Incomplete information and diagnosis traps in social-distancing epidemic games

Timothy C. Reluga

October 6, 2015
started May, 2013

Abstract

Within the epidemic game literature, there is an inconsistency over whether increases in infection risk can discourage prevention practices. In one theory, investments in social distancing and prophylaxis increase monotonely with infection pressure, while in another, there is a turning-point beyond which investments drop off. To-date, the reasons for this difference have been unclear. This article shows that the discrepancies are due to different information assumptions and can be resolved by explicitly modelling diagnosis events. After constructing a unified theory, we find that there is a turning-point in the optimal strength of self-protection as a function of infection pressure. Before this turning point, increases in risk encourage prevention. Beyond this turning point, increased risk discourages prevention. Diagnosis rate controls the location of this turning point. The presence of the turning point can be interpreted as a consequence of decision making under incomplete information, where agents must hedge uncertainties in their current state. With reliable diagnosis, prevention is worth-while. But without reliable diagnosis, individuals may be best-off apathetically assuming they are infected and abandoning prevention once infection pressures are sufficiently high. The resulting Nash equilibrium is a trap with high endemic prevalences that can not be escaped by individual unilateral action. To conclude our analysis, we estimate turning-point conditions for the specific case of HIV in some African nations, and conclude that turning-point effects likely contribute to traps for high-prevalence core-groups found in some sub-Saharan countries.

1 Introduction

Several years ago, an anonymous scholar observed that there appears to be a contradiction within the current theory of epidemiological games. In a well-known paper by Kremer (1996) on behavioral choice in the prevention of HIV infection, the author observed in Section I:

"Increases in prevalence in the pool of available partners may thus cause people
with more than a 50 percent chance of becoming infected over their lifetime to increase their rate of partner change.”

and then a little later,

“It may seem difficult to believe that people could actually increase their activity in response to an increase in prevalence. However, this model suggests that it would be a mistake to assume that intuition developed from experience with people who have a small chance of becoming infected with HIV can be used to generalize about people with high prevalence.”

On the other hand, we recently derived a different result (Li et al., 2015):

“If the relative infection probability is convex and decreasing in individual investment, then there is always a unique best-response, and this best-response increases with both the cost of disease and the infection pressure.”

Our conclusion was a consequence of an earlier mathematical observation that if rates and costs of a state-specific strategy are linear in that strategy, then the payoff in population games takes the form of a monotone rectangular hyperbola a.k.a. linear fractional transform (Reluga, 2009).

Kremer (1996) and Li et al. (2015) agree that increases in prevalence encourage prevention when risk is low, but when risk is high they appear to be in direct disagreement, one claiming prevention always increases in response to increased prevalence, while the other is claiming prevention can decrease in response to increased prevalence. The counter-intuitiveness of the first result when compared to the second suggests to us that the specific context will be important in understanding the logic implied for behavior choice. After a cursory inspection, we find the results do have quite different contexts. Kremer (1996) is considering differences in behavior in a heterogeneous population at risk for human immunodeficiency virus (HIV), a chronic disease without any cure. Li et al. (2015) is primarily investigating policy decisions in the management of infections from which people have a short period of contagiousness,
like flu or diarrheal diseases. These are two quite different emphases that could easily lend themselves to different working hypotheses. But a closer inspection of the two studies reveals that very similar models are being employed, and no obvious a-priori assumption is stated in either paper that would explain the divergence between their results.

Having had our swift heuristics foiled, we have no choice but to buckle down to some hard work. After a diligent analysis and a few supplementary calculations, we propose that the best explanation for this inconsistency is differing assumptions about information. When actors have perfect knowledge, they can tailor their actions to track their states, and the intuitive expectation that increased prevalence should motivate increase prevention holds. But when actors can not rely on knowledge of their own state while making decisions, they must simultaneously balance a variety of possibilities, and expectations of wasted effort can motivate individuals to abandon prevention when the likelihood of infection is large. Li et al. (2015) assume actors have perfect knowledge of their disease state, allowing individuals to abandon useless prevention efforts as soon as they become sick, while Kremer (1996) assumes actors are completely ignorant of their disease state, and thus continue preventative actions long after infection has occurred. These competing hypotheses appear to explain the different results obtained, and highlight the importance of correctly modelling information dynamics in epidemic games in particular and population games more generally.

The analysis that follows provides a formal justification of this interpretation and a new model that explicitly incorporates the effects of delays between infection and diagnosis. Our new model includes both earlier theories as special cases. We partially explore this new model to find conditions under which increased prevalence will demotivate prevention and to determine which situation is most appropriate for the current HIV epidemic.
2 Revisiting earlier theories

To better understand what’s going on with the contradictions between Kremer (1996) and Li et al. (2015), we must delve into the details of both models. We will proceed by transforming both theories into a common language using the classic compartmental $SI$ epidemic model with no recovery as a simple characterization of a chronic infectious disease like HIV. The general formalism we follow is described elsewhere (Reluga and Galvani, 2011). See Appendix A for the correspondence between Kremer’s notation and ours.

Let $N$ be the abundance of people, which we assume to be constant, $S$ be the number of susceptible individuals, and $I$ be the number of infected individuals. Let $\lambda(I)$ be the infection pressure (risk of infection per day per person), depending on the number of infected individuals, $1/\mu$ be the expected lifetime of susceptible individuals, and $1/\delta$ be the expected lifetime of infected individuals. For the moment, we will assume there is no excess death from disease ($\delta = \mu$). Removed individuals are immediately replaced through immigration so the abundance remains constant. We depart from the classic $SI$ model by allowing costly changes in behavior to reduce infection risk. Specifically, let us define the typical relative infection probability $\sigma$ as the probability that an average individual will become infected compared to a baseline with no behavior change. The cost $\bar{c}_s$ is the aggregate per-capita investment rate in that behavior change, and $\sigma(\bar{c}_s)$ is the investment-response of relative infection probability to social distancing costs. Without loss of generality, we assume $0 \leq \sigma \leq 1$. Thus, we have the system of ordinary differential equations

$$\dot{S} = -\sigma(\bar{c}_s)\lambda(I)S + \delta I, \quad (2.1a)$$
$$\dot{I} = \sigma(\bar{c}_s)\lambda(I)S - \delta I, \quad (2.1b)$$
$$\text{where} \quad N = S + I, \quad (2.1c)$$

that describes the population-scale changes in disease prevalence. Closely related models have been studied previously (Kremer, 1996; Chen, 2004; Reluga, 2009, 2010). While conserved
abundance can be used to reduce the dynamics by 1 dimension, we will leave System (2.1) in its current form to facilitate exposition.

At the scale of individuals, the probabilities of being susceptible \( p_S(t) \) or infected \( p_I(t) \) evolve according to the linear system

\[
\begin{bmatrix}
\dot{p}_S \\
\dot{p}_I
\end{bmatrix} =
\begin{bmatrix}
-\sigma(c_s)\lambda - \delta & 0 \\
\sigma(c_s)\lambda & -\delta
\end{bmatrix}
\begin{bmatrix}
p_S \\
p_I
\end{bmatrix},
\]

which we may rewrite in the more compact matrix form \( \dot{p} = Q(c_s)p \). Here, \( c_s \) is the individual’s investment in social distancing, which may differ from the aggregate investment \( c_s \) of the population.

From this Markov process, the lifetime probability of becoming infected is

\[
\psi := \frac{\sigma(c_s)\lambda}{\sigma(c_s)\lambda + \delta}.
\]

The marginal lifetime risk of infection with respect to the relative infection probability

\[
\frac{\partial \psi}{\partial \sigma} = \frac{\delta \lambda}{(\sigma \lambda + \delta)^2}
\]

is positive, meaning that increasing the relative infection probability increases the lifetime likelihood of infection. But as Kremer observed, the marginal lifetime risk of infection does not respond monotonically to changes in infection pressure, as indicated by the fact that the cross-derivative

\[
\frac{\partial^2 \psi}{\partial \lambda \partial \sigma} = \frac{\delta (\delta - \lambda \sigma)}{(\delta + \lambda \sigma)^3}
\]

can change sign. This suggests that increases in infection pressure may discourage prevention at some point. However, to get a precise determination of how the risk geometry influences rational behavior, we must complete the specification of our model.
Let \( \tilde{Q} \) be the corresponding stationary transition rate matrix when prevalence is near equilibrium. In our standard notational style (Reluga and Galvani, 2011), the payoff of a strategy \( c_s \) in a population playing \( \bar{c}_s \) is given by

\[
U(c_s, \bar{c}_s) := \int_{t_0}^{\infty} e^{-ht} v^T p(t) dt = v^T (hI - \tilde{Q})^{-1} p(t_0) \tag{2.6}
\]

where \( p(t_0) \) is the initial state distribution for an individual, \( h \) is the discount rate, and \( v \) is the vector of payoff rates for each state. We may analyze this game using standard methods. The best-response strategy correspondence

\[
c_s^B(\bar{c}_s) := \arg\max_{c_s \geq 0} U(c_s, \bar{c}_s), \tag{2.7}
\]

and the global Nash equilibrium with invasion potential \( c_s^* \) (if it exists) solves

\[
c_s^* \in c_s^B(c_s^*), \quad \forall \bar{c}_s, \quad U(c_s^*, \bar{c}_s) \geq U(\bar{c}_s, \bar{c}_s). \tag{2.8}
\]

Calculation of the payoff function requires specifications of the equilibrium prevalence, initial conditions, and payoff rates. When the infection pressure \( \lambda(I) \) is monotone increasing and concave with linear behavior for small \( I \), System (2.1) has a unique stable steady-state, which we represent as \( (\tilde{S}, \tilde{I}) \). If the disease prevalence is close to the endemic steady state level, as specified by the equilibrium conditions of System (2.1), then

\[
\frac{\tilde{I}}{N - \tilde{I}} = \frac{\sigma(\bar{c}_s)\lambda(\tilde{I})}{\delta}. \tag{2.9}
\]

We still must identify the functional dependence \( \lambda(I) \) of infection pressure on prevalence and the relative infection probability investment-response \( \sigma(c_s) \) to precisely determine the equilibrium prevalence. For \( \lambda(I) \), we consider a standard-incidence hypothesis (Brauer, 2006)
suitable for large populations

\[ \lambda(I) := \frac{\beta I}{N}. \]  

(2.10)

For a relationship between investment in prevention and relative infection probability, we expect a function that is convex, monotone decreasing from 1 but stays positive. Among the several potential choices, a mathematically convenient hypothesis for the investment response is

\[ \sigma(c_s) := \frac{1}{1 + mc_s} \]  

(2.11)

Eqs. (2.9), (2.10), and (2.11) together imply that there is a unique non-negative steady-state infection pressure \( \tilde{\lambda} := \lambda(\tilde{I}) \) calculated from Eq. (2.9) and given as

\[ \tilde{\lambda}(\tilde{c}_s) = \max \left\{ 0, \beta - \frac{\delta}{\sigma(\tilde{c}_s)} \right\}. \]  

(2.12)

and that it is a decreasing function of typical investment \((\partial \tilde{\lambda}/\partial \tilde{c}_s < 0)\).

Up to this point, our model is equally-well suited to the contexts of both Kremer (1996) and Li et al. (2015). The different contexts of the two models could indicate different initial conditions for individuals entering the population. For flu and diarrheal infections, vertical transmission is not usually a concern. For HIV, vertical transmission from mother to child can be a major health concern (Garcia et al., 1999). But since factors associated with vertical transmission are not essential for our risk-management questions and would complicate our analysis, we only consider horizontal transmission and assume all people are initially susceptible in all cases;

\[ \mathbf{p}(t_0) := [p_S(t_0), p_I(t_0)] = [1, 0]. \]  

(2.13)

The key differences between Kremer (1996) and Li et al. (2015) turns out to be the
specification of payoff rates $v$. Li et al. (2015) assumes individuals have full knowledge (perfect information, in game-theory lingo) of their disease state. When individuals have perfect information about their state, the payoff rates are

$$v_{\text{informed}} := [u - c_s, u - c_i],$$

(2.14)

where $u$ is the daily payoff gains, $c_s$ is the daily cost of the individual’s behavioral strategy for preventing infection, and $c_i$ is the daily cost of being infected. Under these payoff gains, a rational individual will abandon costly preventative behaviors once they become infected because preventions no longer offer any benefit. This leads to payoff

$$U_{\text{informed}}(c_s, \bar{c}_s) := \frac{u}{\delta + h} - \frac{c_s + \sigma(c_s)\bar{\lambda}\left(\frac{c_s}{\delta + h}\right)}{\sigma(c_s)\bar{\lambda} + \delta + h},$$

(2.15)

which is a linear fractional transform (a hyperbola with horizontal and vertical asymptotes) in $\sigma\bar{\lambda}$, hence monotone in $\sigma$ and $\bar{\lambda}$ respectively. The best response solves

$$\left(c_s^B - c_i\right) \frac{\partial \sigma}{\partial c_s} = \sigma(c_s^B) + \frac{\delta + h}{\bar{\lambda}}.$$  

(2.16)

Given a specific $\sigma(c_s)$ we can find $c_s^B$ in Eq. (2.16) using root finding routines. But to facilitate geometric analysis of best responses we introduce another method of finding $c_s^B$. First we are looking for a family of functions $f$ which satisfy a tangency condition given by the first order differential equation

$$\left(s - c_i\right) \frac{\partial f}{\partial s} = f(s) + \frac{\delta + h}{\bar{\lambda}}.$$  

(2.17)

The set of exact solutions to this differential equation is

$$f(s; C_1) = C_1 - \frac{s}{c_i} \left[ C_1 + \frac{(\delta + h)}{\bar{\lambda}} \right].$$

(2.18)
where $C_1$ is an integration constant parameterizing the set. Once we have the solution set, we look for a tangent point between $f(c_s; C_1)$ and $\sigma(c_s)$ for some value of the constant $C_1$. That point will satisfy Eq. (2.17) hence is the best response. Any ray emanating from the point $(c_1, -(\delta + h)/\tilde{\lambda})$ belongs to the function family $f$, and since this point moves down as infection pressure increases, it is easy to prove that as long as the relative infection probability investment-response $\sigma(c_s)$ is monotone decreasing, but positive, $c^B_s(\tilde{\lambda})$ must be an increasing function of infection pressure (Li et al., 2015). If we specify $\sigma(c_s)$ according to Eq. (2.11), $c^B_s$ is the unique positive solution of a quadratic equation. Note that the positive value of daily life $u$ does not affect the results because infection cannot shorten life expectancy under this model.

On the other hand, Kremer (1996) assumes

$$v_{\text{ignorant}} := [u - c_s, u - c_s - c_i]. 	ag{2.19}$$

This is consistent with a theory where people are completely ignorant of their disease state and that they will continue their costly preventative behaviors even after they have become infected. It then follows that the payoff under this condition is

$$U_{\text{ignorant}}(c_s, \bar{c}_s) := \frac{u - c_s}{\delta + h} - \frac{\sigma(c_s)\tilde{\lambda}}{\sigma(c_s)\tilde{\lambda} + \delta + h} \left( \frac{c_i}{\delta + h} \right). \tag{2.20}$$

The best response condition is

$$-\frac{\partial \sigma}{\partial c_s} = \left( \frac{\delta + h + \tilde{\lambda} \sigma(c^B_s)}{(\delta + h)\tilde{\lambda}c_i} \right)^2. \tag{2.21}$$

To analyze best response, we first introduce a first-order autonomous ordinary differential
\[- \frac{\partial g}{\partial s} = \frac{\left( \delta + h + \tilde{\lambda}g(s) \right)^2}{(\delta + h)\tilde{\lambda}c_i} , \tag{2.22}\]

which has solutions that are hyperbolic functions of the form

\[g(s; C_2) = \frac{\delta + h}{\tilde{\lambda}} \left( \frac{c_i - s + C_2}{s - C_2} \right) \tag{2.23}\]

where \( C_2 \) is an arbitrary integration constant. The best responses correspond to the tangent point between the true \( \sigma(c_s) \) and one of the convex hyperbolas \( g(c_s) \) from (2.23).

The best-response condition for \( U_{\text{ignorant}} \) in Eq. (2.23) is more flexible than that obtained for \( U_{\text{informed}} \) in Eq.(2.16). For some parameter values, increasing infection pressure decreases the amount of social distancing in the best response. As a specific example, under Eq. (2.11),

\[c^B_s(\tilde{\lambda}) = \max \left\{ 0, \sqrt{\frac{\tilde{\lambda}c_i}{(\delta + h)m}} - \frac{\tilde{\lambda}}{(\delta + h)m} - \frac{1}{m} \right\} . \tag{2.24}\]

The best response is 0 always if \( c_i m < 4 \). When \( c_i m > 4 \), the best response is unimodal in \( \tilde{\lambda} \), and there is a critical infection pressure at which best responses reach their maximum and turn over. We call this value of the infection pressure a “turning point”, and adopt the notation \( \tilde{\lambda}_\tau \), observing that \( \frac{\partial c^B_s(\tilde{\lambda})}{\partial \tilde{\lambda}} = 0 \). For Eq. (2.24), \( \tilde{\lambda}_\tau = (\delta + h)c_i m/4 \).

When comparing the best-response functions for \( U_{\text{informed}} \) and \( U_{\text{ignorant}} \) (see Figure 1), we see that differences in the cost-structures of the strategies, based on information differences, lead to important differences in rational responses when the risk of infection is large. In the case where individuals are ignorant of their disease-state, there is a turning point where risk is so large, behavior change reverses direction. In the case where individuals have perfect information, there is no turning point.
3 Incorporating a diagnosis rate

The comparison of the previous section shows that different assumptions about the information available to the individual can change the qualitative properties of the best response to risks of infection. However, neither assumption seems wholly adequate in general. There will almost always be a lag between infection and diagnosis of the infection. But how slow must diagnosis be, so that costs from wasted prevention efforts reverse the intuitive expectation that increasing prevalence leads to increasing prevention?

To see how diagnosis influences game-play, we introduce a three-compartment model for the analysis of diagnosis rates. In addition to the susceptible compartment \((S)\), we now distinguish between the infected and undiagnosed \((I)\) and the infected and diagnosed \((J)\). Infected and undiagnosed people are diagnosed at a constant rate \(\epsilon\). If a disease shortens lifespan but does not have many symptoms prior to death, it may go undetected. Kremer
Table 1: Important symbols and their definitions

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>Time</td>
</tr>
<tr>
<td>$S(t)$</td>
<td>Density of susceptibles</td>
</tr>
<tr>
<td>$I(t)$</td>
<td>Density of infections</td>
</tr>
<tr>
<td>$N$</td>
<td>Total population density</td>
</tr>
<tr>
<td>$p_S(t)$</td>
<td>probability that an individual is susceptible at $t$</td>
</tr>
<tr>
<td>$p_I(t)$</td>
<td>probability that an individual is infected at $t$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>natural background mortality rate</td>
</tr>
<tr>
<td>$\delta$</td>
<td>infection mortality rate</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>infection pressure</td>
</tr>
<tr>
<td>$\beta$</td>
<td>transmission rate</td>
</tr>
<tr>
<td>$u$</td>
<td>background income rate for life</td>
</tr>
<tr>
<td>$c_i$</td>
<td>cost rate of infection</td>
</tr>
<tr>
<td>$c_a$</td>
<td>individual investment rate in prevention</td>
</tr>
<tr>
<td>$\bar{c}_a$</td>
<td>aggregate per-capital investment in prevention</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>relative infection probability</td>
</tr>
<tr>
<td>$m$</td>
<td>linear efficiency under the hyperbolic prevention hypothesis, Eq. (2.11)</td>
</tr>
<tr>
<td>$h$</td>
<td>discount rate</td>
</tr>
<tr>
<td>$\mathcal{U}$</td>
<td>lifetime utility and the population game payoff</td>
</tr>
<tr>
<td>$\psi$</td>
<td>lifetime chance of being infected</td>
</tr>
<tr>
<td>$J(t)$</td>
<td>Density of diagnosed infections</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>diagnosis rate</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>relative transmission rate after diagnosis</td>
</tr>
<tr>
<td>$c_j$</td>
<td>cost rate of infection after diagnosis</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>mortality rate after diagnosis</td>
</tr>
</tbody>
</table>
Figure 2: The governing equations of the general social distancing game with diagnosis. At the community scale, disease-dynamics are controlled by a nonlinear reaction network (top, diamonds). The dotted edges in the reaction network represent reaction inputs that are not consumed by the reaction. But for an individual in this community, the probabilities of residing in each state are governed by a continuous-time Markov chain (bottom, circles) where linear reaction rates are constant and equal at both scales, while the nonlinear reaction corresponding to disease transmission induces a dependence of the individual-scale infection risk on the community’s state.

(1996) specifically assumes there is no change in lifetime due to illness, but this assumption is easily relaxed. If we allow $\mu$ to be the natural death rate, and $\delta$ to be the death rate of those infected with disease while undiagnosed, and $\zeta$ to be the death rate of those infected with disease after diagnosis, we expect infection to make individuals worse off ($\delta \geq \mu$), but diagnosis to improve outcomes ($\delta \geq \zeta \geq \mu$).

At the population scale,

\[
\begin{align*}
\dot{S} &= -\sigma(\tau_s)\lambda(I,J)S + \delta I + \zeta J, \quad (3.1a) \\
\dot{I} &= \sigma(\tau_s)\lambda(I,J)S - \delta I - \epsilon I, \quad (3.1b) \\
\dot{J} &= \epsilon I - \zeta J, \quad (3.1c) \\
N &= S + I + J. \quad (3.1d)
\end{align*}
\]

Since we are enforcing the constant-population-size condition, births and deaths of susceptible individuals at rate $\mu S$ perfectly offset each other at the population scale. At steady-state,
either $\tilde{\lambda} = 0$ or
\[
\sigma(\bar{c}_s) \tilde{\lambda} = \frac{(\delta + \epsilon)I}{N - I - \frac{\epsilon}{\zeta}I} \quad \text{with} \quad \epsilon I = \zeta J \quad \text{and} \quad S = N - I - J.
\] (3.2)

If the infection pressure has the simple linear standard-incidence form
\[
\lambda(I, J) = \frac{\beta(\sigma(\bar{c}_s)I + \alpha J)}{N}
\] (3.3)
then at steady-state,
\[
\bar{\lambda}(\bar{c}_s) = \max \left\{ 0, \frac{(\beta \sigma(\bar{c}_s)(\alpha \epsilon + \sigma(\bar{c}_s)\zeta) - \zeta(\delta + \epsilon))}{\sigma(\bar{c}_s)(\epsilon + \zeta)} \right\}.
\] (3.4)

However, we will make only minor further use of this result.

When attempting to determine individual payoffs, we should note that costs are also signals. A costly disease will make people sick, which is an automatic signal that something has changed. However, the exact cause of this change may not be apparent. For generality, we consider a disease-cost $c_i$ after infection but before diagnosis, and a disease cost $c_j$ after diagnosis. The relative magnitude of these costs depends on the circumstances. If diagnosis is by chance, and effective treatment begins after diagnosis, we may expect $c_i > c_j$. However, if diagnosis is correlated to an increase in severity of symptoms and no treatment is available, we may expect $c_i < c_j$. Upon diagnosis, we’ll assume individuals abandon their futile investments in social distancing, though sometimes people persist with behaviors even after they have exhausted their usefulness.

For the individual, transition rates are given by
\[
\begin{bmatrix}
\dot{p}_S \\
\dot{p}_I \\
\dot{p}_J
\end{bmatrix} =
\begin{bmatrix}
-\sigma(c_s)\lambda - \mu & 0 & 0 \\
\sigma(c_s)\lambda & -\delta - \epsilon & 0 \\
0 & \epsilon & -\zeta
\end{bmatrix}
\begin{bmatrix}
p_S \\
p_I \\
p_J
\end{bmatrix}
\] (3.5)
with value gains $v = [u - c_s, u - c_s - c_i, u - c_j]^T$. We now apply our standard methods and calculate the generalized utility

$$U_{general}(c_s, \tilde{c}_s) := \frac{u - c_s + \tilde{\lambda}\sigma(c_s) \left( u - c_s - c_i + \epsilon \frac{u - c_j}{\delta + h + \epsilon} \right)}{\lambda\sigma(c_s) + \mu + h}$$ (3.6)

In the limit of very slow diagnosis and no information about infection state, Eq. (3.6) has

$$\lim_{\epsilon \to 0} U_{general} = \frac{u - c_s}{\lambda\sigma(c_s) + \mu + h} + \frac{\tilde{\lambda}\sigma(c_s)(u - c_s - c_i)}{(\tilde{\lambda}\sigma(c_s) + \mu + h)(\delta + h)}$$ (3.7)

which is equal to $U_{ignorant}$ given by Eq. (2.20) when there is no disease-induced mortality ($\delta = \mu$) and diagnosis does not change the cost of infection ($c_i = c_j$). In the limit of rapid diagnosis, when individuals have near-perfect state information,

$$\lim_{\epsilon \to \infty} U_{general} = \frac{u - c_s}{\lambda\sigma(c_s) + \mu + h} + \frac{\tilde{\lambda}\sigma(c_s)(u - c_j)}{(\tilde{\lambda}\sigma(c_s) + \mu + h)(\zeta + h)},$$ (3.8)

which is equal to $U_{informed}$ given by Eq. (2.15) when there is no disease-induced death ($\zeta = \mu$). This limit was studied by Li et al. (2015). Thus, we have a general model that explicitly implements diagnosis rate and obtains both of the pre-existing models as complementary limiting cases of this diagnosis rate.

### 3.1 Generic Analysis

In this section, we will outline how diagnosis changes payoffs, best replies, and turning points in the force of infection. Some of the more mathematical results are presented in Appendix B.

The first innovation of $U_{general}$ is to allow different mortality rates pre- and post-diagnosis ($\delta$ and $\zeta$, respectively). As in the earlier case, the best response for $U_{general}$ exhibits a turning
point in infection pressure when there is no diagnosis ($\epsilon = 0$). This allows the coexistence of multiple Nash equilibria in some cases (Kremer, 1996). For a fixed diagnosis rate, increases in either pre- or post-diagnosis disease mortality will decrease the value of life post-infection, decreasing the total payoff and promoting prevention. The effect of diagnosis on the best-response is contingent on several factors. The best response may increase or decrease in response to faster diagnosis, depending on the risk of infection, the balance of pre- and post-diagnosis costs, and the efficiency of social distancing as specified by $\sigma(c_s)$. To say things more precise, we must calculate the best response (see Figure 3). The best response $c^B_s(\bar{\lambda})$ for $U_{\text{general}}$ is a local solution point of the non-autonomous Riccati equation (B.1), Appendix B. Note that the right hand side is a ratio of linear functions in the diagnosis rate $\epsilon$, and hence, monotone on each side of a vertical asymptote. The direction of this monotone effect switches switches when $\sigma(c_s) = \sigma_{\epsilon}(c_s)$ where

$$\sigma_{\epsilon}(c_s) := \frac{\delta + h}{\bar{\lambda}} \left( \frac{\frac{u - c_j}{\zeta + h} - \frac{u - c_i - c_s}{\delta + h}}{\max \left\{ 0, \frac{u - c_s}{\mu + h} - \frac{u - c_j}{\zeta + h} \right\}} \right).$$

(3.9)

It follows that if $\sigma(c^B_s) > \sigma_{\epsilon}$, then $dc^B_s/d\epsilon > 0$, but otherwise, $dc^B_s/d\epsilon < 0$ (see Figure 4). Some stronger results can also be found. If diagnosis is counter-productive and locks-in higher costs, then at equilibrium it always encourages prevention (see Appendix B).

The effect of diagnosis on payoff is simpler. When diagnosis reduces the costs of infection ($c_i > c_j$ and $\zeta < \delta$), faster diagnoses increases overall payoff ($\partial U_{\text{general}}/\partial \epsilon > 0$). The intuitive reason is that $\epsilon$ appears in $U_{\text{general}}$ only as a weight in a weighted average, and increasing that weight shifts the utility towards the better outcomes. Since the function increases everywhere as $\epsilon$ increases, the utility of the best response must also increase (although we can not say whether the best response itself will increase or decrease). On the other hand, the same argument implies that when diagnosis is leads to costlier infection ($c_i < c_j$ and $\zeta > \delta$), faster diagnoses decreases overall utility ($\partial U_{\text{general}}/\partial \epsilon < 0$) and the utility of the best reply ($dU_{\text{general}}(c^B_s(\epsilon), \bar{c}_s)/d\epsilon < 0$). This all assumes the steady-state infection pressure $\bar{\lambda}$ is constant.
Figure 3: A contour plot of the generalized utility $U_{\text{general}}$ as a function of the infection pressure $\tilde{\lambda}$ and the individual investment $c_s$ based on Eq. (3.6). The black dashed line is the best response as a function of infection pressure. The red dot marks the best-response turning point. Payoffs decrease as infection pressure increases. Note that near the turning point, the Hessian will be diagonal. Parameters: $u = 1$, $c_i = 0$, $c_j = 0.1$, $\mu = 1/50$, $\zeta = \delta = 1/40$, $\epsilon = 1$, $\sigma(c_s) = (1 + 80c_s)^{-1}$. 
Figure 4: The contours of $U_{\text{general}}$ as a function of investment $c_s$ and relative infection risk $\sigma$ when the diagnosis rate $\epsilon = 0.3$ (black, solid) and $\epsilon = 0.55$ (green dashed) for a fixed infection pressure. As diagnosis rate increases, contour lines bend out from the $\sigma,(c_s)$ curve (red dotted line) given by Eq. (3.9) toward becoming straight lines. The best response is determined by the tangent point between $\sigma,(c_s)$ (blue solid thick is one example) and the payoff contours. The net effect of faster diagnosis when $\sigma,(c_s)$ is convex is to pull the best response toward the $\sigma,(c_s)$ curve -- in this case, increasing $c_s^B$. If the tangent point occurred to right of $\sigma,(c_s)$ in the shaded region, then faster diagnosis would decrease the best response. Parameters: $\tilde{\lambda} = 0.9$, $\mu = 1/50$, $\zeta = 1/25$, $\delta = 1/10$, $u = 1$, $c_j = 1/10$, $c_i = 0$, $\sigma(c_s) = (1 + 8c_s)^{-1}$. 
Reductions in infection-pressure resulting from faster diagnosis will increase the payoffs.

Now, what about turning-points of best responses under increasing infection pressure?

From differential geometry (see Figure 3), the turning-point infection pressure $\tilde{\lambda}_r$ must occur where

$$\frac{\partial U}{\partial c_s} = 0 \quad \text{and} \quad \frac{\partial^2 U}{\partial c_s \partial \lambda} = 0.$$  \hspace{1cm} (3.10)

This leads to a nonlinear first-order differential equation

$$\left[ u - c_s^B + \frac{(\mu + h) \left( c_i - \frac{\epsilon(u - c_j)}{\zeta + h} \right)}{\delta + \epsilon - \mu} \right] \frac{d\sigma(c_s^B)}{dc_s} = \frac{(\tilde{\lambda}_r \sigma(c_s^B) + \mu + h) \sigma(c_s^B)}{\left( \mu + h - \tilde{\lambda}_r \sigma(c_s^B) \right)}.$$  \hspace{1cm} (3.11)

Eq. (3.11) has a long closed-form solution

$$K \left( \frac{\mu + h}{\tilde{\lambda}_r} \sigma + 1 \right)^2 = \left[ u - c_s^B + \frac{(\mu + h) \left( c_i - \frac{\epsilon(u - c_j)}{\zeta + h} \right)}{\delta + \epsilon - \mu} \right]$$  \hspace{1cm} (3.12)

where $K$ is a constant of integration. When $\sigma(c_s)$ and $c_s^B(\tilde{\lambda})$ are known, Eq. (3.11) can be solved for the turning-point infection pressure $\tilde{\lambda}_r$ using any standard one-dimensional root-finding technique. Unfortunately, we can not provide a general closed-form solution for the turning-point infection pressure in terms of first-principle parameters, but using Eqs. (B.1) and (3.11) in combination, we can show

$$\tilde{\lambda}_r = \sqrt{(\mu + h)(\delta + \epsilon + h)} \left( \frac{\mu + h}{\sigma(c_s^B(\tilde{\lambda}_r))} \right)$$  \hspace{1cm} (3.13)

where the best response is calculated from Eq. (B.1) and the definition of $\sigma(c_s)$. Considering that infection is expected to accelerate mortality ($\delta \geq \mu$), we can easily obtain Kremer’s
Figure 5: When individuals have no knowledge of their disease state ($\epsilon = 0$), the best response exhibits a turning point. As the disease-induced death rate increases, the best response becomes larger, and the turning point $\tilde{\lambda}_r$ moves to larger forces of infection. These curves provide lower bounds on the best response for all cases of $\epsilon > 0$. Parameters: $u = 1$, $c_i = c_j = 0$, $\zeta = \delta$, $\mu = 1/50$, $\sigma(c_s) = (1 + 80c_s)^{-1}$.

Figure 6: Black curve is the boundary of the region in $\delta \times \lambda$ parameter space where the best response is positive when there is no diagnosis ($\epsilon = 0$). Blue dashed curve is where the best response is largest. Asymptotically, $c_s^B$ is only positive if $\delta > \mu$ and $\mu/(\mu u - 1) < \tilde{\lambda} < \delta/(\mu u - 1)$. When disease mortality ($\delta$) is large, the best response $c_s^B$ is maximized around $\tilde{\lambda} \approx (\mu u + 1)\sqrt{\mu \delta}/2$. Parameters: $u = 1$, $c_i = c_j = 0$, $\zeta = \delta$, $\mu = 1/50$, $\sigma(c_s) = (1 + 80c_s)^{-1}$.

lower bound that infection rates must exceed background mortality rates at the turning point ($\tilde{\lambda}_r > \mu$) and that when the infection pressure reaches this turning point, the background probability of infection over a lifetime without prevention,

$$
\psi(\sigma = 1) = \frac{\tilde{\lambda}_r}{\tilde{\lambda}_r + \mu} > \frac{1}{2},
$$

based on Eq. (2.3).

### 3.2 An example

When $\sigma(c_s)$ is a hyperbola, as suggested by Eq (2.11), we can perform all of our calculations using polynomials and obtain exact (though sometimes algebraically opaque) answers. In the absence of diagnosis ($\epsilon = 0$), the structure of the solution is the same as that found by Kremer.
Figure 7: The turning-point infection-pressure increases as the diagnosis rate increases, and the larger excess death due to infection \( \frac{\delta}{\mu} \), the larger the turning point. Parameters: \( u = 1 \), \( c_i = c_j = 0 \), \( \zeta = \delta \), \( \mu = 1/50 \). \( \sigma(c_s) = (1 + 80c_s)^{-1} \).

(1996), with the extra condition that increasing disease-induced mortality increases the turning infection pressure \( \tilde{\lambda}_r \) and the best-response (Figure 5). As disease-induced mortality increases relative to natural mortality, the best response increases, and the turning point moves to higher infection pressures. Improving the diagnosis rate also increases the turning point (Figure 7), although the best response itself may increase or decrease, depending on the parameter values (Figure 8). The presence of the turning point can allow multiple Nash equilibria to coexist, but faster diagnosis eventually pushes out alternatives and leaves a single Nash equilibrium (Figure 10). In cases with multiple Nash equilibria, the one with least investment has the lowest utility. Thus, slow diagnosis can create a “trap” (Bonds et al., 2010), from which individuals can not unilaterally escape.

For fixed mortality rates, using the best-response condition to eliminate \( c_s \) from Eq. (3.9) shows faster diagnosis promotes prevention \( \left( \frac{dc_s^B}{d\epsilon} > 0 \right) \) if and only if

\[
\left( \frac{u - c_j}{\zeta + h} - \frac{u - c_i}{\delta + h} \right) < \frac{\tilde{\lambda} + \mu + h}{m(\delta + h)(\mu + h)}
\]

(3.15)

(see Figure 8). For example, if diagnosis coincides with a decrease in the value of future life,
Figure 8: The best response depends on the infection pressure $\lambda$ and the diagnosis rate $\epsilon$ if diagnosis improves (left) or worsens (right) outcomes. When infection pressure is low, increases in infection risk can be offset by increased prevention, but when infection pressure is large, the costs of wasted effort after infection out-weigh the benefits prior to infection. The effect of diagnosis is characterized by Eq. (3.15). When diagnosis improves outcomes (left), the sensitivity of the best response to diagnosis rate changes sign -- faster diagnosis leads to less prevention at low infection pressures but more prevention at higher infection pressures. When diagnosis corresponds to worse outcomes (right), faster diagnosis always promotes prevention. Parameters: (left) $\zeta = 1/25$, $\delta = 1/10$, (right) $\zeta = \delta = 1/40$, (both) $u = 1$, $c_i = 0$, $c_j = 1/10$, $\mu = 1/50$, $\sigma(c_s) = (1 + 80c_s)^{-1}$.

such that

$$\frac{u - c_j}{\zeta + h} < \frac{u - c_i}{\delta + h},$$

(3.16)

then faster diagnosis promotes prevention.

3.3 Special case of HIV

Given the spectrum of situations now encompassed by our general model, the question naturally arises, which situation is most appropriate for HIV risk management?

One reasonably appropriate parameter specification for the situation of HIV transmission would be the special case of our general model where disease does not have notable cost during the course of infection, but does shorten the lifespan of infected individuals unless the
Figure 9: Plots of absolute (left) and relative (right) utility contours without diagnosis ($\epsilon = 0$) in Eq. (3.6) under hyperbolic $\sigma$ from Eq. (2.11) and steady-state condition (3.4). There are two Nash equilibria with local invasion potential ($c_s = 0$ and $c_s \approx 0.158$) and one intermediate equilibrium without local invasion potential ($c_s \approx 0.028$). The no-distancing equilibrium ($c_s = 0$) is a trap with low expected utility. Parameters: $h = c_i = c_j = 0$, $u = 1$, $\zeta = 1/25$, $\delta = 1/10$, $\mu = 1/50$, $\beta = 20$, $\alpha = 0.2$, $\sigma(c_s) = (1 + 80c_s)^{-1}$.

individual is diagnosed and receives treatment ($c_i = c_j = 0$ but $\delta > \mu$ and $\zeta = \mu$). Then the expected payoff function

$$u - c_s + \lambda \sigma(c_s) \left( \frac{u - c_s + \epsilon \left( \frac{u}{\mu + h} \right)}{\delta + h + \epsilon} \right) \frac{\lambda \sigma(c_s) + \mu + h}{\lambda \sigma(c_s) + \mu + h}.$$  

For parameter estimates, basic knowledge of public health and HIV epidemiology (Hethcote and Ark, 1992) suggests we measure time in units of years and take $\delta = 1/10$, $\mu = 1/50$, $h = 0.03$, $u = 1$. Based on Eq. (3.13), the turning-point infection pressure $\tilde{\lambda}_r > 0.08$ per year.
Figure 10: Plot of the effect of diagnosis rate per year (\(\epsilon\)) on the Nash equilibria investment in social distancing (left), the expected payoff (middle) and the equilibrium force of infection (right). Calculations use Eq. (3.6) under the hyperbolic \(\sigma\)-hypothesis from Eq. (2.11) and steady-state condition (3.4). There are two equilibria with local invasion potential (blue solid and green dotted), while there is one intermediate equilibrium without local invasion potential (dashed red). As diagnosis rate increases, the lower two equilibria collide and annihilated, leaving only the upper equilibria (left). Note also that the upper Nash equilibrium does not respond monotonely to the diagnosis rate. Interestingly, when the diagnosis rate is very slow, individuals are best off at the high-prevention Nash equilibrium, where the infection pressure also seems to be minimized. A larger infection pressure (and more disease) is tolerated when diagnosis is rapid because treatment is effectively prolonging life (for these parameters) Parameters: \(h = c_i = c_j = 0, u = 1, \zeta = 1/25, \delta = 1/10, \mu = 1/50, \beta = 20, \alpha = 0.2, \sigma(c_s) = (1 + 80c_s)^{-1}\).
Extrapolating based on the very rough approximation that

\[
\text{per-capital prevalence} \approx \frac{\bar{\lambda}}{\lambda + \mu},
\]

and prevalence data from the WHO and UNAIDS (UNAIDS, 2013) infection pressures (risk of infection per person per year) approximated from prevalences may be as high as \(\bar{\lambda} \approx 0.033\) is some nations in Sub-Saharan Africa. This suggests the turning point is not of practical importance for HIV at national spatial scales, no matter what the exact form of \(\sigma(c_s)\) (see Figure 11).

However, populations are not strongly mixed and not homogeneous. It’s long been established that for sexually transmitted diseases, large populations often possess much smaller core groups with sustaining transmission (Yorke et al., 1978; Thomas and Tucker, 1996). Prevalences in some groups have exceeded 50% (Ramjee et al., 1998; Morison et al., 2001), and it seems highly likely that in countries like Botswana and Malawi, infection pressures in core groups have exceeding the turning-point threshold.

4 Conclusion

In this paper, we have developed an epidemic game that explicitly accounts for diagnosis, and shown that this constructively resolves an existing issue in the literature. When diagnosis is fast, individuals effectively have perfect information about their state, and rationally will increase prevention as infection risk increases. But when diagnosis is slow, wasted investments in investment after infection can rationally discourage prevention when risks are high. Our analysis confirms the earlier result that this turning-point only becomes important for HIV management at infections pressures at least 24 times larger than the highest observed infection pressures.

While best-response levels of prevention may be strictly increasing with infection risk for realistic parameter sets, issues of incomplete disease-state information may still have practical
Figure 11: The left plot shows the turning point infection pressure \( \tilde{\lambda}_r \) as a function of diagnosis rate \( \epsilon \) (solid blue) for HIV parameters from Section 3.3, along with bounds where the best response involves no social distancing (dashed red). The right plot shows the expected utility for the corresponding best responses in \( \epsilon \times \tilde{\lambda} \) parameter space. Empirically, diagnosis rates can be weakly bounded to be no faster than a day and no slower than a lifetime (\( \epsilon \in [1/100, 300] \)). The turning-point only comes into play for high infection pressures likely found only in core groups. Note also the counter-intuitive effect that increases in diagnosis rate increase expected utilities despite decreasing the best-response investments in social distancing.
importance for decisions. First we note that while large-scale estimates of risk may not reach
turning-point levels, spatial variations in incidence may be large enough that regional risk
estimates may exceed critical levels. In addition, individual and group perceptions of risks
and costs may diverge significantly from actual values, leading to turning-point estimates
much below the actual value and discouraging preventative behaviors.

While we do suggest a standard-incidence hypothesis for the relationship between preva-
lence and infection pressure, all that is truly needed is that $\lambda(I)$ be a concave, increasing
function with a finite upper bound on the first derivative, so as to ensure the existence and
uniqueness of the disease steady-state.

Acknowledgements

This research was supported by National Science Foundation grants CCF-1215682 and
DMS-0920822, and National Institute of General Medical Sciences NIH grant R01GM089932.
I thank Guoliang Fang for comments on an early draft.

References

M. H. Bonds, D.C. Keenan, P. Rohani, and J. D. Sachs. Poverty trap formed by the ecology

1, 2006. URL http://dx.doi.org/10.3934/mbewu.2006.3.1.

J. I. Bulow, J. D. Geanakoplos, and P. D. Klemperer. Multimarket oligopoly: Strategic


Jing Li, Darla V. Lindberg, Rachel A. Smith, and Timothy C. Reluga. Provisioning of infectious disease public health can be designed to anticipate public policy responses. *In submission*, 2015.


Gita Ramjee, Salim S Abdool Karim, and Adriaan W Sturm. Sexually transmitted infections


Table 2: Translation of symbol names between Kremer (1996) and our paper.

<table>
<thead>
<tr>
<th>Kremer’s Expression</th>
<th>Our Equivalent expression</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y(t) )</td>
<td>( I(t)/N )</td>
<td>Fraction of population infected</td>
</tr>
<tr>
<td>( 1 - Y(t) )</td>
<td>( S(t)/N )</td>
<td>Fraction of population susceptible</td>
</tr>
<tr>
<td>( \beta Y )</td>
<td>( \lambda )</td>
<td>Infection pressure</td>
</tr>
<tr>
<td>( i )</td>
<td>( \sigma(c_s) )</td>
<td>Relative infection probability</td>
</tr>
<tr>
<td>( \delta )</td>
<td>( \delta )</td>
<td>Death rate of infected people</td>
</tr>
<tr>
<td>( u(i) )</td>
<td>( \frac{u-c_s}{c_s} )</td>
<td>Relative value of health per infection cost</td>
</tr>
<tr>
<td>( p(i,Y) )</td>
<td>( \psi(\sigma,\lambda) = \frac{\sigma\lambda}{\sigma\lambda + \delta} )</td>
<td>Lifetime probability of infection</td>
</tr>
</tbody>
</table>

A Model translation

Kremer’s core model for a homogeneous population is formulated in a manner equivalent to a population game. Specifically, he proposes disease dynamics under the law

\[
\dot{Y} = i(Y) \beta Y (1 - Y) - \delta Y
\]  

(A.1)

with a payoff function

\[
U(i,Y) := u(i) - p(i,Y) = u(i) - \frac{i\beta Y}{\delta + i\beta Y}
\]  

(A.2)

where \( u(i) \) is concave increasing, and attains it’s maximum at \( i_{\text{max}} \). For purposes of further development, this formulation is a little awkward, so we change notation to a more explicit style consistent with practices in Li et al. (2015) (see Table 2). Because of some of the necessities of extending the original model to explicitly include diagnosis, this translation only strictly applies for Section 2. Kremer’s original theory does not make mention of discounting or a specification of the lifetime of un-infected individuals, both of which appear in our formulation. The inclusion of discounting alters some formulas in very minor ways. Un-infected mortality will cancel out whenever it is the same as infected mortality; otherwise, it introduces a scaling term that is absorbed into \( u(i) \).
B Best response conditions for the generalized utility

The best response $c^B_s(\tilde{\lambda})$ for the generalized utility given in Eq. (3.6) is a local solution point of the non-autonomous Riccati equation

$$
\tilde{\lambda}\sigma'(c^B_s) = \frac{\left(\tilde{\lambda}\sigma(c^B_s) + \mu + h\right) \left(\tilde{\lambda}\sigma(c^B_s) + \epsilon + \delta + h\right)}{(u - c^B_s)(\mu - \epsilon - \delta) + (h + \mu) \left(\frac{\epsilon (u - c_j)}{(\zeta + h)} - c_i\right)}
$$

(B.1)

and the derivative $\sigma'(c_s)$ is a known function.

Unlike the best responses derived for Eq. (2.15) and (2.20), the daily baseline payoff $u$ remains an important variable in Eq. (B.1) because the course of disease alters the lifetime, and thus the disability-adjusted life years (DALYs) accumulated by a particular strategy.

While I would like to now show you convergent stability of the Nash equilibrium directly from our definition of $U(c_s, \tau_s)$ under appropriate constraints, I have not yet been able to construct this, in part because of the absence of a closed-form representation for the equilibrium $c^*_s$. I can, however, show the following (less-important) result.

**Theorem 1.** If $c^*_s$ is a convergently stable equilibrium strategy for the population game with utility function $U_{general}$, and diagnosis is negatively correlated to outcomes in the sense that $\frac{u - c_j}{\zeta + h} \leq \frac{u - c_i - c^*_s}{\delta + h}$, then $\frac{\partial c^*_s}{\partial \epsilon} \geq 0$

Proof. The sensitivity of an equilibrium to a parameter can be derived using the algebra of infinitesimals and can be expressed in terms of changes in the marginal rate of return (Bulow et al., 1985) --

$$
\frac{\partial c^*_s}{\partial \epsilon} = \frac{\partial^2 U}{\partial c_s \partial \epsilon} - \left(\frac{\partial^2 U}{\partial c^2_s} + \frac{\partial^2 U}{\partial c_s \partial \tau_s}\right)
$$

(B.2)

The denominator is positive whenever the game equilibrium is convergently stable (Eshel,
1983), so the sign of sensitivity is determined by the sign of the numerator,

$$\frac{\partial^2 U}{\partial c_s \partial \epsilon} = \frac{(\delta + h)(\mu + h) \lambda \left[ \left( \frac{u - c_i}{\zeta + h} - \frac{u - c_j - c_s}{\delta + h} \right) \sigma'(c_s) + \frac{\lambda \sigma^2(c_s)}{\delta + h} \right]}{(\delta + h + \epsilon)^2 (\lambda \sigma(c_s) + \mu + h)^2}$$

(B.3)

Since $\sigma'(c_s) \leq 0$, our theorem’s premise implies $\frac{\partial^2 U}{\partial c_s \partial \epsilon} \geq 0$. $\square$