The Behavioral SIR Model, with Applications to the Swine Flu and COVID-19 Pandemics

Samuel Engle* Jussi Keppo† Marianna Kudlyak‡
Elena Quercioli§ Lones Smith¶ Andrea Wilson‖

February 12, 2021
(version 0.8; new version soon)

Abstract

The 1927 *SIR* contagion model assumes an infection passes in random pairwise meetings, and derives a linear dynamical system. We propose and test a log-linear modification that reflects the Nash equilibrium of a costly avoidance game. In our *Behavioral SIR Model* (BSIR), the passing rate falls as a hyperbolic function of the prevalence, and therefore incidence is log-linear in prevalence.

The SIR Model yields extreme predictions for major contagions, not realized, even without systemic social distancing or lockdowns. At breakout, the "curve bends" in the SIR model only with heterogeneous agents. In our BSIR model, increasing avoidance behavior bends the curve in the homogenous agent model. Also, herd immunity happens at lower prevalance in the BSIR Model.

Our model is tractable, and better explains incidence data during the 2009 Swine Flu and the COVID-19 pandemic. In both cases, we statistically reject the SIR model. For Swine Flu, across states, the prevalence elasticity ranges from 0.8 to 0.9. We find a similar slope at breakout in the COVID-19 pandemic, and verify that its curve bending matches our BSIR formula.

The same model — with a similar slope but lower intercept — but with increased losses explains data from national lockdowns for COVID-19.

*Wisconsin, www.sites.google.com/wisc.edu/samuelpengle
†National University of Singapore, www.jussikeppo.com
‡Federal Reserve Bank of San Francisco, www.mariannakudlyak.com. The views expressed here are the authors' and do not reflect those of the Federal Reserve Bank of San Francisco, the Federal Reserve Board or any other institution with which author is affiliated with.
§University of Texas Rio Grande Valley, www.sites.google.com/site/elenaquercioli
¶University of Madison-Wisconsin, www.lonessmith.com
‖Princeton University, www.sites.google.com/site/andreamaviwilson

Contents

1 Introduction

Modeling the transmission of contagious diseases is important, as it informs individual and national policy before and during epidemics. The workhorse SIR contagion model (Kermack and McKendrick, 1927) assumes that infection spreads in random pairwise encounters and transmission between susceptible and infected individuals, and later recovery. This model is tractable, given its linear Markovian structure, and has spawned many useful variants, such as the Sl (eg. polio or HIV), SIS (eg. seasonal flu), SEIR (where E is "exposed", for infections with a long incubation period, like chicken pox).

The transmission dynamics for a contagion impinge on many fields. Geography and culture matter, since proximity and custom impact meeting rates. Social networking impacts who meets whom. Finally, political economy considerations matter, since people variably react to social distancing directives. This paper explores the economics factors — since they can change most rapidly during the course of the contagion.

We derive a tractable modification of SI* contagion models that reflects the Nash equilibrium in a game by susceptible individuals who can exert costly vigilance — by meeting less, or more carefully. Greater vigilance "filters out" more infections, but with diminishing returns. We then venture that everyone minimizes the sum of vigilance costs and expected disease losses. In this risk-compensation setting, positive vigilance balances expected marginal benefits and costs. For low prevalence, zero vigilance is optimal, and SIR dynamics emerge. Above a threshold, a more prevalent or lethal disease elicits greater vigilance. This game has a unique Nash equilibrium vigilance.

Incidence requires an infection passing in a meeting of susceptible $\mathcal{S}$ and infected $\mathcal{I}$:

$$\text{incidence rate} = \mathcal{SI} \text{ meeting rate} \times \text{ passing chance}$$

The passing rate rises in the infectiousness and population density. In the *Behavioral SIR model* (BSIR), the passing chance reflects the filter evaluated at the equilibrium vigilance. So at low prevalence, the SIR passing chance obtains (zero vigilance). For higher prevalence, the passing rate is falling (as equilibrium vigilance is increasing). We assume that the infection filter is a hyperbolic function of vigilance, so that vigilance is effective but with diminishing returns. More strongly, the passing chance is hyperbolic in prevalence, and thus has a constant elasticity in the prevalence. Hence:

$$\textit{prevalence elasticity} \text{ (of incidence rate)} = 1 + \text{prevalence elasticity of the passing rate}$$

Standard contagion models posit a constant passing chance, and thus a zero passing elasticity. But in our BSIR, the prevalence elasticity of incidence is constant and less than one. All told, 1% greater prevalence raises incidence by less than 1%, since the vigilance ramps up, and this percent increment is the same for an initial prevalence.

The prevalence elasticity reflects the ease of proportionately improving the infection filter. This turns on social, cultural and behavioral factors that impact the ease of either eliminating meetings, or meeting more carefully. We nonetheless find in the data that this number ranges around 0.9 across nations of the world, and across states of the USA. We find a similar elasticity for the 2009–10 Swine Flu epidemic for the USA.

To appreciate the importance of closing the loop here with equilibrium, consider the optimization faced by vigilant individuals not wishing to get infected. Assume the disease prevalence rises. An immediate reaction is to become more vigilant. For instance, people wash their hands more, or shy away from crowded stores. But soon it dawns on them that everyone else is more vigilant, and thus everyone relaxes their vigilance: Indeed, prevention efforts are strategic substitutes. Equilibrium accounts for this (infinite) feedback cycle. By contrast, an arbitrary rule for reacting to a riskier environment[1] is almost never a best response to itself. This is an epidemiological version of the 1970s Lucas critique of macroeconomics.

One might then wonder "why only a static Nash equilibrium?" To this end, note that in any dynamic continuum player game, no individual impacts play and so *must myopically best respond to play*: The only question is what disease loss to use. Assuming a constant loss yields our simple BSIR that we empirically justify. It also formally emerges in an infinite horizon equilibrium, with a constant chance of the plague ending.[2]

Our Nash equilibrium has two regimes (Theorems 1 and 2): First, at low prevalence, the SIR model obtains, as people exert no vigilance. But once prevalence surpasses a threshold, people best respond to increases with greater vigilance and thus a lower behavioral passing rate. In this range, the incidence accordingly rises with a constant elasticity less than one. Figure 3.2 depicts this dual finding. The regime shift happens at a lower prevalence for a more deadly or more infectious disease; we expect e.g. that the USA has been in the vigilant regime since news of domestic COVID cases hit.

This equilibrium characterization has some quick useful implications. First, there is

[1] He et al. (2013a) posits a dynamical equation for the public equation of risk and increases in the current deaths. Bootsma and Ferguson (2007) ventures a specific function dictating how much individuals reduce their contacts as the number of deaths in the previous time period increases.

[2] We otherwise ignore future nonstationary equilibria, because the best experts have no agreed models. Plagues are one-offs not suited to standard hyper rational forward-looking agents.

a linear relationship between log incidence and log prevalence. This suggests a natural econometric test of our model, since a slope less than one rejects the SIR model in favor of our richer behavioral model — i.e. a prevalence elasticity less than one. Second, since the behavioral passing rate falls in prevalence, a per capita measure, it must rise in population. This explains the faster rate of spread of COVID in higher population countries, for any given number of cases — and in fact, predicts the precise relationship.

The dynamics of the BSIR model qualitatively resemble those of the SIR, namely, that prevalence is first increasing and then decreasing (Theorem 3). But the peak prevalence is lower and delayed. Moreover, this peak — called *herd immunity* — is smaller the lower is the prevalence elasticity (Theorem 4). This explains why predictions based on the SIR model are so much more dire. This theorem also critically parses between herd immunity, or when the flow turns around, and the actual end of the contagion.

Literature Review. Research into the spread of infectious diseases begins with Kermack and McKendrick (1927, 1932), who model how agents transition over time from susceptible to infected, and then to recovered (or possibly other other states). Bailey et al. (1975), Anderson and May (1992) and Hethcote (2000) and Brauer et al. (2012) are treatises from a mathematical, epidemiological and biological standpoints.[3]

Since then, scientists of epidemiology have enriched the SIR model with avoidance behavior, including social distancing, quarantines, hygiene, masks, travel restrictions, and non-pharmaceutical interventions.[4] Avoidance behavior shed light on pandemics, such as the 1918 Spanish Flu (Markel et al., 2007; He et al., 2013b), and later the 2003 Severe Acute Respiratory Syndrome (SARS) (Lau et al., 2003); and HIV/AIDS (Hyman and Stanley, 1988), and the 2009 Swine flu (Poletti et al., 2009, 2011).

Economists have pointed out the flow of agents from one group to another is not exogenous, but reflects risky choice, and thus disease prevalence, acuteness, or the economic or health care cost. Geoffard and Philipson (1996) and Philipson and Posner (1995) introduced optimal choice, and so pioneered the field of economic epidemiology. We build on their a key idea, the prevalence elasticity, by devising a game built on a constant vigilance elasticity, and then finding that this property implies a constant prevalence elasticity in the Nash equilibrium. In a simple two type model of the spread of AIDS Kremer (1996) introduced equilibrium considerations. Quercioli and Smith

[3]Newman (2002) considered integrated networks.

[4]See Funk et al. (2009, 2010), Ferguson et al. (2006), Liu et al. (1986), and Perra et al. (2011).

(2006) later used the notion of equilibrium for an SIR model.

A group of papers, Chowell and Nishiura (2014), Chowell et al. (2004) and Viboud et al. (2016) discuss the use of geometric (which they call "sub-exponential") or polynomial growth models of spread of infections. They justify the decision to do so by arguing that matching of infected and susceptible individuals happens through clusters rather than random mixing, for certain types of diseases where the infected only meet susceptible people in family gatherings, burial rituals and religious functions. Our paper provides the optimizing equilibrium theoretical framework in which this emerges. This framework makes sense of a flaw in an everywhere geometric growth model makes no sense — for instance, the square root prevalence is unboundedly larger than prevalence as prevalence tends to zero.

Two papers have built on ours: Rann Smorodinsky and David McAdams

information lags can lead to zigzagging

[Omitted: Epidemiologists have reduced the estimated R0, during the COVID pandemic, in an effort to capture the changing behavior exogenously.]

2 Prelude: The SIR and BSIR Contagion Models

Consider a contagion dynamic in continuous time $[0, \infty)$. We assume a continuum $[0, 1]$ of players, and thereby assume no aggregate randomness. The *prevalence* $\pi(t) \in (0, 1)$ is the mass of infected and contagious individuals in this population,[5] while the *susceptible share* $\sigma(t) \in (0, 1)$ is the never-infected fraction of the population.

Fix an infection seed $\pi_0 > 0$ — such as from farmers arriving with Swine Flu in early 2009, or people deplaning off international flights in early 2020 in North America or Europe, infected with COVID-19. A mass $\sigma(0) = 1 - \pi_0$ is initially susceptible.

We consider first the SIR (Susceptible-Infected-Recovered) Model. Each contagious person infects a random number of susceptible people per unit time with mean $\beta > 0$, called the *exogenous passing rate*.[6] That is, β increases in the intrinsic contagiousness of the disease, the population density (so higher in cities), and also reflects the culture and social network. So given independent meetings, the new infection inflow, or *incidence*, equals $\beta\pi(t)\sigma(t)$. Anyone infected "recovers" (or dies, and so is removed from the matching pool) at a fixed *recovery rate* $r > 0$.[7] Altogether, π_t and σ_t evolve over time

[5] In §3.2, we will distinguish between infected and contagious.

[6] In a currently omitted Appendix, this emerges in a discrete random matching story.

[7] More broadly, recovered could mean "removed", and include people who die.

according to:

$$\begin{aligned} \dot{\sigma}(t) &= -\beta\pi(t)\sigma(t) \\ \dot{\pi}(t) &= \beta\pi(t)\sigma(t) - r\pi(t) \end{aligned} \tag{1}$$

The dynamics do not depend on the *recovered mass* $\rho(t)$ of immune individuals, which follow $\dot{\rho}(t) = r\pi(t) \geq 0$. Since $\sigma'(t) \leq 0$, *the prevalence* $\pi(t)$ *either starts falling, or first rises and then falls*, since $\dot{\pi}(t) = [\beta\sigma(t) - r]\pi(t)$ transitions from positive to negative.

The *basic reproduction number* $R0 \equiv \beta/r$ is the number of people a typical contagious person infects before he recovers, if (hypothetically) everyone he met were susceptible. *Herd immunity* occurs when enough people are immune that its spread stops naturally because too few people can transmit it. Thus, the "herd" is immune, even though many individuals within it still are not. Formally, this has been taken to be the point where the *effective reproduction number* $\sigma(t) \cdot R0 \leq 1$. By (1), this yields $\dot{\pi}(t) \leq 0$, and herd immunity starts at the tipping point when $\dot{\pi}(t) = 0$. After this time, recoveries exceed incidence for all later times, and the epidemic dies out.

This paper develops a modified dynamic, called the Behavioral SIR Model (BSIR), described by two additional new variables — a threshold $\underline{\pi} \geq 0$ and an exponent $\varphi \in (0, 1)$. With a small seed $\pi_0 < \underline{\pi}$, dynamics start as (1) — our "chill" regime — and shift into a "vigilant" regime at the first time τ with $\pi(\tau) = \underline{\pi}$. If $\pi_0 \geq \underline{\pi}$, then the vigilant regime starts at time $t = 0$. So long as $\pi(\tau) > \underline{\pi}$, the vigilant regime dynamics apply, namely:

$$\begin{aligned} \dot{\sigma}(t) &= -\beta q(\pi)\sigma(t)\underline{\pi}^{1-\varphi}\pi(t)^{\varphi} \\ \dot{\pi}(t) &= \beta q(\pi)\sigma(t)\underline{\pi}^{1-\varphi}\pi(t)^{\varphi} - r\pi(t) \end{aligned} \tag{2}$$

The SIR model is famously simple, ignoring complications of network interactions, and yet useful predictive model. The BSIR is nearly as simple, nesting the SIR model as a special case. But we argue that it diverges in a way that better matches the data. In particular, while initial exponential growth is the hallmark of the SIR model, the BSIR model implies geometric growth, which we will show fits the data better.

We derive this general model in a fully optimizing equilibrium model of behavior in §3, and then flesh out its dynamic properties in §4. We then show that the BSIR better explains the data for two pandemics, H1N1 in 2009 and COVID-19 in 2020.

3 The Strategic Model of Avoidance

3.1 The Behavioral Contagion Model

The SIR contagion model applies to humans and animals alike. But *homo economicus* will adjust behavior to avoid sickness or death.[8] We flesh out a game played in real time by people in a contagion, to capture how people dislike infection, but find avoiding it costly. The sickness or death stakes in the game amount to a loss $L > 0$. The *vigilance* action reduces passing rates, and is denumerated in its cost $v \geq 0$ per unit time.

Vigilance is very broadly defined as any costly activity that stifles passing. It subsumes extensive margin choices, like the fraction of meetings by Zoom, or longer routes to avoid passing people. Vigilance also includes intensive margin choices, often labelled social distancing.[9] People may sneeze into elbows, washing hands more often, fist-bump rather than handshake, or weak masks. Vigilance should be a personally costly activity — not something that one normally does without thinking. It is therefore idiosyncratic and cultural: not kissing on the cheeks is harder in some countries than others, and for some people than others. Also, with habit persistence or learning by doing, the vigilance required for any passing rate falls over time. It can be history dependent. A handshake might be the normal greeting pre-2020, but post 2020, the lower vigilance greeting might be a fist bump.

Vigilance v scales down passing rates by a multiplicative *filter function* $f(v) \in [0, 1]$ — so that the passing rate is $\beta f(v)f(w)$ if a contagious vigilance v person meets a susceptible vigilance w person. For the very actions we undertake to block others' from infecting us — e.g. wear a mask or skip a meeting — also inhibit us from infecting others. This functional form assumes a symmetric vigilance impact. Next, the two parties' vigilance acts independently, and thus the passing impact is multiplicative. This makes sense even for extensive form vigilance, such as the fraction $f(v)$ of meetings made. Imperfect vaccinations fit, if we view $f(v)$ as the chance it is effective.

We naturally assume that vigilance is effective but with diminishing returns, so that $f' < 0 < f''$, with extreme values $f(0) = 1 > 0 = f(\infty)$. For instance, the in first reducing one's meetings, one will choose the lowest cost actions, at the margin. After these options have been exploited, we turn to higher cost actions. In particular, we posit the hyperbolic $f(v) = (1 + v)^{-\gamma}$, for $\gamma > 0$, with these properties. This ensures a

[8]Indeed, historically, behavior has changed, like quarantines off Venice in the 1300s during Black Death. Or consider how the HIV/AIDS epidemic sparked a culture-changing "safe sex" drive.

[9]See Toxvaerd (2019) and Toxvaerd (2020).

constant *filter elasticity* γ in terms of $V = 1 + v$, which includes an endowed baseline unit vigilance. So one percent more total vigilance costs always leads to a γ-percent infection risk reduction. The SIR model corresponds to a zero filter elasticity. We eventually estimate that $\gamma \approx 1/8$ in the data.[10]

3.2 Vigilance Optimization and Equilibrium Predictions

To make sense of optimal behavior, we parse the infected into two groups: People are first obliviously infected and contagious, and next knowingly infected. We denote by π the mass of unaware infected individuals, and exclude those knowingly infected.[11]

Under the maintained SIR model rules, meetings are random and independent of any traits (like vigilance).[12] Any random population vigilance W therefore induces an expected passing rate $\beta f(v)E[f(W)]$.[13] A *potentially susceptible* person — namely, one who is either susceptible or asymptomatically infected — thinks himself at risk with *susceptible belief* $q(\pi) = \sigma/(\sigma + \pi)$. So his posterior *flow infection chance* is

$$\beta f(v)E[f(W)]q(\pi)\pi \tag{3}$$

Since $f' < 0 < f''$, this chance falls if v rises, $E[f(W)]$ falls, or the prevalence π falls.

We compute the Nash equilibrium with the same loss L each period, with vigilance cost v each period.[14] Players minimize $\beta f(v)E[f(W)]q(\pi)\pi L + v$, their selfish expected total losses:[15] Since the filter function $f(v)$ does not obey an INADA condition near $v = 0$, there is only a corner solution for small $\pi > 0$. But barring a corner solution, marginal analysis identifies a unique interior optimum, because $f' < 0 < f''$. As everyone is identical, the opt for a symmetric pure strategy Nash equilibrium $W = v^* \geq 0$.

Potentially susceptible people play a Nash equilibrium with loss L, minimizing the

[10] This corresponds to a constant health elasticity in the value of life model of Hall and Jones (2007).

[11] These infected may be (resp.) symptomatic (or symptomatic but ignorant) and asymptomatic.

[12] Relaxing this is important, and at the core of any paper that explores a network matching model.

[13] Rowthorny and Toxvaerd (2012), a decentralized SIS model, assumes that a "protection level" π scales down the infection level at constant marginal costs. This would correspond to $f(v) = 1 - v$, and hence their optimal solution is bang-bang; ours is interior in the vigilance regime.

[14] A continuum of agents is crucially unlike finite player games in one crucial respect: For since no one can influence the future, *any forward-looking equilibrium requires that all players play a static Nash equilibrium every period*. The loss L is constant each period if, for instance, people assume that the contagion eventually stops with a constant chance every period. We assumed a static representative agent model. A future Appendix will show that the same equilibrium arises with heterogeneous types.

[15] Adapting the random matching model with verification in Quercioli and Smith (2015), Quercioli and Smith (2006) introduced cost minimization with two way filters for contagious matching games.

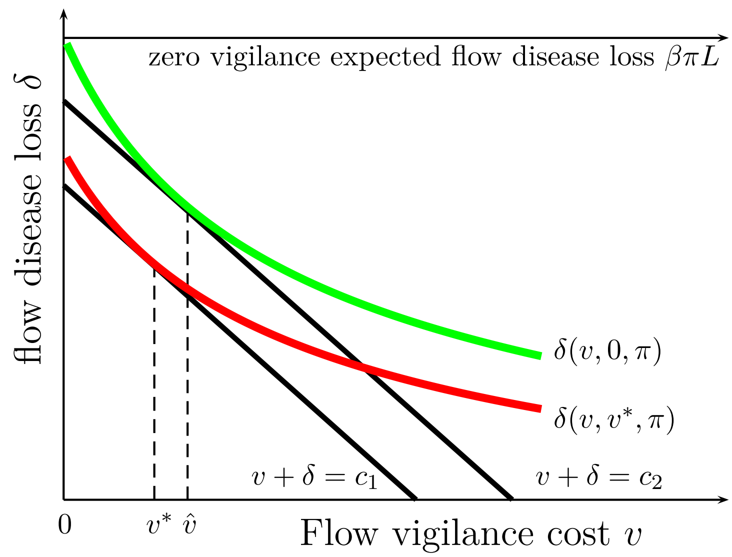


Figure 1: **Optimal Vigilance.** Vigilance v minimizes total expected contagion losses $\delta(v, v^*, \pi) + v$ in (4), equating expected marginal disease losses to one. Equilibrium vigilance is lower than if no one else exerted any vigilance ($v^* < \hat{v}$) due to strategic substitutes. But equilibrium flow disease losses are lower: $\delta(v, v^*, \pi) < \delta(\hat{v}, 0, \pi)$. The SIR model assumes zero vigilance and thus higher expected slow disease losses.

flow disease loss $\delta(v, v^*, \pi) = \beta f(v) f(v^*) q(\pi)\pi L$. At any interior optimum on $[t, t+dt)$:

$$\min_{v\geq 0} \delta(v, v^*, \pi)dt + vdt \quad \Rightarrow \quad 1 = -\delta_v(v, v^*, \pi) \tag{4}$$

This contagion game has the strategic substitutes property — specifically, a higher vigilance v^* by others leads to a lower best reply own vigilance v (so $v^* < \hat{v}$ in Figure 1).

The first order condition in a symmetric pure Nash equilibrium for $v^* > 0$ solves:

$$1 = -\delta_v(v^*, v^*, \pi) = \beta\gamma(1 + v^*)^{-2\gamma-1} q(\pi)\pi L \tag{5}$$

as $f(v) = (1 + v)^{-\gamma}$. Zero vigilance is strictly optimal for small π, since the right side of (5) vanishes if $\pi = 0$. By continuity, $v = 0$ if $q(\pi)\pi \leq (\beta\gamma L)^{-1}$, and thus for all $\pi \leq \underline{\pi}$, where $\underline{\pi}$ slightly exceeds $1/(\beta L\gamma)$.[16] Notably, whenever $\pi > \sigma/(\sigma\gamma\beta L - 1)$,

$$v^* = (\gamma\beta q(\pi)\pi L)^{\frac{1}{2\gamma+1}} - 1$$

and so increases in π, by (5). Intuitively, people grow increasingly vigilant as prevalence rises:

[16] Rewriting $\pi q(\pi) = \frac{1}{\sigma^{-1}+\pi^{-1}}$, the threshold $\underline{\pi}$ shares the monotonicity of the threshold for $\pi q(\pi)$.

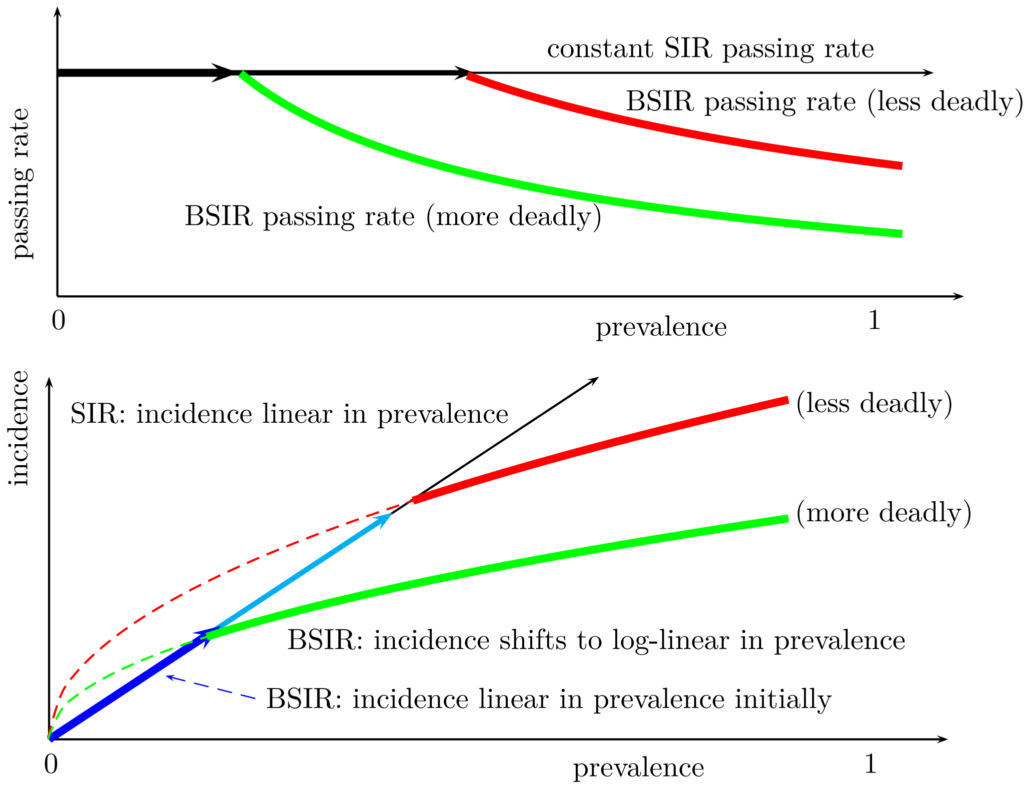


Figure 2: **SIR versus BSIR Passing Rate and Incidence.** The BSIR passing rate is hyperbolic, and so has a constant elasticity — just as hyperbolic demand curves do. The prevalence elasticity is thus constant and less than one.

Theorem 1 (Equilibrium) *In the unique Nash equilibrium, for any prevalence $\pi \geq 0$, players choose a common vigilance $v^*(\pi)$. There is a unique* prevalence threshold $\underline{\pi} > 0$ *— with $v^*(\pi) = 0$ for $\pi \in [0, \underline{\pi}]$, and $v^*(\pi)$ increasing for $\pi \geq \underline{\pi}$. This vigilance $v^*(\pi)$ increases in L, β, and γ, and the prevalence threshold $\underline{\pi}$ falls in L, β, and γ.*

3.3 The Behavioral Passing Rate

We now explore how the Nash equilibrium modifies the SIR model. The key observation is that *the constant filter elasticity in vigilance induces a constant behavioral passing elasticity in prevalence.* To see this, monotonely transform the filter elasticity γ into $\varphi \equiv 1/(2\gamma+1)$. Naturally, $0 < \varphi < 1$ and $\gamma = (1-\varphi)/(2\varphi)$. The *behavioral passing rate* $B(\pi) \equiv \beta f(v^*)^2$ reflects the exogenous passing rate β and the equilibrium filter $f(v^*)$. The effective reproduction number in $B(\pi)/r$ is an equilibrium object.

By Theorem 1, the zero-vigilance SIR model emerges for low vigilance, but vigilance optimally rises in prevalence if $\pi \geq \underline{\pi}$, depressing the behavioral passing rate. We now characterize this relation by raising the FOC (5) to the power $1-\varphi = 2\gamma/(2\gamma+1)$:

$$B(\pi) = \beta(1+v^*)^{-2\gamma} \approx \beta[\beta L(1-\varphi)/(2\varphi)]^{\varphi-1}\pi^{\varphi-1}$$

if $\pi \geq \underline{\pi}$, for the approximate susceptible belief $q(\pi) \approx 1$. So behavior is constant in prevalence with $\varphi = 1$ (namely, the SIR model), and more elastic for lower $\varphi \in (0,1)$.

Theorem 2 *The behavioral passing rate $B(\pi)$ is continuous, with two regimes:*

$$B(\pi) = \begin{cases} \beta & \pi \leq \underline{\pi} \quad \textit{(chill)} \\ q(\pi)\beta(\underline{\pi}/\pi)^{1-\varphi} \approx \beta(\underline{\pi}/\pi)^{1-\varphi} & \pi > \underline{\pi} \quad \textit{(vigilant)} \end{cases} \tag{6}$$

Since the behavioral passing rate (6) is continuous at $\underline{\pi}$, we can also compactly write it as[17]

$$B(\pi) = \min(\beta, \underline{\pi}^{1-\varphi}\beta\pi^{\varphi-1}) \tag{7}$$

As with the derived demand for any "bad", like garbage, this demand for passing implied by Theorem 2 falls in its unit loss L, and in the original level π of the bad.[18]

The falling hyperbolic behavioral passing rate in Theorem 2 for $\pi > \underline{\pi}$ induces a constant coefficient log-linear derived demand for *equilibrium incidence* $B(\pi)\pi\sigma$, adjusting the first term in (1):

$$\log[B(\pi)\pi\sigma] = b + \varphi \log \pi + \log \sigma \tag{8}$$

We call the slope φ the *incidence-prevalence elasticity*. As $q(\pi) \approx 1$, the above intercept is

$$b \approx \log \beta + (1-\varphi)\log \underline{\pi} \equiv \log\left(\beta[\beta L(1-\varphi)/(2\varphi)]^{\varphi-1}\right) \tag{9}$$

and can be positive or negative, depending on L and β. Summarizing:

Corollary 1 (Incidence and Prevalence) *Equilibrium incidence $B(\pi)\pi\sigma$ is log-linear in prevalence π, as in (8), when $\pi > \underline{\pi}$. Also, $\varphi \equiv 1/(2\gamma+1) < 1$, and the intercept b increases in φ and β, and falls in L.*[19]

[17] A geometric passing rate $B(\pi|\varphi) \propto \pi^{\varphi}$ for all prevalence π is impossible, since it is infinitely times the SIR rate as $\pi \downarrow 0$. "Subexponential" formulas (such as Viboud et al. (2016)) cannot globally hold. The fixed SIR passing rate must still obtain for low enough prevalence — our chill regime.

[18] Quercioli and Smith (2015) exploited an interpretation of the counterfeiting rate (analogous to the prevalence here) as the price in an "implicit market". We do not pursue that here.

[19] If the passing rate only reflected one's own vigilance, the incidence-prevalence elasticity would be $\hat{\varphi} = 1/(\gamma+1) = 2\varphi/(1+\varphi) > \varphi$. Consistent with the strategic substitutes property, the equilibrium dual filter is worse than two people optimizing but ignorant of the larger game: $\hat{\varphi}^2 < \varphi$.

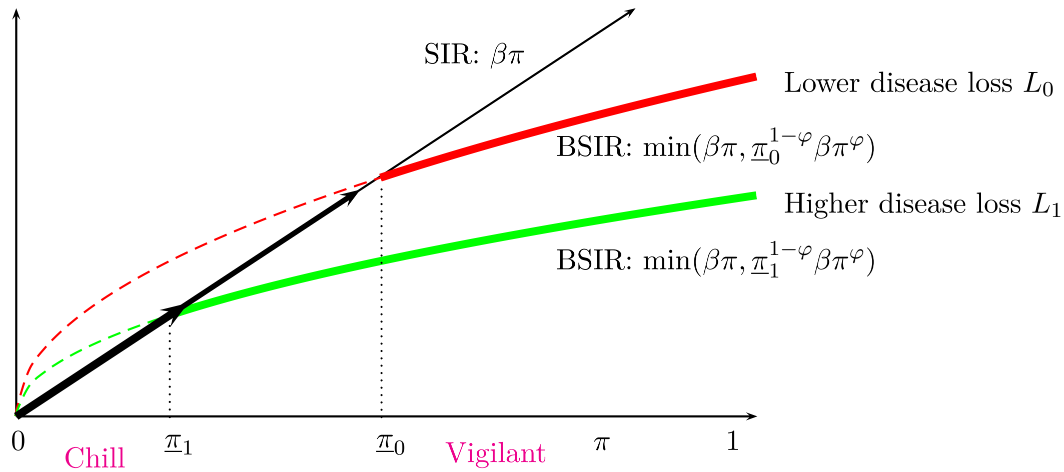


Figure 3: **Equilibrium Infection Chance Under the BSIR**. The SIR model obtains for a smaller prevalence interval $[0, \underline{\pi}]$ if the disease is more dire (loss L_1 versus L_0). For the common flu, SIR dynamics may obtain over a large prevalence interval (Theorem 2).

Proof: To see the last claim, the φ-derivative of b in (8) is $\log[\pi\beta L(1-\varphi)/(2\varphi)] > 0$, using (9), and the L-derivative is $(\varphi - 1)/L < 0$. □

Standard SIR insights about the exogenous passing rate β impact this log-linear relationship (8): For instance, the BSIR intercept b increases in the population density. But the SIR model places no importance of the disease loss on the contagion.

Since predicted vigilance rises in prevalence, which falls in population, we have:

Corollary 2 *Fixing the number of cases, the behavioral passing rate rises in population.*

This result explains why, long before we hit herd immunity, countries with larger populations have more cases. For at high enough case levels, prevalence and so vigilance falls in population in the BSIR model; so the behavioral passing rate rises. This pattern was generally observed across nations in the COVID19 data.

4 The Behavioral SIR Dynamics

Our theory prescribes how prevalence impacts the behavioral passing rate. We now explore the impact of this *contemporaneous* relationship on the contagion dynamics. Theorem 2 naturally yields the dynamics (2) claimed for the BSIR, with a continuous transition to the log-linear dynamics in the vigilant regime at transition point $\pi = \underline{\pi}$.[20]

[20]The SIR model and BSIR continuation has a unique solution, by the Picard–Lindelof theorem.

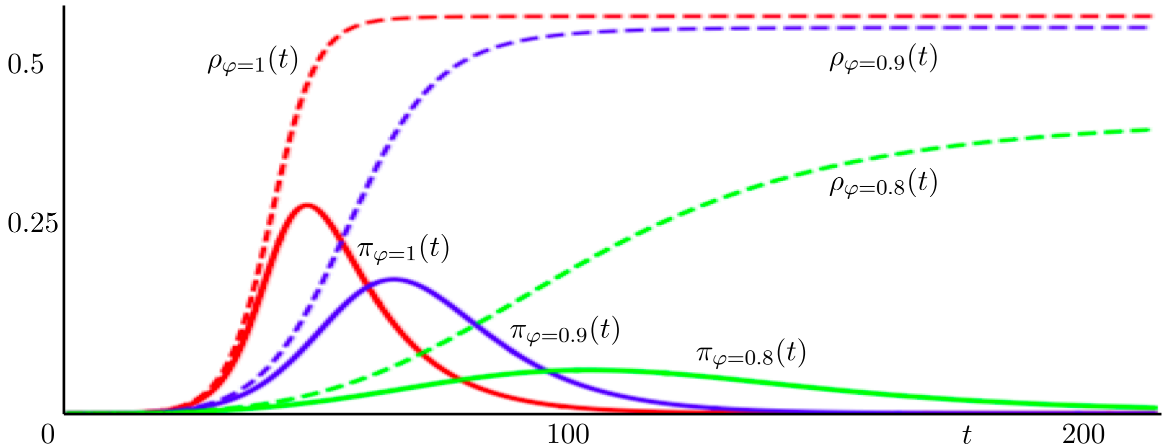


Figure 4: **Contagion Dynamics, for Varying φ.** We illustrate Theorem 4, by plotting the (dashed) shares of past symptomatically infected by that date, for prevalence elasticities $\varphi = 1$, $\varphi = 0.9$, and $\varphi = 0.8$. We assume a passing rate $\beta = 0.9$, and asymptomatic share $\alpha = 0.5$. The solid curves are currently contagious individuals. In fact, this plots the solution for the model with vaccination, so that $\rho_\varphi(t) = (1-\alpha)[1-\sigma_\varphi(t)]$.

There is nothing about equilibrium theory that holds effective R0 below 1 and in fact the infection follows the growth and decline pattern of SIR ... just much more muted

Optimization can in no way holds the growth rate below one, because the static optimization makes no account of recovery rate.

First, prevalence π is still hump-shaped and σ a falling S-shaped function of time.

Theorem 3 (BSIR Dynamics) *The susceptible share $\sigma(t)$ falls monotonically in time, while prevalence $\pi(t)$ either monotonically falls, or first rises and then falls.*

Figure 4 depicts the hump-shaped prevalence π and S-shaped recovered shares ρ. A greater prevalence elasticity $\varphi < 1$ reduces both the susceptible share σ and reduces and delays the peak π. The impact of falling φ is ordinally analogous to that of falling β in the SIR model, but different cardinally (Figure 18).

Of particular interest are *breakout* dynamics — namely, for low times near time $t = 0$ with few infections and recoveries, when approximately $\sigma \approx 1$ and $\pi \approx 0$. Here the analysis is far simpler, since dynamics are formally one-dimensional. Also, it applies where nations find themselves in early 2020.[21] Or just after an effective lockdown ends,

[21]For instance, from an antibody survey of 70,000 people in Spain, one of the hardest hit European

with prevalence down to $\pi \approx 0$, the breakout analysis also holds. But in that case, a positive mass $1 - \bar{\sigma} \in (0,1)$ of people is infected or recovered. Reflecting meetings that encounter susceptible individuals, the passing rate is $\bar{\sigma}\beta$. For times $t < \tau$, the SIR dynamics apply:

$$\dot{\pi}(t) \approx \beta\pi(t) - r\pi(t) \quad \Rightarrow \quad \pi(t) \approx \pi_0 e^{(\beta - r)t}$$

So the homogeneous agent SIR model predicts initial exponential growth at rate $\beta - r$.[22] Provided $\beta > r$, we eventually hit $\pi(t) = \underline{\pi}$. If $\beta \leq r$, then the infection dies out before the vigilant regime starts.

The linear SIR dynamical system is not solvable in closed form, and likewise neither is our log linear behavioral SIR system. But from breakout, we can solve it explicitly. For $t > \tau$, the passing rate is $B(t) \approx \beta[\underline{\pi}/\pi(t)]^{1-\varphi}$, given the susceptible belief $q \approx 1$, by Theorem 2. We have an autonomous first order Bernoulli differential equation:

$$\pi'(t) = \beta\underline{\pi}^{1-\varphi}\pi(t)^{\varphi} - r\pi(t) \quad \Rightarrow \quad \pi(t) = \underline{\pi}\left(\frac{\beta}{r}\left(1 - ke^{-r(1-\varphi)t}\right)\right)^{\frac{1}{1-\varphi}} \tag{10}$$

for the constant $k = (\beta/r - 1)\left(\underline{\pi}/\pi_0\right)^{r(1-\varphi)/(\beta - r)}$. One can check that the prevalence $\pi(t)$ is an increasing, convex and log-concave function of time t.[23]

The *herd immunity locus* are all pairs (σ, π) with stationary prevalence: $\pi' = 0$. That is, the net prevalence flow is at a stationary point, and thus, by Theorem 3, prevalence is thereafter falling. Herd immunity in the SIR model happens because rising immunity chokes off new infections; therefore, the herd immunity locus is independent of π only in the SIR model with prevalence elasticity $\varphi = 1$ in Figure 5. But in the behavioral SIR model, vigilance also chokes off contagions: Since individuals are more vigilant at a higher prevalence, less immunity is needed with more prevalence. In other words, the herd immunity locus is decreasing in $(1-\sigma, \pi)$ space, as depicted in Figure 5.

Herd immunity starts at $(\check{\sigma}_\varphi, \check{\pi}_\varphi)$ when inflow balances outflow: $\pi' = 0$ in (2), if and only if

$$B(\pi)\check{\sigma}_\varphi\check{\pi}_\varphi^{\varphi} = r\check{\pi}_\varphi \quad \Rightarrow \quad \check{\sigma}_\varphi = (r/B(\pi))\check{\pi}_\varphi^{1-\varphi} > r/\beta$$

countries, found that only 5% of Spaniards have been infected with the coronavirus.

[22] We thank Chris Auld (Victoria) for a nice insight: There is a well-known impact of heterogeneity in β via a selection effect: as time goes on the remaining susceptibles are likely low-β types. This by itself flattens the curve.

[23] We can show that this creates a log-concave total case count.

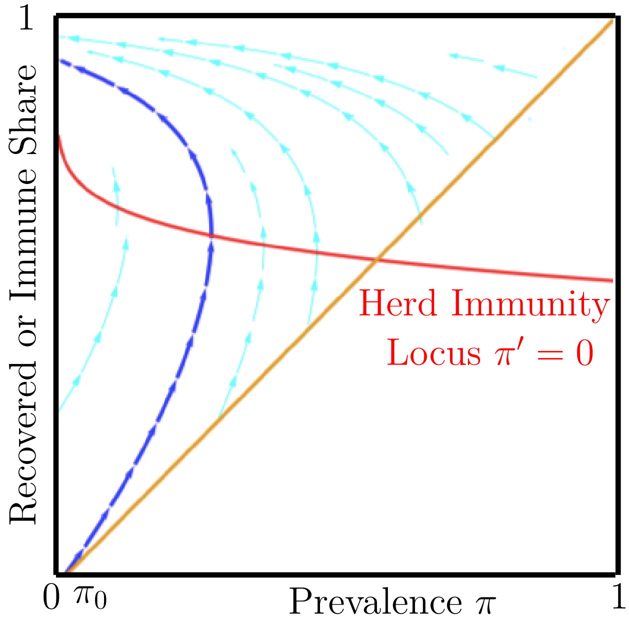


Figure 5: **The Herd Immunity Locus.** We plot the stream path and herd immunity locus for prevalence π and {immune or recovered} $1-\sigma$, assuming $\varphi = 0.85$, $\beta = 0.5$, $r = 1/17$, and $L = 200000$, with no one asymptomatic. Prevalence is stationary ($\pi' = 0$) along the red herd immunity locus, rising below it (fewer immune), falling above it (more immune). This herd immunity locus is falling in $(\pi, 1-\sigma)$ space due to the equilibrium avoidance behavior. Around 90% of people are eventually infected.

Notice that *whereas herd immunity in the SIR model is independent of prevalence (see §2), a tradeoff emerges in the BSIR: more people can be susceptible if more are infected* (Figure 5). For in that case, people are more careful, and thus the passing rate is lower; hence, more can be susceptible. Moreover, herd immunity in the BSIR model allows a larger mass of susceptible people than in the SIR model.

The SIR model predicts more infections. But maximizing avoidance behavior in the BSIR increasingly "flattens" the curve as the prevalence elasticity φ falls (Figure 18). Despite lacking a closed form formula, we indirectly one can show that the herd immunity is increasing in the infectiousness β, and in the behavioral model, falling in the losses L from infection.

Theorem 4 (Herd Immunity) *As the prevalence elasticity φ falls, (i) the herd immunity time τ_φ rises, (ii) peak prevalence π_φ falls, (iii) the herd immunity infection share $1-\sigma_\varphi$ falls, and (iv) its ratio to eventual infections $(1-\sigma_\varphi)/(1-\sigma_\varphi(\infty))$ rises.*

This result and Figure 18 make clear that *infections stop long after we hit herd im-*

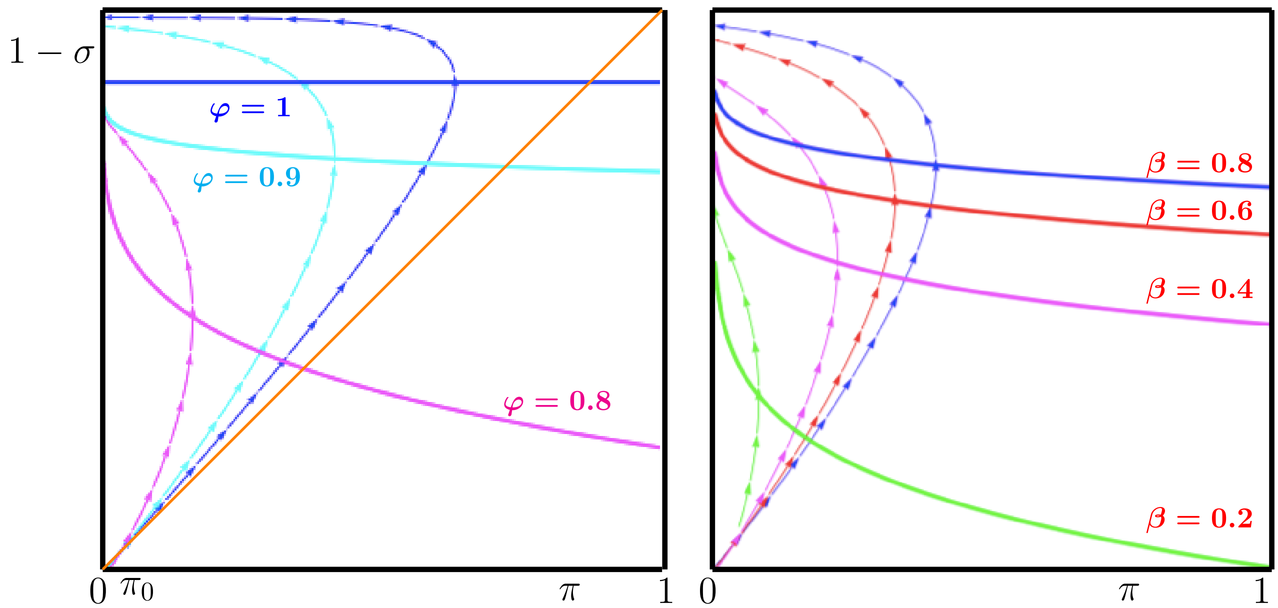


Figure 6: **Contagion Dynamics.** We plot the stream path and herd immunity locus for prevalence elasticities 1 (blue), 0.9 (cyan), and 0.8 (magenta) at left [with passing rate $\beta = 0.5$], and passing rates 0.8 (blue), 0.6 (red), and 0.4 (magenta), and 0.2 (green) [with prevalence elasticity $\varphi = 0.85$], and no asymptomatics. *As φ or β falls, the peak prevalence falls, and the herd immunity locus shifts down and left in $(\pi, 1-\sigma)$-space. Herd immunity is independent of prevalence π only in the SIR model.*

munity![24] Obviously, the true cost of any policy changes reflects the eventual infection numbers, and not number at the moment herd immunity is achieved.

To understand herd immunity, note that it is the tipping point, NOT the endpoint! If the COVID fires start to abate by prevalence =40% (my guess), we eventually will hit say 70%+ by infection or vaccination. Faster vaccination reduces the death toll.

5 COVID19

1. Mitigation
2. Seasonal effects
3. Mutations
4. Vaccinations

Mitigation efforts are game changer, literally: They change the game. Since our

[24] "A note on the derivation of epidemic final sizes"

only parameter is the exogenous passing rate β, we assume that any mitigation effort[25] such as a lockdown scales down β. Since this is a submodular game, the behavioral passing rate scales down by a smaller factor, as vigilance rises, by Corollary 1.

First, we must convert this to a stochastic discrete time model.

To capture the COVID19 pandemic, we enrich our infection model, first toward an SEPSR Model (Susceptible - Exposed - Presymptomatic - Symptomatic - Recovered), and then add two twists. First, we assume that a fraction α of infected individuals are entirely asymptomatic through their infection ("silent spreaders"), and only spread with reduced (possibly zero) chance ζ. Second, we assume two contagious states — a distinction that is inessential for epidemiology, but critical here, because choices must reflect information, and one's symptoms can flag a person that he is sick. Thus, infected individuals pass through a sequence of states (see Figure 7):

We consider an SEPIR model (Susceptible-Exposed-Presymptomatic-Infected-recovered)

- State 1: Exposed, or infected and incubating but not yet contagious
- State 2: Infected, Contagious and Pre-Symptomatic
- State 3: Infected, Contagious and Symptomatic
- State 4: Recovered and no longer Infected

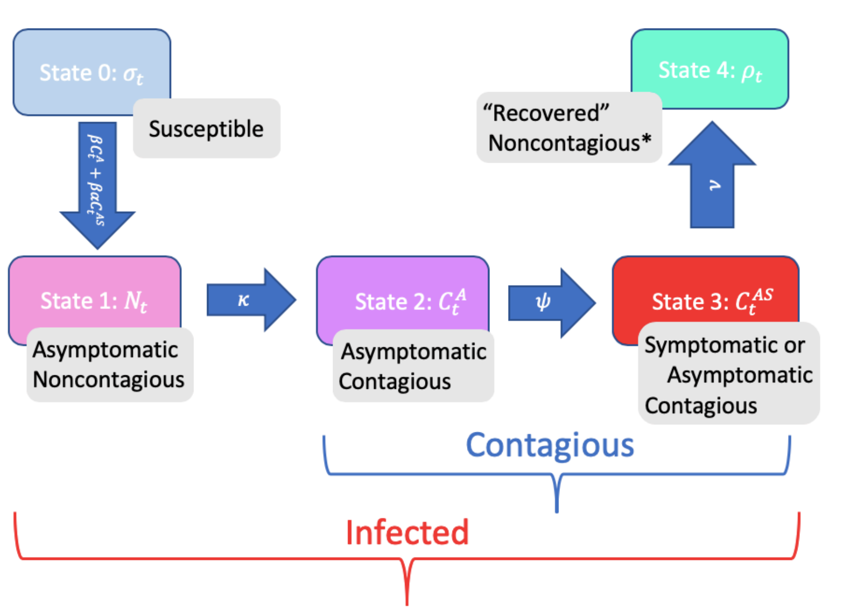
Susceptible individuals, S, enter the exposed class, E, upon infection after contact with infected individuals. Some will remain asymptomatic, A, while the remainder become pre-symptomatic, P. The latter will advance to the symptomatic stage, I, which die from the disease at rate α (referred to as 'virulence'). All others eventually recover, and are assumed immune.

We take primitives from Ferretti et al. (2020), summarized in Figure 7.[26]

We secured data on 4/9/2020 from https://github.com/CSSEGISandData/COVID-19. This data set relies on tests, and thus is perforce incomplete. Day to day variation in the availability of tests data adds volatility to the measured case count.

[25]These are known as NPI, for non pharmaceutical intervention

[26]The incubation period (the time between infection and onset of symptoms) has mean 5.5 days. The relative infectiousness of asymptomatic individuals compared with symptomatic individuals is 0.1. The fraction of infected individuals who are asymptomatic is 0.4, based on media reports from the Diamond Princess.



Parameter	*Description*	*Value*
κ	State 1 $\to$ 2	$1/(5-2.5) = 0.4^a$
ψ	State 2 $\to$ 3	$1/2.5 = 0.4^b$
ν	State 3 $\to$ 4	$1/12 \approx 0.08312^c$

Figure 7: **The SI3R Model.**

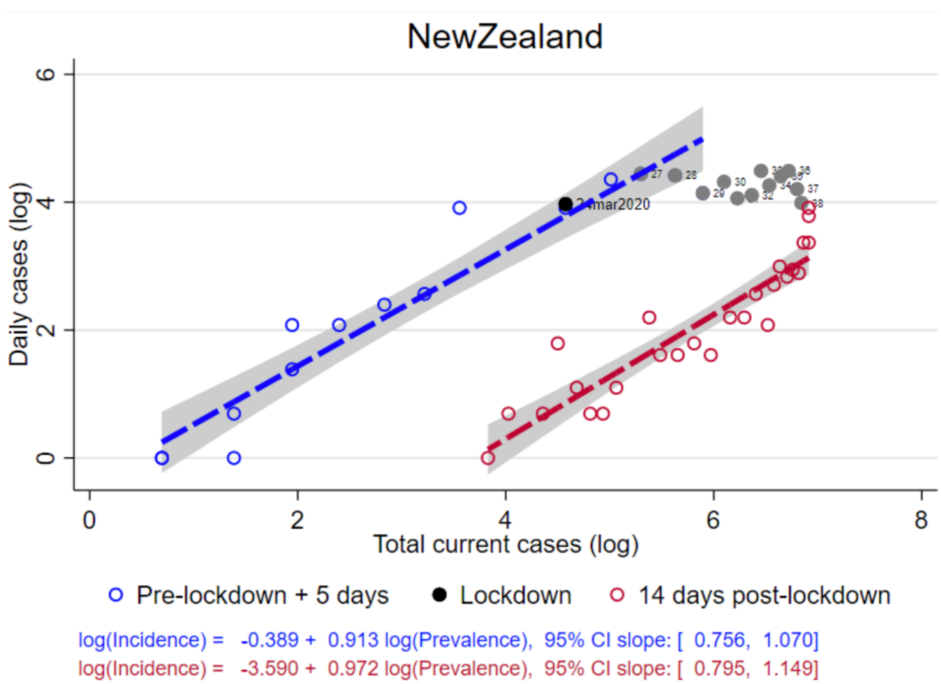


Figure 8: **Pre- and Post Lockdown New Zealand.** We depict log daily cases as a function of the total currently contagious.

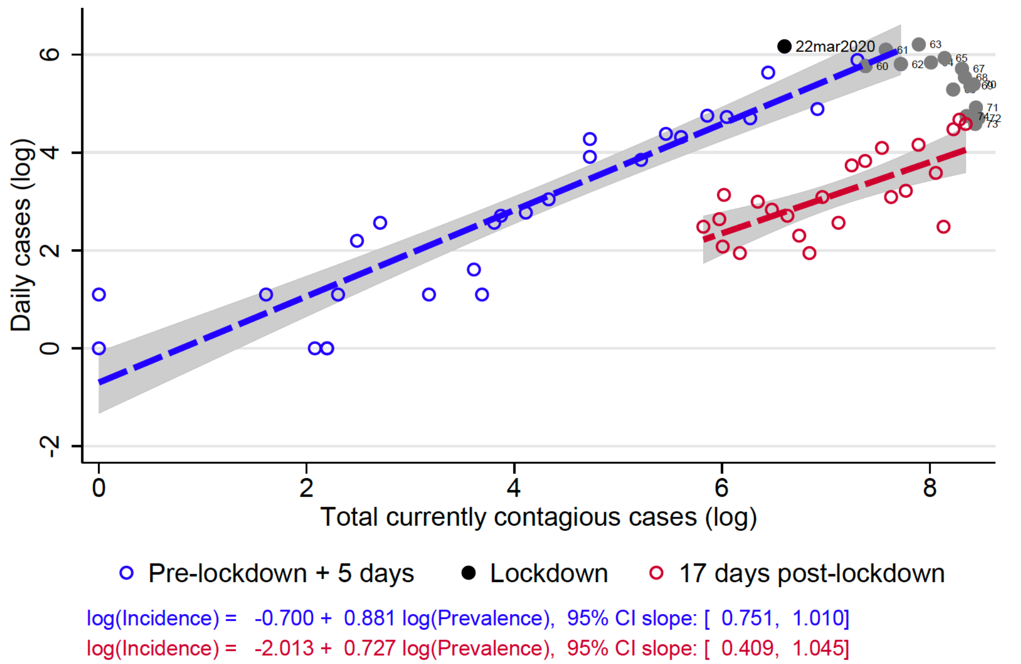


Figure 9: **Pre- and Post Lockdown in Australia.** We depict log daily cases as a function of the total currently contagious.

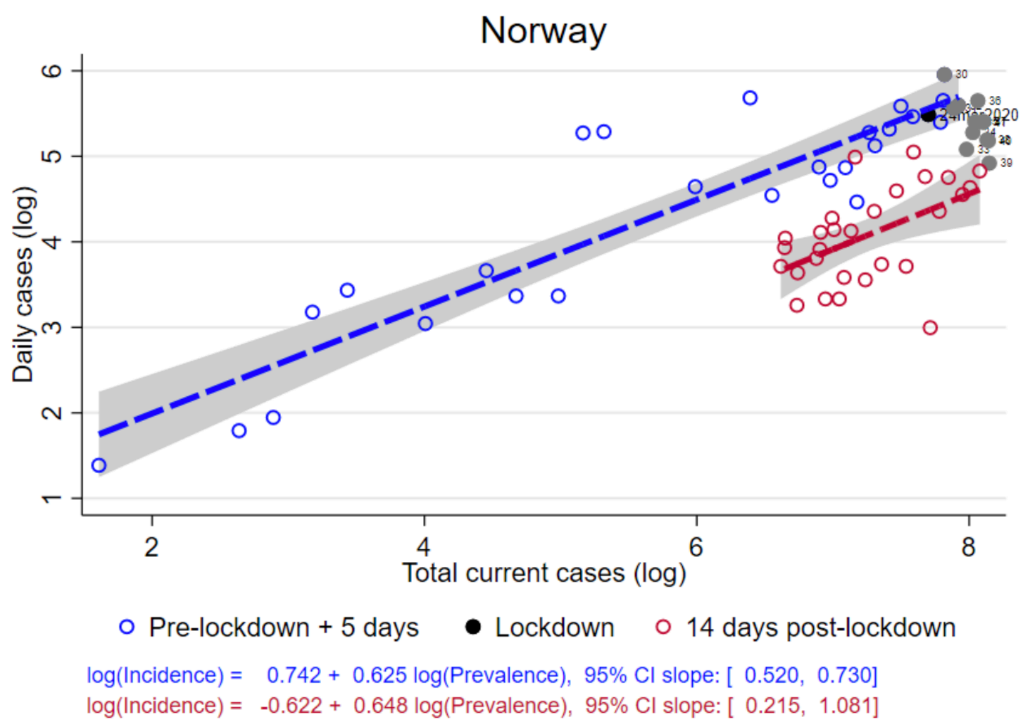


Figure 10: **Pre- and Post Lockdown in Norway.** We depict log daily cases as a function of the total currently contagious.

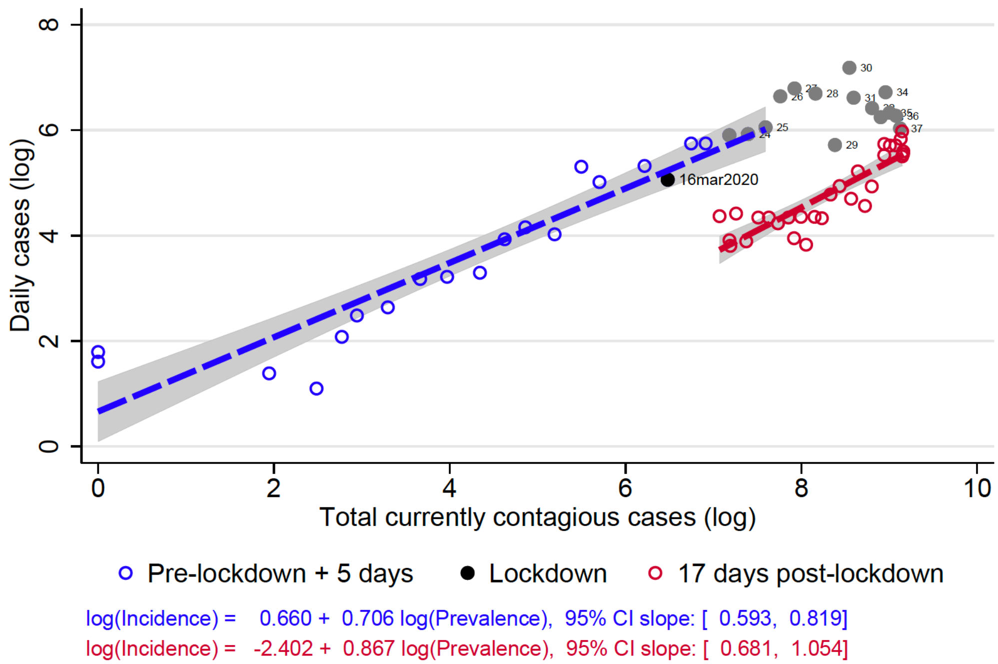


Figure 11: **Pre- and Post Lockdown in Austria.** We depict log daily cases as a function of the total currently contagious.

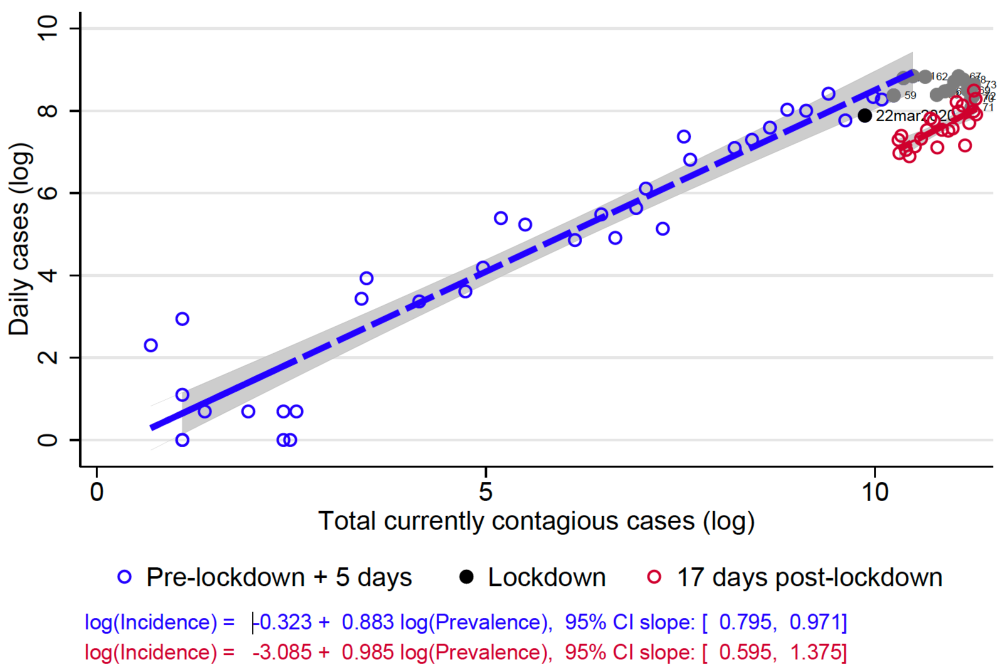


Figure 12: **Pre- and Post Lockdown in Germany.** We depict log daily cases as a function of the total currently contagious.

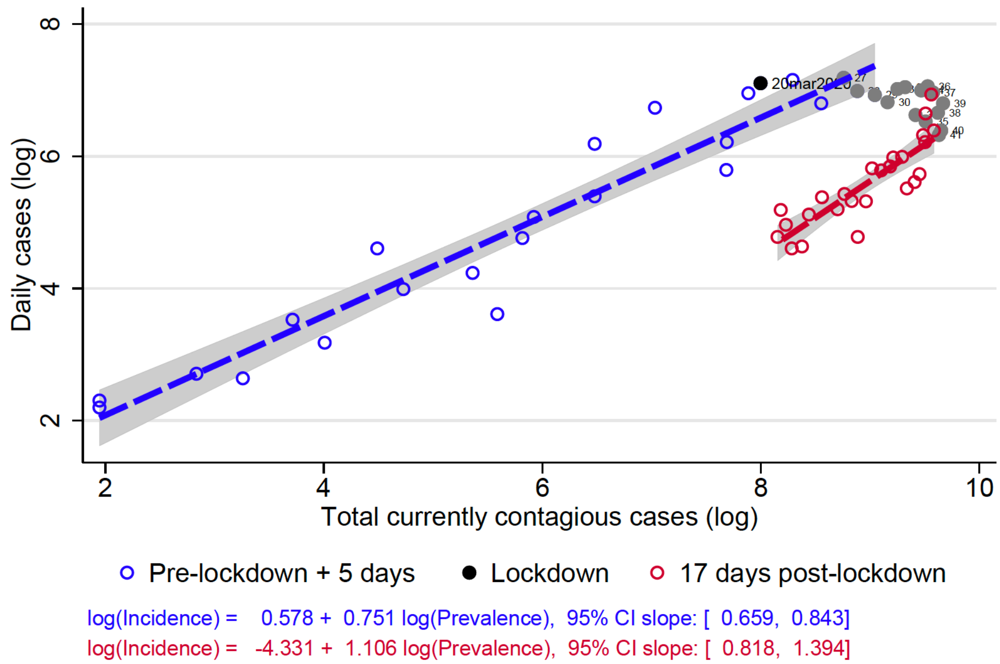


Figure 13: **Pre- and Post Lockdown in Switzerland.** We depict log daily cases as a function of the total currently contagious.

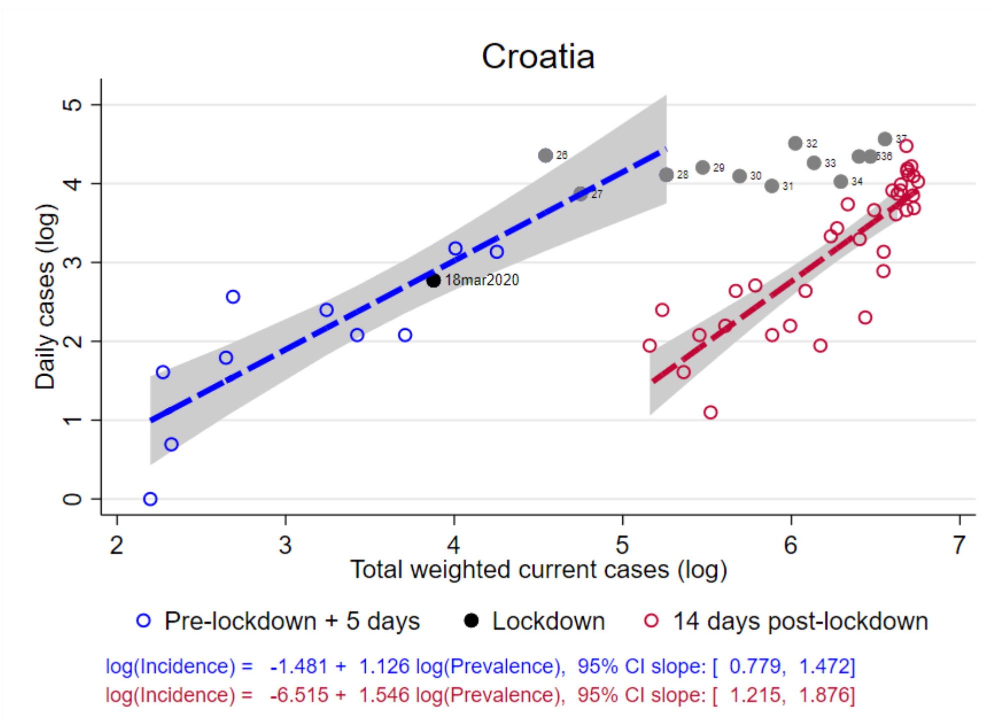


Figure 14: **Pre- and Post Lockdown in Cambodia.** We depict log daily cases as a function of the total currently contagious.

We report regressions of log of incidence of new cases daily nc_t against currently contagious cc_t for various countries. We assume a 14 day contagion window, so that

$$\log(nc_t) = \hat{b} + \varphi \log(cc_t)$$

This corresponds to the regression (8), except that it does not scale for the population.

Lockdowns. Let us interpret mitigation, lockdown, or stay-in-place orders as an exogenous reduction in the passing rate β. Then by Corollary 1

5.1 Age dependent infection rates

NY and NJ were caught off guard with a sudden rise in infectiousness due to the time lags. Now, each state is following the model, with lockdown influencing β in each case.

6 Swine Flu

- Estimate this with the same machinery
- What is the difference between pandemics in β and loss L?
- Despite being more contagious, COVID ha spread less after 11 months than H1N1!

The Swine Flu lasted January 2009 to August 2010, and in the USA starting in April 2009, and lasting about a year. We have acquired a unique data set of this pandemic, with weekly or daily data from the state health departments of each of 41 states in the USA. Our total case count is 150,023 people in 41 states representing 260.1M people. In fact, the CDC estimates that around 61.8M Americans succumbed to H1N1.

H1N1 profited from a standard seasonal effect in summer 2009, akin to a falloff in contagiousness, with a major emergence in the fall of 2009. The pandemic ended at herd immunity, helped by a vaccination that emerged in fall 2009. This affords us an essential glimpse of the future for COVID.

Swine Flu had a clear seasonal component to the transmission rate, with β falling in the summer of 2009.

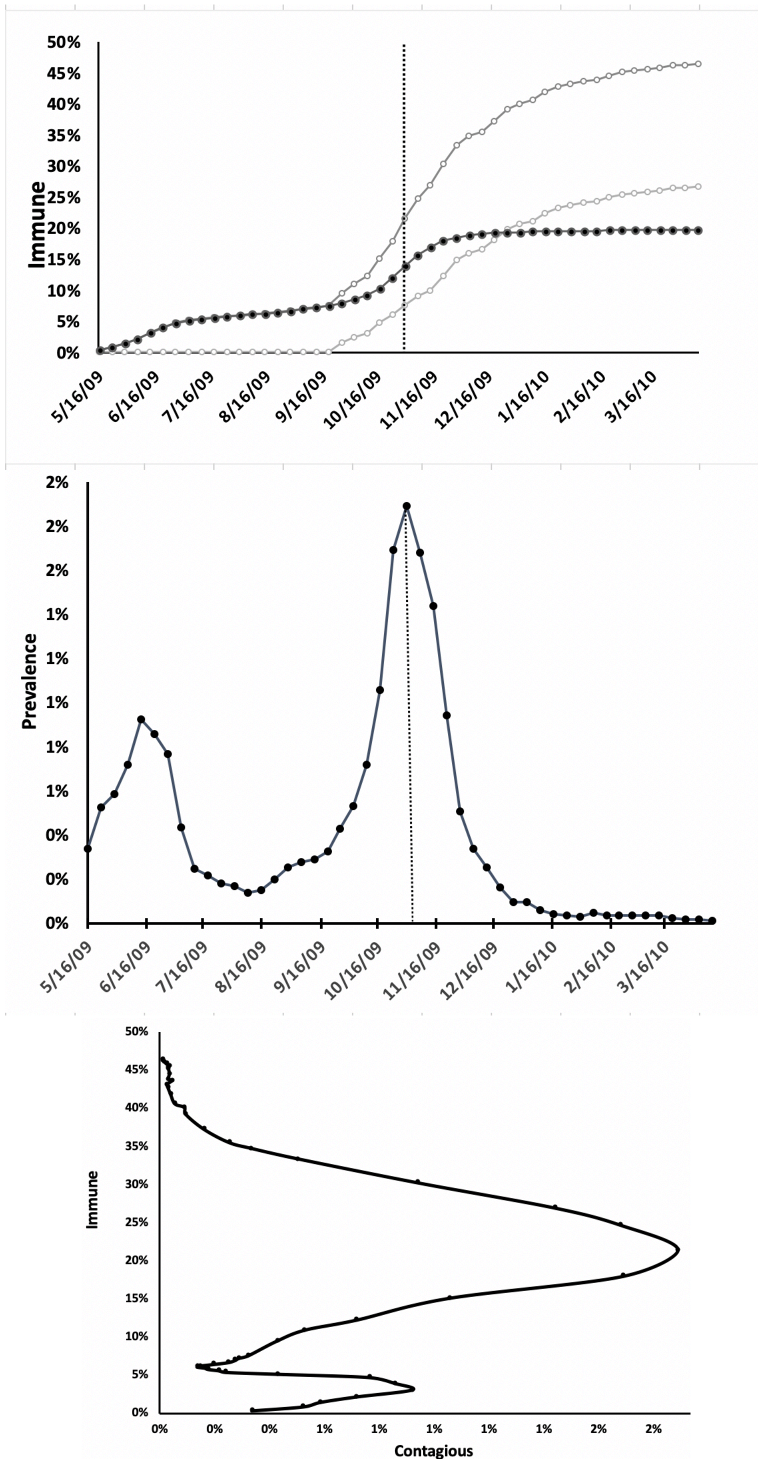
We now modify the model to account for vaccinations. Assume that a fraction $v(t)$ of people are vaccinated by period t. Of course, only the uninfected get vaccinated, and thus we now interpret $\sigma(t)$ as the fraction who have not yet gotten infected, and the true mass of susceptible people is $\sigma(t) - v(t)$. the theoretical dynamics are:

$$\dot{\sigma}(t) = -B(t)\pi(t)[\sigma(t) - v(t)]$$

Here we see the timeline of Swine Flu pandemic in the USA. Herd immunity begin on 10/31/09 with 70.1% infected at that time. In other words, much of the battle still lay ahead.

Herd immunity has been in the air this year, and has been used too casually. Fauci has adapted to the casual usage in the press of the endgame of the pandemic. But the relevant definition is when R0=1, which happens long before the endgame. Best metaphor is that a pandemic is a wildfire and after the fires start to dim, there is still a helluva lot of forest yet to burn. Here is Swine Flu. Rough story is that after herd immunity hit, another 1/3 of infections were yet to occur. I miswrote in my earlier comment: Swine Flu hit herd immunity around 21% on Halloween 2009. Of this, around 7% were vaccinated and 14% were infected. We ended at around 25% vaccinated and 20% infected.

Herd immunity is being abused here and by Fauci. It is a tipping point, and not the endpoint. If the pandemic ends with 2/3 immune (H1N1 ended with about 45%; upper curve in the plot; empty circle plot was vaccinated plot), then the tipping point was likely around 30Secondly, we reject the linear SIR model in favor the log-linear behavioral SIR model. And its tipping point is not an invariant percent of the population. When the pandemic is raging out of control, people are more vigilant, and tipping occurs at a lower percentage immune. The latter insight speaks to an advantage of rapid vaccination: fewer need be infected and die before we hit herd immunity.



7 Conclusion

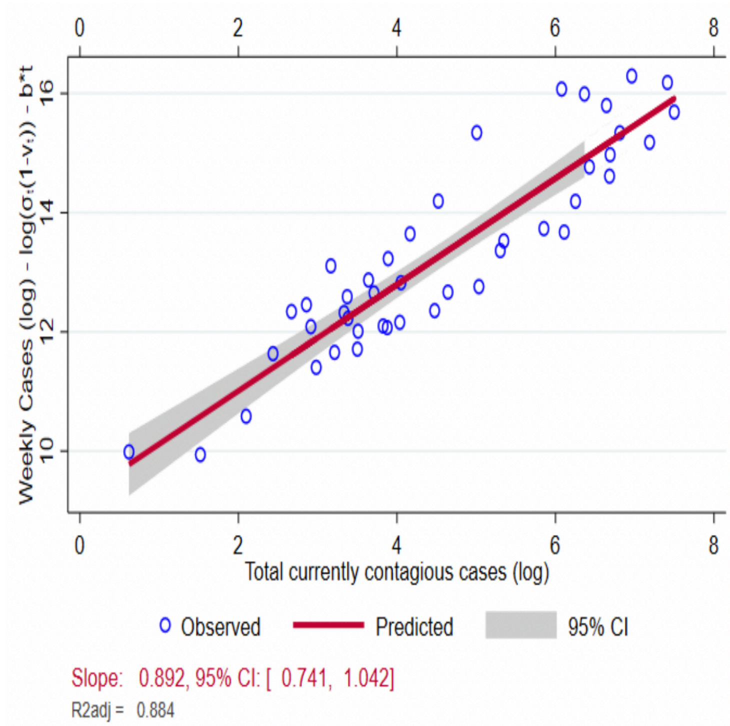


Figure 15: **Log Incidence on Log-Prevalence: CT, ME, MA, NH, RI and VT.** We depict the estimation of the number of new cases N_t on the currently contagious H1N1 cases C_t, accounting for the susceptible fraction σ_t

We ignore the possibility of lags. This makes the model SIR in the short run, if people do not react. This would allow temporary spikes.

If individuals respond to announced cases or deaths, then this creates lags and thus naturally cycles that we observe!

We have developed the behavioral SIR model, and applied it to the two pandemics post 2000. Its behavioral respond also can be applied to SI contagions, like AIDS or herpes, and should predict a log-log rule.

We now depict the monthly ratios of COVID-19 infection rates for adjacent age groups, for the state of Massachusetts in 2020. Notice that

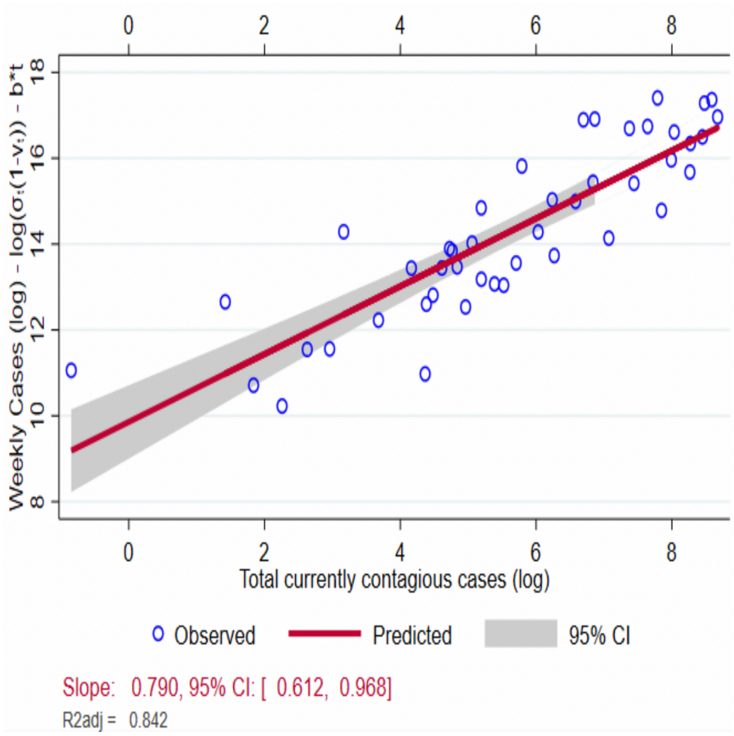
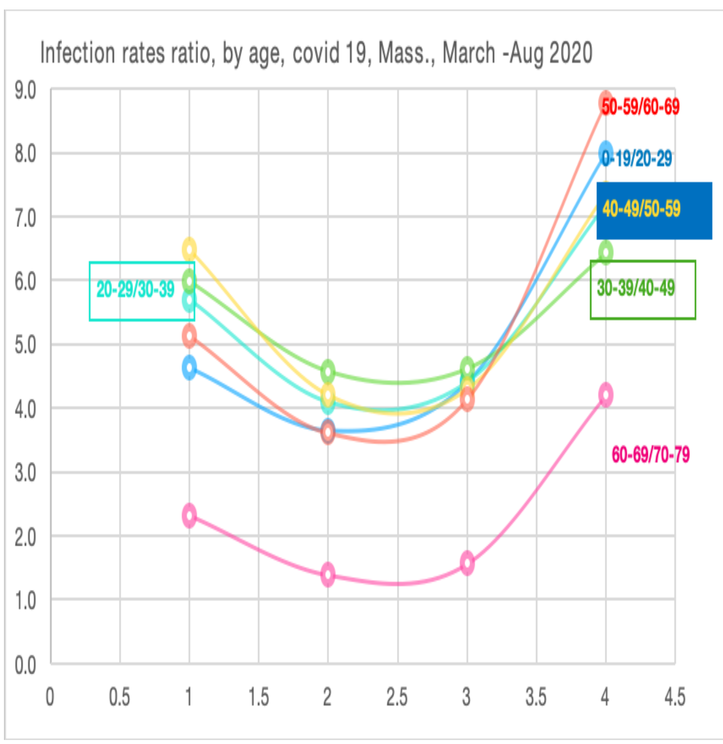


Figure 16: **Log Incidence on Log-Prevalence: NJ and NY.**



Corollary 1 rationalizes the more cavalier behavior of younger less at-risk groups in a contagion (see §5.1).

A Increasing Avoidance Behavior Pre-Lockdown

We document increasing avoidance behavior, before any government recommended social distancing or mandated pre-lockdown! See Figure 17.

B Omitted Plots

C Omitted Proofs

C.1 Derivation of Breakout Equation (10)

If $\pi' = C\pi^{\varphi} - r\pi$, where $C = \beta\underline{\pi}^{1-\varphi}$, then $\pi'/\pi^{\varphi} = C - r\pi^{1-\varphi}$. Define $y = \pi^{1-\varphi}/(1-\varphi)$. Then $y' = C(1-\varphi) - r(1-\varphi)y$, and thus $\log[C(1-\varphi) - r(1-\varphi)y] = -r(1-\varphi)t + k'$, whence $C/r - y = ke^{-r(1-\varphi)t}$, and so $\pi(t) = y(t)^{1/(1-\varphi)} = \left(C/r - ke^{-r(1-\varphi)t}\right)^{1/(1-\varphi)}$.

C.2 Prevalence is Hump-shaped: Proof of Theorem 3

It suffices to prove that $\pi'(t)$ is downcrossing; i.e. that if $\pi'(t) \geq 0$, then this remains true as t falls. In the vigilance regime, (**??**) implies $\pi'(t) \geq 0$ iff $\beta q(\pi)\sigma(t)\underline{\pi}^{1-\varphi} \geq r\pi(t)^{1-\varphi}$. If this holds at time t, then it holds strictly for slightly lower t. For that both increases the LHS (σ monotonically falls) and weakly decreases the RHS (by the premise that π is nondecreasing at t). So $\pi'(t) \geq 0 \Rightarrow \pi'(\tau) > 0$ for $\tau < t$, as desired.□

C.3 Herd Immunity: Proof of Theorem 4

(omitted for now)

References

Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

Norman TJ Bailey et al. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

Martin C. J. Bootsma and Neil M. Ferguson. The effect of public health measures on the 1918 influenza pandemic in u.s. cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593, 2007.

Fred Brauer, Carlos Castillo-Chavez, and Carlos Castillo-Chavez. *Mathematical models in population biology and epidemiology*, volume 2. Springer, 2012.

Gerardo Chowell and Hiroshi Nishiura. Transmission dynamics and control of ebola virus disease (evd): a review. *BMC medicine*, 12(1):196, 2014.

Gerardo Chowell, Nick W Hengartner, Carlos Castillo-Chavez, Paul W Fenimore, and Jim Michael Hyman. The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda. *Journal of theoretical biology*, 229 (1):119–126, 2004.

Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.

Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368:7588–7593, 2020.

Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent AA Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16):6872–6877, 2009.

Sebastian Funk, Marcel Salathé, and Vincent AA Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.

Pierre-Yves Geoffard and Tomas Philipson. Rational epidemics and their public control. *International economic review*, pages 603–624, 1996.

Daihai He, Jonathan Dushoff, Troy Day, Junling Ma, and David J. D. Earn. Inferring the causes of the three waves of the 1918 influenza pandemic in england and wales. *Proceedings of the Royal Society B: Biological Sciences*, 280:1–7, 2013a.

Daihai He, Jonathan Dushoff, Troy Day, Junling Ma, and David JD Earn. Inferring the causes of the three waves of the 1918 influenza pandemic in england and wales. *Proceedings of the Royal Society B: Biological Sciences*, 280(1766):2013–1345, 2013b.

Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4): 599–653, 2000.

James M Hyman and E Stanley. Using mathematical models to understand the aids epidemic. *Mathematical Biosciences*, 90(1-2):415–473, 1988.

William Ogilvy Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society*, 115:700–721, 1927.

William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 138(834):55–83, 1932.

Michael Kremer. Integrating behavioral choice into epidemiological models of aids. *The Quarterly Journal of Economics*, 111(2):549–573, 1996.

JTF Lau, X Yang, H Tsui, and JH Kim. Monitoring community responses to the sars epidemic in hong kong: from day 10 to day 62. *Journal of Epidemiology & Community Health*, 57(11):864–870, 2003.

Wei-min Liu, Simon A Levin, and Yoh Iwasa. Influence of nonlinear incidence rates upon the behavior of sirs epidemiological models. *Journal of mathematical biology*, 23(2):187–204, 1986.

Howard Markel, Harvey B Lipman, J Alexander Navarro, Alexandra Sloan, Joseph R Michalsen, Alexandra Minna Stern, and Martin S Cetron. Nonpharmaceutical interventions implemented by us cities during the 1918-1919 influenza pandemic. *Jama*, 298(6):644–654, 2007.

Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1): 016128, 2002.

Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. Towards a characterization of behavior-disease models. *PloS one*, 6(8), 2011.

Tomas Philipson and Richard A Posner. The microeconomics of the aids epidemic in africa. *Population and Development Review*, pages 835–848, 1995.

Piero Poletti, Bruno Caprile, Marco Ajelli, Andrea Pugliese, and Stefano Merler. Spontaneous behavioural changes in response to epidemics. *Journal of theoretical biology*, 260(1):31–40, 2009.

Piero Poletti, Marco Ajelli, and Stefano Merler. The effect of risk perception on the 2009 h1n1 pandemic influenza dynamics. *PloS one*, 6(2), 2011.

Elena Quercioli and Lones Smith. Contagious matching games. mimeo, Penn Search and Matching Seminar, 2006.

Elena Quercioli and Lones Smith. The economics of counterfeiting. *Econometrica*, 83 (5):1211–1236, 2015.

Robert Rowthorny and Flavio Toxvaerd. The optimal control of infectious diseases via prevention and treatment. CEPR Discussion Paper No. DP8925, 2012.

Flavio Toxvaerd. Rational disinhibition and externalities in prevention. *International Economic Review*, 60:1737–1755, 2019.

Flavio Toxvaerd. Equilibrium social distancing. March mimeo, Cambridge, 2020.

Cécile Viboud, Lone Simonsen, and Gerardo Chowell. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics*, 15:27–37, 2016.

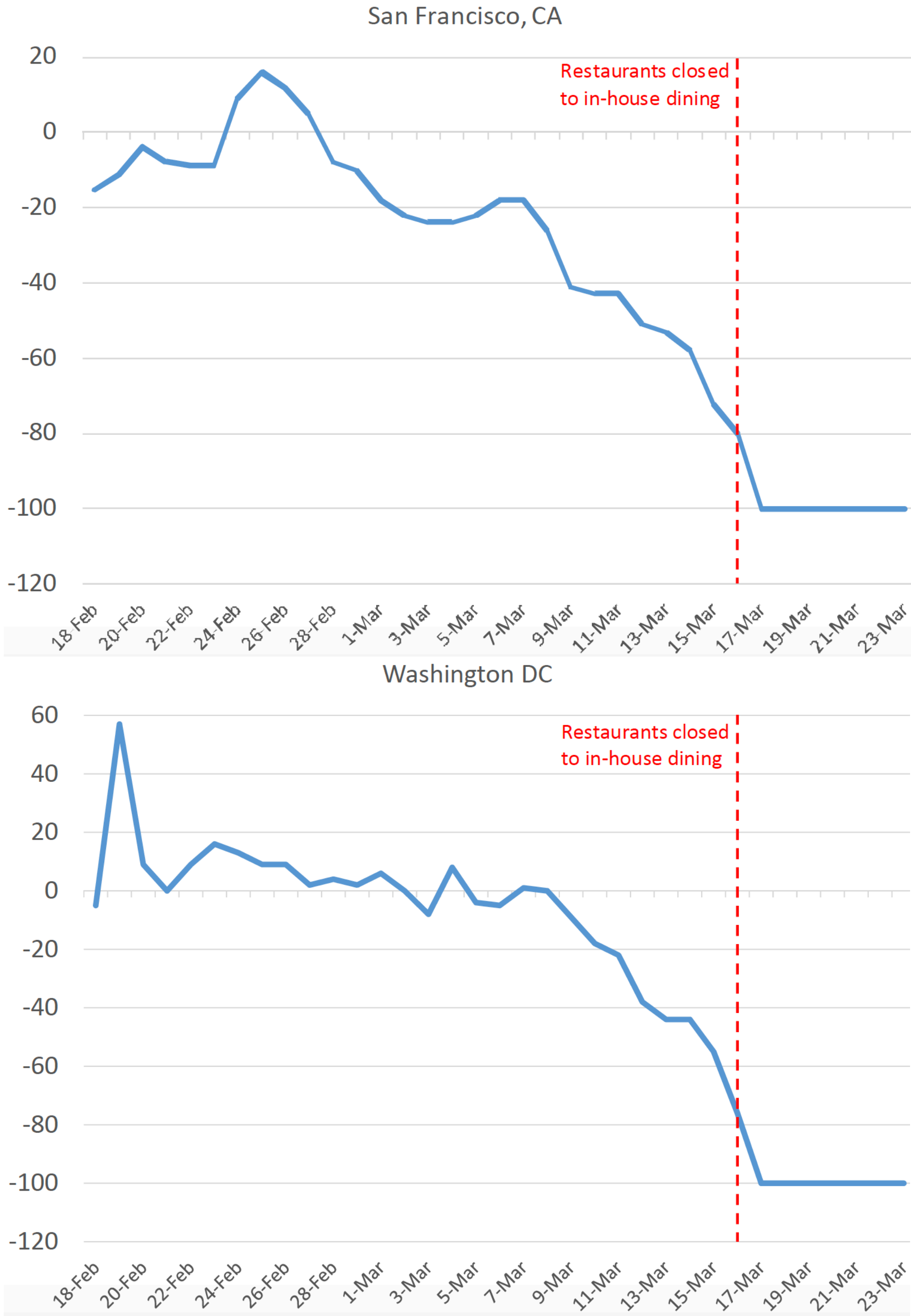


Figure 17: Year-over-year decline of seated diners at restaurants prior to locally-mandated closures (from OpenTable)

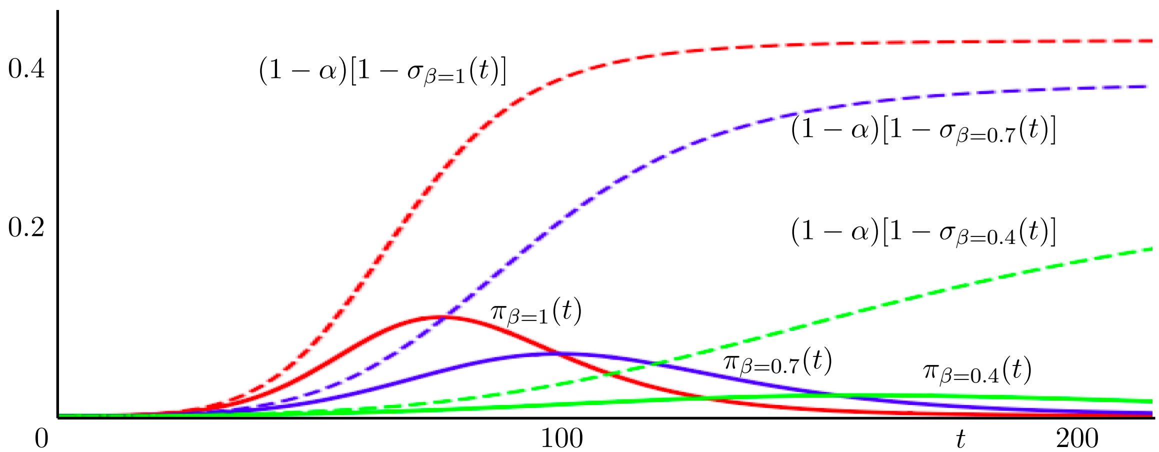


Figure 18: **Contagion Dynamics, for Varying $\boldsymbol{\beta}$.** For prevalence elasticity $\varphi = 0.85$ and $\alpha = 0.5$ asymptomatic (never contagious), we plot the (dashed) shares of past symptomatically infected, for three passing rates β. The solid curves are currently contagious individuals. Prevalence is hump-shaped (Theorem 3). Note to us: must add BSIR ODE with asymptomatic infections.