account of equilibrium variables such as price, quantity, profit, and welfare across a variety of market structures ranging from perfect competition

to Cournot to monopoly and to study how those variables changed in response to parameter changes.

A key variable was the equilibrium marginal externality. We consistently found—across both long- and short-run analyses, across a range of market structures, and across models with either homogeneous or heterogeneous consumers—that the equilibrium marginal externality is nonmonotonic in the infectiousness of the disease as measured by $\mathcal{R}_0$. For low levels of $\mathcal{R}_0$, one consumer's vaccination provides little benefit to others because there is little chance the consumer would have infected them anyway. For high levels of $\mathcal{R}_0$, one consumer's vaccination provides little benefit to others because they will most likely contract it from a different source anyway. The marginal externality is greatest for intermediate values of $\mathcal{R}_0$. This nonmonotonicity carries over to other outcome variables such as GS^{**} (the minimal subsidy necessary to obtain the first-best vaccine quantity) and $\Delta\Pi_m^*$ (the difference between monopoly profit from a drug that does not exert the epidemiological externality and from a vaccine that does). Diseases with moderate infectiousness may exhibit the greatest distortions and be prime targets for subsidy.

In calibrations, we found that strong cases can be made for subsidizing competitively supplied technologies such as the example of circumcisions to prevent HIV. However, with a monopoly supplier, the per-unit subsidy needed to obtain the first best can be prohibitively expensive. For example, in a calibration of the long-run analysis to measles, we found that the subsidy required to obtain the first best would be fifteen times the harm from certainly contracting the disease. With even a small deadweight loss of taxation, such a subsidy would fail a cost-benefit test. In the short-run analysis, Proposition 16 showed that in the limit of a costless and perfectly effective vaccine, if a monopolist would not produce the first best in equilibrium without government intervention, it cannot be induced to produce the first best by any finite per-unit subsidy. Our analysis thus suggests that, in many cases, governments may either need to give up on the first best as a feasible target for a per-unit subsidy or need to consider other policies such as purchasing vaccines in bulk at a negotiated price building in a lump-sum subsidy.

If the negotiated price reflects the threat point of decentralized vaccine sales to individuals on the private market, our results on the relative profitability of vaccines versus drugs gain relevance. If the positive epidemiological externality causes a vaccine to be less lucrative than a drug lacking that externality, the monopoly's threat point may be worse with a vaccine, implying that the hold-up problem and consequent underinvestment may be worse with a vaccine than a drug. To address the hold-up problem, governments may have to consider negotiating with vaccine manufacturers ex ante, prior to substantial investment in R&D and production capacity.

While the long- and short-run analyses shared many key qualitative findings, there were points of contrast. Many of these points of contract can be traced to the presence of the $\hat{I}_0$ infected individuals had yet recovered at vaccine rollout and thus capable of transmitting disease to unvaccinated. This group is irrelevant in the long-run analysis because recovery and population turnover removes them from the steady state. The presence of this group leads to the possibility that universal vaccination of susceptibles with a perfectly effective vaccine can be a viable business strategy. Indeed, in the limiting case of a costless vaccine, Proposition 16 showed this strategy is optimal for a monopoly for both sufficiently low and sufficiently high $\mathcal{R}_0$. The presence of this group raises the possibility that a vaccine with a positive epidemiological externality can be welfare-dominated by a drug without it: if the externality is small, welfare may be driven by the advantage of the drug in treating the $\hat{I}_0$ infecteds for whom the vaccine arrives too late to help, assuming the vaccine must be administered prior to infection to be effective.

The short-run analysis provides fertile ground for understanding when vaccination exhibits increasing social returns. According to Proposition 13, a vaccine exhibits initially increasing social returns if $\mathcal{R}_0 \hat{S}_0 \geq 2$ and everywhere increasing social returns if $\mathcal{R}_0 \hat{S}_0 \geq 2/(1-\theta)$. If the first condition holds, a small capacity should be concentrated in a single region; and if the second condition holds, a first region should be completely served before moving to a second regardless of capacity size. These stark implications for concentrating supplies hinge on the homogeneity of consumers in the model but raise the possibility o equitable allocation leading to inefficiency.

In our calibration to a Covid-19 vaccine, we found that the first best would be approached by a competitively supplied vaccine but a monopoly (selling to individuals on the private market, not bulk sales to governments) would set such a high price that only 29% of susceptibles would buy. A vaccine would have increasing social returns not everywhere but through 36% of the susceptible population. The presence of increasing social returns argues for subsidizing aggressive investment to boost capacity beyond this point if concentrating supplies in few countries is either unpalatable or outweighed in by the benefit of vaccinating vulnerable subpopulations in every country.

Appendix A. Proofs

This appendix supplies proofs not included in the text.

Proof of Proposition 1

Suppose $I^* = 0$ and $Q^* > 0$. We will show $P^* = 0$. Substituting $I^* = 0$ into (8) and rearranging yields $Q^* \geq (\mu/\theta)(1 - 1/\mathcal{R}_0)$. This inequality together with $Q^* > 0$ implies $Q^* \geq \mathcal{Q}_0$ by (12), implying $MPB^* = 0$ by (16). Since $MPB(Q)$ is weakly decreasing in Q, $MPB(\mu) \leq MPB^* = 0$. Then $Q^* = D(P^*)$ cannot be determined by the third branch of equation (18); if it were, $P^* < MPB(\mu) = 0$, violating the nonnegativity of prices. Since $Q^* > 0$, Q^* cannot be determined by the first branch of (18) either. Therefore,

$$Q^* = d(P^*) = \frac{\mu}{\theta}\left[1 - \frac{1}{(1 - P/\theta H)\mathcal{R}_0}\right]. \tag{A1}$$

Since $0 = MPB^* = MPB(Q^*)$, we have

$$Q^* \geq \mathcal{Q}_0 \geq \frac{\mu}{\theta}\left(1 - \frac{1}{\mathcal{R}_0}\right), \tag{A2}$$

where the first inequality follows from equation (16) and the second from (12). Combining (A1) with (A2) and rearranging yields $P^* \leq 0$, implying $P^* = 0$ by the nonnegativity of prices. *Q.E.D.*

Verification of Table 1 Entries

As a preliminary step, we will verify that the intervals in the column headings are ordered as given. The fact that $c > 0$ implies $\tilde{c} > 0$. Assumption (23) implies $\tilde{c} < 1$. Hence $\tilde{c} \in (0, 1)$, implying $1/(1 - \tilde{c}) > 1$, implying the case (LR2) interval for $\mathcal{R}_0$ is to the right of the case (LR1) interval. The ordering of the cases is obvious if $\theta < 1$. If $\theta = 1$, then cases (LR4) and (LR5) fail to exist.

Turning to the individual cases, in case (LR1), $\mathcal{R}_0 \leq 1$ implies $\mathcal{Q}_0 = 0$ by equation (12). Hence, $MPB(Q) = MSB(Q) = MEX(Q) = 0$ for all $Q \geq 0$ by equations (16), (21), and (22). The fact that $MPB(Q) = 0$ for all $Q \geq 0$ implies $Q_c^* = D(c) = 0$ for $c > 0$ by (18). Equilibrium welfare is $W_c^* = SB_c^* - cQ_c^* = (\alpha + \mu)H$ by (19).

In case (LR2), $\mathcal{R}_0 > 1$ implies $\mathcal{Q}_0 > 0$ by (12). Thus, by (16),

$$MPB(0) = \theta H(1 - 1/\mathcal{R}_0) < \theta H\tilde{c} = c, \tag{A3}$$

where the inequality follows from $\mathcal{R}_0 < 1/(1 - \tilde{c})$. We can verify the claim in the text that $MPB(Q)$ is weakly decreasing in Q by differentiating (16):

$$-\frac{\theta^2 H}{\mu \mathcal{R}_0}\left(1 - \frac{\theta Q}{\mu}\right)^{-2} < 0, \tag{A4}$$

implying that $MPB(Q)$ is weakly decreasing in Q. Therefore, $MPB(Q) \leq MPB(0) < c$, implying $Q_c^* = D(c) = 0$ by (18).

Skipping over case (LR3) to cases (LR4) and (LR5), assume $\theta < 1$, so that these cases exist. We

have

$$\mathcal{R}_0 > 1/(1-\theta)(1-\tilde{c}) > 1/(1-\tilde{c}), \tag{A5}$$

implying $\mathcal{Q}_0 > \mu$ by (12). Since $Q \leq \mu < \mathcal{Q}_0$, $MPB(Q)$ is given by the second branch in (16), implying

$$MPB(Q) \geq MPB(\mu) \geq \theta H \left[1 - \frac{1}{(1-\theta)\mathcal{R}_0} \right] \geq \theta H[1-(1-\tilde{c})] = c, \tag{A6}$$

where the first step follows from the fact that $MPB(Q)$ is weakly decreasing, the second from substituting $Q = \mu$ into (16), the third step from (A5), and the last from of $\tilde{c} = c/\theta H$. Thus, $Q_c^* = \mu$ by (18). Since $Q < \mathcal{Q}_0$, the relevant branch in equations (19)–(22) for $SB(Q)$, $MSB(Q)$, and $MEX(Q)$ is the second. Substituting $Q_c^* = \mu$ into these equations gives SB_c^*, MSB_c^*, and MEX_c^*. Substituting Q_c^* and SB_c^* into the welfare formula gives W_c^*.

In case (LR3), $\mathcal{R}_0 \geq 1/(1-\tilde{c})$ implies $MPB(0) \geq c$ by (A3). Further, $\mathcal{R}_0 \leq 1/(1-\theta)(1-\tilde{c})$ implies $MPB(\mu) \leq c$ by (A6). Hence, by (18), $Q_c^* = d(c)$. Substituting c into (17) yields the table entry for Q_c^*. Using (12), one can show $Q_c^* < \mathcal{Q}_0$, implying that the relevant branch in equations (19)–(22) for $SB(Q)$, $MSB(Q)$, and $MEX(Q)$ is the second. The equilibrium value of the remaining variables can be derived as in the previous paragraph.

Since $P_c^* = c$ in all cases, $\Pi_c^* = 0$ in all cases. *Q.E.D.*

Proof of Proposition 2

The proof for variables P_c^*, Π_c^*, and MSB_c^*, which are constant over relevant intervals of $\mathcal{R}_0$, are obvious from Table 1.

The reader can verify that Q_c^* is continuous at the boundaries between cases in Table 1. The table entries for Q_c^* are constant for all cases except (LR3). In this non-empty case, Q_c^* is strictly increasing in $\mathcal{R}_0$. Variables MPB_c^* and W_c^* are analyzed similarly.

The reader can verify that I_c^* is continuous at the boundary between cases in the table and further verify that I_c^* is constant in $\mathcal{R}_0$ in case (LR1), increasing in case (LR2), decreasing in case (LR3), and increasing in cases (LR4) and (LR5). A local optimum is thus attained at the boundary between cases (LR2) and (LR3). Substituting $\mathcal{R}_0 = 1/(1-\tilde{c})$ in the case (LR2) table entry yields $I_c^* = \tilde{c}\mu/(\alpha+\mu)$. The only other candidate for a global maximum is

$$\lim_{\mathcal{R}_0 \uparrow \infty} I_c^* = \frac{(1-\theta)\mu}{\alpha+\mu}. \tag{A7}$$

This is less than the local optimum on the boundary between cases (LR2) and (LR3) if and only if $\tilde{c} > 1-\theta$.

The reader can verify that MEX_c^* is continous at the case (LR2)–(LR3) boundary and at the case (LR3)–(LR4) boundary and further verify that MEX_c^* is decreasing or constant in $\mathcal{R}_0$ in cases (LR2)–(LR5). Hence, MEX_c^* approaches its supremum $\lim_{\mathcal{R}_0 \downarrow 1} MEX_c^* = 1$ at the boundary between cases (LR1) and (LR2). *Q.E.D.*

Proof of Proposition 3

Preliminaries: The proof for MSB_m^*, which is constant within relevant intervals of $\mathcal{R}_0$, is obvious from Table 2. We omit the comparative-static analysis of Q_m^* as it is similar to that for Q_c^* in the

previous proof. We also omit the comparative-static analysis of MPB^*_m and Π^*_m, which are similar to that for Q^*_m. We also omit the comparative-static analysis of MEX^*_m, which is similar to that for MEX^*_c in the previous proof.

Results for P^*_m: Form a price function out of the correspondence in Table 2 by taking the determinate prices in cases (LR3)–(LR5) and extending this function by taking $P^*_m = c$ in cases (LR1) and (LR2) in which the equilibrium monopoly price is indeterminate. One can then proceed to analyze the comparative statics of this price function in a similar way to variables in the preceding paragraph.

Results for I^*_m: The reader can verify that I^*_m is continuous at the boundary between cases in Table 2 and further verify that I^*_m is constant in $\mathcal{R}_0$ in case (LR1), increasing in case (LR2), and increasing in case (LR5). This leaves cases (LR3) and (LR4). Differentiating the table entry for those cases,

$$\frac{\partial I^*_m}{\partial \mathcal{R}_0} = \frac{\mu}{(\alpha+\mu)\mathcal{R}_0^2}\left(1-\frac{1}{2}\sqrt{\frac{\mathcal{R}_0}{1-\tilde{c}}}\right). \tag{A8}$$

Equation (A8) is decreasing in $\mathcal{R}_0$. It is negative for all $\mathcal{R}_0$ in cases (LR3) and (LR4) if it is nonpositive at the lower boundary of case (LR3). Evaluating (A8) at this lower boundary $\mathcal{R}_0 = 1/(1-\tilde{c})$, we see it is nonpositive if and only if $\tilde{c} \geq 1/2$. But then I^*_m is nonmonotonic, for it is increasing in case (LR2) and decreasing in cases (LR3) and (LR4). It reaches a local maximum of $\tilde{c}\mu/(\alpha+\mu)$ at the boundary between cases (LR2) and (LR3). This weakly exceeds $\lim_{\mathcal{R}_0\uparrow\infty} I^*_m = (1-\theta)\mu/(\alpha+\mu)$ if $\tilde{c} \geq 1-\theta$, in which case the local maximum is a global maximum.

Assume $\tilde{c} < 1/2$. Then (A8) is nonnegative for all $\mathcal{R}_0$ in cases (LR3) and (LR4) if and only if (A8) is nonnegative at the upper boundary of case (LR4). Evaluating (A8) at this upper boundary $\mathcal{R}_0 = 1/(1-\theta)^2(1-\tilde{c})$, we see it is nonnegative if and only if $\tilde{c} \leq (1-2\theta)/2(1-\theta)$. Under these conditions, I^*_m is nondecreasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$.

For the remaining parameters, I^*_m reaches a local maximum in the interior of cases (LR3) and (LR4). Setting (A8) to 0 and solving implies that the local maximum is at $\mathcal{R}_0 = 4(1-\tilde{c})$. This local maximum equals $\mu/4(\alpha+\mu)(1-\tilde{c})$, which exceeds $\lim_{\mathcal{R}_0\uparrow\infty} I^*_m = (1-\theta)\mu/(\alpha+\mu)$ if $\tilde{c} \geq (3-4\theta)/4(1-\theta)$, in which case the local maximum is a global maximum.

Results for W^*_m: The reader can verify that W^*_m is continuous at the boundary between cases in Table 2 and further verify that W^*_m is constant in $\mathcal{R}_0$ in case (LR1), decreasing in case (LR2), and decreasing in case (LR5). This leaves cases (LR3) and (LR4). Differentiating the table entry for those cases,

$$\frac{\partial W^*_m}{\partial \mathcal{R}_0} = \frac{\mu H}{\mathcal{R}_0^2}\left(\frac{1}{2}\sqrt{(1-\tilde{c})\mathcal{R}_0}-1\right). \tag{A9}$$

The sign of (A9) depends on the factor in parentheses, which is increasing in $\mathcal{R}_0$. It is nonpositive for all $\mathcal{R}_0$ in cases (LR3) and (LR4) if and only if it is nonpositive at the upper boundary of case (LR4). Evaluating (A8) at this upper boundary $\mathcal{R}_0 = 1/(1-\theta)^2(1-\tilde{c})$, we see it is nonpositive if and only if $\theta \leq 1/4$. Otherwise, W^*_m reaches a local minimum in the interior of cases (LR3) and (LR4). Setting (A9) to 0 and solving yields a local minimum at $\mathcal{R}_0 = 4/(1-\tilde{c})$. This local minimum equals $H[\alpha+(3/4)(1-\tilde{c})\mu]$, which weakly exceeds $\lim_{\mathcal{R}_0\uparrow\infty} W^*_m = H[\alpha+\mu\theta(1-\tilde{c})]$ if $\theta \geq 3/4$, in which case the local minimum is a global minimum. *Q.E.D.*

Proof of Proposition 4

Suppose $\tilde{c} \geq 1/2$. Then I_m^* achieves an interior maximum regardless of θ by Proposition 3. So suppose $\tilde{c} < 1/2$. Then Proposition 3 states that I_m^* achieves an interior maximum if $\tilde{c} > (1-2\theta)/2(1-\theta)$, or rearranging, $\theta > (1-2\tilde{c})/2(1-\tilde{c})$. The right-hand side of this last inequality is decreasing in $\tilde{c}$. Substituting $\tilde{c} = 0$ gives the sufficient condition $\theta > 1/2$ for I_m^* to achieve an interior maximum. Proposition 3 states that $\theta > 1/4$ is sufficient for W_m^* to achieve an interior minimum. Thus, $\theta > 1/2$ is sufficient for both I_m^* and W_m^* to achieve interior optima.

Suppose $\tilde{c} \geq 1/2$. Proposition 3 states that I_m^* achieves a interior global maximum if $\tilde{c} \geq 1-\theta$, or rearraning, if $\theta \geq 1-\tilde{c} \geq 1/2$. Suppose instead that $\tilde{c} < 1/2$. Proposition 3 then states that I_m^* achieves an interior global optimum if $\tilde{c} \geq (3-4\theta)/4(1-\theta)$, or rearranging, if $\theta \geq (3-4\tilde{c})/4(1-\tilde{c})$. The right-hand side of this last inequality is decreasing in $\tilde{c}$. Substituting $\tilde{c} = 0$ gives the sufficient condition $\theta > 3/4$ for I_m^* to achieve an interior global maximum in this case. Combining the sufficient condition derived supposing $\tilde{c} \geq 1/2$ with that derived supposing $\tilde{c} < 1/2$, $\theta \geq 3/4$ is a sufficient condition for I_m^* to attain an interior global maximum. By Proposition 3, $\theta \geq 3/4$ is also sufficient for W_m^* to achieve an interior global minimum. *Q.E.D.*

Proof of Proposition 5

The proof of Proposition 3 showed that when $\tilde{c} < 1/4$, the interior global maximum of I_m^* over $\mathcal{R}_0$ equals $\mu/4(\alpha+\mu)(1-\tilde{c})$. The proof also showed that $\lim_{\mathcal{R}_0\uparrow\infty} I_m^* = (1-\theta)\mu/(\alpha+\mu)$. Dividing the two expressions yields the ratio

$$\frac{1}{4(1-\theta)(1-\tilde{c})} \geq \frac{1}{4(1-\theta)}, \tag{A10}$$

which grows arbitrarily high as $\theta \uparrow 1$.

The proof of Proposition 3 showed that the interior global minimum of W_m^* over $\mathcal{R}_0$ equals

$$H[\alpha + (3/4)\mu(1-\tilde{c})]. \tag{A11}$$

The proof also showed

$$\lim_{\mathcal{R}_0\uparrow\infty} W_m^* = H[\alpha + \theta\mu(1-\tilde{c})]. \tag{A12}$$

Dividing (A11) by (A12) yields the ratio

$$\frac{\alpha + (3/4)\mu(1-\tilde{c})}{\alpha + \theta\mu(1-\tilde{c})}. \tag{A13}$$

Equation (A13) is minimized over $\theta \in [0,1]$ by setting $\theta = 1$. After substituting $\theta = 1$, the resulting ratio can be shown to be increasing in $\tilde{c}$ and decreasing in α. It is thus minimized for $\tilde{c} = 0$ and minimized over $\alpha \in (0,1]$ for $\alpha = 1$. Substituting these values into (A13), the minimized ratio reduces to 3/4. *Q.E.D.*

Proof of Proposition 6

Suppose $\mathcal{R}_0 \leq 1$. Then $Q^{**} = \min(\mathcal{Q}_0, \mu) = \min(0, \mu) = 0$, where the first equality follows from equation (24), the second from substituting $\mathcal{R}_0 \leq 1$ into (12), and the third from $\mu > 0$. No subsidy is needed to generate zero quantity, implying $GS^{**} = 0$.

Suppose $1 < \mathcal{R}_0 < 1/(1-\theta)$. Substituting $1 < \mathcal{R}_0$ into (12) yields $\mathcal{Q}_0 > 0$, implying $Q^{**} = \min(\mathcal{Q}_0, \mu) > 0$. Substituting $\mathcal{R}_0 < 1/(1-\theta)$ into (12) yields $\mathcal{Q}_0 < \mu$, implying $Q^{**} = \min(\mathcal{Q}_0, \mu) = \mathcal{Q}_0$, in turn implying $I^{**} = 0$ by (8). By Proposition 1, $I^{**} = 0$ and $Q^{**} > 0$ imply $P^{**} = 0$.

Suppose $\mathcal{R}_0 \geq 1/(1-\theta)$. Then $\mathcal{Q}_0 \geq \mu$ by (12), implying $Q^{**} = \min(\mathcal{Q}_0, \mu) = \mu$. To find the lowest subsidy delivering $Q^{**} = \mu$, we need to find the highest P^{**} satisfying $D(P^{**}) = \mu$ since the equilibrium price is weakly decreasing in the subsidy.

Consider a price P' such that $P' > MPB(\mu)$. Then $D(P') \leq d(P')$ since only the first two branches of (18) are relevant in the computation of $D(P')$. Combining the inequality

$$P' > MPB(\mu) = \frac{\theta H}{1-\theta}\left(1 - \theta - \frac{1}{\mathcal{R}_0}\right) \tag{A14}$$

with (17) yields $d(P') < \mu$. Thus, $D(P') \leq d(P') < \mu = Q^{**} = D(P^{**})$, implying $P^{**} \neq P'$. But since P' was an arbitrary price greater than $MPB(\mu)$, we have $P^{**} \leq MPB(\mu)$. The highest price satisfying $P^{**} \leq MPB(\mu)$ is $P^{**} = MPB(\mu)$. To verify that this price yields the desired quantity, we have $\mu = d(MPB(\mu)) = D(MPB(\mu)) = D(P^{**})$, where the first equality follows from substituting from (A14) into (17) and the second equality from the fact that only the middle branch is relevant in (18) at a price of $MPB(\mu)$.

The previous paragraph shows $d(P^{**}) = d(MPB(\mu)) = \mu$, implying $P^{**} = d^{-1}(\mu)$. Inverting (17) yields equation (26). *Q.E.D.*

Proof of Proposition 7

We will prove the second bullet point regarding comparative statics of GS^{**} in c and θ. Gleaning the remaining results from Table 3 is relatively straightforward. In case (LR1) of the table, $GS_c^{**} = GS_m^{**} = 0$, in which case the comparative statics hold trivially. The remaining cases can be combined in a single expression for each market structure,

$$GS_c^{**} = \min\left\{c, \max\left\{0, c + \theta H\left[\frac{1}{(1-\theta)\mathcal{R}_0} - 1\right]\right\}\right\} \tag{A15}$$

$$GS_m^{**} = \min\left\{c + \mathcal{R}_0 - 1, \max\left\{0, c + \theta H\left[\frac{1}{(1-\theta)^2\mathcal{R}_0} - 1\right]\right\}\right\}, \tag{A16}$$

which are obviously weakly increasing in c and θ. *Q.E.D.*

Proof of Lemma 1

We begin by proving the claims about $I_t(Q)$. Substituting (39) into (31) yields

$$\dot{I}_t(Q) = \frac{1}{\alpha}I_t(Q)[\mathcal{R}_0 S_t(Q) - 1], \tag{A17}$$

57

or, rearranging,

$$\frac{\dot{I}_t(Q)}{I_t(Q)} = \frac{1}{\alpha}[\mathcal{R}_0 S_t(Q) - 1]. \tag{A18}$$

Recognizing the left-hand side as $\partial \ln I_t(Q)/\partial t$ and integrating yields

$$\int_0^t \frac{\partial \ln I_\tau(Q)}{\partial \tau} d\tau = \int_0^t \frac{1}{\alpha}[\mathcal{R}_0 S_\tau(Q) - 1] d\tau. \tag{A19}$$

Invoking the Fundamental Theorem of Calculus, taking exponentials, and rearranging yields, for all $t \geq 0$,

$$I_t(Q) = I_0(Q) \exp\left(\int_0^t \frac{1}{\alpha}[\mathcal{R}_0 S_\tau(Q) - 1] d\tau\right). \tag{A20}$$

Since $I_0(Q) = \hat{I}_0 > 0$ by assumption, $I_t(Q)$ is the product of two positive factors.

Turn next to proving the claims about $S_t(Q)$. Rearranging (32), $I_t(Q) = \dot{R}_t(Q)/\alpha$. Substituting into (33) and rearranging yields $\dot{S}_t(Q)/S_t(Q) = -(\beta/\alpha)\dot{R}_t(Q) = -\mathcal{R}_0 \dot{R}_t(Q)$ by (39). Recognizing $\dot{S}_t(Q)/S_t(Q) = \partial \ln S_t(Q)/\partial t$ and integrating between $t' \geq 0$ and $t'' \geq t'$ yields

$$\int_{t'}^{t''} \frac{\partial \ln S_\tau(Q)}{\partial \tau} d\tau = -\int_{t'}^{t''} \frac{1}{\alpha}\dot{R}_\tau(Q) d\tau. \tag{A21}$$

Invoking the Fundamental Theorem of Calculus, taking exponentials, and rearranging yields

$$S_{t''}(Q) = S_{t'}(Q)e^{\mathcal{R}_0[R_{t'}(Q) - R_{t''}(Q)]}. \tag{A22}$$

Substituting $t' = 0$ and $t'' = t$ into (A22) yields

$$S_t(Q) = S_0(Q)e^{\mathcal{R}_0[R_0(Q) - R_t(Q)]}. \tag{A23}$$

Now $S_0(Q) = \hat{S}_0 - \theta Q > \hat{S}_0 - Q \geq 0$, where the first step holds by (35), the second by $\theta < 1$, and the third by $Q \in [0, \hat{S}_0]$. The right-hand side of (A23) is thus the product of two positive factors. *Q.E.D.*

Proof of Lemma 2

Substituting $I_t(Q) > 0$ into (32) yields $\dot{R}_t(Q) > 0$, implying $R_{t''}(Q) > R_{t'}(Q)$ for $t'' > t'$, implying $e^{\mathcal{R}_0[R_{t'}(Q) - R_{t''}(Q)]} < 1$. Since $S_{t'}(Q) > 0$ by Lemma 1, $S_{t''}(Q) \leq S_{t'}(Q)$ by (A22). *Q.E.D.*

Proof of Lemma 3

Since $I_t(Q) > 0$ by Lemma 1, by (A17), the sign of $\dot{I}_t(Q)$ is determined by the value of $\mathcal{R}_0 S_t(Q)$ relative to 1.

First, suppose $\mathcal{R}_0 S_0(Q) \leq 1$. Consider any $t > 0$. Lemma 2 implies $S_t(Q) < S_0(Q)$, in turn implying $\mathcal{R}_0 S_t(Q) < \mathcal{R}_0 S_0(Q) \leq 1$. Substituting $\mathcal{R}_0 S_t(Q) < 1$ into (A17) implies $\dot{I}_t(Q) < 0$.

Next, suppose $\mathcal{R}_0 S_0(Q) > 1$. Substititing into (A17) for $t = 0$ implies $\dot{I}_0(Q) > 0$. By Martcheva (2015, p. 13), $I(Q) = 0$. Therefore, we must have $\dot{I}_t(Q) < 0$ for some $t > 0$. Thus, by continuinity, $\dot{I}_T(Q) = 0$ for some $T > 0$. Since $S_t(Q)$ is strictly decreasing by Lemma 1, T is unique, given by the value of t for which (A17) equals 0, implying T satisfies $S_T(Q) = 1/\mathcal{R}_0$. Since $S_t(Q)$ is strictly

decreasing we have $\dot{I}_t(Q) > 0$ for all $t \in [0, T)$, $\dot{I}_t(Q) < 0$ for all $t > T$, and $I_T(Q)$ is the maximum infection rate. *Q.E.D.*

Proof of Lemma 4

See Martcheva (2015, p. 13) for a proof that $I(Q) = 0$. Martcheva (2015, p. 12) argues that the fact that $S_t(Q)$ is positive and montone implies that the limit $S(Q)$ exists.

To prove the remaining claim in the lemma, take the limit $t \uparrow \infty$ in (A23):

$$S(Q) = S_0(Q)e^{\mathcal{R}_0[R_0(Q) - R(Q)]}. \tag{A24}$$

By Lemma 1, $S_0(Q) > 0$. The proof of Lemma 2 showed that $R_t(Q)$ is strictly increasing in t. Thus $R(Q) > R_0(Q)$, implying $S(Q) < S_0(Q)$ by (A24). *Q.E.D.*

Proof of Lemma 5

First, suppose $S_0(Q) \leq 1/\mathcal{R}_0$. Then $S(Q) < 1/\mathcal{R}_0$ because $S_t(Q)$ is strictly decreasing in t by Lemma 2.

Next, suppose $S_0(Q) > 1/\mathcal{R}_0$. In the last paragraph of the proof of Lemma 3, we proved the existence of $T > 0$ such that $S_T(Q) = 1/\mathcal{R}_0$. Since $S_t(Q)$ is strictly decreasing by Lemma 2, we have $S(Q) < S_T(Q) = 1/\mathcal{R}_0$. *Q.E.D.*

Proof of Lemma 6

Substituting (33) and (39) into (31) yields

$$\dot{I}_t(Q) = \frac{\dot{S}_t(Q)}{\mathcal{R}_0 S_t(Q)} - \dot{S}_t(Q). \tag{A25}$$

Integrating (A25) over $t \in [0, \infty)$ and applying the Fundamental Theorem of Calculus,

$$I(Q) - I_0(Q) = \frac{1}{\mathcal{R}_0}[\ln S(Q) - \ln S_0(Q)] - S(Q) + S_0(Q). \tag{A26}$$

Substituting $I_0(Q) = \hat{I}_0$ by (36), noting $I(Q) = 0$ by Lemma 4, and rearranging yields

$$\ln S(Q) - \mathcal{R}_0 S(Q) = \ln S_0(Q) - \mathcal{R}_0[\hat{I}_0 + S_0(Q)]. \tag{A27}$$

Further substituting $S_0(Q) = \hat{S}_0 - Q$ from (35) yields (41).

To derive (42), exponentiating both sides of (A27) and rearranging yields

$$S(Q) = \left\{ S_0(Q)e^{-\mathcal{R}_0[\hat{I}_0 + S_0(Q)]} \right\} e^{\mathcal{R}_0 S(Q)}, \tag{A28}$$

or, equivalently,

$$x = be^{ax}, \tag{A29}$$

where $x = S(Q)$, $a = \mathcal{R}_0$, and $b = S_0(Q)e^{-\mathcal{R}_0[\hat{I}_0+S_0(Q)]}$. It is well-known that (A29) has solution $x = -\bar{L}(-ab)/a = |\bar{L}(-ab)|/a$, where the second equality holds if $a, b > 0$ implying $\bar{L}(-ab) < 0$. Substituting the values of x, a, and b into this solution and further substituting $S_0(Q) = \hat{S}_0 - Q$ from (35) yields (42).

Equation (A29) also has a solution in terms of the lower branch of the Lambert W function, $x = -\underline{L}(-ab)/a$. We reject this solution because it exceeds 1, which is out of bounds for $S(Q)$. Q.E.D.

Proof of Lemma 8

We will prove $\lim_{\mathcal{R}_0\uparrow\infty}[\mathcal{R}_0 S(Q)] = 0$. The claim that $\lim_{\mathcal{R}_0\uparrow\infty} S(Q) = 0$ follows as a direct consequence (and a separate proof was also proved in the text). We have

$$\lim_{\mathcal{R}_0\uparrow\infty} [\mathcal{R}_0 S(\hat{S}_0)] = \lim_{\mathcal{R}_0\uparrow\infty} \left| \bar{L}\left(-\mathcal{R}_0(1-\theta)\hat{S}_0 e^{-\mathcal{R}_0[\hat{I}_0+(1-\theta)\hat{S}_0]} \right) \right| \tag{A30}$$

$$= \left| \bar{L}\left(-(1-\theta)\hat{S}_0 \lim_{\mathcal{R}_0\uparrow\infty} \frac{\mathcal{R}_0}{e^{\mathcal{R}_0[\hat{I}_0+(1-\theta)\hat{S}_0]}} \right) \right| \tag{A31}$$

$$= |\bar{L}(0)|. \tag{A32}$$

Equation (A30) follows by taking limits in (42), (A31) is a simple rearrangement, and (A32) follows from application of l'Hôpital's Rule. Standard results for the Lambert W function imply $\bar{L}(0) = 0$. Q.E.D.

Verification of Table 5 Entries

The equilibrium condition is $P_c^* = c$. Firms earn no profit under perfect competition: $\Pi_c^* = 0$. No consumers purchase in case (SR1), implying $Q_c^* = 0$. All susceptibles purchase in case (SR3), implying $Q_c^* = \hat{S}_0$. In case (SR2), Q_c^* can be found by substituting $P_c^* = c$ in equation (47).

To find R_c^*, note $R_c^* = 1 - I_c^* - S_c^* - \theta Q_c^* = 1 - S(Q_c^*) - \theta Q_c^*$ since $I_c^* = 0$. Substituting $Q_c^* = 0$ gives the entry for R_c^* in case (SR1), and substituting $Q_c^* = \hat{S}_0$ gives the entry for R_c^* in case (SR3). To find R_c^* in case (SR2), set $c = MPB(Q_c^*)$ in equation (45) and rearrange, yielding

$$S(Q_c^*) = (1-\tilde{c})(\hat{S}_0 - \theta Q_c^*) \tag{A33}$$

Substituting (A33) into $R_c^* = 1 - S(Q_c^*) - \theta Q_c^*$ and rearranging yields $R_c^* = 1 - (1-\tilde{c})\hat{S}_0 - \tilde{c}\theta Q_c^*$. Substituting from the table entry for $Q_c^* = \hat{S}_0\tilde{Q}_c^*$ yields the table entry for R_c^*.

Substituting $Q_c^* = 0$ in (45) gives MPB_c^* in case (SR1), and substituting $Q_c^* = \hat{S}_0$ in (45) gives MPB_c^* in case (SR3). For some but not all consumers to purchase in case (SR2) requires $MPB_c^* = c$.

Substituting $Q_c^* = 0$ in (49) gives MSB_c^* in case (SR1), and substituting $Q_c^* = \hat{S}_0$ in (49) gives MSB_c^* in case (SR3). Substituting from (A33) into (49) yields MSB_c^* in case (SR2).

The table entries for MEX_c^* can be obtained by subtracting other table entries: $MEX_c^* = MSB_c^* - MPB_c^*$. To derive the table table entries for W_c^*, by definition $W_c^* = SB_c^* - cQ_c^* = H(1 - R_c^*) - cQ_c^*$, where the second equation follows from (40). Substituting other table entries into this equation gives the table entries for W_c^*. Q.E.D.

Proof of Proposition 9

Preliminaries: For conciseness, let $\mathcal{R}_0'$ denote the boundary value of $\mathcal{R}_0$ between cases (SR1) and (SR2) and $\mathcal{R}_0''$ between cases (SR2) and (SR3), i.e.,

$$\mathcal{R}_0' = \frac{|\ln(1-\tilde{c})|}{\hat{I}_0 + \tilde{c}\hat{S}_0} \tag{A34}$$

$$\mathcal{R}_0'' = \frac{|\ln(1-\tilde{c})|}{\hat{I}_0 + (1-\theta)\tilde{c}\hat{S}_0} \tag{A35}$$

The results for P_c^* and Π_c^* are obvious from Table 5. To show Q_c^* is weakly increasing, it can be verified that it is continuous at thresholds $\mathcal{R}_0'$ and $\mathcal{R}_0''$. In case (SR2), $\partial Q_c^*/\partial \mathcal{R}_0 = -\ln(1-\tilde{c})/\theta\tilde{c}\mathcal{R}_0^2 > 0$. Hence, Q_c^* is weakly increasing in $\mathcal{R}_0$ for all $\mathcal{R}_0 > 0$ and strictly increasing for $\mathcal{R}_0$ in the interior of case (SR2).

Results for MPB_c^*: To show MPB_c^* is weakly increasing, start with case (SR1). Differentiating the table entry,

$$\frac{\partial MPB_c^*}{\partial \mathcal{R}_0} = -\left(\frac{\theta H}{\hat{S}_0}\right)\frac{\partial S(0)}{\partial \mathcal{R}_0}. \tag{A36}$$

Applying the Implicit Function Theorem to (41) yields derivative

$$\frac{\partial S(Q)}{\partial \mathcal{R}_0} = \frac{-S(Q)}{1 - \mathcal{R}_0 S(Q)}[\hat{I}_0 + S_0(Q) - S(Q)]. \tag{A37}$$

The first factor is negative by Lemma 5. The factor in square brackets is positive since $\hat{I}_0 + S_0(Q) - S(Q) > S_0(Q) - S(Q) > 0$, where the first inequality follows from $\hat{I}_0 > 0$ and the second by Lemma 4. Since equation (A37) is negative for all $Q \in [0, \hat{S}_0]$, including $Q_c^* = 0$, (A36) is positive. In case (SR2), MPB_c^* is constant. Differentiating the table entry in cases (SR3) and (SR4),

$$\frac{\partial MPB_c^*}{\partial \mathcal{R}_0} = -\left[\frac{\theta H}{(1-\theta)\hat{S}_0}\right]\frac{\partial S(\hat{S}_0)}{\partial \mathcal{R}_0}, \tag{A38}$$

which is negative since (A37) is negative for all $Q \in [0, \hat{S}_0]$, including $Q_c^* = \hat{S}_0$.

The last step in deriving comparative statics for MPB_c^* is to show it is continuous at both endpoints of case (SR2). Now $MPB(Q)$ is continuous in Q because it is differentiable in Q by (46). Further, $MPB(Q)$ is continuous in $\mathcal{R}_0$ because $S(Q)$ is differentiable in $\mathcal{R}_0$ by (A37). Since Q_c^* is continuous at both endpoints of case (SR2) as argued in the first paragraph of this proof, we have that MPB_c^* is continous at $\mathcal{R}_0'$ and $\mathcal{R}_0''$.

Results for R_c^*: To derive the comparative statics for R_c^*, combining the table entries with (A37) shows R_c^* is increasing in $\mathcal{R}_0$ in case (SR1) as well as cases (SR3) and (SR4). The table entry is obviously decreasing in $\mathcal{R}_0$ in case (SR2). We thus have that R_c^* attains a local maximum at $\mathcal{R}_0'$ if we

can establish that R_c^* is continuous at $\mathcal{R}_0'$. Using the table entry for R_c^* in case (SR1),

$$\lim_{\mathcal{R}_0\uparrow\mathcal{R}_0'} R_c^* = 1 - \lim_{\mathcal{R}_0\uparrow\mathcal{R}_0'} S(0) = 1 - \left(1 - \lim_{\mathcal{R}_0\uparrow\mathcal{R}_0'} \frac{MPB_c^*}{\theta H}\right)\hat{S}_0 = 1 - (1-\tilde{c})\hat{S}_0. \tag{A39}$$

The second equality follows from the table entry for MPB_c^* in case (SR1): $MPB_c^* = \theta H\Phi_I(0) = \theta H[1 - S(0)/\hat{S}_0]$ by (44). The third equality follows from the continuity of MPB_c^* at $\mathcal{R}_0'$, allowing us to substitute the table entry for $MPB_c^* = c$ in case (SR2). Using the table entry for R_c^* in case (SR2),

$$\lim_{\mathcal{R}_0\downarrow\mathcal{R}_0'} R_c^* = 1 - \hat{S}_0 - \hat{I}_0 + \frac{1}{\mathcal{R}_0'}|\ln(1-\tilde{c})| = 1 - (1-\tilde{c})\hat{S}_0. \tag{A40}$$

The equality between (A39) and (A40) proves the continuity of R_c^* at $\mathcal{R}_0'$.

Since R_c^* is increasing in $\mathcal{R}_0$ in cases (SR3) and (SR4), the other candidate for a supremum is

$$\lim_{\mathcal{R}_0\uparrow\infty} R_c^* = 1 - \theta\hat{S}_0. \tag{A41}$$

This equality follows from taking the limit $\mathcal{R}_0 \uparrow \infty$ of the table entry in cases (SR3) and (SR4) and noting that $\lim_{\mathcal{R}_0\uparrow\infty} S(\hat{S}_0) = 0$ by Lemma 8. The local maximum is thus a global maximum if and only if $1 - (1-\tilde{c})\hat{S}_0 \geq 1 - \theta\hat{S}_0$. Rearranging gives $\tilde{c} \geq 1 - \theta$.

Results for MSB_c^*: To provide a roadmap for the comparative-statics analysis of MSB_c^*, we first look at cases (SR1) and (SR2) and show that MSB_c^* has a unique local maximum over $\mathcal{R}_0 \leq \mathcal{R}_0''$. Furthermore, this restricted local maximum is the restricted global maximum over $\mathcal{R}_0 \leq \mathcal{R}_0''$. We then look at cases (SR3) and (SR4) and show that MSB_c^* has at most one restricted local maximum over $\mathcal{R}_0 > \mathcal{R}_0''$. If no restricted local maximum exists there, then we show that the restricted maximum over $\mathcal{R}_0 \leq \mathcal{R}_0''$ is the global maximum over all $\mathcal{R}_0 > 0$. If a restricted local maximum exists in cases (SR3) and (SR4), then either it or the restricted local maximum over $\mathcal{R}_0 \leq \mathcal{R}_0''$ is the global maximum. This establishes, in sum, that MSB_c^* has at most two local maxima, one of which is the global maximum.

To prove that MSB_c^* has a unique local restricted maximum over $\mathcal{R}_0 \leq \mathcal{R}_0''$, we will show that MSB_c^* is increasing in a neighborhood around 0, quasiconcave for all $\mathcal{R}_0 \in (0, \mathcal{R}_0']$, continuous at $\mathcal{R}_0'$, and decreasing for all $\mathcal{R}_0 \in (\mathcal{R}_0', \mathcal{R}_0'')$. The arguments are made in reverse order. It is clear from inspection of Table 5 that MSB_c^* is decreasing for $\mathcal{R}_0 \in (\mathcal{R}_0', \mathcal{R}_0'')$. To show MSB_c^* is continuous at $\mathcal{R}_0'$, using the table entry for MSB_c^* in case (SR1),

$$\lim_{\mathcal{R}_0\uparrow\mathcal{R}_0'} MSB_c^* = \frac{\lim_{\mathcal{R}_0\uparrow\mathcal{R}_0'} MPB_c^*}{1 - \mathcal{R}_0'(1 - \lim_{\mathcal{R}_0\uparrow\mathcal{R}_0'} MPB_c^*/\theta H)\hat{S}_0} = \frac{\theta H\tilde{c}(\hat{I}_0 + \tilde{c}\hat{S}_0)}{\hat{I}_0 + \tilde{c}\hat{S}_0 + (1-\tilde{c})\hat{S}_0\ln(1-\tilde{c})}. \tag{A42}$$

The first equality follows from substituting the table entry for MPB_c^* in case (SR1) directly as well as substituting the implication of that table entry that $S(0) = \hat{S}_0(1 - MPB_c^*/\theta H)$. The second equality follows from $\lim_{\mathcal{R}_0\uparrow\mathcal{R}_0'} MPB_c^* = c$ by continuity and from substituting from (A34). Using the table

62

entry for MSB_c^* in case (SR2),

$$\lim_{\mathcal{R}_0 \downarrow \mathcal{R}_0'} MSB_c^* = \frac{\theta H \tilde{c}^2}{\tilde{c} + (1-\tilde{c})[\ln(1-\tilde{c}) + \mathcal{R}_0' \hat{I}_0]} = \frac{\theta H \tilde{c}(\hat{I}_0 + \tilde{c}\hat{S}_0)}{\hat{I}_0 + \tilde{c}\hat{S}_0 + (1-\tilde{c})\hat{S}_0 \ln(1-\tilde{c})}. \tag{A43}$$

The equality of (A42) and (A43) proves the continuity of MSB_c^* at $\mathcal{R}_0'$.

We next show MSB_c^* is quasiconcave for all $\mathcal{R}_0$ in case (SR1). Differentiating the relevant table entry, substituting from (A37), and eliminating positive constants shows that $\partial MSB_c^* / \partial \mathcal{R}_0$ has the same sign as

$$\left[\hat{S}_0 - S(0)\right]\left[1 - \mathcal{R}_0 S(0)\right] + \left[\hat{I}_0 + \hat{S}_0 - S(0)\right]\left(1 - \mathcal{R}_0 \hat{S}_0\right). \tag{A44}$$

The second derivative of (A44) with respect to $\mathcal{R}_0$—after substituting from (A37), and rearranging considerably—can be shown to equal

$$2\frac{\partial S(0)}{\partial \mathcal{R}_0}\left[\hat{I}_0 + \hat{S}_0 - S(0)\right], \tag{A45}$$

which is negative—as can be shown using arguments similar to those used to prove (A37) is negative. Hence, (A44) is concave. In the limit $\mathcal{R}_0 \downarrow 0$, (A44) approaches $2[\hat{S}_0 - S(0)] + \hat{I}_0$, which is positive by Lemma 4 and $\hat{I}_0 > 0$. Having established that (A44) is concave throughout (SR1) and initially positive, we have that (A44) can change sign at most once. Therefore, $\partial MSB_c^* / \partial \mathcal{R}_0$ is either nonnegative throughout case (i) or positive then negative. In either event, this proves that MSB_c^* is quasiconcave in (SR1). We have already established MSB_c^* is increasing in a neighborhood of $\mathcal{R}_0$ above 0, the last step needed to prove that MSB_c^* has a unique restricted local maximum over $\mathcal{R}_0 \le \mathcal{R}_0''$, which is a global maximum on that restricted set.

We next look at the behavior of MSB_c^* in cases (SR3) and (SR4), showing it has a most one restricted local maximum over $\mathcal{R}_0 > \mathcal{R}_0''$. Similar calculations used in the previous paragraph can be used here to establish the concavity of the following function,

$$\left[(1-\theta)\hat{S}_0 - S(\hat{S}_0)\right]\left[1 - \mathcal{R}_0 S(\hat{S}_0)\right] + \left[\hat{I}_0 + (1-\theta)\hat{S}_0 - S(\hat{S}_0)\right]\left[1 - (1-\theta)\mathcal{R}_0 \hat{S}_0\right], \tag{A46}$$

which determines the sign of $\partial MSB_c^* / \partial \mathcal{R}_0$ in (SR3) and (SR4). Thus, (A46) has at most two roots in those cases, which cannot both be local maxima, implying that MSB_c^* has at most one local maximum over $\mathcal{R}_0 > \mathcal{R}_0''$. The limit as $\mathcal{R}_0 \uparrow \infty$ of (A46) equals

$$\hat{I}_0 + 2(1-\theta)\hat{S}_0 - (1-\theta)\hat{S}_0[\hat{I}_0 + (1-\theta)\hat{S}_0] \lim_{\mathcal{R}_0 \uparrow \infty} \mathcal{R}_0 \tag{A47}$$

after substituting $\lim_{\mathcal{R}_0 \uparrow \infty} S(\hat{S}_0) = \lim_{\mathcal{R}_0 \uparrow \infty}[\mathcal{R}_0 S(\hat{S}_0)] = 0$ by Lemma 8. Expression (A47) approaches $-\infty$ since it involves $\mathcal{R}_0$ multiplied by negative constant. Since the limit $\mathcal{R}_0 \uparrow \infty$ cannot produce a restricted supremum over $\mathcal{R}_0 > \mathcal{R}_0''$, the restricted supremum is either the lower boundary of (SR3), i.e., $\mathcal{R}_0''$, or is the interior restricted maximum. If $\mathcal{R}_0''$ provides the restricted supremum over $\mathcal{R}_0 > \mathcal{R}_0''$, this cannot be a global maximum since MSB_c^* is decreasing in (SR2); the restricted maximum over $\mathcal{R}_0 \le \mathcal{R}_0''$ must then be the global maximum.

Results for MEX_c^*: We use the same roadmap for the comparative-statics analysis of MEX_c^* as for MSB_c^*. We begin by proving that MEX_c^* has a unique local restricted maximum over $\mathcal{R}_0 \le \mathcal{R}_0''$.

We do this by showing that MEX_c^* is increasing in a neighborhood around 0, quasiconcave for all $\mathcal{R}_0 \in (0, \mathcal{R}_0']$, continuous at $\mathcal{R}_0'$, and decreasing for all $\mathcal{R}_0 \in (\mathcal{R}_0', \mathcal{R}_0'')$. The arguments are made in reverse order. Differentiating the table entry for case (SR2) yields

$$\frac{\partial MEX_c^*}{\partial \mathcal{R}_0} = \frac{-\theta H \tilde{c}^2 (1-\tilde{c}) \hat{I}_0}{\left\{ \tilde{c} + (1-\tilde{c})[\ln(1-\tilde{c}) + \mathcal{R}_0 \hat{I}_0] \right\}^2}, \tag{A48}$$

which is negative. The proof that MEX_c^* is continuous at $\mathcal{R}_0'$ is similar to that for MSB_c^* and omitted.

We next show MEX_c^* is quasiconcave for all $\mathcal{R}_0$ in case (SR1). Differentiating the relevant table entry, substituting from equation (A37), and eliminating positive constants shows that $\partial MEX_c^* / \partial \mathcal{R}_0$ has the same sign as

$$\left[\hat{S}_0 - S(0) \right] \left[1 - \mathcal{R}_0 (\hat{I}_0 + \hat{S}_0) \right] + \mathcal{R}_0 S(0) \left[1 - \mathcal{R}_0 S(0) \right] \left[\hat{I}_0 + \hat{S}_0 - S(0) \right]. \tag{A49}$$

All of the factors in (A49) are definitively positive except for $1 - \mathcal{R}_0 (\hat{I}_0 + \hat{S}_0)$. If this is also nonnegative, then $\partial MEX_c^* / \partial \mathcal{R}_0$ is positive in (SR1), implying MEX_c^* is quasiconcave in (SR1), as desired.

So suppose instead that

$$\mathcal{R}_0 (\hat{I}_0 + \hat{S}_0) > 1. \tag{A50}$$

We will show that (A50) implies that (A49) is concave. The second derivative of (A49) with respect to $\mathcal{R}_0$—after substituting from (A37), rearranging considerably, and removing positive factors—can be shown to have the same sign as

$$S(0) \left[1 - \mathcal{R}_0 (\hat{I}_0 + \hat{S}_0) \right] - \left[1 - \mathcal{R}_0 S(0) \right] (\hat{I}_0 + \hat{S}_0) - S(0) \left[2(\hat{I}_0 + \hat{S}_0) - S(0) \right]$$
$$- S(0) \left\{ 1 - \mathcal{R}_0 \left[\hat{I}_0 + \hat{S}_0 - S(0) \right] \right\} \left[\hat{I}_0 + \hat{S}_0 - 2S(0) \right]. \tag{A51}$$

By (A50) and familiar arguments, all the terms in (A51) are negative except possibly the last. If the last term is also nonpositive, the whole expression is negative, establishing (A49) is concave. So suppose instead that the last term is positive. For this to be the case, one of its last two factors must be positive and the other negative. That is, one of the following two sets of conditions must hold:

$$1 - \mathcal{R}_0 \left[\hat{I}_0 + \hat{S}_0 - S(0) \right] > 0, \qquad \hat{I}_0 + \hat{S}_0 - 2S(0) < 0 \tag{A52}$$

$$1 - \mathcal{R}_0 \left[\hat{I}_0 + \hat{S}_0 - S(0) \right] < 0, \qquad \hat{I}_0 + \hat{S}_0 - 2S(0) > 0 \tag{A53}$$

Suppose (A52) holds. Then

$$(A51) < -S(0) \left[2(\hat{I}_0 + \hat{S}_0) - S(0) \right] - S(0) \left\{ 1 - \mathcal{R}_0 \left[\hat{I}_0 + \hat{S}_0 - S(0) \right] \right\} \left[\hat{I}_0 + \hat{S}_0 - 2S(0) \right] \tag{A54}$$

$$< -S(0) \left[2(\hat{I}_0 + \hat{S}_0) - S(0) \right] - S(0) \left[\hat{I}_0 + \hat{S}_0 - 2S(0) \right] \tag{A55}$$

$$= -3S(0) \left[\hat{I}_0 + \hat{S}_0 - S(0) \right]. \tag{A56}$$

Equation (A54) follows from eliminating the first two negative terms of (A51). Equation (A55) follows from substituting 1, which is greater than the factor in braces, for the factor in braces. The fact that this substitution results in an increase in (A55) follows from (A52). Straightforward algebra yields (A56), which is negative by familiar arguments.

Suppose (A53) holds. Then

$$\text{(A51)} < S(0)\left[1-\mathcal{R}_0(\hat{I}_0+\hat{S}_0)\right]-S(0)\left\{1-\mathcal{R}_0\left[\hat{I}_0+\hat{S}_0-S(0)\right]\right\}\left[\hat{I}_0+\hat{S}_0-2S(0)\right] \tag{A57}$$

$$< S(0)\left[1-\mathcal{R}_0(\hat{I}_0+\hat{S}_0)\right]-S(0)\left\{1-\mathcal{R}_0\left[\hat{I}_0+\hat{S}_0-S(0)\right]\right\} \tag{A58}$$

$$= -\mathcal{R}_0 S(0)^2. \tag{A59}$$

Equation (A57) follows from eliminating the second and third two negative terms from (A51). Equation (A58) follows from substituting 1 for the last factor, $\hat{I}_0+\hat{S}_0-2S(0)$. To see that this increases the expression, note $\hat{I}_0+\hat{S}_0-2S(0) < \hat{I}_0+\hat{S}_0 \leq 1$, where the last inequality holds since the size of the infected and susceptible subpopulations at date 0, $\hat{I}_0+\hat{S}_0$, cannot exceed the size of the entire population, normalized to 1. The fact that substituting 1 for $\hat{I}_0+\hat{S}_0-2S(0)$ increases (A57) follows from (A53). Straightforward algebra yields (A59), which is obviously negative.

In sum, we have shown (A51) is negative for $\mathcal{R}_0 < \mathcal{R}_0'$, implying (A49) is concave. In the limit $\mathcal{R}_0 \downarrow 0$, (A49) approaches $\hat{S}_0-S(0)$, which is positive by Lemma 4. These facts are sufficient to establish that MEX_c^* is quasiconcave in case (SR1) by the same arguments used for MSB_c^* above. These are all the facts needed to prove that MEX_c^* has a unique restricted local maximum over $\mathcal{R}_0 \leq \mathcal{R}_0''$, which is a global maximum on that restricted set.

We next investigate the behavior of MEX_c^* in (SR3) and (SR4). Calculations similar to those used above can be used to show a function determining the sign of $\partial MEX_c^*/\partial\mathcal{R}_0$ in cases (SR3) and (SR4),

$$\left[(1-\theta)\hat{S}_0-S(0)\right]\left\{1-\mathcal{R}_0[\hat{I}_0+(1-\theta)\hat{S}_0]\right\}+\mathcal{R}_0 S(\hat{S}_0)\left[1-\mathcal{R}_0 S(\hat{S}_0)\right]\left[\hat{I}_0+(1-\theta)\hat{S}_0-S(\hat{S}_0)\right]. \tag{A60}$$

is concave. Thus, (A60) has at most two roots in cases (SR3) and (SR4), at most one of which is a local maximum for MEX_c^*. Taking the limit of the table entry for MEX_c^* in cases (SR3) and (SR4) and substituting the limit $\lim_{\mathcal{R}_0\uparrow\infty} S(\hat{S}_0) = \lim_{\mathcal{R}_0\uparrow\infty}[\mathcal{R}_0 S(\hat{S}_0)] = 0$ by Lemma 8 yields $\lim_{\mathcal{R}_0\uparrow\infty} MEX_c^* = 0$. Since the limit $\mathcal{R}_0 \uparrow \infty$ produces an infimum for MEX_c^*, not a supremum, the restricted supremum of MEX_c^* over $\mathcal{R}_0 > \mathcal{R}_0''$ is either the lower boundary of case (SR3), i.e., $\mathcal{R}_0''$, or the interior restricted maximum. If $\mathcal{R}_0''$ provides the restricted supremum over $\mathcal{R}_0 > \mathcal{R}_0''$, this cannot be a global maximum since MSB_c^* is decreasing in case (SR2); the restricted maximum over $\mathcal{R}_0 \leq \mathcal{R}_0''$ must then be the global maximum.

Final Results: The comparative statics for W_c^* are obvious from inspection of the table in view of (A37). *Q.E.D.*

Proof of Proposition 10

Results for Π_m^*: The result is a consequence of the Envelope Theorem. Monopoly profit can be written

$$\Pi_m^* = \theta H\left[1-\frac{S(Q_m^*)}{S_0(Q_m^*)}-\tilde{c}\right]Q_m^*. \tag{A61}$$

This is a function of $\mathcal{R}_0$ indirectly through its dependence on Q_m^*, which in turns depends on $\mathcal{R}_0$. It also depends on $\mathcal{R}_0$ because $S(Q)$ is a function of $\mathcal{R}_0$ (although the argument is omitted for brevity). If Q_m^* is an interior solution, as in case (SR2) and (SR3), the first-order condition ensures that the

indirect effect of $\mathcal{R}_0$ on Π_m^* through Q_m^* equals 0. Only the direct effect remains. Hence,

$$\frac{\partial \Pi_m^*}{\partial \mathcal{R}_0} = \left[\frac{-\theta H Q_m^*}{S_0(Q_m^*)} \right] \frac{\partial S(Q_m^*)}{\partial \mathcal{R}_0}, \tag{A62}$$

which is positive since the derivative on the right-hand side is negative by (A37).

Results for R_m^*:   Since $R_m^* = R_c^*$ for all $\mathcal{R}_0$ in case (SR1),

$$\lim_{\mathcal{R}_0 \downarrow \mathcal{R}_0'} R_m^* = \lim_{\mathcal{R}_0 \downarrow \mathcal{R}_0'} R_c^* = 1 - (1 - \tilde{c}) \hat{S}_0, \tag{A63}$$

where the first equality follows by continuity since $\mathcal{R}_0'$ is the upper bound on case (SR1) by (A34) and the second equality follows from (A39).

We proceed to compare (A63) to the limits of R_m^* for extreme values of $\mathcal{R}_0$. We have

$$\lim_{\mathcal{R}_0 \downarrow 0} R_m^* = \lim_{\mathcal{R}_0 \downarrow 0} R_c^* = 1 - \lim_{\mathcal{R}_0 \downarrow 0} S(0) = 1 - \hat{S}_0, \tag{A64}$$

where the first equality follows since $R_m^* = R_c^*$ for all $\mathcal{R}_0$ in case (SR1), the second equality follows from the entry for R_c^* in case (SR1) in Table 5, and the third equality follows from Lemma 7. Equation (A64) is less than (A63). At the other extreme,

$$\lim_{\mathcal{R}_0 \uparrow \infty} R_m^* = \lim_{\mathcal{R}_0 \uparrow \infty} R_c^* = 1 - \theta \hat{S}_0, \tag{A65}$$

where the first equality follows by continuity since $R_m^* = R_c^*$ for all $\mathcal{R}_0$ in case (SR4), the second equality follows from the entry for R_c^* in cases (SR3) and (SR4) in Table 5, and the third equality follows from (A41). Equation (A65) is weakly less than (A63) if $\tilde{c} \geq 1 - \theta$. Since R_m^* is greater at the interior $\mathcal{R}_0'$ than at extreme values of $\mathcal{R}_0$, R_m^* must have an interior maximum.

Results for MSB_m^*:   We show that the limits of MSB_m^* for extreme values of $\mathcal{R}_0$ are exceed by interior values. We have

$$\lim_{\mathcal{R}_0 \downarrow 0} MSB_m^* = \lim_{\mathcal{R}_0 \downarrow 0} MSB_c^* = \theta H \left[1 - \frac{1}{\hat{S}_0} \lim_{\mathcal{R}_0 \downarrow 0} S(0) \right] = \theta H \left(1 - \frac{\hat{S}_0}{\hat{S}_0} \right) = 0, \tag{A66}$$

where the first equality follows since $MSB_m^* = MSB_c^*$ for all $\mathcal{R}_0$ in case (SR1), the second equality follows from the entry for MSB_c^* in case (SR1) in Table 5, and the third equality follows from Lemma 7. To examine the upper limit, the proof of Proposition 9 showed that MSB_c^* asymptotes downward toward $\lim_{\mathcal{R}_0 \uparrow \infty} MSB_c^* = 1$. Since $MSB_m^* = MSB_c^*$ in case (SR4), and all $\mathcal{R}_0$ above a sufficiently high value are contained in case (SR4), MSB_m^* must also slope downward toward its asymptote. Thus MSB_m^* is higher at interior values of $\mathcal{R}_0$ than the extremes.

Results for MEX_m^*:   Arguments similar to the preceding can be used to show $\lim_{\mathcal{R}_0 \downarrow 0} MEX_m^* = \lim_{\mathcal{R}_0 \uparrow \infty} MEX_c^* = 0$. Hence, MEX_m^* is higher for interior values of $\mathcal{R}_0$ than extreme values and thus attains an interior maximum. *Q.E.D.*

Proof of Proposition 11

The sketch of the proof in the text omitted two details filled in here. We first prove Q^{**} for $\mathcal{R}_0$ in a neighborhood above 0. Taking limits in (49),

$$\lim_{\mathcal{R}_0 \downarrow 0} MSB(Q) = \theta H \left[1 - \frac{S_0(Q)}{S_0(Q)} \right] = 0. \tag{A67}$$

Hence, there exists $\mathcal{R}_0$ in a neighborhood above 0 and $\epsilon \in (0,c)$ such that $MSB(Q) < \epsilon$. For $\mathcal{R}_0$ in this neighborhood, $W(Q) = \int_0^Q [MSB(x) - c]dx < (\epsilon - c)Q < 0 = W(0)$. Thus, $Q^{**} = 0$ for $\mathcal{R}_0$ in this neighborhood.

We next prove $GS^{**} > 0$ for some $\mathcal{R}_0 \in (0,\infty)$. Since $Q^{**} = 0$ for all $\mathcal{R}_0$ in neighborhood of 0, $Q^* \leq Q^{**} = 0$ implies $Q^* = 0$ for all $\mathcal{R}_0$ in a neighborhood of 0. The text argued $Q_m^* = \hat{S}_0$ for sufficiently high $\mathcal{R}_0$, implying $\hat{S}_0 = Q_m^* \leq Q^{**} \leq h S_0$, implying $Q^{**} = \hat{S}_0$ for sufficiently high $\mathcal{R}_0$. By the Theorem of the Maximum, since Q^{**} is a maximizer of continuous function $W(Q)$, Q^{**} is continuous, implying the existence of $\mathcal{R}_0 \in (0,\infty)$ such that $Q^* \in (0,\hat{S}_0)$. This Q^{**} must satisfy the first-order condition $MSB(Q^{**}) = c$, implying $MPB(Q^{**}) + MEX(Q^{**}) = c$, implying $MPB(Q^{**}) < c$ since $MEX(Q) > 0$ for all $Q \in (0,\hat{S}_0)$ by (50). *Q.E.D.*

Proof of Proposition 14

Suppose $\mathcal{R}_0 \hat{S}_0 > 2$. Then $1 < \mathcal{R}_0 \hat{S}_0 / 2 < \mathcal{R}_0 [\hat{S}_0 + S(0)]/2 = \mathcal{R}_0 [S_0(0) + S(0)]/2$, where the second step follows from $S(0) > 0$ by Lemma 4. This chain of inequalities implies that (59) holds at $Q = 0$ and thus that the vaccine exhibits initially increasing social returns.

At a general output level $Q \in (0,\hat{S}_0)$,

$$\mathcal{R}_0 \left[\frac{S_0(Q) + S(Q)}{2} \right] > \mathcal{R}_0 \left(\frac{S_0(Q)}{2} \right) = \mathcal{R}_0 \left(\frac{\hat{S}_0 - \theta Q}{2} \right) \geq \mathcal{R}_0 \left(\frac{(1-\theta)\hat{S}_0}{2} \right). \tag{A68}$$

If $\mathcal{R}_0 \hat{S}_0 \geq 2/(1-\theta)$, then the last expression weakly exceeds 1, implying (59) holds for all feasible Q, implying the vaccine exhibits everywhere increasing social returns. *Q.E.D.*

Proof of Proposition 15

Universal vaccination with a perfectly effective vaccine implies $S_0(\hat{S}_0) = (1-\theta)\hat{S}_0 = 0$. We thus have

$$\mathcal{R}_0 \left[\frac{S_0(\hat{S}_0) + S(\hat{S}_0)}{2} \right] = \mathcal{R}_0 \left(\frac{S(\hat{S}_0)}{2} \right) < \frac{1}{2}. \tag{A69}$$

Hence, (59) does not hold for $Q = \hat{S}_0$, implying the vaccine does not exhibit increasing social returns for $Q = \hat{S}_0$ by Proposition 13. *Q.E.D.*

Proof of Proposition 16

Results for Perfect Competition: As noted in the text, after substituting $c = 0$ (implying $\tilde{c} = 0$) into the intervals of $\mathcal{R}_0$ determining cases, only cases (SR3) and (SR4) in the last column of Table 5

remain. Many of these entries involve $S(\hat{S}_0)$. To compute $\lim_{\theta \uparrow 1} S(\hat{S}_0)$, note $0 \leq S(\hat{S}_0) < S_0(\hat{S}_0) = (1-\theta)\hat{S}_0$ where the inequality follows from Lemma 4. Thus, $0 \leq \lim_{\theta \uparrow 1} S(\hat{S}_0) \leq \lim_{\theta \uparrow 1}(1-\theta)\hat{S}_0 = 0$, implying

$$\lim_{\theta \uparrow 1} S(\hat{S}_0) = 0. \tag{A70}$$

The limits of some table entries follow from direct substitution of (A70). This is not the case for MPB_c^* since the limit involves a $0/0$ form. We will compute this limit with the aid of (A24). We have

$$\frac{S(\hat{S}_0)}{S_0(\hat{S}_0)} = e^{\mathcal{R}_0[R_0 - R(\hat{S}_0)]} = e^{-\mathcal{R}_0[\hat{I}_0 + (1-\theta)\hat{S}_0 - S(\hat{S}_0)]}. \tag{A71}$$

The first equality follows from rearranging (A24) and substituting $Q = \hat{S}_0$. The second equality follows from substituting for R_0 from (37) and for $R(\hat{S}_0)$ from the entry for R_c^* in Table 5. Taking the limit of (A71) using (A70) gives

$$\lim_{\theta \uparrow 1} \frac{S(\hat{S}_0)}{S_0(\hat{S}_0)} = e^{-\mathcal{R}_0 \hat{I}_0}, \tag{A72}$$

implying $\lim_{\theta \uparrow 1} MPB_c^* = \theta H(1 - e^{-\mathcal{R}_0 \hat{I}_0})$. The limit of MSB_c^* can then be computed as

$$\lim_{\theta \uparrow 1} MSB_c^* = \lim_{\theta \uparrow 1} \left[\frac{1}{1 - \mathcal{R}_0 S(\hat{S}_0)} \right] \lim_{\theta \uparrow 1} MPB_c^* = \left(\frac{1}{1-0} \right) \theta H \left(1 - e^{-\mathcal{R}_0 \hat{I}_0} \right) = \theta H \left(1 - e^{-\mathcal{R}_0 \hat{I}_0} \right). \tag{A73}$$

Subtracting, show $\lim_{\theta \uparrow 1} MEX_c^* = \lim_{\theta \uparrow 1} MSB_c^* - \lim_{\theta \uparrow 1} MPB_c^* = 0$.

Condition for First Best under Monopoly: Monopoly attains the first best if and only if (54) holds. Substituting $c = 0$ and taking the limit $\theta \uparrow 1$ using (A70) and (A72), the condition becomes

$$\left(1 - e^{-\mathcal{R}_0 \hat{I}_0}\right) \left[1 - \mathcal{R}_0 \lim_{\theta \uparrow 1} \frac{S(\hat{S}_0)}{1-\theta} \right] \geq 0. \tag{A74}$$

The limit on the left-hand side can be computed using (A72). Noting $S_0(\hat{S}_0) = (1-\theta)\hat{S}_0$ in the denominator of (A72), we have $\lim_{\theta \uparrow 1}[S(\hat{S}_0)/(1-\theta)\hat{S}_0] = e^{-\mathcal{R}_0 \hat{I}_0}$, implying $\lim_{\theta \uparrow 1}[S(\hat{S}_0)/(1-\theta)] = \hat{S}_0 e^{-\mathcal{R}_0 \hat{I}_0}$. Substituting into (A74), the limiting condition for first best under monopoly becomes

$$\left(1 - e^{-\mathcal{R}_0 \hat{I}_0}\right) \left(1 - \mathcal{R}_0 \hat{S}_0 e^{-\mathcal{R}_0 \hat{I}_0}\right) \geq 0. \tag{A75}$$

The first factor is positive for all $\mathcal{R}_0 > 0$. Hence, (A75) is nonnegative if and only if the second factor is nonnegative, which upon manipulating is equivalent to

$$-\mathcal{R}_0 \hat{I}_0 e^{-\mathcal{R}_0 \hat{I}_0} \geq -\frac{\hat{I}_0}{\hat{S}_0}. \tag{A76}$$

We have manipulated the condition to facilitate application of Lambert W function. The principal branch of the Lambert W function is increasing, so applying it to both sides of (A76) does not change

the direction of the inequality:

$$-\mathcal{R}_0 \hat{I}_0 \geq \bar{L}(-\hat{I}_0/\hat{S}_0), \tag{A77}$$

equivalent to $\mathcal{R}_0 \leq |\bar{L}(-\hat{I}_0/\hat{S}_0)|/\hat{I}_0$. The lower branch of the Lambert W function is decreasing, so applying it to both sides of (A76) changes the direction of the inequality:

$$-\mathcal{R}_0 \hat{I}_0 \leq \underline{L}(-\hat{I}_0/\hat{S}_0), \tag{A78}$$

$\mathcal{R}_0 \geq |\underline{L}(-\hat{I}_0/\hat{S}_0)|/\hat{I}_0$. Combining these two inequalities in $\mathcal{R}_0$ gives (60).

Interior Monopoly Optimum: If (60) does not hold, then Q_m^* is the interior solution to (53). Equation (53) can only hold for positive price at the limiting value $c = 0$ if and only if the factor in braces equals 0. Setting that factor equal to 0 and solving for $S(Q_m^*)$ yields

$$S(Q_m^*) = \frac{S_0(Q_m^*)}{\mathcal{R}_0[Q_m^* + S_0(Q_m^*)]}. \tag{A79}$$

Taking the limit $\theta \uparrow 1$ and simplifying,

$$S(Q_m^*) = \frac{\hat{S}_0 - Q_m^*}{\mathcal{R}_0 \hat{S}_0}. \tag{A80}$$

Substituting from (42) and simplifying,

$$\bar{L}\left(-\mathcal{R}_0(\hat{S}_0 - Q_m^*)e^{-\mathcal{R}_0(\hat{I}_0 + \hat{S}_0 - Q_m^*)}\right) = -\frac{\hat{S}_0 - Q_m^*}{\hat{S}_0}, \tag{A81}$$

which implies, by definition of the Lambert W function,

$$-\mathcal{R}_0(\hat{S}_0 - Q_m^*)e^{-\mathcal{R}_0(\hat{I}_0 + \hat{S}_0 - Q_m^*)} = -\frac{\hat{S}_0 - Q_m^*}{\hat{S}_0}e^{-(\hat{S}_0 - Q_m^*)/\hat{S}_0}. \tag{A82}$$

Solving,

$$Q_m^* = \frac{\hat{S}_0}{\mathcal{R}_0 \hat{S}_0 - 1}[\mathcal{R}_0(\hat{I}_0 + \hat{S}_0) - 1 - \ln(\mathcal{R}_0 \hat{S}_0)]. \tag{A83}$$

Substituting from (A79) into the relevant entries in the middle column of Table 6 yields $P_m^* = MPB_m^* = \theta H(\mathcal{R}_0 \hat{S}_0 - 1)/\mathcal{R}_0 \hat{S}_0$. Implicit in this expression is a necessary condition for an interior monopoly solution:

$$\mathcal{R}_0 \hat{S}_0 > 1. \tag{A84}$$

It can be shown that (A84) holds if (60) is violated, which must be the case for an interior monopoly optimum. If (60) does not hold, then $\mathcal{R}_0 \geq |\bar{L}(-\hat{I}_0/\hat{S}_0)|/\hat{I}_0$, implying $-\mathcal{R}_0 \hat{I}_0 \leq \bar{L}(\hat{I}_0/\hat{S}_0)$, implying $-\mathcal{R}_0 \hat{I}_0 e^{-\mathcal{R}_0 \hat{I}_0} \leq -\hat{I}_0/\hat{S}_0$, implying $\mathcal{R}_0 \hat{S}_0 \geq e^{-\mathcal{R}_0 \hat{I}_0} > 1$.

Other equilibrium variables can be computed by substituting for Q_m^* and $P_m^* = MPB_m^*$ into the

other relevant table entries and further substituting for $S(Q_m^*)$ from (A79). We have

$$\Pi_m^* = P_m^* Q_m^* = \theta H \hat{S}_0 \left\{ \hat{I}_0 + \hat{S}_0 - \frac{1}{\mathcal{R}_0}[1 + \ln(\mathcal{R}_0 \hat{S}_0)] \right\} \tag{A85}$$

$$MSB_m^* = \frac{MPB_m^*}{1 - \mathcal{R}_0 S(Q_m^*)} = \frac{\theta H (\mathcal{R}_0 \hat{S}_0 - 1)^2}{\mathcal{R}_0 \hat{S}_0 [\mathcal{R}_0(\hat{I}_0 + \hat{S}_0) - 1 - \ln(\mathcal{R}_0 \hat{S}_0)]} \tag{A86}$$

$$MEX_m^* = \mathcal{R}_0 S(Q_m^*) MSB_m^* = \theta H \frac{\mathcal{R}_0 \hat{S}_0 - 1}{\mathcal{R}_0 \hat{S}_0} \left[\frac{\ln(\mathcal{R}_0 \hat{S}_0) - \mathcal{R}_0 \hat{I}_0}{\mathcal{R}_0(\hat{I}_0 + \hat{S}_0) - 1 - \ln(\mathcal{R}_0 \hat{S}_0)} \right]. \tag{A87}$$

Since $Q^{**} = \hat{S}_0$ in the limiting case under consideration, by Proposition 12,

$$\lim_{\theta \uparrow 1} GS_m^{**} = c - MPB(\hat{S}_0) + \lim_{\theta \uparrow 1} \left(\frac{\theta}{1-\theta} \right) MEX(\hat{S}_0) = \infty. \tag{A88}$$

Q.E.D.

Proof of Proposition 17

It remains to show that the limits of $\partial \Phi_I(Q^*)/\partial Q$ as $\mathcal{R}_0$ approaches either extreme equal 0—and to show this for both market structures and for both long- and short-run analyses.

Completing Proof for Long-Run Analysis: Return for the moment to the model with homogeneous consumers in which $\theta H \Phi_I(Q) = MPB(Q)$, implying $\Phi_I(Q) = MPB(Q)/\theta H$. By equation (16),

$$\Phi_I(Q) = \begin{cases} 0 & Q \geq \mathcal{Q}_0 \\ 1 - \dfrac{1}{(1 - \theta Q/\mu)\mathcal{R}_0} & Q < \mathcal{Q}_0. \end{cases} \tag{A89}$$

The same expression could have been obtained (with more work) by direct computation in the model with heterogeneous consumers. For $\mathcal{R}_0 \leq 1$, $\mathcal{Q}_0 = 0$, implying $\Phi(Q) = 0$ for all Q, implying $\partial \Phi_I(Q)/\partial Q = 0$ for all Q, implying $\lim_{\mathcal{R}_0 \downarrow 0} \partial \Phi_I(Q^*)/\partial Q = 0$ for both $Q^* = Q_c^*$ and $Q^* = Q_m^*$.

For $\mathcal{R}_0 \uparrow \infty$, $\mathcal{Q}_0 = \mu/\theta > \mu$ for $\theta < 1$, implying the second branch of (A89) is the relevant one. Differentiating,

$$\frac{\partial \Phi_I(Q)}{\partial Q} = \frac{\theta}{\mu \mathcal{R}_0 (1 - \theta Q/\mu)^2}. \tag{A90}$$

Since $Q \in [0, \mu]$,

$$\frac{\theta}{\mu \mathcal{R}_0} \leq \frac{\partial \Phi_I(Q)}{\partial Q} \leq \frac{\theta}{\mu \mathcal{R}_0 (1-\theta)^2}. \tag{A91}$$

Taking limits,

$$\lim_{\mathcal{R}_0 \uparrow \infty} \frac{\theta}{\mu \mathcal{R}_0} \leq \lim_{\mathcal{R}_0 \uparrow \infty} \frac{\partial \Phi_I(Q)}{\partial Q} \leq \lim_{\mathcal{R}_0 \uparrow \infty} \frac{\theta}{\mu \mathcal{R}_0 (1-\theta)^2}. \tag{A92}$$

The limits on the far left and far right-hand sides of (A92) equal 0, implying $\lim_{\mathcal{R}_0 \uparrow \infty} \partial \Phi_I(Q)/\partial Q = 0$ for all $Q \in [0, \mu]$, including $Q = Q_c^*$ and $Q = Q_m^*$.

Completing Proof for Short-Run Analysis: Differentiating (44) and substituting from (43) yields

$$\frac{\partial \Phi_I(Q)}{\partial Q} = \frac{-\theta \Phi_I(Q) \mathcal{R}_0 S(Q)}{S_0(Q)[1 - \mathcal{R}_0 S(Q)]}. \tag{A93}$$

Lemma 7 states $\lim_{\mathcal{R}_0 \downarrow 0}[\mathcal{R}_0 S(Q)] = 0$, implying $\lim_{\mathcal{R}_0 \downarrow 0} \partial \Phi_I(Q)/\partial Q = 0$ by (A93). Lemma 8 states $\lim_{\mathcal{R}_0 \uparrow 0}[\mathcal{R}_0 S(Q)] = 0$, implying $\lim_{\mathcal{R}_0 \uparrow \infty} \partial \Phi_I(Q)/\partial Q = 0$ by (A93). These limits both hold for all Q, including $Q = Q_c^*$ and $Q = Q_m^*$. *Q.E.D.*

Proof of Proposition 18

Suppose $\mathcal{R}_0 \leq 1$. Then $Q_{md}^* = Q_{mv}^* = 0$, implying $\Pi_{md}^* = \Pi_{mv}^* = 0$ and $W_{md}^* = W_{mv}^* = W_d^{**} = W_d^{**} = (\alpha + \mu)H$, implying $\Delta \Pi_m^* = \Delta W_m^* = \Delta W^{**} = 0$. The fact that $\Delta \Pi_m^* = 0$ for all $\mathcal{R}_0 < 1$ implies $\lim_{\mathcal{R}_0 \downarrow 0} \Delta \Pi_m^* = 0$.

Suppose $\mathcal{R}_0 > 1$. The assumption $c = 0$ implies $\tilde{c} = 0$, leaving two cases in Table 1: (LR3)–(LR4) and (LR5). In case (LR3)–(LR4), defined by

$$\mathcal{R}_0 \in \left(1, \frac{1}{(1-\theta)^2}\right), \tag{A94}$$

substituting from (66) for Π_{md}^* and from the table entry for Π_{mv}^* yields, after rearranging,

$$\Delta \Pi_m^* = \frac{\sqrt{\mathcal{R}_0} - 1}{\theta \mathcal{R}_0}\left[1 + \theta - (1-\theta)\sqrt{\mathcal{R}_0}\right]. \tag{A95}$$

The first factor is positive since $\sqrt{\mathcal{R}_0} > 1$ by (A94); the second is positive since $\sqrt{\mathcal{R}_0}(1-\theta) < 1$ by (A94). In case (LR5), $\Delta \Pi_m^* = \theta/(1-\theta)\mathcal{R}_0$, implying $\lim_{\mathcal{R}_0 \uparrow \infty} \Delta \Pi_m^* = 0$. Combined with the results for $\mathcal{R}_0 \leq 1$, we have that $\Delta \Pi_m^* \geq 0$ for all $\mathcal{R}_0 > 0$ with strict inequality if and only if $\mathcal{R}_0 > 1$. Further, $\inf_{\mathcal{R}_0 > 0} = 0$.

To verify the quasiconcavity of $\Delta \Pi_m^*$, $\Delta \Pi_m^*$ is a constant 0 in (LR1). In case (LR3)–(LR4),

$$\frac{\partial \Delta \Pi_m^*}{\partial \mathcal{R}_0} = \frac{H}{\mathcal{R}_0^2}\left(1 + \theta - \sqrt{\mathcal{R}_0}\right), \tag{A96}$$

implying that $\Delta \Pi_m^*$ is first increasing, reaches a critical point at $\mathcal{R}_0 = (1+\theta)^2$, and then is decreasing. One can verify that the critical point is in the interior of case (LR3)–(LR4), as $(1+\theta)^2 < 1/(1-\theta)^2$. At the boundary of (LR5), $\Delta \Pi_m^*$ is continuous and continues to decline throughout case (LR5), proving $\Delta \Pi_m^*$ is quasiconcave for all $\mathcal{R}_0 > 0$.

Turning to ΔW_m^*, in case (LR5), substituting from (67) for W_{md}^* and from the relevant table entry for W_{mv}^* yields $\Delta W_m^* = -\theta h/\mathcal{R}_0$, which is negative. In case (LR3)–(LR4), substituting the relevant table entry for W_{mv}^* in case (LR3)–(LR4) yields

$$\Delta W_m^* = \frac{h}{\mathcal{R}_0}\left[(\theta - 1)(\sqrt{\mathcal{R}_0})^2 + \sqrt{\mathcal{R}_0} - \theta\right]. \tag{A97}$$

The sign is determined by the factor in brackets, a quadratic equation in $\sqrt{\mathcal{R}_0}$, which is negative if $\sqrt{\mathcal{R}_0}$ lies outside the roots 1 and $\theta/(1-\theta)$. Since $\mathcal{R}_0 > 1$, the relevant condition is $\sqrt{\mathcal{R}_0} > \theta/(1-\theta)$,

implying $\mathcal{R}_0 > [\theta/(1-\theta)]^2$, the stated condition for $\Delta W_m^* < 0$.

Turning to ΔW^{**}, the first-best quantity is sold in equilibrium with a drug, implying $W_d^{**} = W_{md}^*$. By (24), the first-best vaccine quantity is $Q^{**} = \mathcal{Q}_0 = (\mu/\theta)(1-1/\mathcal{R}_0)$ if $\mathcal{R}_0 \in (1, 1/(1-\theta)]$ and $Q^{**} = \mu$ if $\mathcal{R}_0 > 1/(1-\theta)$. Suppose $\mathcal{R}_0 \in (1, 1/(1-\theta)]$. Then $W_v^{**} = (\alpha + \mu)H$. Substituting this value along with the value of $W_d^* = W_d^{**}$ from (67) and rearranging yields $\Delta W^{**} = W_d^{**} - W_v^{**} = -\mu H(1-\theta)(1-1/\mathcal{R}_0) < 0$. Next, suppose $\mathcal{R}_0 > 1/(1-\theta)$. Then, according to the case (LR5) entry in Table 2 setting $\tilde{c} = 0$, $W_v^{**} = H(\alpha + \theta\mu + \mu/\mathcal{R}_0)$. Substituting this value along with the value of $W_d^* = W_d^{**}$ from (67) and rearranging yields $\Delta W^{**} = W_d^{**} - W_v^{**} = -\theta\mu H/\mathcal{R}_0$. Thus, $\Delta W^{**} \leq 0$ for all $\mathcal{R}_0 > 0$ with strict inequality for $\mathcal{R}_0 > 1$. *Q.E.D.*

Proof of Proposition 19

The assumption $c = 0$ implies $\tilde{c} = 0$, leaving two cases in Table 1: (SR2)–(SR3) and (SR4). Nesting those cases, we can write

$$\Delta\Pi_m^* = \theta H \left\{ \hat{I}_0 + \hat{S}_0 \Phi_I(0) - Q_{mv}^* \Phi_I(Q_{mv}^*) \right\}, \tag{A98}$$

where Q_{mv}^* solves $\max_{Q \in [0,\hat{S}_0]} Q\Phi_I(Q)$. Since $Q_{mv}^* > 0$, we have $Q_{mv}^* \Phi_I(Q_{mv}^*) < Q_{mv}^* \Phi_I(0) \leq \hat{S}_0 \Phi_I(0)$, where the first inquality follows from $\partial\Phi_I(Q)/\partial Q < 0$ by (A93) and the second inequality from $Q_{mv}^* \in [0,\hat{S}_0]$. Substituting the preceding inequality into (A98) yields $\Delta\Pi_m^* > \theta H \hat{I}_0$. Thus, $\Delta\Pi_m^* > 0$ for all $\mathcal{R}_0 > 0$.

To derive the results on limits of $\Delta\Pi_m^*$, we have that $\lim_{\mathcal{R}_0 \downarrow 0} \Phi_I(Q) = \lim_{\mathcal{R}_0 \downarrow 0}[1 - S(Q)/S_0(Q)] = 1$ for all $Q \in [0,\hat{S}_0]$ since $\lim_{\mathcal{R}_0 \downarrow 0} S(Q) = \hat{S}_0 - \theta Q = S_0(Q)$ by Lemma 7. Hence, $\lim_{\mathcal{R}_0 \downarrow 0} \Delta\Pi_m^* = \theta H \hat{I}_0$. For all $Q \in [0,\hat{S}_0]$, $\lim_{\mathcal{R}_0 \uparrow \infty} \Phi_I(Q) = 1$ since $\lim_{\mathcal{R}_0 \uparrow \infty} S(Q) = 0$ by Lemma 8. Hence,

$$\lim_{\mathcal{R}_0 \uparrow \infty} Q_{mv}^* \Phi_I(Q_{mv}^*) = \lim_{\mathcal{R}_0 \uparrow \infty} \left\{ \max_{\mathcal{R}_0 \in [0,\hat{S}_0]} Q\Phi_I(Q) \right\} = \max_{Q \in [0,\hat{S}_0]} Q \lim_{\mathcal{R}_0 \uparrow \infty} \Phi_I(Q) = \hat{S}_0 \cdot 1. \tag{A99}$$

from (A99) into (A98) along with $\lim_{\mathcal{R}_0 \uparrow \infty} \Phi_I(0) = 1$ yields $\lim_{\mathcal{R}_0 \uparrow \infty} \Delta\Pi_m^* = \theta H \hat{I}_0$. Now $\Delta\Pi_m^* > \theta H \hat{I}_0$ for all $\mathcal{R}_0 > 0$ implies $\theta H \hat{I}_0 \leq \inf_{\mathcal{R}_0 > 0} \Delta\Pi_m^* \leq \lim_{\mathcal{R}_0 \downarrow 0} \Delta\Pi_m^* = \theta H \hat{I}_0$, which in turn implies $\inf_{\mathcal{R}_0 > 0} \Delta\Pi_m^* = \theta H \hat{I}_0$.

Combining the results from the previous paragraph, $\lim_{\mathcal{R}_0 \downarrow 0} \Delta\Pi_m^* = \lim_{\mathcal{R}_0 \uparrow \infty} \Delta\Pi_m^* = \inf_{\mathcal{R}_0 > 0} = \theta H \hat{I}_0$. But the first paragraph showed $\Delta\Pi_m^* > \theta H \hat{I}_0$. Hence, $\Delta\Pi_m^*$ must be nonmonotonic in $\mathcal{R}_0$, higher in the interior than for either limiting value of $\mathcal{R}_0$.

Turning to limiting values of ΔW_m^* as $\mathcal{R}_0 \downarrow 0$ and $\mathcal{R}_0 \uparrow \infty$, one can show that (54) holds in these limits. Thus, the relevant case for computing W_{mv}^* is (SR4). Substituting $\tilde{c} = 0$ into the relevant entry of Table 6 and multiplying by $\theta H \hat{S}_0$ to reverse the rescaling yields $W_{mv}^* = H[S(\hat{S}_0) + \theta\hat{S}_0]$. Subtracting from (69) and rearranging yields

$$\Delta W_m^* = H \left[\theta\hat{I}_0 + (1-\theta)S(0) - S(\hat{S}_0) \right]. \tag{A100}$$

By Lemma 7, $\lim_{\mathcal{R}_0 \downarrow 0}[(1-\theta)S(0)] = (1-\theta)\hat{S}_0$. The lemma also implies $\lim_{\mathcal{R}_0 \downarrow 0} S(\hat{S}_0) = (1-\theta)\hat{S}_0$. Substituting these limits into (A100) yields $\lim_{\mathcal{R}_0 \downarrow 0} \Delta W_m^* = \theta H \hat{I}_0$. By Lemma 8, $\lim_{\mathcal{R}_0 \uparrow \infty} S(0) = \lim_{\mathcal{R}_0 \uparrow \infty} S(\hat{S}_0) = 0$. Substituting these limits into (A100) yields $\lim_{\mathcal{R}_0 \uparrow 0} \Delta W_m^* = \theta H \hat{I}_0$.

The final step is to provide parameters for which $\Delta W_m^* < 0$. Using Matlab, we verified that for $\mathcal{R}_0 = 2$, $\theta = 0.5$, $\hat{I}_0 = 0.1$, $\hat{S}_0 = 0.8$, (54) holds, implying that the vaccine monopolist supplies first-best quantity $\hat{S}_0$, putting us in case (SR4). Subtracting the relevant Table 6 entry from (69) and simplifying yields $\Delta W_m^* = H[(1-\theta)S(0)+\theta\hat{I}_0 - S(\hat{S}_0)]$, which Matlab calculations show equals -0.09 for the specified parameters. *Q.E.D.*

Appendix B. Cournot Competition

This appendix provides further details behind the analysis under Cournot competition. Under this market structure, the vaccine is manufactured by $n \geq 1$ homogeneous Cournot firms, which choose quantities each period simultaneously.

B1. Long-Run Analysis

Cases (LR1) and (LR2) from Table 1, which involved no sales under perfect competition, will also involve no sales under Cournot since firms mark up marginal costs. Thus the entries in cases (LR1) and (LR2) from both Tables 1 and 2 will also apply to Cournot.

For the remainder of this section, suppose $\mathcal{R}_0 > 1/(1-\tilde{c})$. Letting q_i denote firm i's output and Q_{-i} is the output of i's rivals, i's profit equals $[P(q_i+Q_{-i})-c]q_i = [MPB(q_i+Q_{-i})-c]q_i$. Taking the first-order condition with respect to q_i and then imposing symmetry by substituting $q_i^* = Q^*/n$ and $Q_{-i}^* = (n-1)Q^*/n$ yields equilibrium market output

$$Q_n^* = \frac{\mu}{\theta}\left[1 - \sqrt{\frac{\psi}{(1-\tilde{c})\mathcal{R}_0}}\right], \tag{B1}$$

where

$$\psi = \frac{(n-1)^2}{4n^2(1-\tilde{c})\mathcal{R}_0}\left[1 + \sqrt{1 + \frac{2n(1-\tilde{c})\mathcal{R}_0}{(n-1)^2}}\right]^2. \tag{B2}$$

Substituting Q_n^* for Q in the relevant equations yields

$$P_n^* = MPB_n^* = \theta H\left(1 - \sqrt{\frac{1-\tilde{c}}{\psi\mathcal{R}_0}}\right) \tag{B3}$$

$$I_n^* = \frac{\mu}{\alpha+\mu}\left[\sqrt{\frac{\psi}{(1-\tilde{c})\mathcal{R}_0}} - \frac{1}{\mathcal{R}_0}\right] \tag{B4}$$

$$\Pi_n^* = \frac{\mu H(1-\tilde{c})}{\theta}\left[1 - \sqrt{\frac{1}{\psi(1-\tilde{c})\mathcal{R}_0}}\right]\left[1 - \sqrt{\frac{\psi}{(1-\tilde{c})\mathcal{R}_0}}\right] \tag{B5}$$

$$MEX_n^* = \mu H\sqrt{\frac{1-\tilde{c}}{\psi\mathcal{R}_0}} \tag{B6}$$

$$W_n^* = H\left[\alpha + \mu\left(1 - \tilde{c} + \frac{1}{\mathcal{R}_0} - \sqrt{\frac{1-\tilde{c}}{\psi\mathcal{R}_0}}\right)\right]. \tag{B7}$$

The preceding analysis is valid if

$$\mathcal{R}_0 \leq \frac{\psi}{(1-\theta)^2(1-\tilde{c})}. \tag{B8}$$

Otherwise, $Q_n^* > \mu$ for the Q_n^* in (B1). Producing more than the number of consumers would result in a market price of zero and zero profits for all firms. Instead, firms produce an equal share of industry output $Q_n^* = \mu$. The rest of the equilibrium variables have the same formula as in case (LR5) of Table 2. Note that the threshold between cases (LR4) and (LR5) is different, given by the right-hand side of (B8).

It is easily seen that $\psi = 1$ for $n = 1$, and thus that the preceding expressions for the equilibrium variables collapse to their monopoly values given in Table 2. It is also easily seen that $\lim_{n\uparrow\infty} \psi = 1/(1-\tilde{c})\mathcal{R}_0$, and thus that the preceding expressions for the equilibrium variables collapse to their values under perfect competition given in Table 1.

B2. Short-Run Analysis

Case (SR1) from Table 5, which involved no sales under perfect competition, will also involve no sales under Cournot since firms mark up marginal costs. Thus the entries in case (SR1) from both Tables 5 and 6 will also apply to Cournot.

For the remainder of this section, suppose $\mathcal{R}_0 > |\ln(1-\tilde{c})|/(\hat{I}_0+\tilde{c}\hat{S}_0)$. As in the long-run analysis, firm i's profit equals $[P(q_i+Q_{-i})-c]q_i = [MPB(q_i+Q_{-i})-c]q_i$. Taking the first-order condition with respect to q_i and then imposing symmetry by substituting $q_i^* = Q^*/n$ yields an equation for the interior equilibrium market output analogous to (53) for a monopoly:

$$MPB(Q_n^*) = P(Q_n^*) = \left\{1 - \frac{\theta Q_n^* \mathcal{R}_0 S(Q_n^*)}{n S_0(Q_n^*)[1-\mathcal{R}_0 S(Q_n^*)]}\right\}^{-1} c. \tag{B9}$$

The only difference is the appearance of n in the denominator of the factor in curly braces. One can see this nests monopoly, setting $n = 1$, and perfect competition, taking the limit $n \uparrow \infty$.

This interior solution is the equilibrium market output under Cournot if $Q_n^* < \hat{S}_0$. Otherwise, $Q_n^* = \hat{S}_0$, and all firms produce an equal share $q_n^* = \hat{S}_0/n$ in the symmetric equilibrium.

Appendix C. Consumer-Group Heterogeneity

This appendix provides the formal analysis behind the results in Section 4.3 on the conditions under which the researcher recovers an unbiased estimate of mean *MEX* using aggregate data and which mean (arithmetic or harmonic) if so. The the long- and short-run analyses are divided into two subsections.

C1. Long-Run Analysis

We begin by discussing the renormalization of the SIR model to ensure that mere separation of a population into subgroups does not drive a change in $\mathcal{R}_0$ or *MEX*. Consider a subgroup $g \in G$ having a population of size $w_g \leq 1$. Modify SIR equations (1)–(5) determining disease epidemiology within

the subgroup as follows:

$$S_{tg} + I_{tg} + R_{tg} = w_g \tag{C1}$$

$$\dot{I}_{tg} = \frac{\beta_g I_{tg} S_{tg}}{w_g} - (\alpha + \mu) I_{tg} \tag{C2}$$

$$\dot{R}_{tg} = \alpha I_{tg} - \mu R_{tg} \tag{C3}$$

$$\dot{S}_{tg} = w_g \mu - \frac{\beta_g I_{tg} S_{tg}}{w_g} - \mu S_{tg}. \tag{C4}$$

This system of equations exhibits several differences from the original. First, the compartments evolve independently for each group so are indexed by the group g. Group size w_g and disease transmissibility β_g are also indexed by g. The remaining parameters (μ, α) are common across groups so not subscripted. Another difference is that there is no vaccine in this thought experiment, so $Q = V_t = 0$. Another difference is that the population size on the right-hand side of (C1) is w_g, not 1. This shows up again in (C4) in that the inflow of newborn susceptibles is $w_g \mu$ not μ as before. The key difference is the division of the $\beta_g I_{tg} S_{tg}$ term by w_g in (C2) and (C4), in effect increasing the per-person contact rate in a smaller population so that total contacts remains constant for a person separated from a larger population into a subgroup.

Let $\mathcal{R}_{0g} = \beta_g / (\alpha + \mu)$. Setting $\dot{I}_{tg} = \dot{R}_{tg} = \dot{S}_{tg} = 0$ in (C1)–(C4) and solving for the steady-state infection rate yields

$$I_g = \max\left[0, \frac{\mu}{\alpha + \mu}\left(1 - \frac{1}{\mathcal{R}_{0g}}\right)\right] w_g, \tag{C5}$$

implying

$$I_g = \frac{w_g \mu}{\alpha + \mu}\left(1 - \frac{1}{\mathcal{R}_{0g}}\right) \tag{C6}$$

if $\mathcal{R}_{0g} > 1$.

Tables 1 and 2 state that $MEX^* = \theta H / \mathcal{R}_0$ under both perfect competition and monopoly for $\mathcal{R}_0 \in (1, 1/(1-\tilde{c})]$. In fact, one can show that the different values MEX^* takes on for yet higher $\mathcal{R}_0$ are due to changes in vaccine consumption behavior at higher $\mathcal{R}_0$, not the increase in $\mathcal{R}_0$ itself. Constraining $Q = 0$, which shuts down changes in consumer behavior, one can show via direct calculation that $MEX(0) = \theta H / \mathcal{R}_0$ for all $\mathcal{R}_0 > 1$. This relationship can also be shown to hold at the group level. Hence, for all $\mathcal{R}_{0g} > 1$,

$$MEX_g(0) = \frac{\theta H}{\mathcal{R}_{0g}}. \tag{C7}$$

The researcher believes (C5) applies to the population as a whole with a homogeneous basic reproductive ratio $\mathcal{R}_0$. Dropping the group subscript, setting $w_g = 1$, and assuming that the research observes a positive aggregate infection rate gives the following estimating equation for $\mathcal{R}_0$:

$$\mathcal{R}_0 = \left[1 - \left(\frac{\alpha + \mu}{\mu}\right) \sum_{g \in G} I_g\right]^{-1}. \tag{C8}$$

Combining this with the relationship $MEX(0) = \theta H / \mathcal{R}_0$ that the researcher believes applies at the

aggregate level yields the following estimating equation for the aggregate marginal externality:

$$MEX(0) = \theta H \left[1 - \left(\frac{\alpha + \mu}{\mu}\right) \sum_{g \in G} I_g\right] = \sum_{g \in G} \left(w_g \frac{\theta H}{\mathcal{R}_{0g}}\right) = \sum_{g \in G} w_g MEX_g(0). \tag{C9}$$

The second equality follows from substituting from (C6), and the last equality follows from substituting from (C7). Thus, if $\mathcal{R}_{0g} > 1$ for all $g \in G$, the naïve estimate of aggregate $MEX(0)$ is the weighted arithmetic mean of $MEX_g(0)$.

Suppose $\mathcal{R}_{0g} \leq 1$ for $g \in G'$ and $\mathcal{R}_{0g} > 1$ for $g \in G''$, where $G', G'' \neq \emptyset$. Then $I_g = 0$ for $g \in G'$, implying (C9) becomes

$$MEX(0) = \theta H \left[1 - \left(\frac{\alpha + \mu}{\mu}\right) \sum_{g \in G''} I_g\right] \tag{C10}$$

$$= \theta H \left[1 - \sum_{g \in G''} w_g \left(1 - \frac{1}{\mathcal{R}_{0g}}\right)\right] \tag{C11}$$

$$= \theta H \left(1 - \sum_{g \in G''} w_g\right) + \sum_{g \in G''} MEX_g(0) \tag{C12}$$

$$= \theta H \sum_{g \in G'} w_g + \sum_{g \in G} MEX_g(0). \tag{C13}$$

Equation (C13) shows that the researcher's estimate of $MEX(0)$ is biased upward by a term proportional to the share of the population in groups with $\mathcal{R}_{0g} \leq 1$. Implicit in the researcher's ability to compute $MEX(0)$ from an estimate of $\mathcal{R}_0$ is knowledge of θH. If the researcher also knows $\sum_{g \in G'} w_g$, an unbiased estimate of $\sum_{g \in G} MEX_g(0)$ can be computed simply by subtracting $\theta H \sum_{g \in G'} w_g$ from the naïve estimator $MEX(0)$.

C2. Short-Run Analysis

For the short-run analysis, we again need to renormalize the basic SIR model to ensure that mere separation of a population into subgroups does not drive a change in $\mathcal{R}_0$ or MEX. Modify the relevant SIR equations (29)–(33) as follows:

$$S_{tg} + I_{tg} + R_{tg} = w_g \tag{C14}$$

$$\dot{I}_{tg} = \frac{\beta_g I_{tg} S_{tg}}{w_g} - \alpha I_{tg} \tag{C15}$$

$$\dot{R}_{tg} = \alpha I_{tg} \tag{C16}$$

$$\dot{S}_{tg} = -\frac{\beta_g I_{tg} S_{tg}}{w_g}. \tag{C17}$$

These are similar to (C1)–(C4) except that terms involving μ capturing births and deaths are omitted from short-run model. See the previous section for an explanation of how the original system of equations maps into the renormalized system. Since $Q = 0$ because there is no vaccine in this thought experiment, we have $S_{0g} = \hat{S}_{0g}$. Consistent with our approach in the long-run analysis,

where we assumed all parameters were common across groups but β_g, here assume the initial share of susceptibles and infecteds are common across groups: $\hat{S}_{0g} = w_g \hat{S}_0$ and $I_{0g} = \hat{I}_{0g} = w_g \hat{I}_0 = w_g I_0$.

For the short-run analysis, redefine $\mathcal{R}_{0g} = \beta_g / \alpha$. Since $Q = 0$, $S_{0g} = \hat{S}_{0g}$. Substituting these equations into (C15) expressed at $t = 0$ yields

$$\frac{\dot{I}_{0g}}{I_{0g}} = \alpha \left(\frac{\mathcal{R}_{0g} \hat{S}_{0g}}{w_g} - 1 \right), \tag{C18}$$

approximated by the discrete change in I_{tg} over the short interval from $t = 0$ to $t = 1$ by

$$\frac{I_{1g} - I_{0g}}{I_{0g}} = \alpha \left(\frac{\mathcal{R}_{0g} \hat{S}_{0g}}{w_g} - 1 \right). \tag{C19}$$

Substituting $I_{0g} = w_g I_0$ and rearranging yields

$$I_{1g} = w_g I_0 (1 - \alpha + \alpha \mathcal{R}_{0g} \hat{S}_0). \tag{C20}$$

The same argument used to show that (C7) holds in the long-run analysis can be used to show it holds in the short-run analysis as well.

The researcher believes (C19) applies to the population as a whole with a homogeneous basic reproductive ratio $\mathcal{R}_0$. Dropping the group subscript, setting $w_g = 1$, and rearranging gives the following estimating equation for $\mathcal{R}_0$:

$$\mathcal{R}_0 = \frac{1}{\hat{S}_0} \left[1 + \frac{1}{\alpha} \left(\frac{I_1 - I_0}{I_0} \right) \right]. \tag{C21}$$

Combining this with the relationship $MEX(0) = \theta H / \mathcal{R}_0$ that the researcher believes, in accordance with (C7), applies at the aggregate level yields the following estimating equation for the aggregate marginal externality:

$$MEX(0) = \theta H \hat{S}_0 \left[1 + \frac{1}{\alpha} \left(\frac{I_1 - I_0}{I_0} \right) \right]^{-1} = \theta H \hat{S}_0 \left[1 + \frac{1}{\alpha} \left(\frac{1}{I_0} \sum_{g \in G} I_{1g} - 1 \right) \right]^{-1}. \tag{C22}$$

Substituting from (C20) and rearranging yields

$$MEX(0) = \left(\sum_{g \in G} w_g \frac{\mathcal{R}_{0g}}{\theta H} \right)^{-1} = \left[\sum_{g \in G} \frac{w_g}{MEX_g(0)} \right]^{-1}, \tag{C23}$$

where the second equality follows from substituting from (C7). This last expression is the weighted harmonic mean of $MEX_g(0)$.

References

Ahituv, Avner, V. Joseph Holz, and Tomas Philipson. (1996) "The Responsiveness of the Demand for Condoms to the Local Prevalence of AIDS," *Journal of Human Resources* 31(4): 869–897.

Althouse, Benjamin M., Theodore C. Bergstrom, and Carl T. Bergstrom. (2010) "A Public Choice Framework for Controlling Transmissible and Evolving Diseases," *Proceedings of the National Academy of Sciences* 107(1): 1696.

Anderson, Roy and Robert May. (1991) *Infectious Diseases of Humans*. Oxford: Oxford University Press.

Bailey, Norman. (1975) *The Mathematical Theory of Infectious Diseases and Its Application*. London: Griffen.

Bailey, Robert C., Stephen Moses, Corette B. Parker, Kawango Agot, Kian Maclean, John N. Krieger, Carolyn F. M. Williams, Richard T. Campbell, and Jeckoniah O. Ndinya-Achola. (2007) "Male Circumcision for HIV Prevention in Young Men in Kisumu, Kenya: A Randomised Controlled Trial," *Lancet* 369: 643–656.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Dhruva Kothari (2010) "Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives," *British Medical Journal* 340.

Bauch, Chris T., and David J. D. Earn. (2004) "Vaccination and the Theory of Games," *Proceedings of the National Academy of Sciences* 101(36): 13391-13394

Berndt, Ernst, Rachel Glennerster, Michael Kremer, Jean Lee, Ruth Levine, Georg Weisacker, and Heidi Williams. (2007) "Advance Market Commitments for Vaccines Against Neglected Diseases: Estimating Costs and Effectiveness," *Health Economics* 16: 491–511.

Bethune, Zachary and Anton Korinek (2020) "Covid-19 Infection Externalities: Trading Off Lives vs. Livelihoods," National Bureau of Economic Research Working Paper. (No. w27009).

Boulier, Bryan, Tejwant Datta, and Robert Goldfarb. (2007) "Vaccination Externalities," *The B.E. Journal of Economic Analysis & Policy* 7: www.bepress.com/bejeap/vol7/iss1/art23.

Brennan, Michael J. (2009) "The US Food and Drug Administration Provides a Pathway for Licensing Vaccines for Global Diseases," *PLoS Medicine* 6(7): e1000095.

Brito, Dagobert L., Eytan Sheshinski, and Michael D. Intrilligator. (1991) "Externalities and Compulsory Vaccination," *Journal of Public Economics* 45: 69–90.

Chen, Frederick and Flavio Toxvaerd. (2014) "The Economics of Vaccination," *Journal of Theoretical Biology* 363: 105–117.

Chowell, Gerardo, Paul W. Fenimore, Melissa A. Castillo-Garsow, and Carlos Castillo-Chavez. (2003) "SARS Outbreaks in Ontario, Hong Kong and Singapore: The Role of Diagnosis and Isolation As a Control Mechanism," *Journal of Theoretical Biology* 224: 1–8.

Cook, Joseph, Marc Jeuland, Brian Maskery, Donald Lauria, Dipika Sur, John Clemens, and Dale Whittington. (2009) "Using Private Demand Studies to Calculate Socially Optimal Caccine Subsidies in Developing Countries," *Journal of Policy Analysis and Management* 28 (1): 6–28.

Diekmann, O. , J. A. P. Heesterbeek, and J. A. J. Metz. (1990) "On the Definition and the Computation of the Basic Reproduction Ratio R_0 in Models for Infections Diseases in Heterogeneous Populations," *Journal of Mathematical Biology* 28: 365–382.

Enayati, Shakiba and Osman Y. ʿOzaltin. (2020) "Optimal Influenza Vaccine Distribution with Equity," *European Journal of Operational Research* 283: 714–725.

Francis, Peter J. (1997) "Dynamic Epidemiology and the Market for Vaccinations," *Journal of Public Economics* 63: 383–406.

Funk, Sebastian, Marcel Salathé, and Vincent A. A. Jansen. (2010) "Modelling the Influence of Human Behaviour on the Spread of Infectious Diseases: A Review," *Journal of The Royal Society Interface* 7 (50): 1247–56.

Galeotti, Andrea, and Brian W. Rogers. (2013) "Strategic Immunization and Group Structure," *American Economic Journal: Microeconomics* 5(2): 1–32.

Geoffard, Pierre-Yves and Tomas Philipson. (1997) "Disease Eradication: Public vs. Private Vaccination," *American Economic Review* 87: 222–230.

Gersovitz, Mark (2003) "Births, Recoveries, Vaccinations, and Externalities," in R. Arnott (ed.) *Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz*, pp. 469–483.

Gersovitz, Mark and Jeffrey S. Hammer. (2004) "The Economical Control of Infectious Diseases," *Economic Journal* 114: 1–27.

Gersovitz, Mark and Jeffrey S. Hammer. (2005) "Tax/Subsidy Policy Toward Vector-Borne Infectious Diseases," *Journal of Public Economics* 89: 647–674.

Hoerger, Thomas J., Sarah Schillie, John S. Wittenborn, Christina L. Bradley, Fangjun Zhou, Kathy Byrd, and Trudy V. Murphy. (2013) "Cost Effectiveness of Hepatitis B Vaccination in Adults with Diagnosed Diabetes," *Diabetes Care* 36: 63–69.

Keeling, Matt J. and J. V. Ross. (2015) "Optimal Prophylactic Vaccination in Segregated Populations: When Can We Improve on the Equalising Strategy?" *Epidemics* 11: 7–13.

Keeling, Matt J. and Andrew Shattock. (2012) "Optimal but Inequitable Prophylactic Distribution of Vaccine," *Epidemic* 4: 78–85.

Kennedy, Caitlin E., Ping T. Yeh, Kaitlyn Atkins, Virginia A. Fonner, Michael D. Sweat, Kevin R. O'Reilly, George W. Rutherford, Rachel Baggaley, and Jluia Samuelson. (2020) "Economic Compensation Interventions to Increase Uptake of Voluntary Medical Male Circumcision for HIV Prevention: A Systematic Review and Meta-Analysis," *PLoS ONE* 15(1): e0227623.

Kermack, William O. and A. G. McKendrick. (1927) "A Contribution to the Mathematical Theory of Epidemics," *Proceedings of the Royal Society A (Mathematical, Physical and Engineering Sciences)* 115: 700–721.

Kissler, Stephen M., Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. (2020) "Projecting the Transmission Dynamics of SARS-CoV-2 Through the Postpandemic Period" *Science* 368(6493): 860–868.

Kremer, Michael. (1996) "Integrating Behavioral Choice into Epidemiological Models of AIDS," *Quarterly Journal of Economics* 111(2): 549–573.

Kremer, Michael and Christopher M. Snyder. (2015) "Preventives vs. Treatments," *Quarterly Journal of Economics* 130(3): 1167–1239

Kremer, Michael and Christopher M. Snyder. (2018) "Preventives versus Treatments Redux: Tighter bounds on Distortions in Innovation Incentives with an Application to the Global Demand for HIV Pharmaceuticals," *Review of Industrial Organization* 53 (1): 235–273

Manfredi, Piero, and Alberto D'Onofrio (ed.) (2013) *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*. New York: Springer-Verlag.

Manski, Charles F. (2017) "Mandating Vaccination with Unknown Indirect Effects," *Journal of Public Economic Theory* 19: 603–619.

Marseille Elliott, Bruce Larson, Dhruv S. Kazi, James G. Kahn, and Sydney Rosen. (2015) "Thresholds for the Cost-Effectiveness of Interventions: Alternative Approaches," *Bulletin of the World Health Organization* 93: 118—124.

Martcheva, Maia. (2015) *An Introduction to Mathematical Epidemiology*. New York: Springer.

May, Robert. (2000) "Simple Rules with Complex Dynamics," *Science* 287(5453): 601–602.

Mechoulan, Stéphane. (2007) "Market Structure and Communicable Diseases," *Canadian Journal of Economics* 40: 468–492.

Nguyen, Chantal and Jean M. Carlson. (2016) 'Optimizing Real-Time Vaccine Allocation in a Stochastic SIR Model," *PLoS One* 11(4): e0152950.

Philipson, Tomas. (2000) "Economic Epidemiology and Infectious Disease," in J. Newhouse and T. Culyer (ed.) *Handbook of Health Economics, Volume IB*: 443–465.

Tasslimi, Azadeh, Mari M. Nakamura, Orin Levine, Maria D. Knoll, Louise B. Russell, and Anushua Sinha. (2011) "Cost Effectiveness of Child Pneumococcal Conjugate Vaccination in GAVI-Eligible Countries," *International Health* 3: 259–269.

Thomas, Patricia. (2002) "The Economics of Vaccines," Harvard Medical International (HMI) World. September/October.

Ubel, Peter A., Richard A. Hirth, and Michael E. Chernew. (2003) "What is the Price of Life and Why Doesn't It Increase at the Rate of Inflation?" *Archives of Internal Medicine* 163: 1637–1641.

U.K. Government Office for Science. (2020) "The R Number and Growth Rate in the UK." [Internet.] Downloaded September 28, 2020 from https://www.gov.uk/guidance/the-r-number-in-the-uk#latest-r-number-and-growth-rate.

U.K. Office for National Statistics. (2020) "Coronavirus (COVID-19) Infection Survey Pilot: England, Wales and Northern Ireland, 25 September 2020." [Internet.] Downloaded September 28, 2020 from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/englandwalesandnorthernireland25september2020#number-of-people-in-england-who-had-covid-19.

UNAIDS/WHO/SACEMA Expert Group on Modelling the Impact and Cost of Male Circumcision for HIV Prevention. (2009) "Male Circumcision for HIV Prevention in High HIV Prevalence Settings: What Can Mathematical Modelling Contribute to Informed Decision Making?" *PLoS Medicine* 6(9): e1000109.

U.S. Centers for Disease Control and Prevention. (2020a) "CDC Vaccine Price List." [Internet.] Downloaded October 8, 2020 from https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html.

U.S. Centers for Disease Control and Prevention. (2020b) "Vaccines and Preventable Diseases: Vaccines by Disease." [Internet.] Vaccine efficacies downloaded August 28, 2020 from https://www.cdc.gov/vaccines/pubs/pinkbook/hepb.html for hepatitis B, from https://www.cdc.gov/vaccines/vpd/mmr/public/index.html for measles, from https://www.cdc.gov/vaccines/vpd/dtap-tdap-td/hcp/about-vaccine.html for pertussis and tetanus, from https://www.cdc.gov/vaccines/vpd/polio/index.html for polio, and from from https://www.cdc.gov/vaccines/vpd/shingles/index.html for shingles.

U.S. General Accounting Office. (2002) "Childhood Vaccines: Ensuring an Adequate Supply Poses Continuing Challenges," Report no. GAO-02-987.

Table 1: Long-Run Equilibrium Variables under Perfect Competition as Functions of $\mathcal{R}_0$

	Case			
	(LR1)	(LR2)	(LR3)	(LR4), (LR5)
Variable	$\mathcal{R}_0 \in [0,1]$	$\mathcal{R}_0 \in \left(1, \frac{1}{1-\tilde{c}}\right]$	$\mathcal{R}_0 \in \left(\frac{1}{1-\tilde{c}}, \frac{1}{(1-\theta)(1-\tilde{c})}\right]$	$\mathcal{R}_0 \in \left(\frac{1}{(1-\theta)(1-\tilde{c})}, \infty\right)$
P_c^*	c	c	c	c
Q_c^*	0	0	$\frac{\mu}{\theta}\left[1-\frac{1}{(1-\tilde{c})\mathcal{R}_0}\right]$	μ
Π_c^*	0	0	0	0
I_c^*	0	$\frac{\mu}{\alpha+\mu}\left(1-\frac{1}{\mathcal{R}_0}\right)$	$\frac{\mu}{\alpha+\mu}\left[\frac{\tilde{c}}{(1-\tilde{c})\mathcal{R}_0}\right]$	$\frac{\mu}{\alpha+\mu}\left(1-\theta-\frac{1}{\mathcal{R}_0}\right)$
MPB_c^*	0	$\theta H\left(1-\frac{1}{\mathcal{R}_0}\right)$	c	$\theta H\left[1-\frac{1}{(1-\theta)\mathcal{R}_0}\right]$
MSB_c^*	0	θH	θH	θH
MEX_c^*	0	$\frac{\theta H}{\mathcal{R}_0}$	$\theta H(1-\tilde{c})$	$\frac{\theta H}{(1-\theta)\mathcal{R}_0}$
W_c^*	$H(\alpha+\mu)$	$H\left(\alpha+\frac{\mu}{\mathcal{R}_0}\right)$	$H[\alpha+\mu(1-\tilde{c})]$	$H\left[\alpha+\theta\mu(1-\tilde{c})+\frac{\mu}{\mathcal{R}_0}\right]$

Notes: The distinction between cases (LR4) and (LR5) in the last column, relevant for monopoly in the next table, is irrelevant for perfect competition here.

Table 2: Long-Run Equilibrium Variables under Monopoly as Functions of $\mathcal{R}_0$

			Case	
	(LR1)	(LR2)	(LR3), (LR4)	(LR5)
Variable	$\mathcal{R}_0 \in [0,1]$	$\mathcal{R}_0 \in \left(1, \frac{1}{1-\tilde{c}}\right]$	$\mathcal{R}_0 \in \left(\frac{1}{1-\tilde{c}}, \frac{1}{(1-\theta)^2(1-\tilde{c})}\right]$	$\mathcal{R}_0 \in \left(\frac{1}{(1-\theta)^2(1-\tilde{c})}, \infty\right)$
P_m^*	†	†	$\theta H \left(1 - \sqrt{\frac{1-\tilde{c}}{\mathcal{R}_0}}\right)$	$\theta H \left[1 - \frac{1}{(1-\theta)\mathcal{R}_0}\right]$
Q_m^*	0	0	$\frac{\mu}{\theta}\left[1 - \frac{1}{\sqrt{(1-\tilde{c})\mathcal{R}_0}}\right]$	μ
Π_m^*	0	0	$\frac{\mu H(1-\tilde{c})}{\theta}\left[1 - \frac{1}{\sqrt{(1-\tilde{c})\mathcal{R}_0}}\right]^2$	$\mu\theta H\left[1 - \tilde{c} - \frac{1}{(1-\theta)\mathcal{R}_0}\right]$
I_m^*	0	$\frac{\mu}{\alpha+\mu}\left(1 - \frac{1}{\mathcal{R}_0}\right)$	$\frac{\mu}{\alpha+\mu}\left[\frac{1}{\sqrt{(1-\tilde{c})\mathcal{R}_0}} - \frac{1}{\mathcal{R}_0}\right]$	$\frac{\mu}{\alpha+\mu}\left(1 - \theta - \frac{1}{\mathcal{R}_0}\right)$
MPB_m^*	0	$\theta H\left(1 - \frac{1}{\mathcal{R}_0}\right)$	$\theta H\left(1 - \sqrt{\frac{1-\tilde{c}}{\mathcal{R}_0}}\right)$	$\theta H\left[1 - \frac{1}{(1-\theta)\mathcal{R}_0}\right]$
MSB_m^*	0	θH	θH	θH
MEX_m^*	0	$\frac{\theta H}{\mathcal{R}_0}$	$\theta H\sqrt{\frac{1-\tilde{c}}{\mathcal{R}_0}}$	$\frac{\theta H}{(1-\theta)\mathcal{R}_0}$
W_m^*	$H(\alpha+\mu)$	$H\left(\alpha + \frac{\mu}{\mathcal{R}_0}\right)$	$H\left[\alpha + \mu\left(1 - \tilde{c} + \frac{1}{\mathcal{R}_0} - \sqrt{\frac{1-\tilde{c}}{\mathcal{R}_0}}\right)\right]$	$H\left[\alpha + \mu\theta(1-\tilde{c}) + \frac{\mu}{\mathcal{R}_0}\right]$

Notes: The distinction between cases (LR3) and (LR4) in the penultimate column, relevant for perfect competition in previous table, is irrelevant for monopoly here. †Any value $P_m^* \geq c$ is consistent with zero sales in equilibrium.

Table 3: Long-Run Optimal Subsidy GS^{**} as a Function of $\mathcal{R}_0$

			Case		
	(LR1)	(LR2′)	(LR3′)	(LR4)	(LR5)
Variable	$\mathcal{R}_0 \in [0,1]$	$\mathcal{R}_0 \in \left(1, \frac{1}{1-\theta}\right]$	$\mathcal{R}_0 \in \left(\frac{1}{1-\theta}, \frac{1}{(1-\theta)(1-\bar{c})}\right]$	$\mathcal{R}_0 \in \left(\frac{1}{(1-\theta)(1-\bar{c})}, \frac{1}{(1-\theta)^2(1-\bar{c})}\right]$	$\mathcal{R}_0 \in \left(\frac{1}{(1-\theta)^2(1-\bar{c})}, \infty\right)$
GS_c^{**}	0	c	$c+\theta H\left[\frac{1}{(1-\theta)\mathcal{R}_0}-1\right]$	0	0
GS_m^{**}	0	$c+\theta H(\mathcal{R}_0-1)$	$c+\theta H\left[\frac{1}{(1-\theta)^2\mathcal{R}_0}-1\right]$	$c+\theta H\left[\frac{1}{(1-\theta)^2\mathcal{R}_0}-1\right]$	0

Notes: Cases (LR1), (LR4), and (LR5) are identical to those in Tables 1 and 2. Cases (LR2′) and (LR3′) differ from (LR2) and (LR3) in the previous tables in that the boundary between (LR2′) and (LR3′) is $1/(1-\theta)$ rather than $1/(1-\bar{c})$.

Table 4: Estimates of $\mathcal{R}_0$ from the Epidemiology Literature

Disease	Time period	Location	$\mathcal{R}_0$	Source
SARS	2002	Singapore	1.1	Chowell *et al.* (2003)
HIV	1981–85	United Kingdom	2–5	Anderson and May (1991)
Rubella	1960–70	United Kingdom	6–7	Anderson and May (1991)
Chicken pox	1944–68	United Kingdom	10–12	Anderson and May (1991)
Mumps	1960–80	United Kingdom	11–14	Anderson and May (1991)
Measles	1950–68	United Kingdom	16–18	Anderson and May (1991)
Pertussis	1944–78	United Kingdom	16–18	Anderson and May (1991)

Notes: All Anderson and May (1991) estimates are from their Table 4.1, reporting estimates from studies of England and Wales within the United Kingdom. The HIV estimate is for Type 1 in the subpopulation of men who have sex with men (MSM).

Table 5: Short-Run-Equilibrium Variables under Perfect Competition as Functions of $\mathcal{R}_0$

|	Case		
	(SR1)	(SR2)	(SR3), (SR4)
Variable	$\mathcal{R}_0 \in \left(0, \frac{\lvert \ln(1-\tilde{c})\rvert}{\hat{I}_0 + \tilde{c}\hat{S}_0}\right]$	$\mathcal{R}_0 \in \left(\frac{\lvert \ln(1-\tilde{c})\rvert}{\hat{I}_0 + \tilde{c}\hat{S}_0}, \frac{\lvert \ln(1-\tilde{c})\rvert}{\hat{I}_0 + (1-\theta)\tilde{c}\hat{S}_0}\right]$	$\mathcal{R}_0 \in \left(\frac{\lvert \ln(1-\tilde{c})\rvert}{\hat{I}_0 + (1-\theta)\tilde{c}\hat{S}_0}, \infty\right)$
P_c^*	c	c	c
Q_c^*	0	$\frac{1}{\theta}\left\{\hat{S}_0 + \frac{1}{\tilde{c}}\left[\frac{1}{\mathcal{R}_0}\ln(1-\tilde{c}) + \hat{I}_0\right]\right\}$	$\hat{S}_0$
Π_c^*	0	0	0
R_c^*	$1 - S(0)$	$1 - \hat{S}_0 - \hat{I}_0 + \frac{1}{\mathcal{R}_0}\lvert\ln(1-\tilde{c})\rvert$	$1 - S(\hat{S}_0) - \theta\hat{S}_0$
MPB_c^*	$\theta H \Phi_I(0)$	c	$\theta H \Phi_I(\hat{S}_0)$
MSB_c^*	$\frac{\theta H \Phi_I(0)}{1 - \mathcal{R}_0 S(0)}$	$\frac{\theta H \tilde{c}^2}{\tilde{c} + (1-\tilde{c})[\ln(1-\tilde{c}) + \mathcal{R}_0 \hat{I}_0]}$	$\frac{\theta H \Phi_I(\hat{S}_0)}{1 - \mathcal{R}_0 S(\hat{S}_0)}$
MEX_c^*	$\frac{\theta H \Phi_I(0)\mathcal{R}_0 S(0)}{1 - \mathcal{R}_0 S(0)}$	$\frac{\theta H \tilde{c}(1-\tilde{c})[\lvert\ln(1-\tilde{c})\rvert - \mathcal{R}_0 \hat{I}_0]}{\tilde{c} + (1-\tilde{c})[\ln(1-\tilde{c}) + \mathcal{R}_0 \hat{I}_0]}$	$\frac{\theta H \Phi_I(\hat{S}_0)\mathcal{R}_0 S(\hat{S}_0)}{1 - \mathcal{R}_0 S(\hat{S}_0)}$
W_c^*	$H S(0)$	$H \hat{S}_0 (1-\tilde{c})$	$H\left[\theta \hat{S}_0(1-\tilde{c}) + S(\hat{S}_0)\right]$

Notes: Computable expressions for $S(0)$ and $S(\hat{S}_0)$ can be derived from equation (42): defining $\hat{\epsilon}_I = e^{-\mathcal{R}_0 \hat{I}_0}$ and $\hat{\epsilon}_S = e^{-\mathcal{R}_0 \hat{S}_0}$, we have $S(0) = \lvert\bar{L}(-\mathcal{R}_0 \hat{S}_0 \hat{\epsilon}_I \hat{\epsilon}_S)\rvert/\mathcal{R}_0$ and $S(\hat{S}_0) = \lvert\bar{L}(-\mathcal{R}_0(1-\theta)\hat{S}_0\hat{\epsilon}_I\hat{\epsilon}_S^{1-\theta})\rvert/\mathcal{R}_0$. Infection probabilities can be computed using (44): $\Phi_I(0) = 1 - S(0)/\hat{S}_0$ and $\Phi_I(\hat{S}_0) = 1 - S(\hat{S}_0)/(1-\theta)\hat{S}_0$. The distinction between cases (SR3) and (SR4) in the last column, relevant for monopoly in the next table, is irrelevant for perfect competition here.

Table 6: Short-Run-Equilibrium Variables under Monopoly as Functions of $\mathcal{R}_0$

| Variable	Case		
	(SR1) $\mathcal{R}_0 \in \left(0, \frac{|\ln(1-\tilde{c})|}{\hat{I}_0 + \tilde{c}\hat{S}_0}\right]$	(SR2), (SR3) $\mathcal{R}_0 > |\ln(1-\tilde{c})|/(\hat{I}_0 + \tilde{c}\hat{S}_0)$ but does satisfy (54)	(SR4) $\mathcal{R}_0$ satisfies (54)
P_m^*	†	$\theta H \left[1 - \frac{S(Q_m^*)}{\hat{S}_0 - \theta Q_m^*}\right]$	$\theta H \left[1 - \frac{S(\hat{S}_0)}{(1-\theta)\hat{S}_0}\right]$
Q_m^*	0	‡	$\hat{S}_0$
Π_m^*	0	$\theta H \left[1 - \frac{S(Q_m^*)}{\hat{S}_0 - \theta Q_m^*} - \tilde{c}\right] Q_m^*$	$\theta H \left[(1-\tilde{c})\hat{S}_0 - \frac{S(\hat{S}_0)}{(1-\theta)}\right]$
R_c^*	$1 - S(0)$	$1 - S(Q_m^*) - \theta Q_m^*$	$1 - S(\hat{S}_0) - \theta \hat{S}_0$
MPB_m^*	$\theta H \Phi_I(0)$	$\theta H \Phi_I(Q_m^*)$	$\theta H \Phi_I(\hat{S}_0)$
MSB_m^*	$\frac{\theta H \Phi_I(0)}{1 - \mathcal{R}_0 S(0)}$	$\frac{\theta H \Phi_I(Q_m^*)}{1 - \mathcal{R}_0 S(Q_m^*)}$	$\frac{\theta H \Phi_I(\hat{S}_0)}{1 - \mathcal{R}_0 S(\hat{S}_0)}$
MEX_m^*	$\frac{\theta H \Phi_I(0) \mathcal{R}_0 S(0)}{1 - \mathcal{R}_0 S(0)}$	$\frac{\theta H \Phi_I(Q_m^*) \mathcal{R}_0 S(Q_m^*)}{1 - \mathcal{R}_0 S(Q_m^*)}$	$\frac{\theta H \Phi_I(\hat{S}_0) \mathcal{R}_0 S(\hat{S}_0)}{1 - \mathcal{R}_0 S(\hat{S}_0)}$
W_m^*	$HS(0)$	$H\left[(1-\tilde{c})\theta \tilde{Q}_m^* + S(Q_m^*)\right]$	$H\left[\theta \hat{S}_0(1-\tilde{c}) + S(\hat{S}_0)\right]$

Notes: Equation (42) provides a formula for computing $S(Q_m^*)$ given Q_m^*, and the previous table provides simpler expressions for computing $S(0)$ and $S(\hat{S}_0)$. Infection probabilities can be computed using (44): $\Phi_I(0) = 1 - S(0)/\hat{S}_0$, $\Phi_I(\hat{S}_0) = 1 - S(\hat{S}_0)/(1-\theta)\hat{S}_0$, and $\Phi_I(Q_m^*) = 1 - S(Q_m^*)/(\hat{S}_0 - \theta Q_m^*)$. The distinction between cases (SR2) and (SR3) in the middle column, relevant for perfect competition in previous table, is irrelevant for monopoly here. The set of $\mathcal{R}_0$ satisfying equation (54) need not form an interval. †Any value $P_m^* \geq c$ is consistent with zero sales in equilibrium. ‡Entry is the Q_m^* solving (53).

Figure 1: Graphs of Long-Run Equilibrium Variables as Functions of $\mathcal{R}_0$
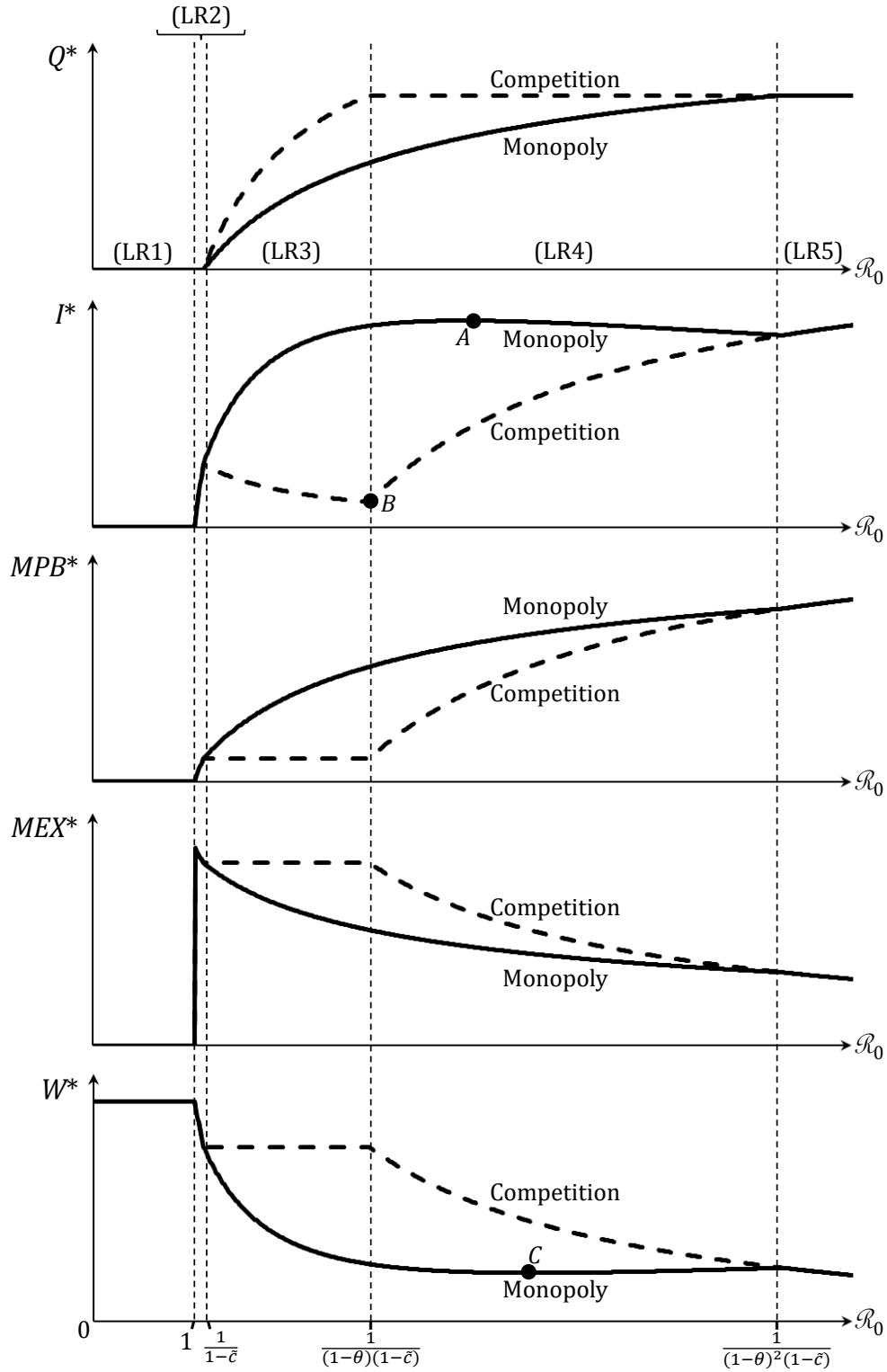
Notes: Graph of formulas provided in Tables 1 and 2, illustrated for specific parameter values ($\theta = 0.6$, $\alpha = 0$, $\mu = 1$, $c = 0.1$, $H = 1$). Where the dotted and solid curves overlap, the solid curve represents both industry structures. Panel illustrating equilibrium prices omitted since $P_m^* = MPB_m^*$ in all cases in which the monopolist serves some consumers and since $P_c^* = c$ does not require illustration. Also omitted is a panel illustrating equilibrium profit since the graph for Π_m^* resembles that for Q_m^* or MPB_m^* and since $\Pi_c^* = 0$ does not require illustration.
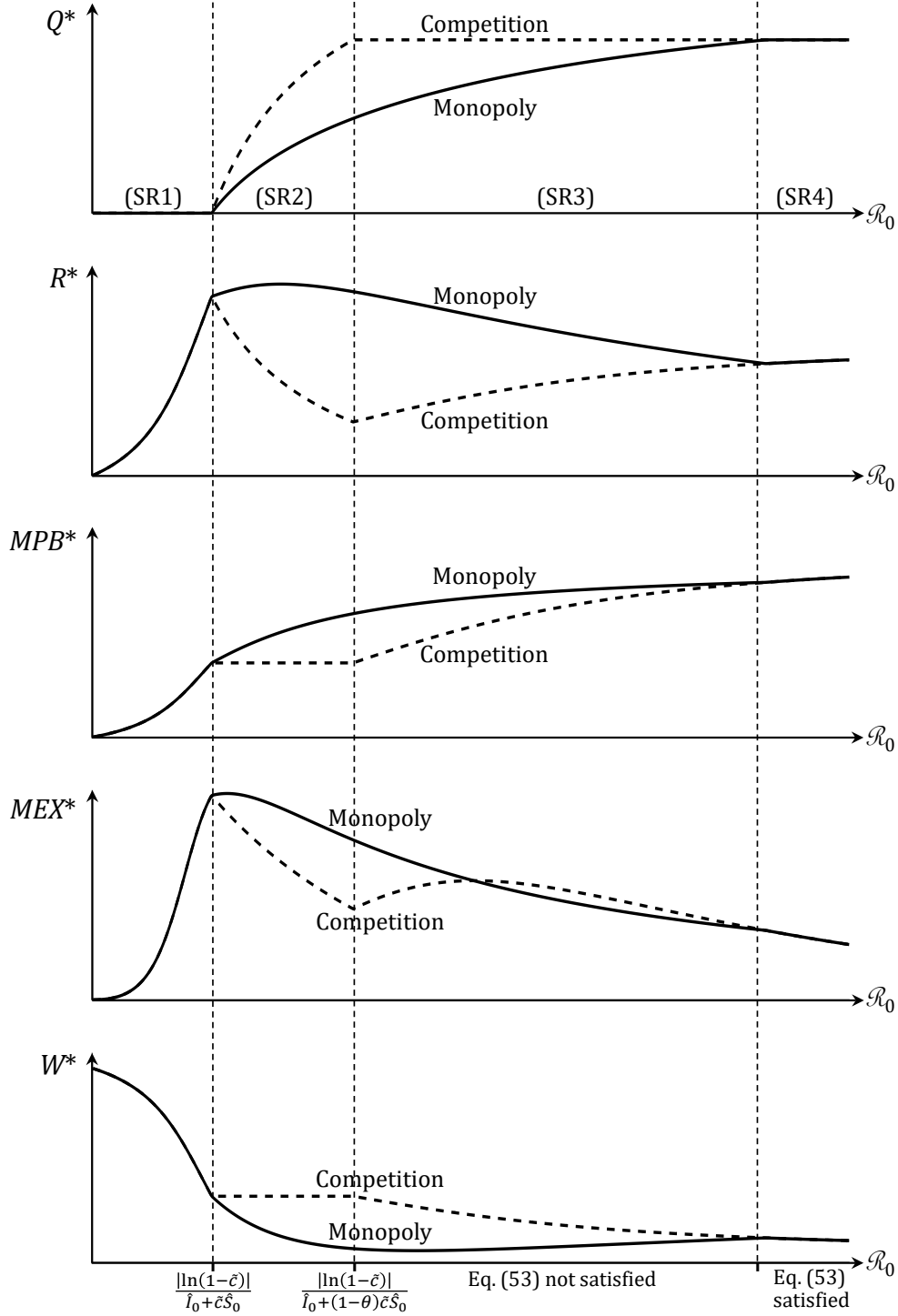
Figure 2: Graphs of Long-Run Optimal Subsidy GS^{**} as Function of $\mathcal{R}_0$
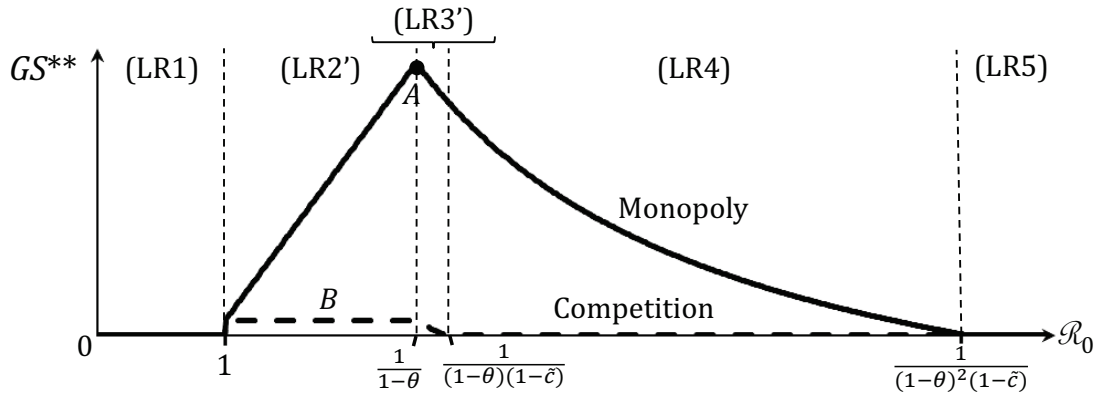
Notes: Graph of formulas provided in Table 3, illustrated for the specific parameter values indicated in previous figure.

Online Appendix D: Supplementary Figures

Figure D1: Graphs of Short-Run Equilibrium Variables as Functions of $\mathcal{R}_0$

Notes: Graph of formulas provided in Tables 5 and 6, illustrated for specific parameter values ($\theta = 0.7$, $\alpha = \mu = 0$, $c = 0.3$, $H = 1$, $\hat{I}_0 = 0.1$, $\hat{S}_0 = 0.8$). Where the dotted and solid curves overlap, the solid curve represents both industry structures. See Figure 1 for additional notes.