Title
Essays on Behavior and Infectious Disease

Author
Carpenter, Anne Laura

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays on Behavior and Infectious Disease

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Anne Carpenter

Dissertation Committee:
Professor Michael T. McBride, Chair
Professor Stergios Skaperdas
Associate Professor Daniel Bogart

2016

DEDICATION

To My Husband

Because you got to class first every day just so that you could sit in the back. Because you answered all the in-class questions correctly but found that 0*1=1 on an exam. Because you always know what the answer to problems should look like, but you do not know the first step. Because you have always looked out for me from the very beginning. I love you.

To My Father

For sitting me down in the first grade and insisting that I could, in fact, understand math. Even today women are underrepresented in math oriented fields. Thank you for refusing to allow me to believe that my gender had anything to do with my mathematical abilities. Without you, I would have quit math in the first grade.

To My Mother

Because you are always there when I need somebody to talk to. Because your experience in graduate school paved the way for mine. Because you do not care what others think. Because you always overcome life's obstacles. Because you are the only person I know brave enough to try to get rocks past the TSA agents.

To PoPo

For all of the discussions about politics and finance. Those discussions encouraged my interest in economics. Also, thank you for your patience when you repeatedly tried to explain to me how the stock market works. I get it now.

To Granny Mary

Because you attended college as a woman in the 1940s. Because you started your own sorority when you decided that you did not like the sororities available. Because you hated to cook. Because you got a job as a teacher the day your sons jokingly told you that you couldnt work since you were "just a mom." Because you loved to read. Most importantly because you were never the cookie baking type of grandmother.

To Atlas and Lily

Without you, I would have been done with my dissertation 2 months ago.

In the words of the immortal Taylor Swift, "Haters gonna hate, hate, hate, hate, hate."

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the chair of my committee, Michael McBride, and the other members on my committee, Stergios Skaperdas and Daniel Bogart. Each of them has provided support from the idea stage through the analysis stage of this dissertation. I am grateful for the opportunity to pursue my interdisciplinary interests. I appreciate all of their helpful advice and encouragement.

I would like to thank Garret Ridinger for all of his suggestions and support.

I acknowledge and appreciate support from the Experimental Social Science Lab (ESSL) at the University of California Irvine, the Center for Economics and Public Policy at the University of California Irvine, the Economics Department at the University of California Irvine, and the Graduate Division at the University of California Irvine.

Permission to use copyrighted material has been granted by Springer International Publishing and was previously published as follows: Carpenter, A., 2014. Social Computing, Behavioral-Cultural Modeling and Prediction. Vol. 8393 of Lecture Notes in Computer Science. Springer International Publishing, Ch. Behavior in the Time of Cholera: Evidence from the 2008-2009 Cholera Outbreak in Zimbabwe, pp. 237244.

CURRICULUM VITAE

Anne Carpenter

EDUCATION

Doctor of Philosophy in Economics — 2016
University of California, Irvine

Bachelor of Science in Economics, Minor in Finance, Summa Cum Laude — 2009
Missouri State University

ACADEMIC POSITIONS

Part-time Faculty — 2015-2016
Orange Coast College

Teaching Assistant — 2011-2016
University of California, Irvine

Teaching Assistant — 2009
Missouri State University

PUBLICATIONS

Carpenter, Anne."Behavior in the Time of Cholera: Evidence from the 2008-2009 Cholera Outbreak in Zimbabwee." In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer International Publishing: pp.237-244. — 2014

CONFERENCE PRESENTATIONS

Experimental Social Science Workshop — 2015
University of California, Irvine

ESA North American Conference. — 2015
Dallas

ESA North American Conference — 2014
Ft. Lauderdale

Social Computing, Behavioral-Cultural Modeling, and Prediction Conference Poster Session. Washington D.C. — 2014

Graduate Student Poster Session, Department of Economics. — 2014
University of California, Irvine

Experimental Social Science Graduate Student Workshop. 2014
University of California, Irvine

Center for Economic Public Policy Donor Briefing. 2014
University of California, Irvine

Institute for Mathematics in the Behavioral Sciences Lunch Seminar. 2013
University of California, Irvine

Graduate Student Poster Session, Department of Economics. 2013
University of California, Irvine

Graduate Division Summer Research Conference. 2011
University of California, Irvine

ABSTRACT OF THE DISSERTATION

Essays on Behavior and Infectious Disease

By

Anne Carpenter

Doctor of Philosophy in Economics

University of California, Irvine, 2016

Professor Michael T. McBride, Chair

Although individuals have the potential to significantly impact the magnitude of epidemics, understanding individual responses to disease remains relatively understudied. One problem with studying behavioral responses to disease is obtaining individual level data on actions taken to protect against disease prior to, during, and after an outbreak. Using an evolutionary game with the replicator dynamic, I endogenize the decision to treat water in an epidemiological model of cholera. I calibrate the model to generate aggregate predictions for the cholera epidemic magnitude and the share of the population treating their water during the 2008-2009 Zimbabwe cholera epidemic. I show that the model captures both the 2008-2009 outbreak magnitude and the share of the population treating their water after the outbreak. To examine individual level disease prevention decisions, I use this model to simulate a disease environment in an economics laboratory experiment using students as human subjects. In the experiment, subjects are told that an infectious disease has been discovered in the environment. Subjects must decide whether to invest in a disease prevention technology. Subjects that choose to invest in the disease prevention technology are fully protected against infection. Subjects that choose not to invest in the disease prevention technology are at risk of infection. The payoffs from the experiment are taken from the epidemiological model of cholera with endogenous water treatment. I compare the model predictions with observed experimental data. In this way, I generate the individual level panel data necessary

to explore the impact of prevention costs, outbreak information, and social exclusion costs on the probability of investing in disease prevention.

Chapter 1

Behavior in the Time of Cholera: Evidence from the 2008-2009 Cholera Outbreak in Zimbabwe

Water transmitted diseases are the fifth leading cause of global mortality. According to World Health Organization (WHO) estimates, 94% of water transmitted diseases could be prevented with access to clean water and sanitation (World Health Organization, 2007). Despite the promising returns to investment in clean water and sanitation, individual use of disease prevention products, like water chlorination tablets, is low in developing countries (Dupas, 2011b). If on average individuals in developing countries were forced to spend their entire income on subsistence, then this would explain the low investment in disease prevention products. However, individuals living on less than $1 a day spend an average of 10%-18% of their income on alcohol, tobacco, and local festivals (Banerjee and Duflo, 2007). This suggests that even the very poor have some discretion in the way they spend their income. So, why is investment in water treatment products so low despite the potential for water transmitted disease? While many factors likely explain low water treatment levels, some

researchers have indicated that individuals may exhibit prevalence dependent behavior, and therefore, may base decisions to treat water on the incidence of water transmitted diseases in the environment (Geoffard and Philipson, 1997; Philipson, 2000).

This paper explores the dynamic relationship between cholera incidence and water treatment behavior. The hypothesis is that if individuals exhibit prevalence dependent water treatment behavior, then there should be an increase in the share of the population treating their water during a cholera outbreak. However, theoretical research on prevalence dependent behavior suggests that outbreak induced behavior change will not be sufficient to eliminate an outbreak (Geoffard and Philipson, 1997; Philipson, 2000). The key finding of this paper is that prevalence dependent water treatment behavior is a factor contributing to endemic cholera. Furthermore, in absence of the WHO interventions during the 2008-2009 cholera outbreak in Zimbabwe, the share of the population treating their water would have converged to a level that would have allowed a high cholera incidence to persist. Therefore, these results support theoretical model predictions regarding the inability of individuals exhibiting prevalence dependent behavior to coordinate their individual actions to eliminate disease.

1.1 Related Work

Microeconomic research has focused on the individual level barriers to water treatment that may explain low investment in water treatment products in developing countries. Through the use of randomized controlled trials factors such as information, credit constraints, and product pricing have been varied to explore the role of such factors in individual water treatment decisions (Madajewicz et al., 2007; Devota et al., 2012; Ashraf et al., 2007). However, this research typically ignores the fact that water treatment decisions take place in a dynamic disease environment. The idea that individuals might exhibit prevalence dependent behavior with regard to prevention of certain diseases has been demonstrated in several empirical

studies (Philipson, 2000). However, these studies do not explore the role of prevalence dependent behavior in the decision to treat water. Although some mathematical models have incorporated behavioral responses to disease (Geoffard and Philipson, 1996; Hyman and Li, 1997; Klein et al., 2007; Kremer, 1996; Auld, 2003; Valle et al., 2005), the primary focus of these models is sexually transmitted diseases, with a few exceptions. This model is the first of my knowledge to explore the dynamic relationship between water treatment behavior and cholera incidence.

Furthermore epidemiologists have developed mathematical models of cholera that incorporate biological realism in an effort to provide more accurate forecasts of epidemic magnitudes (Anderson and May, 1991; Andrews and Basu, 2011; Mukandavire et al., 2011). These models are used to explore potential epidemics both in the absence of and in the presence of potential interventions in order to determine the best policy response to an outbreak. However, these models assume that individual decisions, including water treatment decisions, are constant over time and exogenous to the disease environment.

This paper contributes to the literature in both economics and mathematical epidemiology by combining the behavioral focus of economic models with the biological focus of the mathematical epidemiology models. By modeling the decision to treat water as a function of the share of the cholera infected population, I endogenize the decision to treat water. This enables me to explore the influence of the dynamic disease environment on water treatment decisions and the ability of individuals to coordinate prevention measures to eliminate an outbreak.

1.2 Model

1.2.1 Water Treatment Game

To model changes in population water treatment behavior, I use an evolutionary game (Sandholm, 2010). In the game, agents have two possible strategies: treat water (T) or do not treat their water (NT). Table 1.1 provides the game.

Table 1.1: Water Treatment Game

	T	NT
T	$\beta - C$	$\beta - C(I(t))$
NT	$\beta - C(I(t))$	$-C(I(t))$

In this evolutionary game, α represents the share of the population not treating their water, and $(1 - \alpha)$ represents the share of the population treating their water. Here, β is the benefit from treating water, and C is the cost of treating water. The variable $I(t)$ represents the number of infected individuals which changes over time, t. Additionally, $C(I(t))$ is the cost of not treating water and is an increasing function of $I(t)$. Therefore, the cost of not treating water increases as the number of cholera infected individuals increases.

Because individuals' payoffs depend upon the aggregate share of individuals making each choice, this model has the nice property that it can account for the role that externalities might play in behavior change. In table 1.1, it is easy to see that if $C > C(I(t))$, then the game is a hawk-dove game. In this case, a portion of the population will free-ride on the water treating group in equilibrium. When cholera infection is unlikely, a larger portion of the population will not treat water in equilibrium. However, if $C < C(I(t))$, then the game is a coordination game where the dominant strategy is to treat water. Therefore, at higher infection levels, the positive externality imposed by the water treating group is not sufficient to allow free-riding to exist in the population.

The replicator dynamic presented in equation 1 captures the evolution of the population water treatment over time, t.

$$\frac{d\alpha(t)}{dt} = \alpha\left[\Pi(NT) - \bar{\Pi}\right]. \tag{1.1}$$

Here, α represents the share of the population not treating their water, $\Pi(NT)$ is the payoff from not treating water, and $\bar{\Pi}$ is the average payoff for all strategies. Equation 1 demonstrates that the change in the share of the population not treating their water moves according to the changes in the underlying payoffs. If the payoff from not treating water is larger than the average payoff, then the share of the population not treating their water will grow. Conversely, if the average payoff is larger than the payoff for not treating water, then the share of the population not treating their water will shrink. Equation 2 provides the final derivation for the change in the share of the population not treating their water (proof see S.I. 1.1).

$$\frac{d\alpha(t)}{dt} = \alpha(1-\alpha)\left\{(1-\alpha)\left[C - C(I)\right] - \alpha\beta\right\}. \tag{1.2}$$

1.2.2 Cholera Outbreak System

Following standard epidemiological practice, I use an SIR model to model the cholera outbreak (Anderson and May, 1991; Andrews and Basu, 2011; Mukandavire et al., 2011). The population can be divided into 3 categories: susceptible individuals, $S(t)$; infected individuals, $I(t)$; and recovered individuals with temporary cholera immunity, $R(t)$. The change in

the cholera susceptible population is

$$\frac{dS(t)}{dt} = \mu N + \omega R(t) - q\alpha(t)S(t)\frac{B(t)}{k + B(t)} - \mu S(t). \tag{1.3}$$

Here, N represents the entire population, μ represents the birth and deaths rates, μN represents the number of individuals born into the population, and $\mu S(t)$ the number of deaths in the susceptible population unrelated to cholera. In this model, the number of recovered individuals with temporary cholera immunity is $R(t)$, ω is the rate at which people lose their temporary immunity, and $\omega R(t)$ is the number of individuals that lose their temporary immunity and again become susceptible to cholera. Here, k is the level of bacteria in the environment necessary for a 50% probability of cholera infection, $B(t)$ is the level of bacteria that exists in the aquatic environment, and $\frac{B(t)}{k+B(t)}$ is the probability that consumption of cholera contaminated water leads to cholera infection (Codeco, 2001).

Standard models of cholera outbreak assume the fraction of the population that consumes contaminated water, $q\alpha(t)$, to be a constant. To incorporate changes in water treatment behavior, I break up the consumption of cholera contaminated water into the share of the population not treating their water, $\alpha(t)$, times the probability that untreated water is cholera contaminated, q. This model assumes that the probability that untreated water is cholera contaminated is constant and exogenous. Although q is assumed to be constant, it is possible that the infection rate influences the probability that untreated water is cholera contaminated. Unfortunately, no data exists at this time to explore this relationship. In this model, the assumption of a constant and exogenous q provides a considerable simplification of the model analysis; however, future research should seek to determine how the infection rate influences the probability that untreated water is cholera contaminated. It follows from the variable definitions that the product $q\alpha(t)S(t)\frac{B(t)}{k+B(t)}$ is the number of susceptible individuals

that become cholera infected in time, t. The change in population infected with cholera is

$$\frac{dI(t)}{dt} = \alpha(t) * q * S(t)\frac{B(t)}{k + B(t)} - (\mu_c + \mu + \gamma)I(t) \tag{1.4}$$

Here, μ_c is the cholera death rate, γ is the cholera recovery rate, $\mu_c I(t)$ is the number of infected individuals that die from cholera, $\gamma I(t)$ is the number of infected individuals that recover and receive temporary immunity to cholera, and $\mu I(t)$ is the number of infected individuals that die from causes unrelated to cholera. The change in the recovered and temporarily cholera immune population is

$$\frac{dR(t)}{dt} = \gamma I(t) - (\mu + \omega)R(t). \tag{1.5}$$

The change in V. cholerae bacteria in the aquatic reservoir is

$$\frac{dB(t)}{dt} = \theta I(t) - \delta B(t). \tag{1.6}$$

Here, θ is the rate at which infected individuals shed V. cholerae bacteria into the environment through defecation, δ is the death rate of V. cholerae bacteria in the aquatic environment, $\theta I(t)$ is the number of bacteria that are shed into the aquatic environment, and $\delta B(t)$ is the number of bacteria in the aquatic environment that die. The water treatment behavior discussed in section 3.1 is the new dynamic being explored in this paper. Equation 2 provides the change in the population share not treating their water. For simplicity, I assume that

$C(I(t)) = \lambda \frac{I(t)}{N}$, a linear function of the infection rate. Here, $\lambda \epsilon [0, \infty)$ is a scalar that can be interpreted as the intensity with which the outbreak is perceived. Therefore, the change in the population share not treating their water is

$$\frac{d\alpha(t)}{dt} = \alpha(1-\alpha)\left\{(1-\alpha)\left[C - \lambda\frac{I(t)}{N}\right] - \alpha\beta\right\}. \tag{1.7}$$

1.3 Selected Calibration

To test the ability of the model to fit water treatment behavior during a cholera outbreak, I calibrated the model to the 2008-2009 cholera outbreak in Zimbabwe. To do so I use data on cholera incidence, outbreak induced changes in water treatment behavior, and cholera biology (see S.I. 2.1 for data description).

The calibrations for Mashonaland West illustrate the ability of the model to fit observe cholera incidence and water treatment behavior (see S.I. 2.2 for all provinces). In figure 1.1, the first graph provides the observed weekly cholera incidence and the calibrated incidence for the same period. The second graph demonstrates that the model can capture changes from the pre-outbreak behavior to the 2010 post-outbreak behavior.

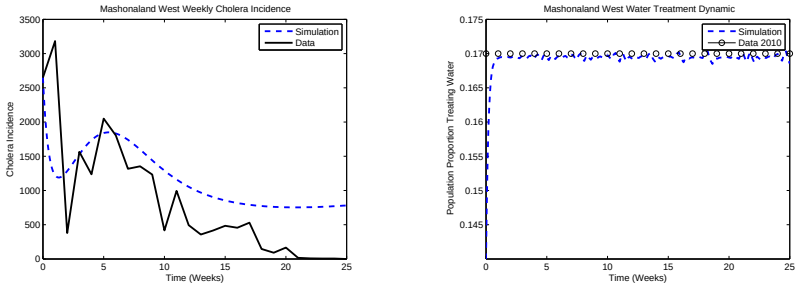


Figure 1.1: Mashonaland West Cholera Incidence and Water Treatment Behavior: Model Fit vs. Observed Data

To calibrate the model, I chose the parameters β, C, and λ that best fit the observed data (for parameters see S.I. 2.2). I vary the costs and benefits for each province separately to determine calibration sensitivity to parameters chosen (see S.I. 3). I find that when the costs of water treatment are higher than the perceived benefits, the share of the population treating their water converges to a level that allows cholera to persist in the environment. However, when the costs of water treatment are sufficiently low relative to the benefits, cholera will be eliminated due to a greater increase in water treatment within the population. Prior research has explored the ability of V. cholerae bacteria to survive in the aquatic reservoir as an explanation for endemic cholera. This paper demonstrates that even when incorporating the persistence of bacteria in the aquatic environment, a cholera outbreak will be minimal if the costs of water treatment are sufficiently low. Thus, an explanation of endemic cholera should include prevalence dependent water treatment behavior as a contributing factor.

A drop in cholera incidence occurs in the data around week 7 that is not matched by the model predictions. However, around week 7, the WHO had partnered with multiple organizations working in Zimbabwe to provide clean water, water purification tablets, and cholera prevention information free of charge (UN OCHA, 2009b,a). This intervention effectively reduced the cost of water treatment to zero, which allowed benefits of water treatment to outweigh costs of treatment thereby ending the outbreak. Thus, the model prediction should be viewed as the level of cholera incidence that would have persisted in absence of the intervention. This model suggests that because the costs of water treatment were higher than the perceived benefits, the increases in cholera incidence would not have induced a sufficient share of the population to treat their water. Therefore the outbreak would have persisted. This finding is consistent with theoretical research that suggests that if individuals exhibit prevalence dependent behavior, then they will be unable to coordinate individual actions to eliminate outbreaks (Geoffard and Philipson, 1997).

1.4 Discussion

This paper explores the issue of low water treatment in developing countries considering the disease environment in which these choices take place. I find that high water treatment costs and low perceived water treatment benefits cause the share of the population treating their water to be low despite the persistence of cholera. Model calibrations of the 2008-2009 cholera outbreak in Zimbabwe demonstrate two important implications of this modeling approach. First, cholera incidence can be eliminated with low water treatment costs relative to perceived benefits despite the presence of V. cholerae bacteria in the aquatic environment. Second, without the WHO interventions in Zimbabwe during the 2008-2009 cholera outbreak, the share of the population treating their water would have converged to a level that would have enabled a high cholera incidence to persist.

Previous research has focused on the ability of V. cholerae bacteria to survive in the aquatic environment as an explanatory factor in endemic cholera (Codeco, 2001). However, this paper demonstrates that cholera incidence can be eliminated when the costs of water treatment are sufficiently low relative to the benefits of water treatment regardless of the level of V. cholerae bacteria in the aquatic environment. Thus, prevalence dependent water treatment behavior coupled with high costs of water treatment can allow the population to converge to a level of water treatment that enables cholera to persist. Therefore, policy focused on eliminating endemic cholera should not only consider persistence of V. cholerae bacteria in the aquatic environment, but also should acknowledge that changes in water treatment behavior are likely prevalence dependent. Furthermore, policy makers should also consider that it may be very difficult, if not impossible, for individuals acting independently to coordinate actions to eliminate water transmitted diseases.

During the 2008-2009 cholera outbreak in Zimbabwe, the Zimbabwean government was reluctant to acknowledge or act to end the cholera epidemic (UN OCHA, 2009a). Fortunately,

the WHO acted to coordinate actions to provide clean water, water purification tablets, and cholera prevention information free of charge to individuals in affected regions (UN OCHA, 2009a). This model demonstrates that although individuals may respond to outbreaks by changing outbreak related behaviors, the changes in behavior may not be sufficient to eliminate the outbreak. In the case of Zimbabwe, without the interventions by the WHO, the share of the population treating their water would have converged to a level that would have enabled a high cholera incidence to persist. Therefore, policy makers should not rely on individual behavioral responses to end epidemics. Instead, there is significant scope for policy makers to aid individuals in coordinating actions to eliminate outbreaks.

While this model provides a first pass at incorporating outbreak related behavior into a model of cholera outbreak, it does not provide a full biological analysis of a cholera epidemic. Future research should seek to incorporate more biological realness into the model of cholera incidence to provide better analysis of potential epidemic magnitudes. Specifically, incorporating the highly infectious cholera state and the associated health behavior, post-defecation hand-washing, will provide a more thorough analysis of the impacts of behavior change on cholera epidemics.

Chapter 2

Modeling Behavior During Epidemics: Understanding Behavioral Responses to Disease

2.1 Introduction

It seems likely that individuals respond to disease outbreaks by taking various measures to prevent themselves and others from contracting a particular disease. These measures have been reported anecdotally in the media during many different epidemics: the wearing of face masks during the 2003 SARS epidemic in China, the avoidance and ultimate closure of public places during the 2009 H1N1 epidemic in Mexico, and the self-imposed quarantines of returning U.S. health workers during the 2014 Ebola epidemic. If individuals engage in self-protective or other-protective behaviors during epidemics, then epidemiological models that ignore behavioral responses to disease could overestimate the epidemic magnitude. However, individuals could exhibit behavioral biases that prevent them from accurately assessing their

own risk during an epidemic. If epidemiological models ignore the unwillingness to engage in self or other protective behaviors, then predictions of epidemic magnitude could be underestimated. Despite the potential impacts of such changes in behavior on disease dynamics, incorporating endogenous behavioral responses into mathematic models of disease is a relatively new area of study (Klein et al. (2007); Ferguson (2007); Carpenter (2014); Funk et al. (2015)).

One potential reason that incorporating individual responses to disease into epidemiological models is relatively new is the lack of individual level data on epidemic induced behavior. The ideal dataset would be available at the individual level, and it would include all potential actions taken by individuals to protect themselves and others from illness. We would need data both prior to an outbreak and during an outbreak to determine how the outbreak changed behavior. Furthermore, we would like data collection to extend beyond the cessation of the outbreak to determine the persistence of behavioral responses to disease. While it is useful to think about the ideal dataset for the study of behavior during epidemics, collection of such data is typically not the primary goal of public health officials during an outbreak. Instead, public health efforts typically focus energy and resources on preventing the spread of disease through information and disease prevention products. So, although data on behavior during epidemics may be available in some instances, it typically is not the ideal individual level panel including all disease prevention behaviors over time. This makes it difficult to test models incorporating endogenous behavioral responses to disease against actual individual behavior during epidemics (Funk et al. (2010); Manfredi and d'Onofrio (2013); Carpenter (2014); Wang et al. (2015)).

Researchers engaged in modeling endogenous behavioral responses to disease recognize this challenge and have provided some solutions to test model assumptions for behavior against observed behavior during epidemics. First, some researchers have suggested testing the models against data on historical epidemics when such data does exist (Ferguson (2007)).

Second, some researchers have proposed using non-traditional sources such as digital media to gather information on individual behavior during epidemics (Funk et al. (2015)). While both of these methodologies can provide useful data for the testing of model assumptions on individual behavior, the non experimental nature of this data means that researchers may not always be able to obtain the specific types of data desired. I provide an additional method for obtaining data on epidemic induced individual behavior: a laboratory experiment.

Conducting a laboratory experiment provides advantages over traditional data collection, because it enables researchers to obtain data that directly answers their research questions with minimal cost. Using an experiment to study disease induced behavior allows researchers to randomize exposure to different policy treatments, capture individual level data on behavior over time, and collect detailed demographic information on experiment participants. While the laboratory experiment does provide benefits over other types of data collection, a potential drawback of the method is the issue of external validity. It is unclear whether a simulated disease environment induces the same behavioral responses as would be observed in an actual disease environment. In this paper, I use a laboratory experiment to create a simulated disease environment in which to test an epidemiological model with endogenous behavioral response to disease. I address the issue of external validity by comparing individual behavior in the experiment with reported real world engagement in disease prevention behaviors to determine whether the abstract laboratory environment captures real world attitudes towards disease prevention.

I use a modified version of the evolutionary game presented in Carpenter (2014) to endogenize disease prevention behavior in a mathematical model of disease. Existing research suggests that demand for disease prevention is prevalence elastic (Geoffard and Philipson (1996), Geoffard and Philipson (1997), Philipson (2000), Ahituv et al. (1996), Philipson (1996), Goldstein et al. (1996), and Philipson (2000)). Prevalence elastic demand is characterized by an increase in demand for disease prevention technologies when disease incidence increases and

a decrease in demand for disease prevention technologies when disease incidence decreases. In this model, the dynamic for behavior is derived from underlying individual payoffs which are a function of the infection level. In this way, I endogenize behavioral response to disease by incorporating current research on prevalence elastic demand for disease prevention. I calibrate the model in the experiment using parameters taken from the infectious disease modeling literature. Building on the experimental work of Chen et al. (2013), experiment subjects are framed to consider an infectious disease environment, and they are faced with a decision each round: invest in disease prevention or do not invest in disease prevention. The payoffs from each choice are taken from the evolutionary game and investing in disease prevention provides full protection against disease. By comparing choices made by subjects in the experiment with model predictions for investment in disease prevention, I test the model predictions for investment in disease prevention.

There are two primary hypotheses from this model. First, I expect subjects to exhibit prevalence elastic demand for disease prevention. This hypothesis is an assumption of the mathematical model, but it is based on the limited theoretical and empirical work on the behavioral responses to disease outbreaks. Second, I expect that when the cost of investing in disease prevention is sufficiently low, investment in disease prevention will be high enough to eliminate an outbreak. This hypothesis is a direct prediction of the mathematical model of disease with endogenous investment in disease prevention. I find support for both hypotheses: subjects exhibit prevalence elastic demand for disease prevention and investment in disease prevention is high enough to eliminate an outbreak when the cost of investing is sufficiently low. Although I do find that subjects exhibit prevalence elastic demand for disease prevention, I find that this effect diminishes as infection levels rise. Furthermore, I find that there is persistence in the impact of disease incidence on investment in disease prevention. Interestingly, I find that being infected in the previous round significantly reduces the probability that an individual invests in disease prevention. To address the issue of external validity, I compare subjects reported real-world engagement in disease prevention

behaviors with behavior in the experiment. I find that subjects who report free-riding on the disease prevention investments of others in reality are significantly less likely to invest in disease prevention in the experiment. This suggests that subjects behavior in the experiment is correlated with real world disease prevention behavior.

This paper proceeds as follows: Section 2 provides the model used in the experiment. Section 3 describes the experimental design, Section 4 provides the results, and Section 5 concludes.

2.2 Epidemiological Model

2.2.1 Model

The purpose of this model is twofold: to endogenize individual responses to disease in an SIR model in a manner consistent with evidence on behavior during disease outbreaks and to develop a model that captures aggregate investment in disease prevention. In other words, the goal here is to use information on individual decisions during an outbreak to develop a model that captures aggregate investment in disease prevention. While it seems reasonable that individuals may exhibit sophisticated behavior regarding investment in disease prevention, it is not clear how sophisticated a model of individual behavior needs to be in order to capture aggregate investment in disease prevention.

While behavior could be endogenized using N-player game with the standard Nash solution concepts, there are a couple of problems with this approach. First, the susceptible population for infectious diseases is usually large meaning that computing the strategies for each individual player is prohibitively time consuming without necessarily adding insight into the aggregate effects of individual behavior. Second, standard game theory assumes that individuals have full knowledge of the game and the equilibrium. Players in a standard game

are assumed to know the strategies available, the payoffs from every strategy profile, and to correctly anticipate the actions of all other players in the game. In the context of disease, this means an individual would have full knowledge of the disease dynamic, and the individual would know the prevention strategies taken by all individuals susceptible to a disease. Even epidemiologists do not have full knowledge of all the factors that influence disease dynamics; instead, they approximate the underlying dynamic using SIR models. Furthermore, individuals do not know every person who is susceptible to a particular disease let alone every action taken by each susceptible person to prevent a disease. So, it is unreasonable to assume that they can correctly anticipate the actions of all the players in the game. While these assumptions may be reasonable in smaller games, they require a high level of knowledge that individuals generally do not possess in a disease context.

To avoid unreasonable assumptions on players' knowledge in the game, I use an evolutionary game with the replicator dynamic to endogenize individual responses to disease. Evolutionary games were developed to incorporate strategic interaction in large populations where traditional game theoretic assumptions may be too strong. The replicator dynamic requires two assumptions: an infinite population and random interaction between individual players in the game. This particular dynamic determines how the share of the population using a strategy evolves over time, but it does not provide information about the way in which individual players make decisions. Although the replicator dynamic does not explicitly state how the players determine their strategies, it can be derived from several types of imitative behavior.

Using an evolutionary game with the replicator dynamic efficiently captures aggregate behavior in large populations, and it relaxes the assumptions regarding individual knowledge of the game and the equilibrium. Since the population is assumed to be infinite with the replicator dynamic, the simple dynamic captures aggregate strategies without requiring the computation of each individual strategy, which is cumbersome in large populations. For

populations that are large but finite, the replicator dynamic can be seen as an approximation of the aggregate strategies. This dynamic does not require individual players to have knowledge of the game or knowledge of the equilibrium; instead, individual players do not need to have any information about the game. Therefore, the replicator dynamic determines how the share of the population engaged in a strategy evolves over time, but it does not provide information about the way in which individual players make decisions. The dynamic simply captures the fact that a given strategy will grow in the population when the payoff from that strategy is higher than the average payoff from all strategies.

A drawback of the replicator dynamic is that individual players are not modeled as exhibiting sophisticated behavior. In addition to knowledge of the game and knowledge of the equilibrium, information regarding the efficacy of each strategy is not important for individual decisions in the game. While individuals may exhibit more sophisticated behavior like using information on strategy efficacy to make their decisions, this dynamic does not require this. While this is a drawback of modeling individual behavior, it is not clear whether it is necessary to model individuals as sophisticated decision makers to capture the share of the population engaging in each strategy in aggregate. In other words, using the replicator dynamic enables me to test whether it is necessary to model sophisticated individual behavior to capture aggregate investment in disease prevention.

In this paper, I use a model based on the model presented in Carpenter (2014) where the dynamic for aggregate investment in disease prevention is derived from individual payoffs that underlie the decision to invest in disease prevention. While some models of disease have started to incorporate endogenous individual response to disease, theory is largely unclear on the ways in which individuals might respond to disease. One channel through which disease has been shown to alter behavior is through the disease incidence or prevalence elastic demand for disease prevention. Prevalence elastic demand for disease prevention is characterized by an increase in demand for disease prevention as disease incidence increases

and a decrease in demand for disease prevention as disease incidence decreases. While there are many different ways that this behavior could be endogenized, this model assumes that individuals change their behavior in response to their payoffs which are influenced by the infection level. In this way, I incorporate individual response to disease in a manner consistent with the evidence that suggests individuals exhibit prevalence elastic demand for disease prevention.

I modified the Carpenter (2014) model to simplify the experimental design and analysis. Due to the limited size of the experimental lab, it was necessary to have all subjects making a disease prevention investment decisions each round. Three simplifications of the original model allow the same subjects to make decisions each round for the duration of the experiment. First, I omit the possibility of a natural birth and death cycle in the population. Second, I remove the possibility that an infection could "kill" an infected individual. Instead, subjects can be infected but infection does not cause subjects to leave the population through death. These two simplifications ensure that the population of subjects is fixed throughout the experiment. Finally, although many diseases confer temporary immunity, I remove the possibility of temporary immunity to the disease. This enables subjects to make decisions each round thereby increasing the amount of data I was able to collect. Future research can and should explore the impacts of changes in the population, infection induced removal from the experiment, and infection induced immunity on the likelihood of investment in disease prevention. Equations 1-4 present the modified version of Carpenter (2014).

$$\frac{dS(t)}{dt} = I(t) - q\alpha(t)S(t)\frac{B(t)}{k + B(t)} \tag{2.1}$$

The change in the susceptible population, $\frac{dS(t)}{dt}$, is provided in equation 1. The susceptible population, $S(t)$, is increased by the number of individuals that were disease infected, $I(t)$,

and have recovered. The susceptible population is decreased by the number of individuals that become disease infected, $q\alpha(t)S(t)\frac{B(t)}{k+B(t)}$. The share of the population not investing in disease prevention is $\alpha(t)$, and the probability that not investing in disease prevention leads to disease exposure is q. The quantity $q\alpha(t)$ is the share of the population that is actually exposed to the disease. This exposure occurs through consumption of bacteria, but it does not necessarily result in infection. The probability that consuming bacteria results in infection depends upon the amount of bacteria consumed. The number of bacteria that must be consumed to give a 50% probability of infection is given by k, and $B(t)$ is the number of bacteria in the environment. Thus, the value $\frac{B(t)}{k+B(t)}$ is the time dependent probability of infection. So, the quantity $q\alpha(t)\frac{B(t)}{k+B(t)}$ is the share of the population that becomes disease infected.

$$\frac{dI(t)}{dt} = q\alpha(t)S(t)\frac{B(t)}{k+B(t)} - I(t) \tag{2.2}$$

The change in the infected population, $\frac{dI(t)}{dt}$, is given by equation 2. The infected population is increased by the number of individuals that become disease infected, $q\alpha(t)S(t)\frac{B(t)}{k+B(t)}$. These infected individuals are removed from the susceptible population and included in the infected population. The infected population is decreased by the number of individuals that were disease infected, $I(t)$, and have recovered. These recovered individuals are removed from the infected population and included in the susceptible population.

$$\frac{dB(t)}{dt} = \theta I(t) - \delta B(t) \tag{2.3}$$

The change in the bacteria population in the environment, $\frac{dB(t)}{dt}$, is displayed in equation

3. The rate at which infected individuals shed bacteria into the environment is given by θ. The population of bacteria in the environment is increased by the number of bacteria shed by infected individuals, $\theta I(t)$. The death rate of bacteria in the environment is δ. The population of bacteria in the environment is decreased by the number of bacteria that die in the environment, $\delta B(t)$.

$$\frac{d\alpha(t)}{dt} = \alpha(t)(1-\alpha(t))\left[c - c(I(t)) - \alpha(t)b\right] \tag{2.4}$$

The change in the share of the population not investing in disease prevention, $\frac{d\alpha(t)}{dt}$, is shown in equation 4. The cost of investing in disease prevention is given by c, and investing in prevention provides full protection against disease. The benefit from avoiding infection is represented by b. The cost of not investing in disease prevention is $c(I(t))$. In this paper, $c(I(t))$ is assumed to be a linear function of the infection level. For a description of the underlying evolutionary game and a step-by-step derivation of the dynamic see appendix A.1.

2.2.2 Model Calibrations

To generate model predictions, I calibrate the model using the biological parameters given in table 2.1. Since the original model in Carpenter (2014) is a model of cholera incidence, the biological parameters used here represent the biology of a cholera epidemic. Although the biological parameters of a cholera outbreak typically fall within a range of values, I assume the biological parameters for the SIR model are fixed at a value that falls within the biological range. I make this assumption, because I am interested in individual behavior during an outbreak and not the biology of the outbreak.

Table 2.1: Model Parameters

	Treatments		
Parameters	High Cost	Low Cost	Footnotes
N	40	40	1
k	10^3	10^3	2
q	.99	.99	3
θ	70	70	3
δ	$\frac{7}{30}$	$\frac{7}{30}$	3
Start Values			
S(t)	40	40	1
$B(t)$	10^3	10^3	2
α(t)	0.525	0.1125	4
$I(t)$	0	0	5
Payoff Values			
b	$\frac{10}{10}$	$\frac{10}{10}$	6
c	$\frac{20}{10}$	$\frac{1}{10}$	6
c(I(t))	$\frac{I(t)}{10}$	$\frac{I(t)}{10}$	6

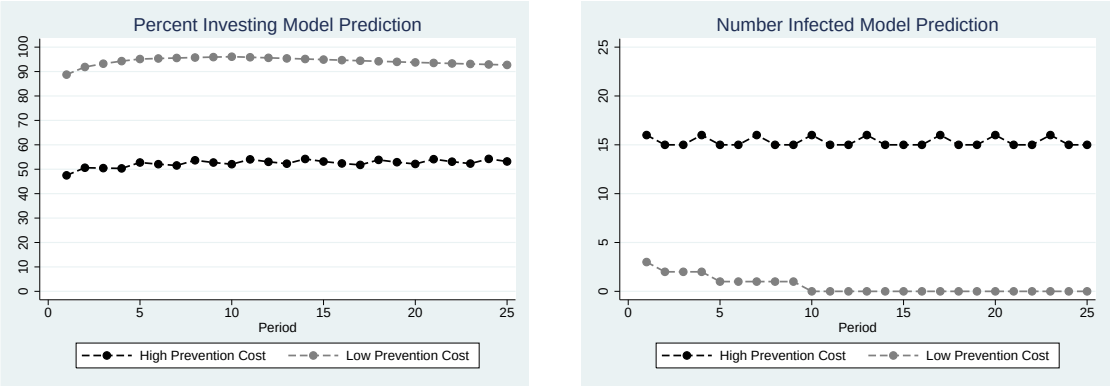
1. The number of subjects that participated in each experiment session. 2. These numbers give a high probility of infection if not investing to induce subjects to care about infection. 3. Based on models of cholera outbreak taken from Carpenter (2014). 4. Based on subjects initial experiment investment level. 5. Nobody was initially infected with the disease. 6. Experimental units were converted to money at a rate of 10 experimental units per 1 USD

The predictions for individual behavior are derived by varying the parameters that influence the payoffs for the two available strategies: invest in prevention and do not invest in prevention. Specifically, I explore the role of the cost of investing in prevention in the decision to invest in disease prevention. Policy makers typically attempt to change prevention investment behavior during an outbreak by changing the cost of investment. For example, during a cholera outbreak public health officials might intervene by providing clean water free of charge to individuals in the affected area. Policy makers are less likely to to be able to influence the benefit from prevention investment during an outbreak since the benefits from disease prevention products are typically fixed. For this reason, I fix the benefit from investing in disease prevention and vary the cost of investing in disease prevention.

Table 2.1 provides the parameters used to calibrate the payoffs for the evolutionary game under different scenarios. The only difference between the two calibrations occurs for the value c. The value c represents the cost of investing in disease prevention. The high cost column provides the model parameter values used to calibrate the model when the cost of investing in disease prevention is high. The low cost column provides the model parameter values used to calibrate the model when the cost of investing in disease prevention is low. Thus, the two different model calibrations provide predictions for prevention investment and disease incidence under different costs of prevention investment.

Figure 2.1 gives the model predictions when the model is calibrated using different costs of investing in disease prevention. It is easy to see that the model predicts that investment in disease prevention will be high and disease incidence low under low costs of investing in disease prevention. Conversely, under high costs of investing in disease prevention, the model indicates that investment in disease prevention will be low and disease incidence high. Using the model calibrations, I generate predictions for individual behavior in an infectious disease laboratory experiment.

Figure 2.1: Model Predictions Under Different Costs of Prevention Investment



2.2.3 Hypotheses

I identified 2 hypotheses to test in this experiment based on the calibrated SIR model.

Hypothesis 1. *High costs of prevention investment will have a significant negative impact on the probability that an individual invests in disease prevention.*

This hypothesis comes from the difference in the percent of the population investing in disease prevention when the cost of investing in prevention is high rather than low. When calibrating the model using high costs of prevention investment, the model predicts that approximately 50% of the population will invest in disease prevention. However, when calibrating the model using low costs of prevention investment, the model predicts that approximately 90% of the population will invest in disease prevention. Thus, on an individual level, I expect that experiencing high costs of prevention investment will significantly reduce the probability that an individual invests in disease prevention.

Hypothesis 2. *Subjects will exhibit prevalence elastic demand for disease prevention.*

The model presented in section 2.2.1 assumes that subjects will exhibit prevalence elastic demand for disease prevention, a behavior that has been observed empirically. Prevalence elastic demand for disease prevention products is characterized by an increase in demand for prevention products as disease incidence increases and a decrease in demand for prevention products as disease incidence decreases. One purpose of the following experiment is to test the model assumption that subjects exhibit prevalence elastic demand for disease prevention.

2.3 Experiment Design

2.3.1 Basics

I conducted two experiment sessions in a computer laboratory in a large public university with undergraduate students as human subjects. Students learned of the experimental subject pool through classroom advertisements. Students registered to be a part of the subject pool through an online registration system. Several days before each session an email was sent to randomly selected subjects in the pool notifying them of the upcoming experiment session. Students interested in participating were instructed to register for a specific session in the online registration system. Those who signed up received an e-mail reminder both the day before and the day of the planned session. Subjects were not allowed to participate in more than one session.

The susceptible population size for many diseases is typically very large. Unfortunately, the lab in this is experiment is much smaller than the susceptible population for most diseases with a total of 40 computers. To obtain the largest possible population size for this experiment, I had to ensure that the lab was filled to capacity during each experimental session.

In order to ensure the computer lab was filled to the maximum capacity of 40 subjects, more than 40 subjects were recruited for each session. Subjects received $7 for showing up regardless of their participation in the experiment. Subjects that showed-up but were not needed for a given session were paid the show-up fee and told they were eligible to participate in future sessions. In this way, I obtained the largest possible population size given the laboratory's computer constraints.

2.3.2 Design

To test the two hypotheses presented in section 2.2.3, I conducted a within subject experiment. The treatment variable was the cost of investing in disease prevention (cost of prevention). The cost of prevention variable took on two values: high cost and low cost. Subjects participated in a total of 50 experiment rounds. In the first session, subjects participated in the high cost treatment for the first 25 rounds and participated in the low cost treatment for the last 25 rounds. To control for order effects, I conducted an additional session where the order of exposure to the treatment variables was reversed. In the second session, subjects participated in the low cost treatment for the first 25 rounds and participated in the high cost treatment for the last 25 rounds.

I used the calibrated SIS model presented in section 2 to generate the disease outbreak dynamic. To program the SIS model presented in section 2.2 and conduct the experiment, I used the z-tree software package (Fischbacher (2007)). Since the maximum susceptible population size was 40 subjects and I conducted a total of 2 sessions, I had a total of 80 participants in the experiment. Each subject participated in a disease environment with 40 potentially susceptible individuals; so, each experiment session was its own independent disease environment and, therefore, outbreak. The parameters presented in Table 2.1 were used to calibrate the SIS model at the start of each experiment session.

Although the model and parameters used were consistent with a model of a cholera outbreak, the instructions presented the disease environment in general terms such that the disease could have been any infectious disease. At the start of the experiment, subjects were told that there was the possibility of contracting an "infectious disease". They were informed that there was a "disease prevention technology" that would provide full protection against disease. Subjects were told that each round they would choose whether to "invest" in the "disease prevention technology." Subjects were instructed that if they chose not to invest in the disease prevention technology, then they would be susceptible to infection. Screen shots of the full experiment, including instructions, are available in the appendix A.4.

Each round, subjects chose between investing in a disease prevention technology and not investing in a disease prevention technology. Subjects had 30 seconds to make a decision in the first two rounds of each of the two treatments. Subjects were given 15 seconds to make a decision for the remaining 23 rounds. If a subject failed to make the decision within the allotted time, then her default choice was "do not invest" in disease prevention.

Subjects' investment decisions were fed into the SIS model equation for $\alpha(t)$ to generate an aggregate infection level. The aggregate infection level was used to determine the number of subjects that were selected as "infected." Only subjects that did not invest in disease prevention during the round could be selected as "infected." Since not all subjects that chose not to invest in disease prevention became "infected," a random number generator was used to select "infected" subjects. Infection lasted only for the current round after which subjects were again susceptible to infection.

After all subjects had made their decisions, subjects received information about the round in the form of a results screen. The results screen provided subjects with their own decision, their own infection status, their own round payoff, the total number of subjects infected, the population percentage investing in disease prevention, the payoff from investing in disease prevention, the payoff from not investing in disease prevention and not becoming infected,

Table 2.2: Experimental Payoffs- Variables

	Not Infected	Infected
Invest	E+b-c	—
Do Not Invest	E+b-c(I(t))	E-c(I(t))

Table 2.3: Experimental Payoffs- Values

	Not Infected	Infected
Invest	40+10-20(1)	—
Do Not Invest	40+10-I(t)	40-I(t)

and the payoff from not investing in disease prevention and becoming infected. Beginning in round 2, subjects were provided with a history box that provided the information from the results screen. It is important to note that subjects were not provided with their objective probability of infection in any of the treatments. This feature of the experiment mimics the environment in which individuals make disease prevention investment decisions in reality. Individuals do not know their objective probability of infection for any particular disease; instead, they form a belief about their probability of infection. This is capture by the information environment in this experiment.

Decisions in the experiment were incentivized to encourage subjects to seriously consider their choices. Tables 2.2 and 2.3 provide the experimental payoffs. A subject's payoff in the round was jointly determined by their own strategy and by their own infection status. The left hand side (LHS) of the tables indicates the two potential strategies: invest in disease prevention or do not invest in disease prevention. The column headers indicate the potential outcomes for each strategy: not infected with disease or infected with disease. A subject's payoff in the experiment was based off the evolutionary disease prevention game and the subject's own endowment, E. Each round every subject was endowed with 40 experimental units to prevent them from earning a negative payoff.

Since investment in disease prevention provides full protection against disease, subjects that invest in disease prevention can not become infected. If a subject chose to invest in disease prevention, she receive a benefit of avoiding infection, b, of 10 experimental units. The cost of investing in disease prevention, c, was 20 experimental units in the high cost treatment and 1 experimental unit in the low cost treatment. In table 2.3, the cost of investing in

disease prevention in the low cost treatment is given in parentheses.

If a subject chose not to invest in disease prevention, she was susceptible to infection. If a subject chose not to invest in disease prevention and she was not infected, then she received a benefit of avoiding infection, b, of 10 experimental units. If a subject did not choose to invest in disease prevention and she was infected, then she received no benefit. All subjects that did not invest in disease prevention paid a cost of not investing in disease prevention, $c(I(t))$, equal to the number of infected individuals, $I(t)$, in the round.

I interpret the cost of not investing in disease prevention differently based upon a subject's infection status. For the subjects selected for infection, this cost of not investing in the disease prevention technology is the cost of becoming infected. In reality, this would include missing time from work or school, going to the doctor, and paying for medication. For the subjects not selected for infection, this cost of not investing in the disease prevention technology is the cost of social exclusion. Individuals in the real world arguably pay costs for choosing not to invest in disease prevention regardless of whether they become infected. These costs can include increased time spent monitoring the disease, avoidance of places most likely to increase the chance of infection, and being banned from public places. Because such factors are difficult to capture in this environment, I simply assume that subjects choosing not to invest in disease prevention pay a cost regardless of their infection status.

Subject payments were based on the the disease experiment and an incentivized lottery conducted after the disease experiment. One round from the high cost treatment and one round from the low cost treatment were randomly selected for payment for the disease experiment. Payoffs in the rounds were measured in experimental units which were converted to dollars at a rate of 10 experimental units per 1 USD. At the end of the 50 rounds, subjects were asked to participate in a simple incentivized lottery of their choice to measure risk preferences. Following Caldara (2013) who modified the approach used by Eckel and Grossman (2008), each of three lotteries corresponded to a type of risk preferences: risk

loving, risk neutral, or risk averse. Subjects chose one of three lotteries allowing for a rough elicitation of risk preferences. The incentivized lottery was also conducted in experimental units which were converted at a rate of 2 experimental units per 1 USD. Subjects were aware of conversion rates prior to their decisions. The payouts for the rounds and lottery were added to the $7 show-up fee for their total payment. The total payments were rounded up to the nearest quarter. Subjects received on average $17.50 for 60 minutes of participation.

2.4 Results

2.4.1 Descriptive Statistics

Table 2.4 provides general descriptive statistics for subjects in the experiment. There were more female subjects than male subjects in the overall experiment; however, being female is not significant in explaining behavior in the experiment. There were more subjects of Asian decent than subjects of other races which reflects the general student population at the large public university where this experiment was conducted. Similarly, being of Asian decent was not significant in explaining behavior in the experiment.

Table 2.4: Descriptive Statistics

	Statistics
Male	30%
Female	70%
White (not hispanic)	9%
Hispanic	24%
Asian	60%
Black	1%
Mixed Race	4%
Undisclosed	2%
Average Payment	$17.30
Number of Subjects	80

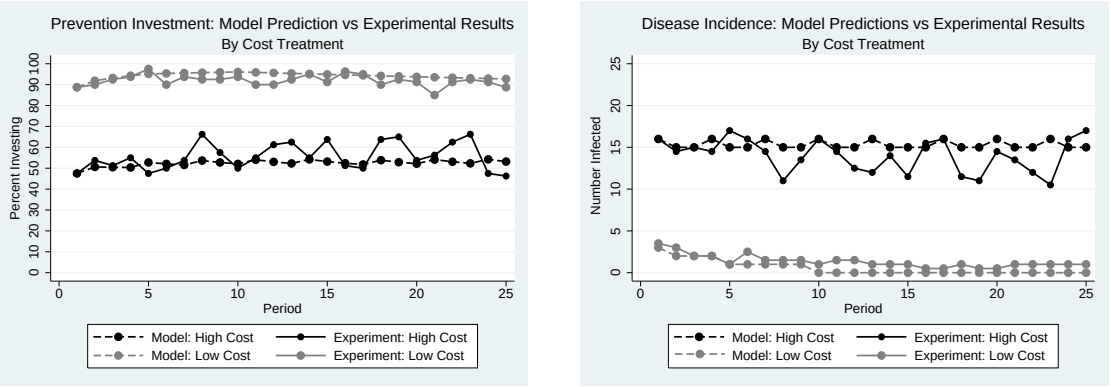
Figure 2.2 demonstrates the differences across treatment in the percent of the population

investing in disease prevention. Hypothesis 1 suggests that when the costs of investing in disease prevention are high, the probability of investing in disease prevention will be low. In the high cost treatment, the percentage of subjects investing in the disease prevention technology ranges from 40%-70%. In the low cost treatment, the percentage of subjects investing in the disease prevention technology ranges from 80%-98%. Average investment is 56% in the high cost treatment and 92% in the low cost treatment. Disease incidence ranges from 10 to 19 subjects in the high cost treatment and from 1 to 4 subjects in the low cost treatment. In the low cost treatment, the epidemic is nearly eliminated. Although regression analysis will show that this difference is statistically significant, the graph clearly demonstrates that high costs of investing in prevention are important in explaining prevention investment decisions.

Figure 2.2 show the relationship between the model prediction and experimental results for the high cost treatment. Though subjects are more responsive to changes in the infection level than the model predicts, the model does a decent job of predicting the experimental data. In the high cost treatment, the model predicts that average investment will be 52% over 25 rounds. In the experiment, I observed an average investment of 56% over 25 rounds. In the high cost treatment, the model predicts that 15.32 people will be infected on average each round. In the experiment, I observed an average infection of 14 people each round. These predictions are closed to the observed values with the observed infection slightly lower than the predicted infection. Although there are small differences between predicted and observed behavior, it is interesting that the replicator dynamic does relatively well at approximating aggregate behavior even in this small, finite population.

Figure 2.2 provides the comparison for the model prediction and experimental results for the low cost treatment. In the low cost treatment, the model predicts that over 25 rounds average investment will be 94%, and I observed an average investment of 92%. The model predicts that over 25 rounds the average infection will be .56 of a person, and I observed an average infection of 1.34 people each round. These predictions are closed to the observed values with

Figure 2.2: Model Prediction vs. Experimental Results



the observed infection slightly higher than the predicted infection. The difference between the observed and predicted values are due to the model prediction that disease prevention investment will approach 100% and infections will fall to 0 beginning in the 10th round. However, in the experiment, at least 1 person each round chose not to invest in disease prevention causing at least 1 infection each round. The motivations of the subjects that choose not to invest in disease prevention are not clear. Since these subjects make up such a small part of the sample, it is difficult to determine whether any factors might explain their behavior. Future research could use larger samples to determine what factors drive such subjects' behavior.

2.4.2 Treatment Results

The results reported in this section use a linear probability model with corrections for error term correlations. Since subjects make repeated decisions over time in this experiment, I have an individual level panel dataset. This means that the error terms are likely not independent, but instead, they are correlated within an individual. I control for this using fixed and random effects model specifications, and I report clustered, robust standard errors. Since the dependent variable of interest is investment in disease prevention, the dependent variable is binary. Although the correct model specification is non-linear, I report a linear

probability model for ease of interpretation. The results reported in this paper are robust to non-linear model specifications. A full discussion and the results of a non-linear model specification are provided in appendix A.2.

Result 1. *The probability that a subject invests in disease prevention is significantly lower in the high cost treatment.*

Hypothesis 1 states that high costs of investment will significantly reduce the probability of investing in the disease prevention product. As the costs of protecting against disease rise, subjects should be less likely to invest in disease prevention. Thus, being in the high cost treatment should reduce the probability that subjects invest in disease prevention relative to the low cost treatment. To test this I control for a dummy variable that indicates whether subjects are in the high cost treatment.

Table 2.5 demonstrates that being in the high cost treatment significantly reduces the probability that a subject invests in disease prevention. Being in the high cost treatment is associated with a 36% reduction in a subject's probability of investment. This effect is significant at the 1% level even when controlling for order and round effects. Similarly, table 2.6 shows that the magnitude of this effect increases when controlling for other factors that impact the probability of investing in the disease prevention technology including infection levels, prior infection status, and historical infection levels. Since subjects are exposed to both high and low cost treatments, this finding demonstrates that an important driver of investment in disease prevention is the cost of investing in disease prevention.

Result 2. *Subjects exhibit prevalence elastic demand for disease prevention.*

Hypothesis 2 suggests that subjects will exhibit prevalence elastic demand for disease prevention. In other words, if disease incidence increases, the probability a subject invests in

Table 2.5: Effect of Treatments on Decision to Invest in Disease Prevention

	Investment in Disease Prevention RE		
	(1)	(2)	(3)
High Cost Treatment	-0.36***	-0.36***	-0.36***
	(0.03)	(0.05)	(0.05)
Order Effects	-0.01	0.00	0.00
	(0.04)	(0.04)	(0.04)
High Cost Treatment * Order Effects		-0.01	-0.01
		(0.06)	(0.07)
Round Effects	No	No	Yes
Observations	4000	4000	4000
R^2	0.17	0.17	0.18

Clustered robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

disease prevention will also increase. Since I am interested in the impact of disease prevalence on the probability of investing in disease prevention, I use a 1 round lag in disease incidence to measure disease prevalence. To show that subjects are responding to the aggregate infection level and not changes in their own infection status, I control for each subjects infection status in the previous round.

Table 2.6 equations 1-6 demonstrate that subjects do exhibit prevalence elastic demand for disease prevention. Equation 1 shows that an increase in disease prevalence (incidence: 1 lag) alone is not significant in explaining the probability of investing in disease prevention. However, a subjects infection status in the prior round, the history of the outbreak, and the possibility that subjects react differently to the outbreak at different aggregate infection levels likely also impact an individual's probability of investing in disease prevention. Equations 2-6 control for these relevant factors and demonstrate that an increase in the disease incidence in the previous round is associated with an increase in the probability of investing in disease prevention. This effect is significant at the 1% level.

Table 2.6: Effect of Changes in Disease Incidence on Decision to Invest in Disease Prevention

	Investment in Disease Prevention					
			FE			RE
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	0.003	0.01	0.01*	0.03***	0.05***	0.05***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Infected: 1 Lag		-0.15***	-0.15***	-0.15***	-0.15***	-0.20***
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Incidence: 2 Lags			0.01***		0.02***	0.02***
			(0.01)		(0.01)	(0.01)
Squared Incidence: 1 Lag				-0.001**	-0.002***	-0.002***
				(0.00)	(0.00)	(0.00)
Free Rider Factor						-0.06**
						(0.02)
High Cost Treatment	-0.43***	-0.43***	-0.62***	-0.55***	-0.84***	-0.83***
	(0.07)	(0.07)	(0.10)	(0.08)	(0.11)	(0.12)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3552	3552	3404	3552	3404	3358
R^2	0.22	0.24	0.24	0.24	0.24	0.19

Clustered robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Result 3. *Being infected in the previous round reduces the probability of investing in disease prevention in the current round.*

It is not immediately clear how subjects should be expected to change their decisions, if at all, in response to becoming infected in the previous round. If the outbreak was worse than anticipated, then a subject might respond to becoming infected by investing in disease prevention in the next round. Similarly, if a subject dislikes the lower payoff from becoming infected, then she might respond by investing in disease prevention in the next round. However, if a subject plays the odds or probability matches, then she might respond to becoming infected by not investing in disease prevention in the next round (Vulkan (2000)). Similarly, if a subject is risk seeking in losses, then she might try to make up for the lower payoff by not investing in disease prevention in the next round (Tversky and Kahneman (1986)). To test the direction of this effect, I used a dummy variable as an indicator for subjects that were infected in the prior round.

Table 2.6 equations 2-6 provides the results for the effect of a prior round infection (infected: 1 lag) on the probability of investing in disease prevention. I find that being infected in the previous round significantly reduces the probability of investing in disease prevention. This effect is significant at the 1% level. While I demonstrate that subjects are less likely to invest in disease prevention after being infected in the previous round, I can not offer a definitive explanation for this behavior. It could be that subjects are probability matching or it could be that in the abstract disease environment, subjects are just trying to make up for receiving a lower payoff.

Future research should seek to determine whether individuals do probability match in an infectious disease context and the relative frequency of this behavior. In a public health context, probability matching is particularly problematic for diseases that do not confer life-long immunity. If individuals probability match after an infection, public health officials could

have difficulty eliminating an outbreak due to reinfection of those individuals. Understanding the frequency of probability matching and the conditions under which it occurs during an outbreak could be useful in eliminating future epidemics.

Result 4. *As disease incidence increases, the probability of investment increases at a decreasing rate.*

Since the probability of becoming infected is higher at higher levels of infection, I expected that the probability of investing in disease prevention would be greater at higher levels of infection. In other words, I expected an increase of 1 infection when 14 people are infected to increase the probability of investing in disease prevention more than an increase of 1 infection when only 1 person is infected. To test this, I create a 1 round lagged squared disease incidence variable.

Table 2.6 equations 4-6 show that I actually find the opposite effect. The lagged, squared disease incidence (squared incidence: 1 lag) has a negative impact on the probability of investing in disease prevention. This means that as infection levels rise, an increase in the infection level has a smaller impact on the probability of investing in disease prevention. This effect is significant at the 1% level. A potential explanation for this result is that some subjects exhibit herd behavior by mimicking the investment decisions of other subjects. While this behavior has been theorized and reported in other experimental contexts (Shiller (1995),Bikhchandani et al. (1998), and Hirshleifer and Teoh (2003)), this experiment demonstrates that herd behavior may be important in explaining prevention investment in a disease context. At the end of each experiment round, subjects were told both the number of subjects infected in the round and the percentage of the population investing in disease prevention during the round. Higher infection levels in the experiment were caused by fewer subjects investing in disease prevention. If subjects placed more weight on information about actions taken by others than they did on the private information about their own infection

status, then higher disease incidence would reduce the probability of investing in disease prevention.

This result is in direct conflict with the model assumption that subjects exhibit prevalence elastic demand for disease prevention. Prevalence elastic demand suggests that higher disease incidence will cause an increase in the probability of investing. While I do find evidence of prevalence elastic demand, I also find evidence that indicates subjects may exhibit herd behavior. This suggests that there is a trade-off between prevalence elastic demand and herd behavior at high levels of infection. Understanding the conditions under which this trade-off exists and the implications of this trade-off for disease dynamics is an important area for future study.

Result 5. *There is persistence in the impact of disease incidence on investment decisions*.

I expected higher past levels of infection to increase the probability of investing in disease prevention. There are a couple possible channels through which historical infection levels may influence current behavior. The first is that higher levels of infection in the past may make a subject less likely to stop investing in disease prevention. Since fewer subjects stop investing in disease prevention, higher historical infection levels may increase the probability that a subjects invests in disease prevention. Another possibility is that a subject may begin to understand the pattern for the disease outbreak after several rounds. Thus, she may recognize that higher infection levels in the past make a current infection more likely. A subject may be more likely to invest in disease prevention in response to this realized pattern. To test whether disease incidence has a persistent impact on the probability of investing in disease prevention, I control for a 2 round lag in disease incidence. This control does not indicate the channel through which historical disease incidence impacts current behavior, but it does illustrate the direction of the effect.

Table 2.6 equations 3, 5, and 6 give the impact of historical infection levels (incidence: 2 lags) on the probability of investing in disease prevention. I confirm that higher historical infection levels do have a significant positive impact on the probability of investing in disease prevention. This effect is significant at the 1% level. Furthermore, although higher infection levels 1 round and 2 rounds ago both increase the probability of investing in disease prevention, the impact of higher infections 2 rounds ago is smaller than the impact of higher infections 1 round ago. This suggests that the most recent infection levels impact prevention investment decisions more than historical infection levels.

Further research should explore the conditions under which disease incidence has a long term effect on the probability of investment in disease prevention and the channels through which this behavior change occurs. From a health policy perspective, it is important to understand how long certain health behaviors persist to determine when a population might again be susceptible to illness. This can aid public health officials in targeting disease response. Additionally, it would be useful to understand why higher historical infection levels increase the probability of investing in disease prevention. If this change occurs due to pattern recognition, then perhaps public health officials can increase individual disease awareness to induce pattern recognition earlier in an outbreak. If this change occurs due to an unwillingness to stop investing in disease prevention, then perhaps public health officials should target initial adoption of disease prevention behaviors knowing that adopters will be less likely to stop. Determining the factors that influence prevention investment persistent and the duration of the persistence is an interesting area for future research.

2.4.3 Disease Attitudes and Investment Decisions

A potential concern with this type of research is that the results may lack external validity. Since the subjects are university students making decisions in an abstract disease environ-

ment, the choices made in the experiment may not accurately capture choices individuals make during disease epidemics. To generalize the behavior in the experiment, it is necessary that the choices made in the experiment reflect behavior during disease epidemics.

I address this concern by comparing reported engagement in real-world disease prevention behaviors and beliefs about others' engagement in disease prevention with behavior in the experiment. After the experiment finished, subjects completed a questionnaire containing questions on demographics, engagement in disease prevention behaviors, and beliefs about others' engagement in disease prevention behaviors. Specifically, I ask subjects to self-report the frequency with which they wash their hands after using the restroom, wash their hands before eating, and cover their mouth when coughing in public. These questions were scaled ranging from 1 to 5 with 1 being every time and 5 being never. Additionally, I ask subjects to report the percentage of the time that most people wash their hands after using the restroom, wash their hands before eating, and cover their mouth when coughing in public. Higher reported percentages indicate higher subjective beliefs about the engagement in disease prevention by others. Using this data, I expect reported real-world behavior and beliefs about others to be correlated with behavior in the experiment if the experiment has external validity.

Interestingly, self-reported engagement in disease prevention behaviors, like hand washing and covering coughs, by themselves are not significantly correlated with the probability of investing in disease prevention in either experimental treatment. A potential explanation for this is that subjects might not accurately report their own engagement in behaviors like hand washing and covering coughs. Subjects likely know the importance of hand-washing and covering coughs for disease prevention, and therefore, are likely to overstate their own engagement in these behaviors. In fact, I find little variation in responses with most subjects reporting frequent engagement in these behaviors (see appendix A.3).

Result 6. *Subjects who report free-riding off the disease prevention investments of others in reality are less likely to invest in disease prevention in the experiment.*

Since I want to isolate attitudes toward disease prevention, I can use factor analysis to weight each question according to the unobserved latent variable: disease prevention attitudes. Using each individual's reported engagement in real-world disease prevention and beliefs about others' real-world engagement in disease prevention, I create a 'Free Rider' factor that captures individual responses to the questions. The six questions had solid internal consistency with Cronbach's alpha of 0.61. The 'Free Rider' factor has an eigenvalue of 1.01. Subjects that scored high on the 'Free Rider' factor report lower engagement in disease prevention and reportedly believe a higher percentage of others engage in disease prevention behavior. If the experiment captures real-world disease prevention behavior, then subjects scoring high on the 'Free Rider' factor should be more likely to free ride off the disease prevention investments of others in the experiment. In other words, a high score on the 'Free Rider' factor should reduce the probability of investing in disease prevention in the experiment.

I find that reported real-world free-riding on disease prevention investments is correlated with free-riding behavior in the experiment. Table 2.6 equation 6 demonstrates that an increase in one standard deviation in reported real-world free riding decreases the probability of investment in disease prevention in the experiment by 6 percentage points. This effect is significant at the 1% level. While free-riding behavior has been well documented with regards vaccine uptake, I demonstrate that reported free-riding behavior extends beyond vaccination to other preventive behaviors such as hand washing and covering coughs. Furthermore, I find that reported real-world engagement in free-riding behavior is consistent with decisions made in the experiment. This suggests that laboratory experiments can capture real-world engagement in disease prevention.

2.5 Conclusion

Understanding individual behavioral responses to disease is important for understanding the best ways in which to prevent or reduce an outbreak. Although individuals likely change their behavior in response to infectious disease outbreaks, research in this area is relatively new. The primary challenge for researchers has been collecting or obtaining data on individual level behaviors during an outbreak. Since efforts are typically focused on prevention interventions during an outbreak, there is little data on the additional actions taken by individuals to protect themselves and others from the outbreak. This makes it difficult to endogenize behavior in mathematical models of disease, because it is not clear how behavior should be endogenized. Furthermore, when behavior is endogenized in mathematical models of disease, the lack of individual level data on decisions during an outbreak makes it difficult to test the model assumptions regarding behavior.

I provide one solution for generating individual level data with which to test behavior assumptions in mathematical models of disease: a laboratory experiment. To do this, I endogenize behavior into a mathematical model of disease using an evolutionary game and existing research that suggests individuals exhibit prevalence elastic demand for disease prevention. In the game, different payoffs correspond to each choice: invest or do not invest in disease prevention. The choice that provides the highest payoff will grow in aggregate. In the lab, I create an artificial disease environment and allow subjects to make their disease prevention investment decisions. Since the dynamic for aggregate investment in disease prevention is derived from the individual payoffs underlying the decision, I am able to provide experimental subjects with payoffs from the mathematical model. I then use subjects' choices during the experiment to compare the model predictions with the observed experimental data.

The mathematical model provides two primary hypotheses. First, high costs of investing in disease prevention will reduce investment in disease prevention. Second, subjects will exhibit

prevalence elastic demand for disease prevention. In other words, as infection levels rise, the probability of investing in disease prevention should rise. I used data from the laboratory experiment to test these two hypotheses. Furthermore, I use the data gathered from the experiment to explore new factors that might be relevant in explaining investment in disease prevention during disease outbreaks.

I find support for both hypotheses, and I show that additional factors are relevant in explaining the probability of investing in disease prevention. High costs of investing in disease prevention have a significant negative impact on the probability of investing in disease prevention. Similarly, higher infection levels significantly increase the probability of investing in disease prevention. However, I find that this effect is larger at lower infection levels than at higher infection levels. Additionally, I find that a subject's own infection status and the outbreak history are significant in explaining the probability of investing in disease prevention. Being infected in the previous period significantly reduces the probability that a subject invests in disease prevention; whereas, higher historical infection levels increase the probability that a subject invests in disease prevention.

A potential issue with this type of research is the external validity of the decisions made in the artificial disease environment. Specifically, this abstract environment may not accurately capture subjects real-world attitudes toward disease prevention. I address this issue by comparing individual behavior in the experiment with reported real-world engagement in disease prevention behaviors like hand washing and covering coughs. I find that subjects who report free-riding off the disease prevention investments of others in reality are more likely to free-ride off the disease prevention investment of others in the experiment. This suggests that the abstract laboratory environment does capture real-world attitudes towards investment in disease prevention.

Several findings in this paper are interesting areas for future research. First, it is important to determine the conditions under which individuals probability match after an infection with

an illness that does not confer immunity. In this context, subjects that probability match continue to forego investing in disease prevention even after a non immunity conferring infection. Understanding how this behavior can be mitigated is important for reducing the magnitude of disease outbreaks.

Second, more work is needed to understand the trade-off between prevalence elastic demand for disease prevention and other behavioral biases. Although I do find evidence of prevalence elastic demand for disease prevention, I show that this behavior diminishes at higher infection levels. Since subjects have information on the aggregate choices of others and higher infection levels are caused by lower investment in disease prevention, I suggest that one potential explanation for this is that subjects may exhibit herd behavior. If subjects exhibit herd behavior in disease environments, then understanding the conditions that lead subjects to follow others may help to increase investment in disease prevention when infection levels are high.

Third, determining the length and reason that higher historical infection levels increase investment in disease prevention can help researchers understand the causes of behavior change. Knowing when individuals are likely to stop engaging in disease prevention can help indicate when a population will again become susceptible to a disease. Additionally, understanding how the information about historical infections impacts current decisions can help researchers identify the best ways to encourage investment in disease prevention.

Finally, although I found that reported engagement in real-world disease prevention behaviors is correlated with disease prevention behavior in the experiment, better questionnaires that illicit health attitudes and beliefs are needed to determine whether behavior in the experiment reflects real-world behavior. There are actually two issues here: question design and types of questions to ask. The challenge in designing the questions is to ask them in such a way that an individual's beliefs or attitudes regarding health are accurately reflected by their response to the question. The challenge in asking the questions is knowing which

questions should be correlated with behavior. In order to provide more tests of the external validity of the experimental disease environment, we need to develop a better understanding of the types of questions and responses that should be correlated with behavior in an infectious disease context.

Chapter 3

The Role of Outbreak Information and Social Exclusion in Investment in Disease Prevention

3.1 Introduction

The risk of infectious diseases constitutes an increasing threat to public health. The ease of global travel enables emerging infectious diseases to be rapidly transmitted to new host populations (Morse (2001)). Furthermore, the global changes in climate are expected to increase the incidence and distribution of infectious diseases by increasing the range and abundance of animal reservoirs and insect vectors, prolonging transmission cycles, and reestablishing previously eliminated endemic, infectious diseases (Greer et al. (2008)). Despite the increasing threat posed by infectious diseases, epidemic induced behavioral responses to disease remain relatively understudied.

The primary challenge associated with studying epidemic induced behavior is collection of

data on individual behaviors during an infectious disease outbreak. Ideally, researchers would have access to an individual level panel dataset containing data on all actions taken by individuals to protect against disease before, during, and after an epidemic. Since the primary goal of public health officials during an epidemic is to end the epidemic, obtaining such detailed data on individual behavior is rare.

To generate data on epidemic induced disease prevention behavior, I build on work by Chen et al. (2013) and **Carpenter (2016)** by conducting a laboratory experiment with a simulated disease environment. The experimental payoffs are given by an epidemiological model with endogenous disease prevention investment developed by Carpenter (2014) and **Carpenter (2016)**. These models endogenize the decision to invest in disease prevention using an evolutionary game with the replicator dynamic. Since the disease prevention behavior dynamic is derived from underlying individual payoffs, I provide experimental subjects with the payoffs from the evolutionary game and compare model predictions with observed experimental behavior.

The purpose of this paper is to explore the two assumptions made regarding outbreak information and social exclusion costs in **Carpenter (2014)** and **Carpenter (2016)** to determine their impacts on individual behavior during infectious disease outbreaks. First, these models assume that behavior evolves according to the replicator dynamic, a technique that allows individuals to exhibit unsophisticated decision making behavior. An implication of this assumption is that providing subjects with outbreak information does not impact their decision to invest in disease prevention. Second, individuals choosing to refrain from disease prevention investment face costs of social exclusion. Social exclusion costs are a type of ostracism or isolation imposed by society on individuals that have potentially been exposed to a disease. These models assume that individuals that avoid investing in disease prevention are subjected to social exclusion costs. A result of this assumption is an increase in the probability of investing in disease prevention.

Empirically, the role of information provision in disease prevention is mixed, but the impact seems to depend crucially on the type of information and to whom it's given (Dupas (2011c)). In the context of HIV/AIDS, Duflo et al. (2015) find that providing HIV/AIDS education increases self reported use of condoms, but it does not actually decrease the incidence of teenage childbearing. However, Dupas (2011a) show that providing teenagers with information that HIV prevalence is higher among adult men than among teenage boys leads to a reduction in the incidence of teenage pregnancies with adult partners. Similarly, in the context of water contamination, Aziz et al. (2006) demonstrate that exposure to information about the effects of arsenic contaminated water does not increase the adoption of behaviors to reduce exposure to contaminated water. However, Madajewicz et al. (2007) find that providing a household with information about the arsenic contamination status of their well water increases the probability that the household switches wells in the event of contamination. Dupas (2011c) provides a review of the literature on information provision and health behavior changes. This paper diverges from prior literature by focusing specifically on the provision of information regarding epidemic magnitude and the prevention investment decisions of others. Since research on information provision and disease protective behaviors is mixed, I base the hypothesized relationship between outbreak information and prevention investment decisions on the mathematical model of disease with endogenous disease prevention behavior. I hypothesize that provision of information on outbreak magnitude and on disease prevention investment decisions of others will not impact the probability of investing in disease prevention.

Research suggests that the presence of social exclusion costs increases investment in disease prevention. Historically, social exclusion has been implemented as a technique to increase adoption of disease prevention behaviors. Examples of social exclusion include disgust toward disease propagating behaviors (Curtis and Biran (2001); Curtis (2011)), quarantining potentially disease exposed individuals (Gensini et al. (2004); Tognotti (2013)), and banning public school attendance for unvaccinated children (Orenstein and Hinman (1999); Wilson

et al. (2005)). For simplicity, I assume that experimental disease exposure comes only from failure to invest in disease prevention, and therefore, individuals that choose not to invest in disease prevention face costs of social exclusion. This paper diverges from prior research by explicitly incorporating social exclusion costs into the payoff for not investing in disease prevention. Like prior research, the model suggests that facing social exclusion costs for foregoing prevention investment increases the probability of investing in disease prevention. So, I hypothesize that enforcing social exclusion costs for individuals who fail to invest in disease prevention will increase the probability that an individual invests in disease prevention.

I use the laboratory experiment with a simulated disease outbreak to explore the impact of outbreak information and social exclusion on the probability of investing in disease prevention. Subjects must choose whether to invest in disease prevention in each round of the experiment. Prevention investment is costly, but provides full protection against disease. Subjects that choose to forego prevention investment are at risk of infection. The cost of foregoing prevention investment varies by treatment. I use a between subjects experimental design. There are 2 treatment variables: outbreak information and exclusion costs. The outbreak information variable takes on two values: outbreak information and no outbreak information. The exclusion cost variable takes on two values: exclusion costs and no exclusion costs. I compare the experiment results across treatments and with model predictions.

I find that outbreak information and social exclusion costs have a significant impact on the probability of investing in disease prevention. I reject the hypothesis that outbreak information does not have a significant effect on the decision to invest in disease prevention. Instead, providing subjects with information on the outbreak magnitude and on the share of the population investing in disease prevention reduces the probability of investing in disease prevention. This effect happens through two channels: a decrease in the probability that a subject starts investing in disease prevention and an increase in the probability that a subject stops investing in disease prevention. I demonstrate support for the hypothesis that

social exclusion costs increase the probability of investing in disease prevention. This effect happens through two channels: an increase in the probability that a subject starts investing in disease prevention and a decrease in the probability that a subject stops investing in disease prevention.

This paper proceeds as follows: Section 2 provides the model used in the experiment. Section 3 describes the experimental design, Section 4 provides the results, and Section 5 concludes.

3.2 Epidemiological Model

The model used in this paper is a modified version of the SIR model of cholera outbreak developed in Carpenter (2014) and modified in Carpenter (2016). The full epidemiological model of disease is characterized by a system of four dynamic equations. This model includes a dynamic equation for changes in the disease susceptible population, changes in the disease infected population, changes in the environmental bacteria population, and changes in the share of the population not investing in disease prevention. Although I present the dynamic equations for the biological components of the model below, the focus of this paper is on disease prevention investment behavior.

To endogenize the decision to invest in disease prevention, I use an evolutionary game with the replicator dynamic. Evolutionary game theory was developed to model strategic interaction in large populations. In an infectious disease setting, populations are generally large. Standard game theory requires that all players have full knowledge of the underlying disease dynamic, equilibria, and all disease prevention actions taken by others. In large populations, like a typical population facing a disease epidemic, it is unlikely that all players have common knowledge of the game. Using an evolutionary game avoids the strong assumptions on players knowledge of the game.

The replicator dynamic does not provide information about the way in which individuals make decisions. Instead, it captures the idea that a given strategy will grow when the payoff from that particular strategy is larger than the average payoff. One way to explain individual behavior in this dynamic is that individuals are randomly matched with another individual each time period. If the matched individuals have different strategies, they can each compare the payoff under the alternative strategy. When the payoff under the alternative strategy is higher, they will adopt the other strategy with some probability. To explain the payoff differences in the two models presented, I will use this interpretation of the replicator dynamic. However, it should be noted that the replicator dynamic can arise under other types of individual behavior, as well.

Like Carpenter (2016), constraints on both the budget size and the experimental lab size made it necessary to have all subjects making a decision in each round of the experiment. I modify the model presented in Carpenter (2014) in three ways to achieve this experimentally. First, I omitted the natural birth and death cycle from the model. Second, I eliminate the possibility that disease infection can "kill" or remove a subject from the population. These two modifications together ensure that the size of the population in the experiment is fixed. Third, I omit the possibility that disease infection can confer temporary immunity. Removing the possibility that the infection can provide temporary immunity to reinfection allows all subjects to make a decision each round. This increases the amount of data that I am able to collect with a small population size. Future research should explore the impact of these changes on the probability of investing in disease prevention.

The change in the susceptible population is given by equation 1. The susceptible population, $S(t)$, is increased by the number of infected individuals that recover from disease infection, $I(t)$. It is decreased by the number of individuals that become disease infected, $q\alpha(t)S(t)\frac{B(t)}{k+B(t)}$. The value q is the probability that not investing in disease prevention leads to disease exposure, and $\alpha(t)$ is the share of the population that does not invest in disease

prevention. In this model, exposure to the disease does not necessarily mean that an individual will become disease infected. Individuals' probability of becoming infected is increasing in the number of bacteria that they are exposed to. The value k is the level of bacteria consumption that would result in a 50% probability of infection, and $B(t)$ is the number of bacteria in the environment at time t. So, the fraction $\frac{B(t)}{k+B(t)}$ is the time dependent probability of infection.

$$\frac{dS(t)}{dt} = I(t) - q\alpha(t)S(t)\frac{B(t)}{k + B(t)} \tag{3.1}$$

The change in the disease infected population is provided by equation 2. The infected population is increased by the number of people that become disease infected, $q\alpha(t)S(t)\frac{B(t)}{k+B(t)}$. Upon becoming infected, individuals leave the susceptible population and join the infected population. The infected population is decreased by the number of infected people that recover from the disease, $I(t)$. When subjects recover, they move from the infected population back into the susceptible population.

$$\frac{dI(t)}{dt} = q\alpha(t)S(t)\frac{B(t)}{k + B(t)} - I(t) \tag{3.2}$$

The change in the bacteria in the environment is given by equation 3. The bacteria population in the environment is increased by the number of bacteria infected individuals shed into the environment. The bacteria population in the environment is decreased by the number of bacteria that die off in the environment. Infected individuals shed bacteria into the environment at a rate of θ. So, the quantity $\theta I(t)$ is the number of bacteria shed into the environment. The value δ is the rate at which bacteria die in the environment. Thus, the

quantity $\delta B(t)$ is the number of bacteria that die in the environment at time, t.

$$\frac{dB(t)}{dt} = \theta I(t) - \delta B(t) \tag{3.3}$$

I derive two different dynamics for changes in the share of the population not investing in disease prevention. The first dynamic is the dynamic for disease prevention investment behavior developed in Carpenter (2016). This dynamic is derived from an evolutionary game where individuals that do not invest in prevention and are matched with an investor pay a cost. I argue that this is the cost of social exclusion or social ostracism that accompanies the decision not to invest in disease prevention. The second dynamic I develop to capture disease prevention investment behavior relaxes this assumption. It is derived from an evolutionary game where individuals that do not invest in prevention and are matched with an investor do not pay a cost. Each dynamic for disease prevention investment behavior combined with the dynamic equations 1-4 presented above represents a separate model for disease outbreak.

3.2.1 Model 1

Table 3.1: Prevention Investment Game: Model 1

	Invest	Do Not Invest
Invest	b-c	b-c
Do Not Invest	b-c(I(t))	-c(I(t))

Table 3.1 provides the payoffs for the evolutionary prevention investment game. As is standard practice for evolutionary games, table 3.1 provides the payoffs for only 1 player, the player on the left hand side. Individuals that choose to invest in disease prevention receive a benefit of avoiding infection, b. The cost of investing in disease prevention is c. Individuals choosing to invest in disease prevention receive this benefit and incur this cost regardless of whether they are randomly matched with another investor.

The payoff for individuals that choose not to invest in disease prevention depends upon whether they are match with an investor or a non investor. Individuals that choose not to invest in disease prevention and are matched with an investor receive a benefit of avoiding infection, b. These individuals pay a cost, $c(I(t))$, which is a function of the number of infected individuals at a given point in time. This cost is interpreted as the cost of social exclusion. As infection levels rise, non-investors face increased costs of social exclusion or social ostracism for choosing not to invest. For this reason, I refer to model 1 as the 'exclusion cost' model. Individuals that choose not to invest in disease prevention and are matched with a non investor receive no benefit. These individuals pay a cost, $c(I(t))$, which is a function of the number of infected individuals at a given point in time. This cost is interpreted as the cost of infection. As infection levels rise, becoming infected becomes more costly as medical resources become strained under increased use.

$$\frac{d\alpha(t)}{dt} = \alpha(t)(1-\alpha(t))\left[c - c(I(t)) - \alpha(t)b\right] \tag{3.4}$$

I used the replicator dynamic with the payoffs for each strategy reported in table 3.1 to derive the dynamic for disease prevention investment. Equation 4 provides the dynamic for the change in the share of the population not investing in disease prevention used in Model 1. For the full step-by-step derivation, please see appendix A.1.1. Model 1 is fully characterized by equations 1-4.

3.2.2 Model 2

Table 3.2 shows the payoffs for the evolutionary prevention investment game. Table 3.2 provides the payoffs for only 1 player, the player on the left hand side. It is easy to see

Table 3.2: Prevention Investment Game: Model 2

	Invest	Do Not Invest
Invest	b-c	b-c
Do Not Invest	b	-c(I(t))

here that the payoff for investing in disease prevention is the same for model 1 and model 2. Individuals that choose to invest in disease prevention receive a benefit from avoiding infection, b. Individuals that choose to invest in disease prevention pay a cost to invest, c.

The payoff for individuals that choose not to invest in disease prevention depends upon whether they are matched with an investor or a non investor. Like model 1, individuals that choose not to invest in disease prevention and are matched with another non-investor pay a cost of infection that is a function of the number of infected individuals, $c(I(t))$, at a given point in time. These individuals receive no benefit. The difference between models 1 and 2 occurs in the payoff for individuals that chose not to invest in disease prevention and are matched with an investor. In both models 1 and 2, these individual receive a benefit of avoiding infection, b. However, in model 2, there are no costs for choosing not to invest in disease prevention and being matched with an investor. In other words, there are no social exclusion costs in model 2. For this reason, I refer to model 2 as the 'no exclusion cost' model.

$$\frac{d\alpha(t)}{dt} = \alpha(t)(1-\alpha(t))\left[c - \alpha(t)(b + c(I(t)))\right] \tag{3.5}$$

I used the replicator dynamic with the payoffs for each strategy reported in table 3.2 to derive the dynamic for disease prevention investment. Equation 5 provides the dynamic for the change in the share of the population not investing in disease prevention used in Model 2. For the full step-by-step derivation of the dynamic, please see appendix A.1.2. Model 2

is fully characterized by equations 1-3 and equation 5 (given above).

3.2.3 Model Calibrations

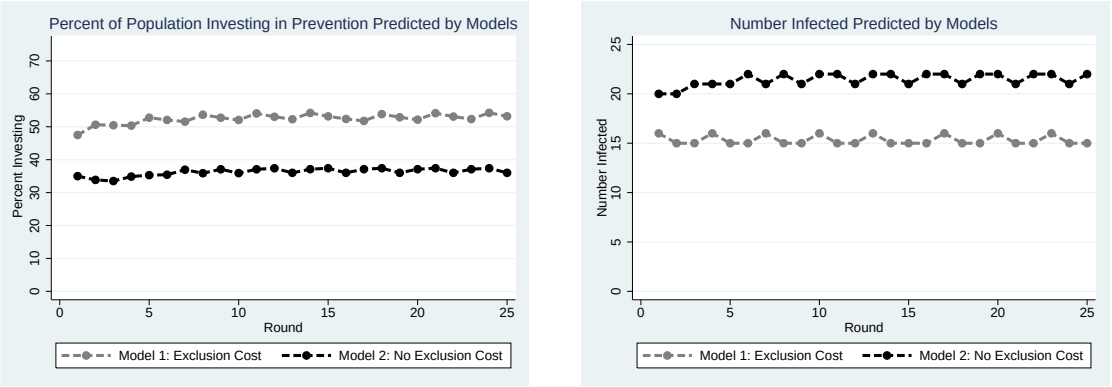
Table 3.3: Model Parameters

Parameters	Values	Footnotes
N	40	1
k	10^3	2
q	.99	3
θ	70	3
δ	$\frac{7}{30}$	3
Start Values		
S(t)	40	1
$B(t)$	10^3	2
α(t)	0.525	4
$I(t)$	0	5
Payoff Values		
b	$\frac{10}{10}$	6
c	$\frac{20}{10}$	6
c(I(t))	$\frac{I(t)}{10}$	6

1. The number of subjects that participated in each experiment session. 2. These numbers give a high probility of infection if not investing to induce subjects to care about infection. 3. Based on models of cholera outbreak taken from Carpenter (2014). 4. Based on subjects initial experiment investment level. 5. Nobody was initially infected with the disease. 6. Experimental units were converted to money at a rate of 10 experimental units per 1 USD

To generate the model predictions, I calibrated each of the equations in the system for each model. Table 3.3 provides both the biological parameters and the payoffs used to calibrate the dynamic system. The biological parameters are based on a model of cholera outbreak since the model was originally developed to capture behavior during a cholera epidemic. The parameter values chosen for the evolutionary game payoff matrix are the values used in

Figure 3.1: Predicted Prevention Investment and Infection Levels by Model



Carpenter (2016).

Figure 3.1 demonstrates the predictions of the calibrated models. By comparing the share of the population not investing in disease prevention under the two models, I generate predictions for behavior when subjects are exposed to social exclusion costs. Model 1 predicts that approximately 50% of the population will invest in disease prevention. Model 2 predicts that approximately 36% of the population will invest in disease prevention. Comparing these predictions indicates that investment in disease prevention should be higher when individuals face costs of social exclusion (model 1) than when individuals do not face costs of social exclusion (model 2). This seems reasonable since prediction is consistent with prior research that suggests the presence of social exclusion costs increase investment in disease prevention.

Although figure 3.1 provides the predictions of the calibrated models, there is no distinct prediction for the presence or absence of outbreak information. The replicator dynamic does not allow for information to play a role in the change in the share of the population investing in disease prevention. In other words, the presence or absence of outbreak information available to individuals does not change the predictions of either model. Regardless of the information available to individuals, model 1 predicts that approximately 50% of the population will invest in disease prediction, and model 2 predicts that approximately 36%

of the population will invest in disease prevention.

3.2.4 Hypotheses

I identified 2 hypotheses to test in this experiment based on the calibrated SIR models.

Hypothesis 1. *The presence of social exclusion costs will increase investment in disease prevention.*

This hypothesis comes from the difference in the percent of the population investing in disease prevention when individuals face social exclusion costs (model 1) versus when individuals do not face social exclusion costs (model 2). When calibrating models 1 and 2, the percent of the population investing in disease prevention is higher under model 1 than under model 2. This makes sense intuitively, because the cost of foregoing investment in disease prevention is higher in model 1. Thus, the alternative, investing in disease prevention, is relatively less costly. This suggests that when individuals face social exclusion costs, they should be more likely to invest in disease prevention.

Hypothesis 2. *The presence of outbreak information will not impact investment in disease prevention.*

This hypothesis is based on the use of the replicator dynamic to model aggregate prevention investment behavior. The replicator dynamic allows for unsophisticated disease prevention investment behavior. An implication of this is that the presence of outbreak information should not influence the decision to invest in disease prevention. Empirically, the impact of information on investment in disease prevention is mixed. So, I hypothesize that providing subjects with outbreak information will not change their disease prevention investment

behavior.

3.3 Experiment Design

3.3.1 Basics

I conducted 4 experiment sessions at a large public university using undergraduate students as human subjects. Students learned of the laboratory through classroom advertisements, and they registered for the experimental pool through an online registration system. Potential subjects were selected at random and notified of the upcoming experiment session through email a few days before the session. If a subject was interested in participating in the experiment, she was instructed to register for the session through an online registration system. If a student registered for a session, she received an email reminder the day before and the day of the planned experiment session. Subjects were not allowed to participate in more than one session.

Although the susceptible population for many diseases is very large, the laboratory where this experiment was conducted has only 40 computers. To ensure the largest possible susceptible population size for this experiment, I had to make sure that the lab was filled to capacity during the experiment. To do this, I recruited more than 40 subjects for each experimental session. Subjects that showed up but were not needed were paid the $7 show-up fee. They were also told that they were eligible to participate in a future experiment session. This allowed for the largest experiment sample size given the laboratory's constraints.

3.3.2 Design

This experiment used a 2x2 factorial design with two treatment variables: social exclusion and information. The social exclusion variable took on two values: exclusion cost and no exclusion cost. The information variable took on two values: outbreak information and no outbreak information. There were a total of 4 treatments each consisting of 25 rounds. Subjects were not allowed to participate in more than one treatment.

To generate the underlying disease dynamic, I calibrated the model presented in section 3.2 using parameter values given in table 3.3. I used the zTree software package (Fischbacher (2007)) to program the artificial disease environment. The model was calibrated at the beginning of each session; so, each session experienced their own independent disease outbreak. I conducted 4 experimental sessions, 1 per treatment. There were 40 subjects that participated in each session, and there was a total of 160 subjects in the overall experiment.

Although I used a model of cholera outbreak to generate the underlying disease dynamic, the experiment instructions were framed using disease neutral language. Subjects were told that the government had discovered the presence of an "infectious disease" in the environment. They were also told that there was a "disease prevention technology" that would provide full protection against the disease. Subjects were then instructed that they must chose whether to "invest" in the "disease prevention technology." Choosing to invest in the disease prevention technology provided full protection against disease. However, the protection lasted only for the current round. Choosing not to invest in the disease prevention technology made subjects susceptible to "infection." If a subject was selected as "infected," the infection lasted only for the current round.

Each round subjects had to chose "invest" or "do not invest" in disease prevention. In the first 2 rounds of each treatment, subjects had 30 seconds to make their decision. In the remaining 23 rounds of each treatment, subjects had 15 seconds to make their decision.

Subjects were informed that if they failed to make their decision in the allotted time, then "do not invest" would be automatically selected.

I used the subjects choices to determine the value for $\alpha(t)$. Using the model in section 3.2, I was then able to generate the aggregate infection level. Only subjects that chose not to invest in disease prevention were at risk of infection in the round. However, not all subjects that chose not to invest in disease prevention actually became infected. I used a random number generator to assign the "infected" status to subjects that chose not to invest in disease prevention. The "infection" lasted only for the round.

Once all subjects had made their decisions, they received information about the round in the form of a results screen. The information provided to subjects varied by information treatment. In the outbreak information treatment, the results screen gave subjects information on their own decision, their own infection status, their own round payoff, the total number of subjects infected, the population percentage investing in disease prevention, the payoff from investing in disease prevention, the payoff from not investing in disease prevention and not becoming infected, and the payoff from not investing in disease prevention and becoming infected. In the no outbreak information treatment, the results screen gave subjects information on their own decision, their own infection status, and their own round payoff. In each treatment, subjects received the information provided on the results screen as a history box beginning in round 2. Subjects were not provided with information about their objective probability of infection in any of the treatments. This was done to better capture real-world disease prevention investment decisions where individuals do not know their objective probability of infection.

Following standard experimental economics practice, subjects' decisions in the experiment were incentivized. Tables 3.4-3.7 provide the payoffs for the experiment. Tables 3.4 and 3.5 give the experimental payoffs for the exclusion cost treatment, model 1. Tables 3.6 and 3.7 show the experimental payoffs for the no exclusion cost treatment, model 2. The left hand

Table 3.4: Exclusion Cost- Payoff (Variables)

	Not Infected	Infected
Invest	E+b-c	—
Do Not Invest	E+b-c(I(t))	E-c(I(t))

Table 3.5: Exclusion Cost- Payoff (Values)

	Not Infected	Infected
Invest	40+10-20	—
Do Not Invest	40+10-I(t)	40-I(t)

Table 3.6: No Exclusion Cost- Payoff (Variables)

	Not Infected	Infected
Invest	E+b-c	—
Do Not Invest	E+b-c(I(t))	E-c(I(t))

Table 3.7: No Exclusion Cost- Payoff (Values)

	Not Infected	Infected
Invest	40+10-20	—
Do Not Invest	40+10-I(t)	40-I(t)

side (LHS) of each table demonstrates the choice available to each subject: invest in disease prevention or do not invest in disease prevention. The column headers of each table show the potential outcomes for each strategy: not infected or infected. Within a given round, a subject's payoff was determined by their own choice, their own infection status, and their endowment, E. Subjects were endowed with 40 experimental units each round.

In the exclusion cost treatment, subjects payoffs depend on their own choice and their own infection status. If a subject chose to invest in disease prevention, she received a benefit, b, of 10 experimental units from avoiding infection. The cost of investing in disease prevention, c, was 20 experimental units. If a subject chose not to invest in disease prevention and she was not infected, she received a benefit, b, of 10 experimental units for avoiding infection. If a subject chose not to invest in disease prevention and she was infected, she received no benefit. The cost of not investing in disease prevention, $c(I(t))$, was equal to the number of infected individuals in the round, $I(t)$, regardless of a subjects infection status. For subjects who did not invest in disease prevention and were not infected, this is the cost of social exclusion. For subjects who did not invest in disease prevention and were infected, this is the cost of infection.

In the no exclusion cost treatment, subjects payoff still depend on their own choice and their own infection status. If a subject chose to invest in disease prevention, she received a

benefit, b, of 10 experimental units from avoiding infection. The cost of investing in disease prevention, c, was 20 experimental units. If a subject chose not to invest in disease prevention and she was not infected, she received a benefit, b, of 10 experimental units from avoiding infection. If a subject chose not to invest in disease prevention and she was not infected, she paid no cost. If a subject chose not to invest in disease prevention and she was infected, she received no benefit from avoiding infection. The cost of not investing in disease prevention and becoming infected, $c(I(t))$, was equal to the number of infected individuals, $I(t)$, in the round. This cost is the cost of becoming infected and includes missing work, missing school, and costs of any medical interventions.

Subjects received payment based on their choices in the disease experiment and in the incentivized lottery conducted after the disease experiment. One round from the experiment was randomly selected for payment. Round payoffs were given in experimental units which were converted to USD at a rate of 10 experimental units per USD. At the end of the 25 rounds, subjects were asked to participate in an incentivized lottery of their choice. Using an approach developed by Eckel and Grossman (2008) and modified by Caldara (2013), each potential lottery corresponded to a type of risk preference: risk loving, risk neutral, or risk averse. So, the lottery served as a rough elicitation of risk preferences. The lottery was also conducted in experimental units with a conversion rate of 2 experimental units per USD. Subjects were aware of all conversion rates prior to their decisions. The payoff for the round and lottery were added to the $7 show-up fee for the final payment. Payments were rounded up to the nearest quarter. Subjects were paid on average $15.00 for 40 minutes of participation.

3.4 Results

3.4.1 Descriptive Statistics

Table 3.8 shows the descriptive statistics for this experiment. There were more females than males that participated in the overall experiment. However, I do not find that gender is significant in explaining behavior in the experiment. There are significanly more asian students than students of other races that participated in the experiment. This reflects the population of the student body at the large public university where this experiment was

Table 3.8: Descriptive Statistics

	Statistics
Male	33%
Female	67%
White (not hispanic)	9%
Hispanic	21%
Asian	60%
Black	1%
Mixed Race	6%
Undisclosed	3%
Average Payment	$15.00
Number of Subjects	160

conducted. I do not find that race is significant in explaining behavior in the experiment.

Figures 3.2 and 3.3 demonstrate the differences across treatments in the percentage of the population investing in disease prevention and disease incidence, respectively. Figure 3.2 shows that the percentage of the population investing in disease prevention was lower when subjects were not exposed to exclusion costs than when subjects were exposed to exclusion costs. Additionally, it appears that average investment levels are slightly higher when subjects are not given outbreak information than when subjects receive outbreak information. Figure 3.3 reveals that disease incidence is lower in the exclusion cost treatment. This is a direct result of the higher investment in disease prevention. Additionally, average infection levels appear to be lower when subjects are not given information about this disease outbreak than when subjects receive outbreak information. This is a direct result of the higher investment in disease prevention in the no outbreak information treatment.

(M vs E): Like Carpenter (2016), I find that the model closely resembles the observed experimental behavior in the treatment with exclusion costs and outbreak information. Figure 3.2 provides evidence that the percentage investing in disease prevention is closely predicted by the model. Interestingly, though the model predicts that information does not impact investment in disease prevention, I find that providing subjects with outbreak information appears to decrease investment in disease prevention, and therefore, increase the magnitude of the outbreak in the exclusion cost treatment.

Figure 3.2: Prevention Investment: Model Prediction vs. Experimental Results

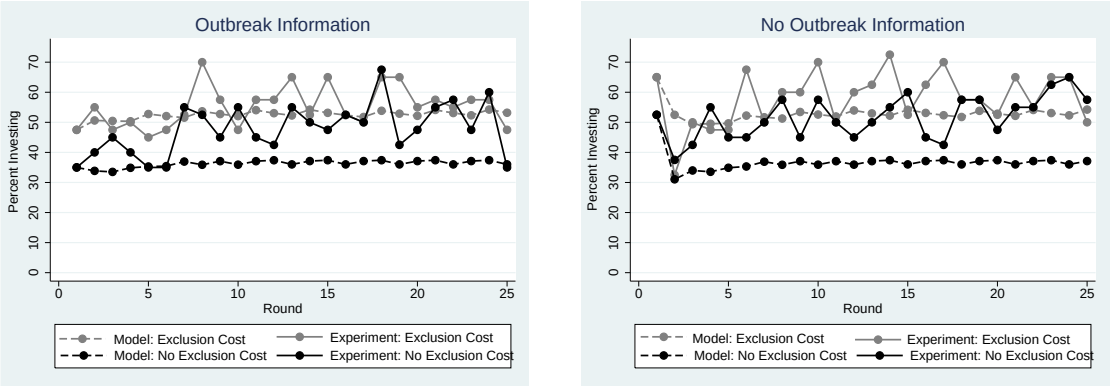
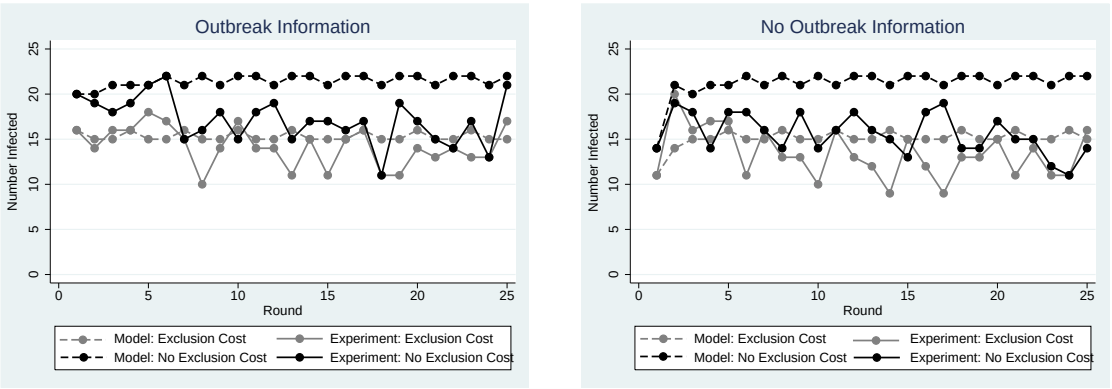


Figure 3.3: Disease Incidence: Model Prediction vs. Experimental Results



Figures 3.2 and 3.3 demonstrate that the no exclusion cost model is not as effective at predicting observed experiment behavior as the model with exclusion costs. Although I do provide evidence that removing costs of social exclusion decreases investment in disease prevention, the magnitude of this effect is not as large as the model predicts. Furthermore, like the exclusion cost treatment, I find that providing subjects with outbreak information decreases investment in disease prevention thereby increasing the infection level.

3.4.2 Treatment Results

I test whether the decision to invest in disease prevention is a stationary variable. Since subjects make repeated choices over time, it is possible that a subject's decision to invest in

disease prevention is not stationary. I use a panel data unit root test **(Levin et al. (2002))** to determine whether the decision to invest in disease prevention is stationary. I fail to reject the null hypothesis that the variable contains a unit root. In other words, I find that the investment in disease prevention variable is not stationary.

To correct for this problem, I transform the variable into a stationary variable. I do this by taking the first difference of the decision to invest in disease prevention by subject. Since the variable is binary, the first difference is also discrete. However, the first difference gives 3 discrete outcomes rather than 2: the decision to starting investing in disease prevention, the decision to stop investing in disease prevention, and the decision to not change the investment decision. This means that I can not look at the probability of investing in disease prevention directly. Instead, I look at the factors that influence a subject's decision to start and stop investing in disease prevention.

I adapt the hypotheses to the newly transformed dependent variable. An increase in the probability of investing in disease prevention can happen in 2 ways: an increase in the probability that a subjects starts investing or a decrease in the probability that a subject stops investing. So, I hypothesize that being in the exclusion cost treatment will increase the probability that a subject starts investing in disease prevention and decrease the probability that a subject stops investing in disease prevention. Information is not hypothesized to impact the probability of investing in disease prevention. So, I hypothesize that information will not impact either the probability that a subject starts investing in disease prevention or the probability that a subject stops investing in disease prevention.

I choose to report a linear model specification with random effects in the main paper for simplicity. Under the linear specification, I explore 2 dependent variables of interest: the probability that a subject starts investing in disease prevention and the probability that a subject stops investing in disease prevention. Since these dependent variables are binary, the correct model specification is actually non-linear. In appendix A.2, I show that the results

Table 3.9: Effect of Treatments on Changes in Disease Prevention Investment Decisions

	Decision: Start Investing			Decision: Stop Investing		
	RE					
	(1)	(2)	(3)	(4)	(5)	(6)
Outbreak Information	-0.01	-0.03	-0.03	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Exclusion Cost	-0.01	-0.02	-0.02	-0.01	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Outbreak Information*		0.02	0.02		0.01	0.01
Exclusion Cost		(0.03)	(0.03)		(0.03)	(0.03)
Round Effects	No	No	Yes	No	No	Yes
Observations	3840	3840	3840	3840	3840	3840
R^2	0.00	0.00	0.01	0.00	0.00	0.01

Clustered robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

reported below are robust to a non-linear specification using random effects probit.

Result 1. Social exclusion costs increase investment in disease prevention.

Hypothesis 1 indicates that social exclusion costs will increase the probability of investing in disease prevention. Since the cost of not investing in disease prevention is higher when subjects are forced to pay social exclusion costs than when subjects are not forced to pay social exclusion costs, being in the social exclusion cost treatment should increase investment in disease prevention. Thus, being in the social exclusion cost treatment should have a statistically significant positive impact on the probability of investing in disease prevention. I hypothesize that this increase in the probability of investing in disease prevention works through two channels: an increase in the probability of starting prevention investment and a decrease in the probability of stopping prevention investment. To test this hypothesis, I use a dummy variable to indicate whether a subject was exposed to the social exclusion cost treatment.

Table 3.9 shows that being in the social exclusion cost treatment (exclusion cost) is not

Table 3.10: Effect of Changes in Disease Incidence on Decision to Start Investing in Prevention

	Decision: Start Investing					
	RE					
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	0.02***	0.02***	0.02***	0.01***	0.04**	0.04**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
Infected: 1 Lag		0.35***	0.32***	0.32***	0.32***	0.32***
		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
# Decision Changes			0.03***	0.03***	0.03***	0.03***
			(0.00)	(0.00)	(0.00)	(0.00)
Incidence: 2 Lags				-0.00		-0.00
				(0.00)		(0.00)
Squared Incidence: 1 Lag					-0.00	-0.00
					(0.00)	(0.00)
Outbreak Information	-0.04**	-0.04**	-0.03**	-0.02*	-0.03**	-0.02*
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Exclusion Cost	0.05***	0.05**	0.06***	0.05***	0.06***	0.04***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3840	3840	3840	3680	3840	3680
R^2	0.02	0.18	0.25	0.26	0.25	0.26

Clustered robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11: Effect of Changes in Disease Incidence on Decision to Stop Investing in Prevention

	Decision: Stop Investing					
	RE					
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	-0.01***	-0.01***	-0.01***	-0.02***	-0.02	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
# Infections: 1 Lag		-0.01***	-0.005***	-0.005***	-0.005***	-0.005***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
# Decision Changes			0.03***	0.03***	0.03***	0.03***
			(0.00)	(0.00)	(0.00)	(0.00)
Incidence: 2 Lags				-0.01***		-0.01***
				(0.00)		(0.00)
Squared Incidence: 1 Lag					0.00	-0.00
					(0.00)	(0.00)
Outbreak Information	0.00	0.00	0.02**	0.03***	0.02**	0.03***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Exclusion Cost	-0.05**	-0.06***	-0.05***	-0.07***	-0.05***	-0.07***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3840	3840	3840	3680	3840	3680
R^2	0.01	0.02	0.08	0.08	0.08	0.08

Clustered robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

significant in explaining either the decision to start or the decision to stop investing in disease prevention when controlling only for treatment effects and round effects. However, once I control for other factors relevant in the decision to invest in disease prevention, I demonstrate that being in the exclusion cost treatment has a significant impact on both the decision to start and the decision to stop investing in disease prevention. Table 3.10 equations 1-6 demonstrate that being in the exclusion cost treatment increases the probability that a subject starts investing in disease prevention. Similarly, table 3.11 equations 1-6 show that being in the exclusion cost treatment decreases the probability that a subject stops investing in disease prevention. Both of these results are significant at the 1% level. This indicates that facing social exclusion costs increases the probability of investing in disease prevention through two channels: an increase in the probability of starting prevention investment and a decrease in the probability of stopping prevention investment.

These results indicate that higher costs of social exclusion can increase the probability of investing in disease prevention. This has policy implications for the prevention of infectious diseases. If policy makers desire to increase investment in preventable diseases, one way in which prevention investment can be increased is through an increase in the costs of non investment. This is likely to be most effective for diseases where prevention investment is easily monitored. However, for diseases where prevention investment is not easily monitored, it may be possible to increase social exclusion by changing norms surrounding the disease prevention behaviors. For more discussion see section 3.5.

Result 2. Providing subjects with outbreak information reduces investment in disease prevention.

Hypothesis 2 states that providing subjects with outbreak information will not impact a subject's probability of investing in disease prevention. Since the replicator dynamic used in the evolutionary model of disease prevention investment allows for subjects to exhibit

unsophisticated decision making behavior, providing subjects with outbreak information should not influence the probability of investing in disease prevention. Thus, being in the outbreak information treatment should not have a statistically significant impact on the probability of investing in disease prevention. To test this, I use a dummy variable to indicate whether a subject participated in the outbreak information treatment.

Table 3.9 demonstrates that when controlling only for treatment effects and round effects, providing subjects with outbreak information (outbreak information) is not significant in explaining either the decision to start or the decision to stop investing in disease prevention. However, when controlling for other relevant factors in disease prevention investment decisions, I find that providing subjects with outbreak information is significant in explaining both the decision to start and the decision to stop investing in disease prevention. Table 3.10 equations 1-6 show that being in the outbreak information treatment actually reduces the probability that a subject starts investing in disease prevention. Furthermore, table 3.11 equations 3-6 reveal that being in the outbreak information treatment increases the probability that a subject stops investing in disease prevention. Taken together these two findings suggest that providing outbreak information decreases the probability of investing in disease prevention through two channels: a decrease in probability of starting prevention investment and an increase in the probability of stopping prevention investment.

These findings suggest that more research should be done to understand how individuals respond to information in disease contexts and how these responses can be endogenized in mathematical models of disease. Although the dynamic used in this paper assumes that providing subjects with outbreak information is not relevant in the decision to invest in disease prevention, the results suggest that subjects use outbreak information when making disease prevention investment decisions. Furthermore, figure 3.2 demonstrates that providing subjects with outbreak information is actually necessary to generate aggregate disease prevention investment that resembles model predictions. Interestingly, when subjects are

not provided with information regarding outbreak magnitude, the ambiguity regarding the disease magnitude appears to dominate the high costs of investing in disease prevention. The result is higher aggregate investment in disease prevention when subjects are not provided with information regarding the outbreak. Future research should seek to determine whether this result holds in a typical disease environment.

Result 3. Subjects exhibit prevalence elastic demand for disease prevention.

Based on prior research, I hypothesized that subjects would exhibit prevalence elastic demand for disease prevention (Geoffard and Philipson (1996); Philipson (1996); Goldstein et al. (1996); Geoffard and Philipson (1997); Ahituv et al. (1996); Philipson (2000); Chen et al. (2013); Carpenter (2016)). Prevalence elastic demand for disease prevention is characterized as an increase in disease prevention investment if disease incidence increases and a decrease in disease prevention investment if disease incidence decreases. Since I can not examine the probability of investing in disease prevention directly, I examine both the probability that a subject starts investing and the probability a subject stops investing in disease prevention. Although these dependent variables have not been analyzed directly in the literature, I hypothesize that prevalence elastic demand works through two channels as disease incidence increases: an increase in the probability a subject starts investing in disease prevention and a decrease in the probability that a subject stops investing in disease prevention. I use a one round lag in the number of disease infected individuals to measure disease prevalence.

Tables 3.10 and 3.11 provide the results for the impact of disease incidence (incidence: 1 lag) on the probability of starting prevention investment and the probability of stopping prevention investment, respectively. Like prior work, I find that subjects exhibit prevalence elastic demand for disease prevention. Table 3.10 equations 1-4 demonstrate that higher disease incidence in the prior round increases the probability that a subject starts investing in prevention in the current round. Table 3.11 equations 1-4 display evidence that a higher

disease incidence in the prior round decreases the probability that a subject stops investing in prevention in the current round. Both of these results are significant at the 1% level. These results contribute to prior work by showing that prevalence elastic demand for disease prevention works through two channels as disease incidence rises: an increase in the probability of starting prevention investment and a decrease in the probability of stopping prevention investment.

Carpenter (2016) finds that there are non-linear effects of disease incidence on the probability of investing in disease prevention. Since I am unable to explore the impact of non-linearity in disease incidence on the probability of investing in disease prevention directly, I hypothesized that the non-linearity would work through two channels as infection levels rise: the probability of starting prevention investment increases at a decreasing rate and the probability of stopping prevention investment increases at a decreasing rate. Tables 3.10 and 3.11 equations 5 and 6 provide the results of controls for non-linear effects of disease incidence on the probability of starting prevention investment and the probability of stopping prevention investment, respectively. In both tables, controlling for a non-linear effect in disease incidence (squared incidence: 1 lag) creates a multicollinearity issue. In table 3.10, inclusion of non-linear effects of disease incidence causes the coefficient on the control for lagged disease incidence (incidence: 1 lag) to lose a level of significance. In table 3.11, including the non-linear effects of disease incidence causes the coefficient on the control for lagged disease incidence (incidence: 1 lag) to lose significance. So, issues with multicollinearity prevent an analysis of the non-linear impacts of disease incidence on both the probability of starting prevention investment and the probability of stopping prevention investment.

Although this paper demonstrates that the absence of outbreak information and presence of social exclusion costs do increase investment in disease prevention, these policies do not eliminate prevalence elastic demand for disease prevention. Furthermore, while I do find that subjects exhibit prevalence elastic demand for disease prevention on average, there are some

subjects that do not change their behavior in response to the disease incidence. Instead, a small group of subjects always choose to invest in prevention or never choose to invest in prevention regardless of the magnitude of the outbreak. More research is need to understand the underlying factors that cause subjects to be of the type that exhibit prevalence elastic demand for disease prevention. With a better understanding of the type of person that exhibits prevalence elastic demand, policies can be better targeted to reduce or eliminate this behavior.

Result 4. Being infected in the previous round increases the probability that a subject starts investing in disease prevention.

It is not obvious if subjects should change their behavior in response to being infected in the previous round. If an outbreak is worse than anticipated by a subject or a subject dislikes the lower payoff received from becoming infected, then she might respond by investing in disease prevention. However, if a subject plays the odds or probability matches, then she might respond to an infection by not changing her decision to not invest in disease prevention (Vulkan (2000)). Similarly, if a subject is risk seeking in losses, then she might respond to an infection by not changing her decision to not invest in disease prevention (Tversky and Kahneman (1986)). Prior research suggests that on average a subject responds to an infection by not changing her decision to not invest in disease prevention; so, being infected in the prior round reduces the probability of investing in disease prevention (Carpenter (2016)). Since I am unable to examine the effect of an infection in the previous on the probability of investing in disease prevention directly, I examine the impact on the probability that a subject starts investing in disease prevention. I do not explore the impact of being infected in the previous round on the probability that a subject stops investing in disease prevention, because investment in disease prevention provides full protection against disease. Based on Carpenter (2016), I hypothesize that an prior period infection works to decrease investment

in disease prevention by decreasing the probability that a subject starts investing in disease prevention. To test this hypothesis, I use a dummy variable to indicate whether a subject was disease infected in the previous round.

Table 3.10 gives the results for the impact of an infection in the prior round (infected: 1 lag) on the probability that a subject starts investing in disease prevention. Unlike Carpenter (2016), I find that being infected in the prior round increases the probability of investing in disease prevention through an increase in the probability of starting prevention investment. Table 3.10 equations 2-6 demonstrate that being infected in the previous round has a positive impact on the probability of starting prevention investment, an effect significant at the 1% level. This result generates debate surrounding the issue of how subjects respond, if at all, to a prior infection.

Carpenter (2016) finds that an infection in the prior round decreases the probability that a subject invests in disease prevention. However, I find that being infected in the previous round increases the probability that a subject invests in disease prevention through an increase in the probability that a subject starts investing in disease prevention. I suggest that these conflicting results are the result of obtaining a sample of subjects that is not representative of the population's average behavioral response to an infection in the prior round despite the random selection of experiment subjects. Since theory is not clear on whether subjects should change their behavior in response to an infection, more empirical research is needed to determine which effect dominates on average. Without more research, it is not possible to determine which of the conflicting results is not representative of the average population behavior. Future research should seek to determine whether subjects change their disease prevention behavior in response to an infection in the prior round.

Result 5. A history of prior infections decreases the probability that a subject stops investing in disease prevention.

It is not immediately clear how subjects should be expected to change their decisions in response to a history of infections, if at all. If an outbreak is continuously worse than anticipated, then a subject might respond to a history of infections by investing in disease prevention. Similarly, if a subject dislikes the history of lower payoffs that result from a history of infections, then she might respond to a history of infections by investing in disease prevention. On the other hand, if a subject probability matches, then she might respond to a history of infections by continuing her strategy: not investing in prevention (Vulkan (2000)). Similarly, if a subject is risk seeking in losses, then she might respond to the history of lower payoffs resulting from a history of infections by continuing her strategy: not investing in prevention (Tversky and Kahneman (1986)). Since I am unable to look at the impact of a history of prior infections on the probability of investing in disease prevention directly, I analyze the impact of a history of prior infections on the probability that a subject stops investing in disease prevention. Although a history of prior infections might be relevant in the decision to start investing in disease prevention, a subject's prior infection status is a proxy for her history of prior infections. Therefore, I only use the actual history of infections variable in analyzing the probability that a subject stops investing in disease prevention. I hypothesize that a higher number of historical disease infections will decrease the probability that a subject stops investing in disease prevention. To test this hypothesis, I construct a count variable that indicates the total number of infections a subject has experienced in the prior rounds.

Table 3.6 demonstrates that a history of prior infections (# infections: 1 lag) is significant in explaining the decision to stop investing in disease prevention. Table 3.11 equations 2-6 show that as the number of infections a subject experiences increases, the probability that she stops investing in disease prevention decreases. This effect is significant at the 1% level.

This result indicates that increases in prior infections increase the overall probability of investing in disease prevention through a reduction in the probability of stopping prevention investment.

I find that the more infections a subject had historically, the greater the probability that she continues to invest in disease prevention. Since subjects respond to a history of infections by declining to stop investing in disease prevention, a history of prior infections works to increase the probability that a subject invests in disease prevention. I suggest two possible explanations for this behavior: either the outbreak is continually worse than anticipated or subjects dislike the lower payoff that results from a history of infections. While I can't determine which explanation, if either, is driving this behavior, future experiments can be designed in a way that allows the researcher to provide an explanation for this behavior.

Result 6. A willingness to change prevention investment strategies is associated with an increase in the probability of starting and stopping investment in disease prevention.

It seems likely that an increased willingness to test out disease prevention investment strategies should result in subjects being more likely to change their decision. In other words, if a subject is more willing to try out different disease prevention strategies, then she should be more likely to start investing in disease prevention than other subjects. Similarly, if a subject is more willing to try out different disease prevention strategies, then she should be more likely to stop investing in disease prevention than other subjects. So, I hypothesize that a willingness to try out prevention investment strategies would increase the probability that a subject both starts and stops investing in disease prevention. To measure the willingness to switch strategies, I use a count variable that indicates the total number of times that a subject changed her disease prevention investment decision.

Tables 3.10 and 3.11 show that a willingness to test different disease prevention strategies (# decision changes) is significant in explaining both the probability of starting prevention investment and the probability of stopping prevention investment, respectively. Table 3.10 equations 3-6 demonstrate that a greater willingness to test different strategies increases the probability that a subject starts investing in disease prevention. Furthermore, table 3.11 equations 3-6 provide evidence that a greater willingness to test different strategies increases the probability that a subject stops investing in disease prevention. Both of these effects are significant at the 1% level. These results support the hypothesis that a willingness to test out disease prevention strategies is associated with both an increase in the probability of starting prevention investment and an increase in the probability of stopping prevention investment.

I show that a willingness to test out disease prevention investment strategies increases the probability of starting prevention investment and the probability of stopping prevention investment. In the context of a disease outbreak, this type of behavior could potentially increase the outbreak magnitude as people switch among potential prevention strategies due to an inability to discern a causal relationship between prevention investment and disease prevention. Specifically, this type of behavior is likely to be problematic when investment in disease prevention does not provide full protection against disease. Future research should attempt to understand the individual factors that influence this type of behavior. By understanding the factors that drive this behavior, policies can be developed to encourage individuals to adopt an effective prevention strategy rather than switching among a variety of ineffective prevention strategies.

Result 7. The persistence of the impact of disease incidence on the probability of investing in disease prevention occurs through a reduction in the probability that subjects stop investing in disease prevention.

It seems reasonable that higher historical disease levels have a persistent positive impact on the probability of investing in disease prevention for a couple of reasons. First, people might recognize a pattern of higher disease levels over time, and therefore, become increasingly willing to invest in prevention. Second, people might remember the negative implications of failing to invest in disease prevention, and therefore, may be less likely to stop investing in disease prevention. In fact, Carpenter (2016) finds that historical disease incidence has a persistent positive impact on the probability of investing in disease prevention. Since I am unable to analyze the probability of investing in disease prevention directly, I explore this effect through two channels: the probability a subject starts investing in prevention and the probability a subject stops investing in prevention. I hypothesize that higher historical disease incidence increases the probability that a subject starts investing in prevention and decreases the probability that a subject stops investing in prevention. I use a two round lag in the infection levels to measure historical disease incidence.

Tables 3.10 and 3.11 report the impact of historical disease incidence (incidence: 2 lags) on the probability of starting prevention investment and the probability of stopping prevention investment, respectively. Table 3.10 equations 4 and 6 provide evidence that historical disease incidence is not significant in explaining the probability that a subject starts investing in disease prevention. Table 3.11 equations 4 and 6 shows that a higher historical disease incidence reduces the probability that a subject stops investing in disease prevention, an effect significant at the 1% level. These results suggest that higher levels of historical disease incidence increase the probability of investing in disease prevention through only one channel: a decrease in the probability that a subject stops investing in disease prevention.

These results indicate that a reduction in the probability that a subject stops investing in disease prevention is responsible for the persistent positive impact of disease incidence on the probability of investing in disease prevention discovered in Carpenter (2016). This suggests that after a real-world outbreak, prevention investment levels should remain elevated due to a reduction in the number of people that stop investing in disease prevention. Future work should seek to determine whether this pattern is observed in real-world disease prevention investment data. Furthermore, understanding the factors that cause subjects to not exhibit investment persistence is important information useful in designing policies to increase investment in disease prevention. Additionally, knowing the duration of the investment persistence is useful in determining when a population will again become disease susceptible. Future research should explore the factors that cause subjects to fail to exhibit persistent prevention investment and the duration of prevention investment persistence.

3.5 Conclusion

The increased ease of travel and changes in climate have increased the risk of global epidemics. Despite the fact that infectious diseases pose an increasing threat to society, individual responses to disease epidemics remain an understudied topic.This is due in large part to a lack of data on individual behavior during epidemics. I avoid this issue by generating data using a laboratory experiment with a simulated disease environment. The disease environment simulated in the experiment is based on a model with endogenous disease prevention investment. This mathematical model of disease provides testable hypotheses regarding the effect of outbreak information and social exclusion costs on investment in disease prevention. The purpose of this paper is to explore the impact of outbreak information and social exclusion on the probability of investing in disease prevention.

The hypotheses tested in this paper are generated from the mathematical model of disease with endogenous disease prevention investment. Previous research on the role of information in disease prevention investment suggests that the impact of information depends critically on the type of information and to whom its given. For this experiment, I explore the impact of information on outbreak magnitude and the percentage of the population investing in disease prevention. The mathematical model of disease used in the experiment suggests that this information should not significantly impact the probability of investing in disease prevention. So, I hypothesize that the provision of outbreak information will not impact disease prevention investment. Prior research on the role of social exclusion costs in disease prevention investment indicates that social exclusion costs significantly increase the probability of investing in disease prevention. The mathematical model used in this experiment demonstrates the same relationship between social exclusion costs and disease prevention investment. So, I hypothesize that the presence of social exclusion costs will significantly increase the probability of investing in disease prevention.

I find that both outbreak information and social exclusion significantly impact the probability of investing in disease prevention. In contrast to the model prediction, I show that providing subjects with information on the outbreak magnitude and the aggregate percent of individuals investing in disease prevention is important in explaining disease prevention investment. This effect is negative and significant. In other words, providing subjects with outbreak information significantly reduces the probability that they invest in disease prevention. This reduction in the probability of investing in disease prevention works through two channels: an increase in the probability of stopping prevention investment and a decrease in the probability of starting prevention investment. Additionally, I demonstrate that the presence of social exclusion costs is important in explaining disease prevention investment. This effect is positive and significant. Individuals are significantly more likely to invest in disease prevention when they face social exclusion costs for foregoing prevention investment. This effect also works through two channels: an decrease in the probability that a subject

stops investing in disease prevention and an increase in the probability that a subject starts investing in disease prevention.

The models developed in Carpenter (2014) and Carpenter (2016) provide a useful method for testing epidemic induced behavior against epidemiological models of disease. Although these models allow individuals to exhibit unsophisticated decision making behavior, the results of this experiment suggest that individuals may exhibit more sophisticated disease prevention behavior than captured by these models. Future research should seek to explore alternatives to the replicator dynamic that may better capture individual decisions in a disease context. Furthermore, this experiment explored only one type of information: outbreak information; however, in a typical disease environment individuals are exposed to many different types of information. Future research should explore the impact of other types of information on the probability of investing in disease prevention. For example, one important information type is information on the efficacy of disease prevention technologies. Providing subjects with information on the efficacy of prevention technologies is likely to increase prevention investment if efficacy is high but decrease prevention investment if efficacy is low. Understanding when to provide information about the efficacy of prevention strategies is another important area of research for the prevention of disease.

This paper provides additional support for the idea that social exclusion costs can increase investment in disease prevention. Thus, a policy implication of this paper is that disease prevention investment can be increased by raising the costs for failing to invest in disease prevention. Admittedly, this policy will work best for diseases for which prevention investment is easily monitored. Diseases that are preventable through vaccination fall into this category, and laws requiring vaccination for public school attendance were effective at achieving high levels of vaccination in the general population. However, it may be possible to raise the costs for failing to invest in disease prevention even for diseases for which prevention investment is not easily monitored. Diseases that are spread through the germs on your hands are an

example of this. Hand washing to prevent the spread of disease has only been practiced since the 19th century. Despite the fact that hand washing is relatively difficult to monitor, hand washing has become a commonly practiced disease prevention strategy. This is likely due to hand washing campaigns that spread the knowledge of the importance of hand washing for disease prevention. These campaigns likely have helped to established hand washing as a norm for disease prevention, and individuals that refuse to wash their hands are subjected to disgust and social pressure to conform to hand washing standards. Therefore, policies that seek to increase investment in disease prevention should explore ways to increase the costs of failing to invest in disease prevention.

Bibliography

Ahituv, A., Hotz, V. J., Philipson, T., 1996. The responsiveness of the demand for condoms to the local prevalence of aids. The Journal of Human Resources 31, 4.

Ai, C., Norton, E. C., 2003. Interaction terms in logit and probit models. Economics Letters 80 (1), 123–129.

Anderson, R. M., May, R. M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press.

Andrews, J. R., Basu, S., 2011. Transmission dynamics and control of cholera in haiti: An epidemic model. Lancet 377, 1248–12455.

Ashraf, N., Berry, J., Shapiro, J. M., July 2007. Can higher prices stimulate product use? evidence from a field experiment in zambia. NBER Working Paper 13247, 1–55.

Auld, M. C., 2003. Choices, beliefs, and infectious disease dynamics. Journal of Health Economics 22, 361–377.

Aziz, S. N., Boyle, K. J., Rahman, M., September 2006. Knowkenya of arsenic in drinking-water: Risks and avoidance in matlab, bangladesh. Journal of Health, Population, and Nutrition 24 (3), 327–335.

Banerjee, A. V., Duflo, E., 2007. The economic lives of the poor. Journal of Economic Perspectives 21 (1), 141–168.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1998. Learning from the behavior of others: Conformity, fads and ininformation cascades. The Journal of Economic Perspectives 12 (3), 151–170.

Buis, M. L., 2010. Stata tip 87: Interpretation of interactions in nonlinear models. The Stata Journal 10 (2), 305–308.

Caldara, M., 2013. Bidding behavior in pay-to-bid auctions: An experimental study. Working Paper.

Carpenter, A., 2014. Social Computing, Behavioral-Cultural Modeling and Prediction. Vol. 8393 of Lecture Notes in Computer Science. Springer International Publishing, Ch. Behavior in the Time of Cholera: Evidence from the 2008-2009 Cholera Outbreak in Zimbabwe, pp. 237–244.

Carpenter, A., 2016. Modeling behavior during epidemics: Understanding behavioral responses to disease.

Chen, F., Griffith, A., Cottrell, A., Wong, Y.-L., January 2013. Behavioral responses to epidemics in an online experiment: Using virtual diseases to study human behavior. PLOS One 8 (1), 1–10.

Codeco, C. T., 2001. Endemic and epidemic dynamics of cholera: The role of the aquatic reservoir. BMC Infectious Diseases 1.

Currie, J., 2000. Handbook of Health Economics. Vol. 1. Elsevier, Ch. Ch. 19 Child Health in Developed Countries, pp. 1053–1090.

Curtis, V., 2011. Why disgust matters 366 (1583), 3478–3490.

Curtis, V., Biran, A., 2001. Dirt, disgust, and disease: Is hygiene in our genes? 44 (1), 17–31.

Devota, F., Duflo, E., Dupas, P., Pariente, W., Pons, V., 2012. Happiness on tap: Piped water adoption in urban morocco. American Economic Journal: Economic Policy 4 (4), 68–99.

Duflo, E., Dupas, P., Kremer, M., September 2015. Education, hiv, and early fertility: Evidence from kenya. American Economic Review 105 (9), 2257–2297.

Dupas, P., January 2011a. Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. American Economic Journa: Applied Economics 3 (1), 1–36.

Dupas, P., September 2011b. Health behavior in developing countries. Annual Review of Economics 3, 425–499.

Dupas, P., September 2011c. Health behavior in developing countries. Annual Review of Economics 3, 425–499.

Eckel, C. C., Grossman, P. J., 2008. Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. Journal of Economic Behavior and Organization 68 (1), 1–17.

Ferguson, N., 2007. Capturing human behavior. Nature 446 (7137), 733.

Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2), 171–178.

Funk, S., Bansal, S., Bauch, C. T., Eames, K. T., Edmunds, W. J., Galvani, A. P., Klepac, P., 2015. Nine challenges in incorporating the dynamics of behavior in infectious diseases models. Epidemics 10, 21–25.

Funk, S., Salathe, M., Jansen, V. A. A., 2010. Modelling the influence of human behavior on the spread of infectious diseases: A review. Journal of the Royal Society Interface 7 (50), 1247–1256.

Gensini, G. F., Yacoub, M. H., Conti, A. A., 2004. The concept of quarantine in history: From plague to sars 49, 257–261.

Geoffard, P.-Y., Philipson, T., August 1996. Rational epidemics and their public control. International Economic Review 37 (3), 603–624.

Geoffard, P.-Y., Philipson, T., March 1997. Disease eradication: Private versus public vaccination. The American Economic Review 87 (1), 222–230.

Goldstein, K. P., Philipson, T., Joo, H., Daum, R., 1996. The effect of epidemic measles on immunization rates. The Journal of the American Medical Association 276 (1), 56–58.

Greene, W. H., 2012. Econometric Analysis, 7th Edition. Prentice-Hall, Upper Saddle River, NJ.

Greer, A., Ng, V., Fisman, D., 2008. Climate change and infectious diseases in north america: The road ahead. CMAJ 178 (6), 715–722.

Hirshleifer, D., Teoh, S. H., 2003. Herd behavior and cascading in capital mmarket: A review and synthesis. European Financial Management 9 (1), 25–66.

Hyman, J. M., Li, J., August 1997. Behavior changes in sis std models with selective mixing. SIAM Journal of Applied Mathematics 57 (4), 1082–1094.

Klein, E., Laxminarayan, R., Smith, D. L., 2007. Economic incentives and mathematical models of disease. Environment and Development Economics 12, 707–732.

Kremer, M., May 1996. Integrating behavioral choice into epidemiological models of aids. Quarterly Journal of Economics 111 (2), 549–573.

Levin, A., Lin, C.-F., Chu, C.-S. J., 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. Journal of Econometrics 108 (1), 1–24.

Madajewicz, M., Pfaff, A., Geen, A. V., Graziano, J., Hussein, I., Momotaj, H., Sylvi, R., Ahsan, H., 2007. Can information alone change behavior? response to arsenic contamination of groundwater in bangladesh. Journal of Development Economics 84, 731–754.

Manfredi, P., d'Onofrio, A. (Eds.), 2013. Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases. Springer Science and Business Media.

Measure DHS, 2005. Available datasets: Zimbabwe. http://www.measuredhs.com/data/available-datasets.cfm.

Measure DHS, 2010. Available datasets: Zimbabwe. http://www.measuredhs.com/data/available-datasets.cfm.

Morse, S. S., 2001. Plagues and Politics: Infectious Disease and International Policy. Palgrave Macmillan UK, Ch. Factors in the Emergence of Infectious Diseases, pp. 8–26.

Mukandavire, Z., Liao, S., Wang, J., Gaff, H., Smith, D. L., Jr., J. G. M., May 2011. Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in zimbabwe. PNAS 108 (21), 8767–8772.

Orenstein, W., Hinman, A. R., 1999. The immunization system in the united states – the role of school immunization laws 17, S19–S24.

Philipson, T., 1996. Private vaccination and public health: An empirical examination for u.s. measles. The Journal of Human Resources 31 (3), 611–630.

Philipson, T., 2000. Chapter 33 economic epidemiology and infectious diseases. Handbook of Health Economics 1, 1761–1799.

Sandholm, W. H., 2010. Population Games and Evolutionary Dynamics. The MIT Press.

Shiller, R. J., May 1995. Conversation, information, and herd behavior. The American Economic Review 85 (2), 181–185.

Tognotti, E., 2013. Lessons from the history of quarantine: from plague to influenza a 19 (2), 254–259.

Tversky, A., Kahneman, D., 1986. Rational choice and the framing of decisions. The Journal of Business 59 (4), S251–S278.

UN OCHA, November 2009a. Evaluation of the wash response to the 2008-2009 zimbabwe cholera epidemic and preparedness planning for future outbreaks. https://zw.humanitarianresponse.info/document/evaluation-wash-response-2008-2009-zimbabwe-cholera-epidemic-and-preparedness-planning.

UN OCHA, January 2009b. Weekly situation report on cholera in zimbabwe no. 10, 21 jan 2009. http://reliefweb.int/sites/reliefweb.int/files/resources/A40A3859537F4E9385257546006E5B01-Full_Report.pdf.

Valle, S. D., Hethcote, H., Hyman, J., Castillo-Chavex, C., 2005. Effects of behavioral changes in a smallpox attack model. Mathematical Biosciences 195, 228–251.

Vulkan, N., February 2000. An economist's perspective on probability matching. Journal of Economic Surveys 14 (1), 101–118.

Wang, X., Gao, D., Wang, J., 2015. Influence of human behavior on cholera dynamics. Mathematical Biosciences 267, 41–52.

Wilson, T. R., Fishbein, D. B., Ellis, P. A., Edlavitch, S. A., 2005. The impact of a school entry law on adolescent immunization rates 37, 511–516.

Wooldridge, J. M., 2012. Econometric Analysis of Cross Section and Panel Data, 2nd Edition. MIT Press.

World Health Organization, 2007. Combating waterborne disease at the household level. http://www.who.int/water_sanitation_health/publications/combating_diseasepart1lowres.pdf.

World Health Organization, 2008. Zimbabwe epidemiological bulletin archive. http://www.who.int/hac/crises/zwe.

World Health Organization, January 2009 b. Zimbabwe health cluster bulletin no 6 - 27 january 2009.
URL http://www.who.int/hac/crises/zmb/sitreps/health_cluster_bulletin_27jan2009/en/index.html

Appendices

A Appendix

A.1 Model

The model presented in this paper is a simplification of the more biologically complex models of cholera (Madajewicz et al., 2007; Andrews and Basu, 2011). The reason for the biological simplification of cholera in this model is that complications of the models presented in this paper involve multiple endogenous health behaviors that are assumed to be exogenous in the mathematical epidemiology literature. As the first model to incorporate endogenous behavior into a model of cholera incidence, simplifying the analysis to include only one dynamic behavior seemed the logical first step in analyzing the role of health behavior in cholera epidemics. However, future models should seek to incorporate the biological complications and endogenize related health behaviors.

Water Treatment Game

Evolutionary game theory can provide insight into how prevalence dependent behavior affects an outbreak at the aggregate level. This model uses the replicator dynamic, because a whole class of strategy revision protocols reduce to this dynamic, including some types of

imitative behavior (Sandholm, 2010). The main assumptions of the replicator dynamic are that individuals interact with each other randomly and that the population is infinite. The general description of the replicator dynamic is that individuals are programmed with a strategy, and those strategies that provide a higher than average payoff will grow while those strategies that provide a lower than average payoff will shrink. Although this dynamic is vague on how the share of the population playing a given strategy changes, as stated it can be the result of certain imitative behaviors.

Another important feature of this model is that agents are not optimizers. This assumption is clearly an abstraction as one would expect that agents do place some weight on the future. This differs from standard economic models were individuals are assumed to be optimizers. However, standard models that assume agents make optimal decisions given their constraints do not find substantial experimental support. Since the purpose of this model is to provide specific analysis the way in which cholera incidence influences water treatment behavior, the replicator dynamic may provide a better approximation of aggregate behavior than standard optimization.

In this model agents begin with an initial strategy: either treat water (T) or do not treat water (NT). Table A.1 provides the game. In this game, β is the benefit received from water treatment. For simplicity, this model assumes that the benefit received from water treatment is the same regardless of whether an agent is part of the water treating population. This assumption captures the fact that households treating their water provide some level of protection for their neighbors. If a household treats their water, then they do not become cholera infection. This, in turn, lowers their neighbors probability of infection. Here, C is the cost of treating water. These costs can include monetary costs, informational costs, time costs, and preference costs. Similarly, $C(I(t))$ represents the cost associated with not treating water. Since it is likely that not treating water becomes more costly as the infection rate increases, I assume that the cost of not treating water is an increasing function of the

infection rate. This assumption endogenizes the choice of water treatment by incorporating prevalence dependence into the choice.

Table A.1: Water Treatment Game

	T	NT
T	$\beta - C$	$\beta - C(I(t))$
NT	$\beta - C(I(t))$	$-C(I(t))$

Using the replicator dynamic, the share of the population not treating their water is

$$\frac{d\alpha}{dt} = \alpha\left(\Pi(NT) - \bar{\Pi}\right). \tag{A.1}$$

Here $\Pi(NT)$ represents the individual's payoff to not treating their water, and $\bar{\Pi}$ represents the average payoff for all strategies. Equation 1 demonstrates that the change in water treatment behavior moves according to the change in underlying payoffs. If the payoff for not treating water is larger than the average payoff, then the share of the population not treating their water will grow. Conversely, if the average payoff is larger than the payoff for not treating water, then the share of the population not treating their water will shrink. To illustrate these points, I must first derive the payoff for not treating water:

$$\Pi(NT) = (1-\alpha)\{\beta - C(I(t))\} + \alpha\{-C(I(t))\} \tag{A.2}$$

Equation 17 demonstrates that an individuals payoff for not treating his water depends upon the fraction of the population that treat their water and the fraction of the population that do not treat their water. Similarly, to determine the final behavior dynamic, I must derive

the average payoff:

$$\begin{aligned}\bar{\Pi} &= (1-\alpha)\left[(1-\alpha)\{\beta - C\} + \alpha\{\beta - C(I(t))\}\right] \\ &+ \alpha\left[(1-\alpha)\{\beta - C(I(t)\} + \alpha\{-C(I(t))\}\right]\end{aligned} \tag{A.3}$$

With a little algebra, equation 18 reduces to:

$$\bar{\Pi} = (1-\alpha)\left[\beta - (1-\alpha)C - \alpha C(I(t))\right] + \alpha\left[(1-\alpha)\beta - C(I(t))\right] \tag{A.4}$$

Now substituting equations 17 and 19 into equation 16 and doing a little algebra, I find:

$$\frac{d\alpha(t)}{dt} = \alpha(1-\alpha)\left\{(1-\alpha)\left[C - C(I(t))\right] - \alpha\beta\right\} \tag{A.5}$$

Because agents payoffs depend upon the aggregate share of agents making each choice, this model has the nice property that it can account for the role that externalities might play in behavior change. In table A.1, it is easy to see that if $C > C(I(t))$, then a portion of the population will free-ride on the water treating group in equilibrium. When cholera infection is unlikely, a larger portion of the population will be free-riders in equilibrium. However, at higher infection levels i.e. $C(I(t)) > C$, the positive externality imposed by the water treating group is not sufficiently high to allow free-riding to exist in the population.

Thus, periods of high infection make investments in water treatment more valuable thereby inducing agents to treat their water.

There are a couple important things to note with this game. First this game is intended to represent only situations where individuals must treat their water each time fresh water is collected. It is not intended to represent situations where a third party treats and supplies clean water through a water and sanitation infrastructure. Second, this game has the implicit assumption that treating water provides full protection against contracting cholera. Although treating water does not fully protect against contracting cholera in reality, this assumption greatly simplifies the model.

A.2 Model Calibrations

Data

To obtain cholera incidence data, I collected weekly cholera incidence by province for Zimbabwe from the WHO epidemiological bulletins (World Health Organization, 2008). The cholera epidemic bulletins are available beginning December 15, 2008 through June 13, 2009. Although the cholera outbreak in Zimbabwe began in August of 2008, weekly data on cholera incidence is not continuously available prior to December 15, 2008.

To obtain data on the share of the population treating their water, I used data from the Demographic Health Surveys (DHS) conducted in Zimbabwe. These surveys administer detailed individual and household level questions on demographic characteristics and health behaviors, including water treatment behavior. Because these surveys are conducted every five years on average, I do not have data during the 2008-2009 cholera outbreak. However, I am able to obtain data on water treatment in 2005 and again in 2010. Thus, data on water treatment in Zimbabwe comes from two snapshots in time: before and after the cholera

epidemic. Since these surveys occur at the household level, I aggregate the total number of households treating their water as a share of the total households surveyed in a given province for both 2005 and 2010 (Measure DHS, 2005, 2010). I used this as my estimate of the share of the population treating their water for each province.

Data on the biological parameter values were taken from the mathematical epidemiology literature. It should be noted that many of the parameters occur for a range of values. However, in the literature it is standard practice to calibrate an outbreak model with a single parameter value from within the biological range. Since the purpose of this paper is to explore the dynamic relationship between cholera infection and water treatment behavior, I assume a single value from within the biological range for each biological parameter. Table A.2 provides the non province specific biological parameters used to simulate the model. This table also provides the sources used to obtain information on the parameter values. Table A.3 gives the province specific biological parameters. Additionally, table A.3 provides the province specific assumed values for the benefit of water treatment, β; the cost of water treatment, C; and the intensity of the outbreak perception, λ. To calibrate the model, I must also use initial values for each of the dyanmic equations.

Table A.4 shows the initial values used for each dynamic equation. Data on the number of individuals with temporary cholera immunity at the start of the outbreak, $R(t)$, is unavailable. I assume that at the start of the outbreak, there are no recovered and temporarily cholera immune individuals. The reason for this assumption is that cholera immunity last less than a full year. At the beginning of an outbreak very few, if any, people will have cholera immunity. For the cholera incidence, $I(t)$, initial values, I use the number of people infected with cholera during the week of December 15, 2008, the first week that continuous weekly cholera incidence data is available. Because data on water treatment behavior is unavailable during the outbreak, I approximate this behavior using the share of the population not treating their water in 2005 as the initial value for $\alpha(t)$. Similarly, data on the level of

V. cholerae bacteria in the aquatic reservoir, $B(t)$, and the number of cholera susceptible individuals, $S(t)$, is unavailable. So, I chose the initial values that provided the best fit between the model calibration and the observed data.

Complete Results

Using the parameters in tables A.2 through A.4, I simulate equations 3-7 (see paper) for each province in Zimbabwe. The first graph for each province provides the actual cholera incidence for each week between December 15, 2008 and June 13, 2009 and the model predicted incidence during the same time period. The second graph for each province demonstrates the change in the proportion of the population treating their water. For the calibrations, I assume that the initial share of the population treating their water during the week of December 15, 2008 is the same as the share of the population treating their water in 2005. I show for each province that incorporating the change in the share of the population treating their water between 2005 and 2010 into the model of cholera incidence provides a calibrated outbreak that closely matches data from the initial outbreak. This model assumes that the full change in the share of the population treating their water occurred during the cholera outbreak. While it is possible that this change in water treatment behavior may have occurred at some point prior to the outbreak, it is unlikely due to the fact that there was no other major catalyst for behavior change during this period.

It should be noted here that the model fails to match the observed drop in cholera cases that occurred in most provinces around week 7 of the outbreak. Instead, these calibrations demonstrate that converging to the share of the population treating their water in 2010 allows cholera to persist in the population. An explanation for the divergence between the observed data and the model calibration will be presented in section A.3. While the benefits and costs assumed in this calibration appear to qualitatively match the initial stages of the observed outbreak, exploring the model under different cost and benefit structures will

Table A.2: Model Parameters

Parameters	Literature Values	Weekly Values	Literature Source
δ	(1/30 days)	(7/30 weeks)	1 2 3
θ	10 cells/ ML/ day	70 cells/ML/ week	2
γ	(1/5 days)	(7/5 weeks)	2 3 4
k	10^5 cells	10^5 cells	2 3 5 6
μ	(1/43 years)	1/(43*52 weeks)	7
q	———	0.99	No Sources Available
ω	(1/0.8 years)	1/(0.8*52 weeks)	2 3 8

1. Hartley, D.M, Morris, J.G., Smith, D.L.: Hyperinfectivity: A critical element in the ability of v. cholerae to cause epidemics. PLOS Medicine 3(1) (January 2006) 0063-0069. 2. Mukandavire, Z., Liao, S., Wang, J., Gaff, H., Smith, D.L., Morris, J.G.: Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in zimbabwe. PNAS 108(21) (May 2011) 8767-8772. 3. Andrews, J.R., Basu, S.: Transmission dynamics and control of cholera in haiti: An epidemic model. Lancet 377(9773) (April 2011) 1248-1255 4. Kaper, J.B., Morris, J.G., Levine, M.M.: Cholera. Clinical Microbiology Reviews 8(1) (January 1995) 48-86. 5. Nelson, E.J., Harris, J.B., Morris, J.G., Calderwook, S.B., Camilli, A.: Cholera transmission: The host, pathogen, and bacteriophage dynamic. Nature 7 (October 2009) 693-702. 6. Codeco, C.T.: Endemic and epidemic dynamics of cholera The role of the aquatic reservior. BMC Infectious Diseases 1(1) (2001) 7. World Health Organization: Life expectancy at birth: Both sexes 2000 (2013) $http://gamapserver.who.int/gho/interactive\_charts/mbd/life\_expectance/atlas.html$. 8. King, A.A., Ionides, E.L., Pascual, M., Bouma, M.J.: Inapparent infections and cholera dynamics. Nature 484 (August 2008) 877-880.

demonstrate the impact of costs and benefits of water treatment on the cholera epidemic magnitude.

Table A.3: Province Specific Model Parameters

	Parameters				
	μ_{c1}	N_1	β	C	λ
Bulawayo	0.041	718,278	150	500	100
Harare	0.035	2,012,784	215	400	100
Manicaland	0.048	1,665451	140	400	100
Mashonaland Central	0.029	1,056,666	140	400	100
Mashonaland East	0.070	1,196,772	180	500	100
Mashonaland West	0.042	1,300,012	102	500	100
Masvingo	0.059	1,401,672	62	500	100
Matabeleland North	0.064	748,317	21	500	100
Matabeleland South	0.030	693,230	56	500	100
Midlands	0.045	1,554,058	110	500	100
Zimbabwe	0.046	12,347,240	135	500	100

1. World Health Organization (2013)

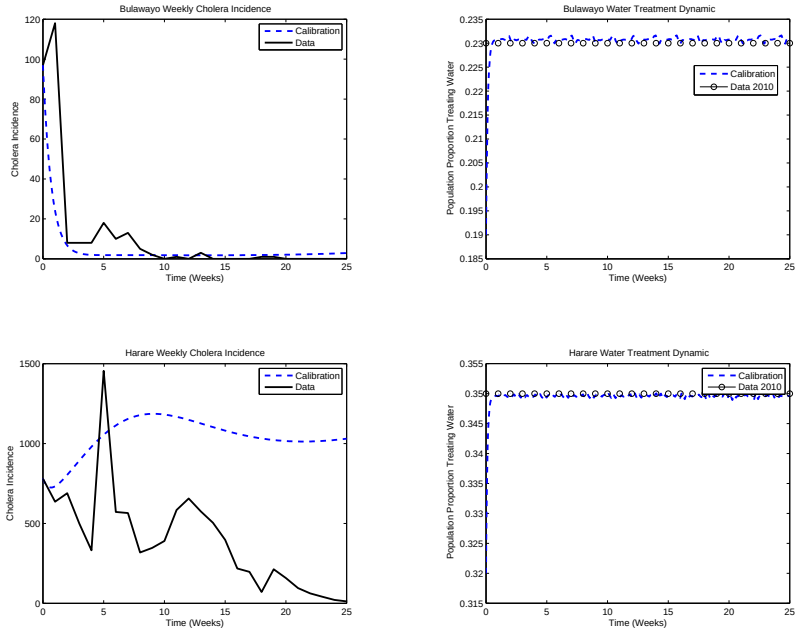
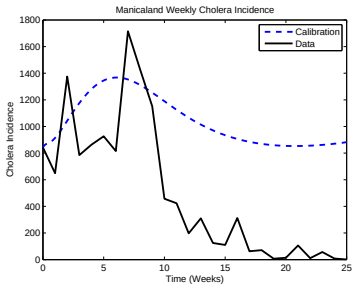
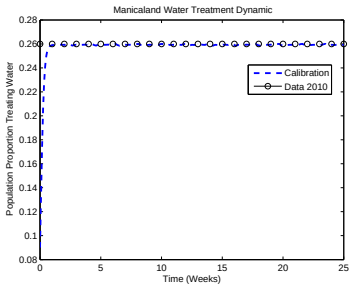


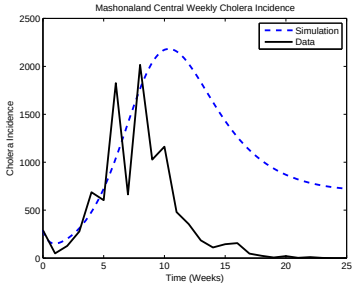
Table A.4: Initial Model Values

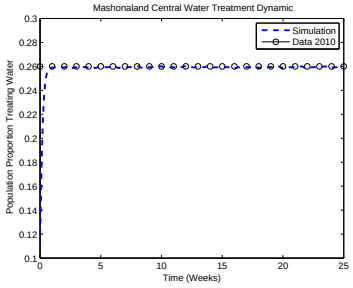
	Initial Parameter Values				
	S(t)	$I(t)_1$	R(t)	B(t)	$\alpha(t)_2$
Bulawayo	50	97	0	10^{-10}	0.81
Harare	15,000	780	0	10^{5}	0.68
Manicaland	16,500	842	0	10^{5}	0.91
Mashonaland Central	25,000	291	0	10^{-10}	0.89
Mashonaland East	3,000	199	0	10^{6}	0.92
Mashonaland West	19,500	2,650	0	10^{-10}	0.86
Masvingo	17,500	945	0	10^{-10}	0.90
Matabeleland North	1,000	0	0	10^{6}	0.95
Matabeleland South	2,000	146	0	10^{6}	0.91
Midlands	11,000	41	0	10^{5}	0.95
Zimbabwe	40,000	5,993	0	10^{-10}	0.87

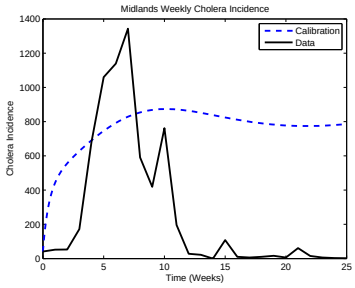
1. World Health Organization (2008)
2. Measure DHS (2005)

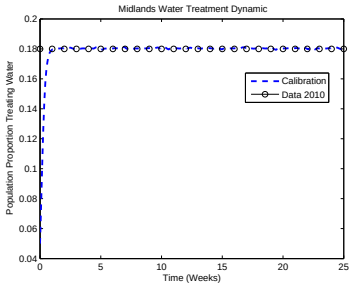


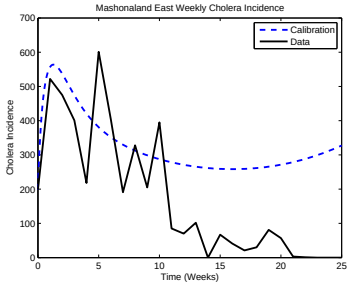


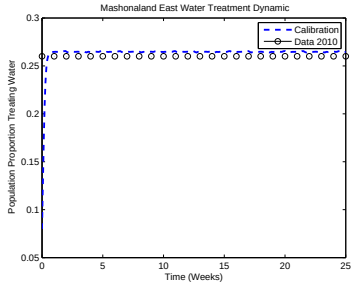


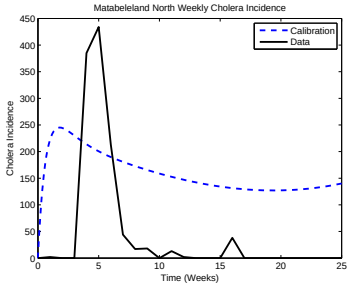


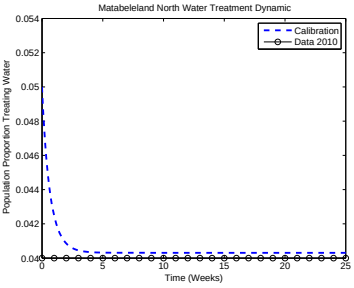


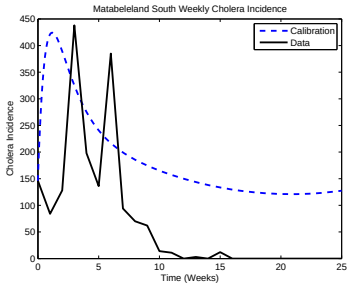


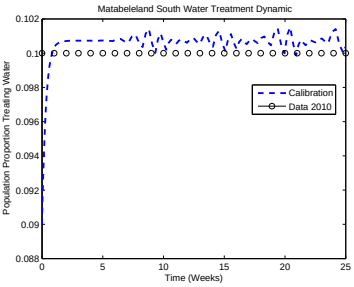

Mashonaland East Weekly Cholera Incidence
Calibration
Data
Cholera Incidence
Time (Weeks)

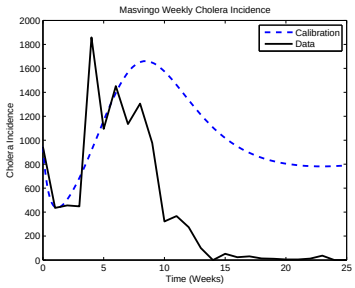

Mashonaland East Water Treatment Dynamic
Calibration
Data 2010
Population Proportion Treating Water
Time (Weeks)

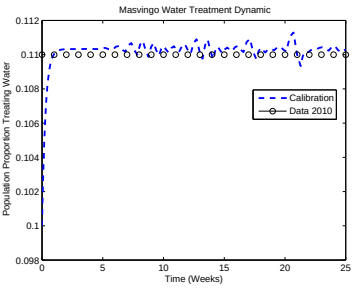

Matabeleland North Weekly Cholera Incidence
Calibration
Data
Cholera Incidence
Time (Weeks)

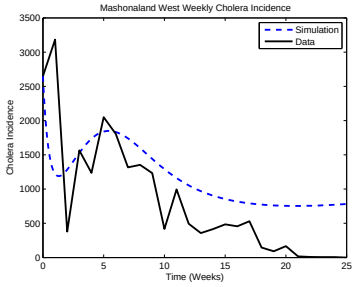

Matabeleland North Water Treatment Dynamic
Calibration
Data 2010
Population Proportion Treating Water
Time (Weeks)

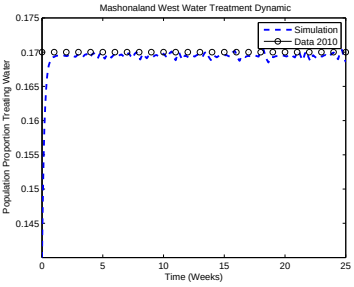

Matabeleland South Weekly Cholera Incidence
Calibration
Data
Cholera Incidence
Time (Weeks)

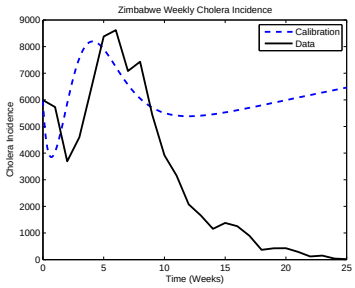

Matabeleland South Water Treatment Dynamic
Calibration
Data 2010
Population Proportion Treating Water
Time (Weeks)

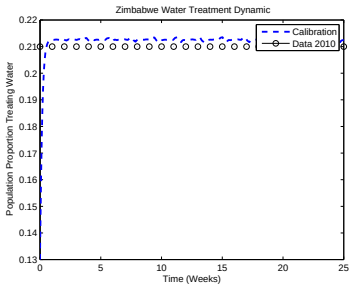

Masvingo Weekly Cholera Incidence
Calibration
Data
Cholera Incidence
Time (Weeks)


Masvingo Water Treatment Dynamic
Calibration
Data 2010
Population Proportion Treating Water
Time (Weeks)


Mashonaland West Weekly Cholera Incidence
Simulation
Data
Cholera Incidence
Time (Weeks)


Mashonaland West Water Treatment Dynamic
Simulation
Data 2010
Population Proportion Treating Water
Time (Weeks)





A.3 Sensitivity Analysis

In this section, I examine the sensitivity of the simulations presented in section A.2 to changes in both the benefits of water treatment, β, and the costs of water treatment, C. Since changing both the benefits and costs simultaneously would create confusion about the catalyst underlying any subsequent changes in cholera incidence or in the share of the population treating their water, I change only one factor while holding all other model parameters constant. Thus, the sensitivity analysis is divided into two sections: benefit sensitivity analysis and cost sensitivity analysis.

Benefit Sensitivity Analysis

This section explores the effect of changes in water treatment benefits on both the magnitude of a cholera outbreak and the share of the population treating their water. As in section A.2, the non province specific model parameters used in the sensitivity analysis are provided in table A.2, and the initial values for each of the dynamic equations used in this analysis are listed in table A.4. The province specific parameters used in each of the calibrations in this section are available in table A.5. From these tables, it is easy to see that the only parameters changed for these calibrations are the values of β. Furthermore, the benefit level used in calibration 1, provided in table A.5 column 3, is the same as the benefit level used for the calibration in section A.2. Thus, calibrations 2-4 demonstrate the outbreak magnitude and

the share of the population treating their water under alternative water treatment benefits.

In each calibration, the benefits of water treatment, β, are greater than the benefits of water treatment used in the previous calibration. As an example, the benefits of water treatment used in calibration 2 are larger than the benefits of water treatment used in calibration 1. However, the relationship between the benefits of water treatment, β, and the cost of water treatment, C, varies by calibration. This occurs because the cost of water treatment is non-changing across calibrations while the benefit of water treatment is increasing across calibrations. In calibrations 1 and 2, the benefit of water treatment, β, is lower than the cost of water treatment, C, for each province. However, the relative difference between the benefits of water treatment and the costs of water treatment varies by calibration and by province. In calibration 3 and 4, the benefits of water treatment, β, exceeds the cost of water treatment, C. Again, however, the relative difference between the benefits of water treatment and the costs of water treatment varies by calibration and by province.

Comparing the calibration 1-4 on each of the graphs, it is easy to see that increasing the benefits of water treatment induces a greater share of the population to treat their water. This, in turn, reduces the overall outbreak magnitude. Therefore, the benefits of water treatment are negatively related to outbreak magnitude. However, an important finding is that simply raising the benefits of water treatment above the cost of water treatment, as in calibration 3 and 4, does not induce the entire population to treat their water. This is partially due to the model structure. In this model, raising the benefits of water treatment, β, increases the payoff for both the share of the population treating their water and the share of the population not treating their water. As a simplifying assumption, I have assumed that everyone recieves the same benefit of water treatment regardless of their own strategy. However, it might be more realistic to assume that individuals treating their water recieve a higher benefit from water treatment than those who do not treat their water. In this case, increasing the benefits of water treatment, β, would induce greater population water

treatment and lower outbreak magnitude than that presented in calibrations 1-4.

The role of perceived water treatment benefits in inducing water treatment behavior is important for this model. Most previous research on low water treatment in developing countries has focused on barriers to water treatment like information costs, time costs, and monetary costs associated with purchasing and using water treatment products. The underlying idea with this research is that significant barriers prevent individuals from treating their water. However, because calibration 1 provides the best fit with the data, the analysis suggests that the perceived benefits of water treatment are significantly lower than the costs of water treatment. Thus, even substantial reductions in the barriers to water treatment may not overcome the low perceived benefits of water treatment. Therefore, raising the perceived benefits of water treatment could be important in inducing individuals to treat their water. This could potentially be achieved through informational campaigns or by tying the benefits of water treatment to other goods with high benefits. However, the percieved benefits of water treatment will need to be substanitally higher than the costs of water treatment in order to induce the population to coordinate on the level of population water treatment necessary to prevent cholera outbreaks.

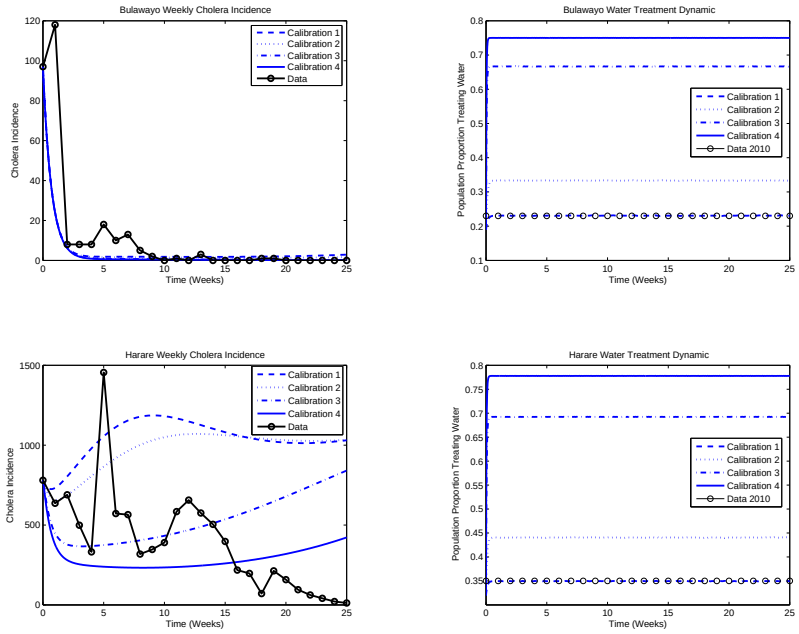
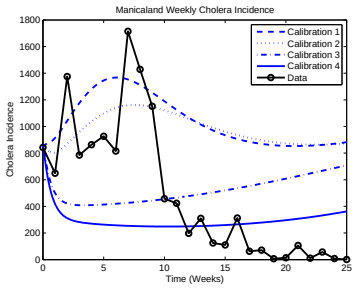
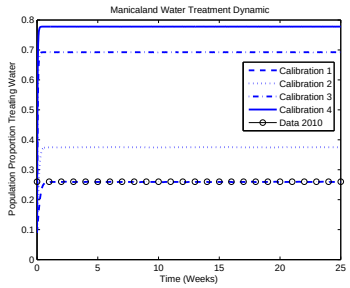


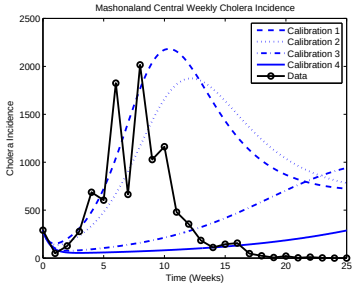
Table A.5: Province Specific Model Parameters

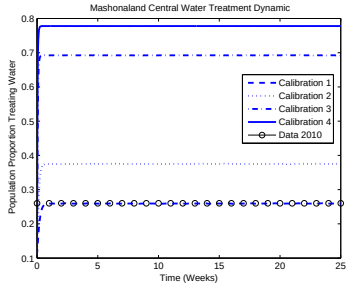
	μ_{c1}	N_1	β Values				C	λ
			Calibration 1	Calibration 2	Calibration 3	Calibration 4		
Bulawayo	0.041	718,278	150	250	1000	1500	500	100
Harare	0.035	2,012,784	215	315	900	1400	400	100
Manicaland	0.048	1,665,451	140	240	900	1400	400	100
Mashonaland Central	0.029	1,056,666	140	240	900	1400	400	100
Mashonaland East	0.070	1,196,772	180	280	1000	1500	500	100
Mashonaland West	0.042	1,300,012	102	202	1000	1500	500	100
Masvingo	0.059	1,401,672	62	162	1000	1500	500	100
Matabeleland North	0.064	748,317	21	121	1000	1500	500	100
Matabeleland South	0.030	693,230	56	156	1000	1500	500	100
Midlands	0.045	1,554,058	110	210	1000	1500	500	100
Zimbabwe	0.046	12,347,240	135	235	1000	1500	500	100

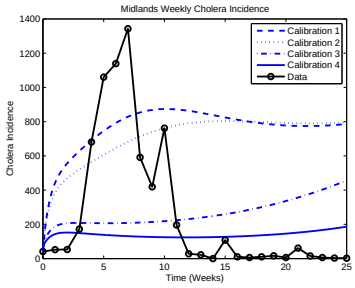
1. World Health Organization (2008)

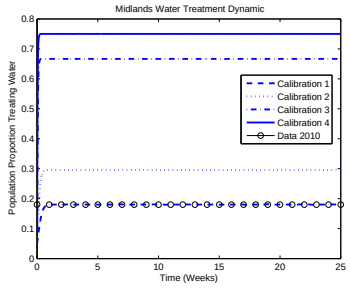

Manicaland Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

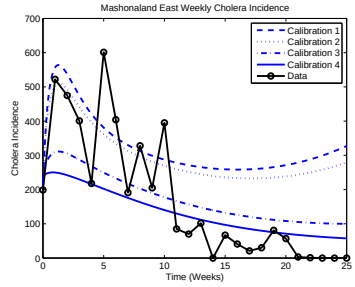

Manicaland Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010

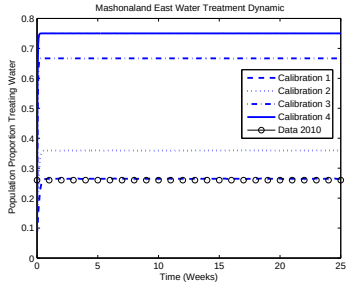

Mashonaland Central Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

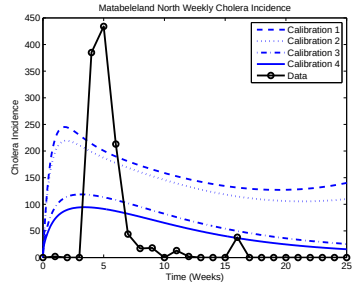

Mashonaland Central Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010

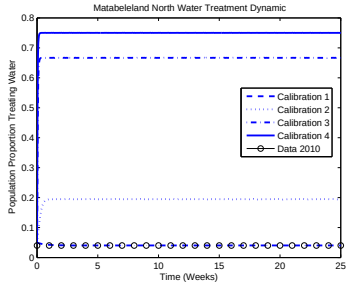

Midlands Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

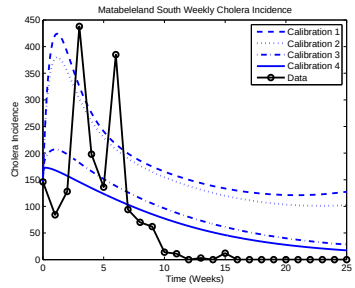

Midlands Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010

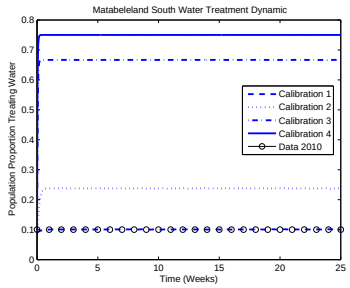

Mashonaland East Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

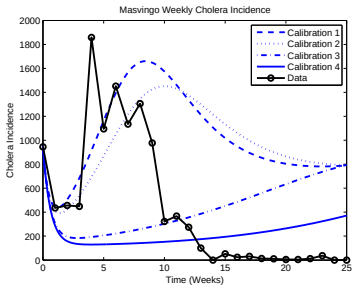

Mashonaland East Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010

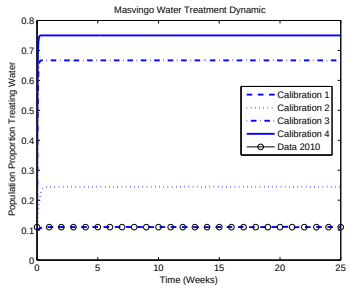

Matabeleland North Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

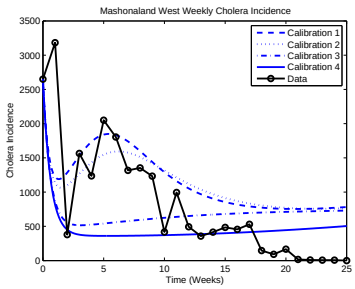

Matabeleland North Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010

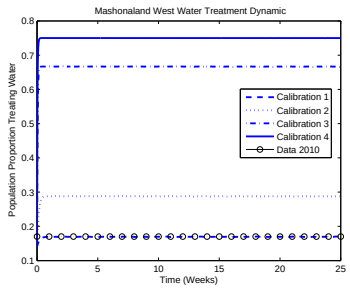


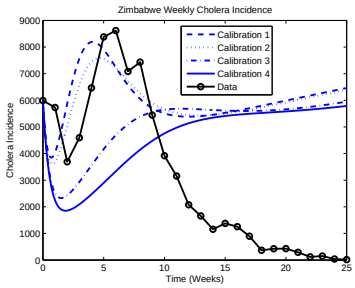


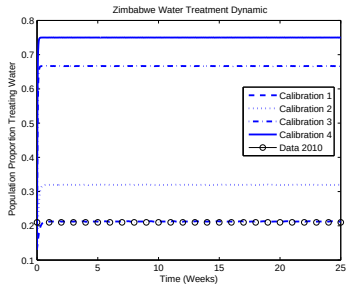












A.4 Cost Sensitivity Analysis

This section investigates the effect of changes water treatment costs on both the magnitude of a cholera outbreak and the share of the population treating their water. Like the Benefit Sensitivity Analysis section , the non province specific model parameters used in the

sensitivity analysis are provided in table A.2, and the initial values for each of the dynamic equations used in this analysis are available in table A.4. The province specific parameters used in each of the calibrations in the section are available in table A.6. From these tables, it is easy to see that the only parameters changed for these calibrations are the values of C. Furthermore, the cost level used in calibration 1, provided in table A.6 column 4, is the same as the cost level used for the calibration in section A.2. Thus, calibrations 2-4 demonstrate the outbreak magnitude and the share of the population treating their water under alternative water treatment costs.

In each calibration, the costs of water treatment, C, are less than the costs of water treatment used in the previous calibration. As an example, the costs of water treatment used in calibration 2 are less than the costs of water treatment used in calibration 1. However, the relationship between the costs of water treatment, C, and the benefits of water treatment, β, varies by calibration. This occurs because the benefit of water treatment is non-changing across calibrations while the cost of water of water treatment is decreasing across calibrations.In calibrations 1 and 2, the cost of water treatment, C, is greater than the benefits of water treatment, β, for each province. However, the relative difference between the costs of water treatment and the benefits of water treatment varies by calibration and by province. In calibration 3, the benefit of water treatment, β, exceeds the cost of water treatment, C, for every province except Masvingo, Matabeleland North, and Matabeleland South. Due to extremely low benefits of water treatment in these provinces, the cost of water treatment, C, exceeds the benefit of water treatment, β, for calibration 3. In calibration 4, the costs of water treatment are zero. Since the benefit of water treatment is positive for each province, the benefit of water treatment exceeds the cost of water treatment for every province in calibration 4.

Comparing calibrations 1-4 on the graphs, it is easy to see that reducing the costs of water treatment induces a greater share of the population to treat their water. This, in turn, re-

duces the overall outbreak magnitude. Therefore, the costs of water treatment are positively related to outbreak magnitude. Thus, as expected, research that focuses on eliminating the potential costs or barriers to water treatment can have a significant impact on the level of cholera present in the population. However, when the benefits of water treatment are close to zero, even large reductions in the cost of water treatment will fail to greatly increase the share of the population treating their water. This is best illustrated by calibration 3 for Masvingo, Matabeleland North, and Matabeleland South. The cost of water treatment is substantially reduced in this calibration. However, due to very low benefits of water treatment, the share of the population treating their water remains substantially below 50%. This could explain why some research has found that adoption of water treatment is significant when the price of water treatment solution is fully subsidized. However, for small price increases, where the price is still highly subsidized, the demand for water treatment products falls significantly (Ashraf et al., 2007). This model suggests that these empirical findings are due to near zero perceived water treatment benefits.

Additionally, calibration 4 provides a characterization of the outbreak magnitude and share of the population treating their water under zero water treatment cost. As can be seen in each of the provincial calibrations, introduction of zero costs of water treatment causes the level of infected individuals to fall quickly. In the model, this is because the cost of infection is higher than the cost of water treatment for any positive level of infection; thus, the water treatment strategy grows. Interestingly, during the 2008-2009 cholera outbreak, the WHO along with other NGOs organized and distributed clean water, water treatment solution, and sanitation information to cholera affected areas beginning around week 7 of the calibrations and lasting for duration of the outbreak (World Health Organization, 2009 b.). This mid-outbreak action effectively reduced the cost of water treatment to zero for individuals located in epidemic areas. This model assumes a positive cost of water treatment that is constant across time. So, the change in the cost of water treatment around week 7 caused by the WHO is not capture by this model. However, using calibration 4 it is easy to see that lowering

the cost of water treatment to near-zero levels would induce the water treatment strategy to immediately grow. This, in turn, would cause the epidemic to drop off as it does around week 7 in the observed data. Therefore, calibration 4 provides support for the argument that the drop-off in cholera cases observed in the data around week 7 is likely due to WHO interventions. Thus, calibration 1 or the calibrations in section A.2 can be thought of as the level of cholera that would have persisted in each province had the WHO and others not intervened.

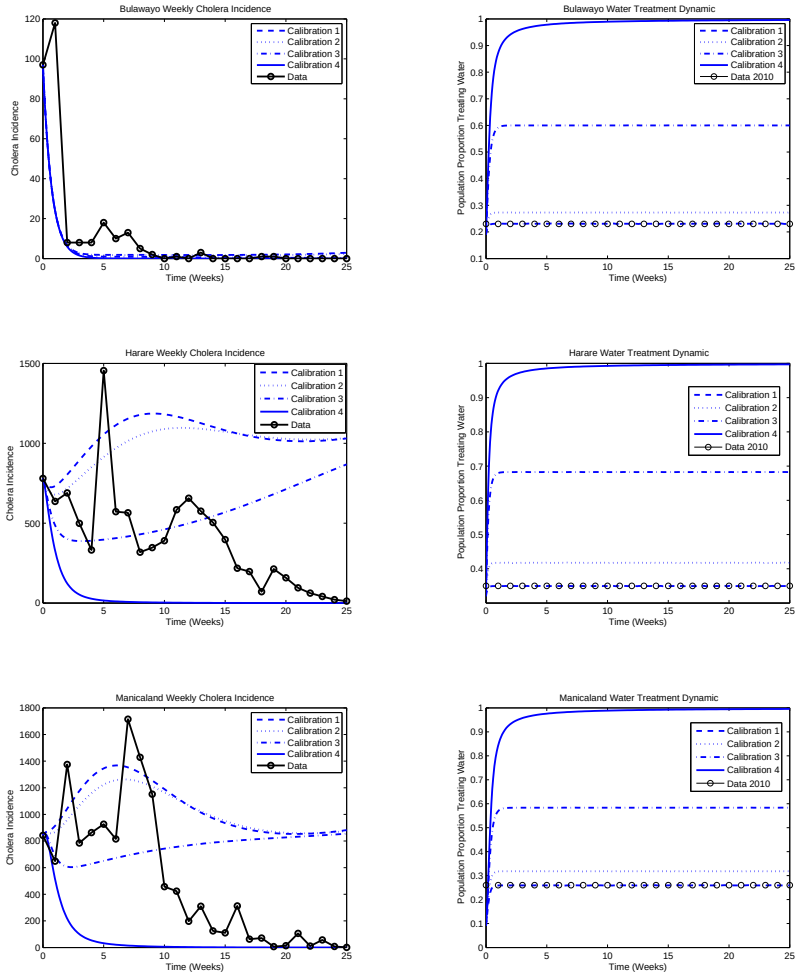
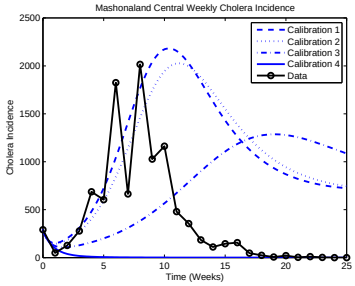
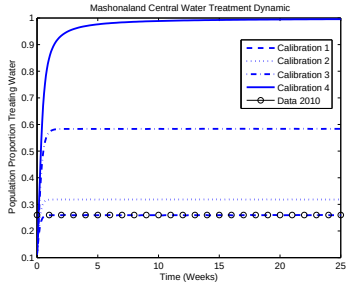


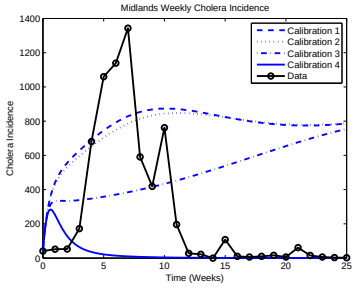
Table A.6: Province Specific Model Parameters

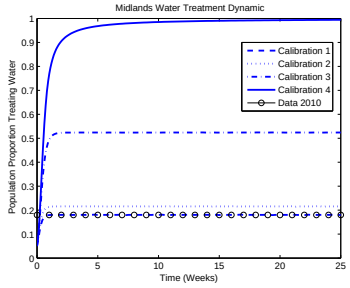
	μ_{c1}	N_1	β	C Values				λ
				Calibration 1	Calibration 2	Calibration 3	Calibration 4	
Bulawayo	0.041	718,278	150	500	400	100	0	100
Harare	0.035	2,012,784	215	400	300	100	0	100
Manicaland	0.048	1,665,451	140	400	300	100	0	100
Mashonaland Central	0.029	1,056,666	140	400	300	100	0	100
Mashonaland East	0.070	1,196,772	180	500	400	100	0	100
Mashonaland West	0.042	1,300,012	102	500	400	100	0	100
Masvingo	0.059	1,401,672	62	500	400	100	0	100
Matabeleland North	0.064	748,317	21	500	400	100	0	100
Matabeleland South	0.030	693,230	56	500	400	100	0	100
Midlands	0.045	1,554,058	110	500	400	100	0	100
Zimbabwe	0.046	12,347,240	135	500	400	100	0	100

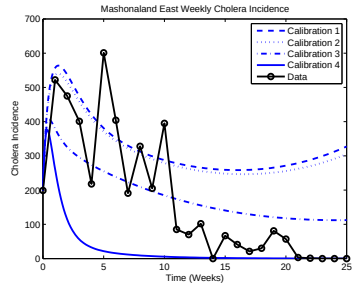
1. World Health Organization (2008)

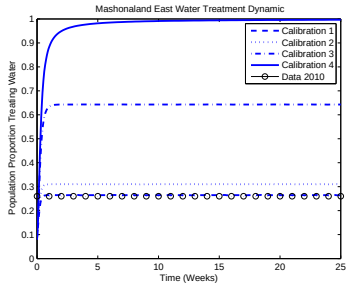

Mashonaland Central Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

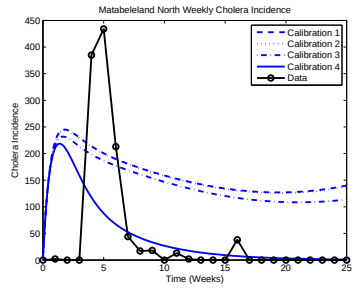

Mashonaland Central Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010

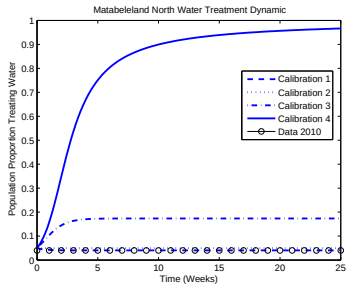

Midlands Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

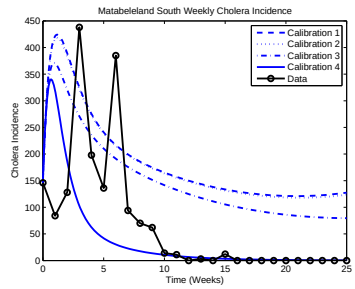

Midlands Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010

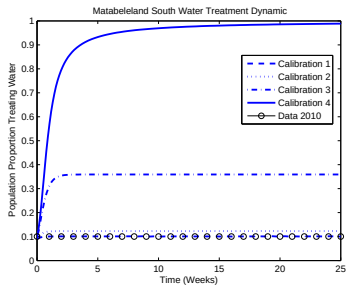

Mashonaland East Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data

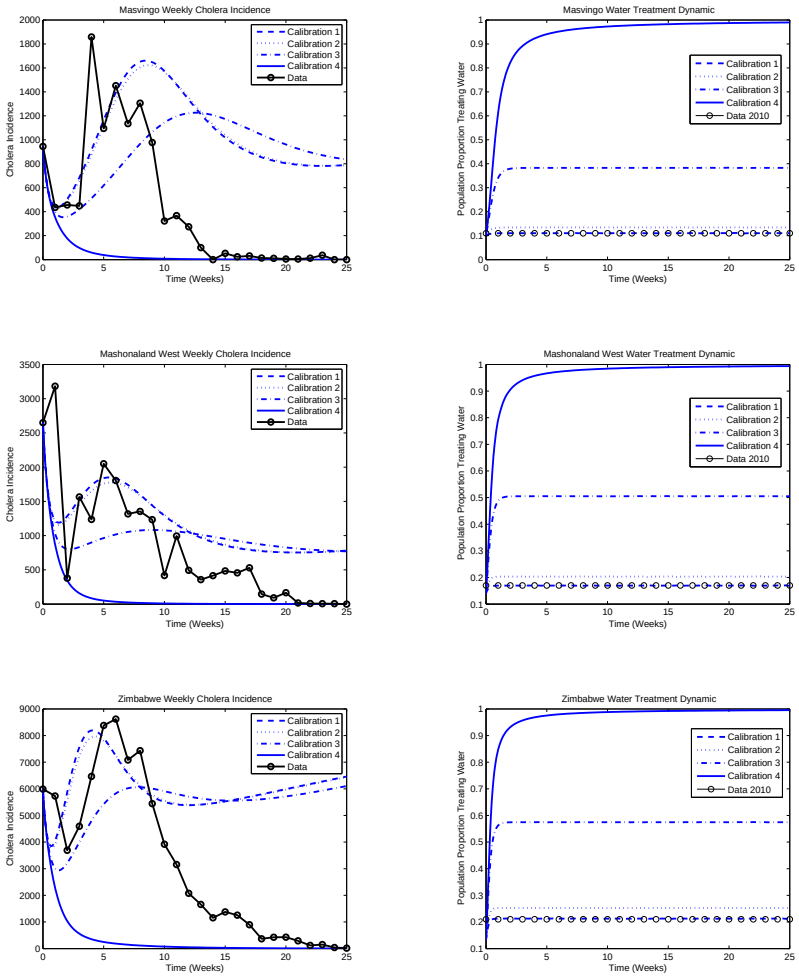

Mashonaland East Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010


Matabeleland North Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data


Matabeleland North Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010


Matabeleland South Weekly Cholera Incidence
Cholera Incidence
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data


Matabeleland South Water Treatment Dynamic
Population Proportion Treating Water
Time (Weeks)
Calibration 1
Calibration 2
Calibration 3
Calibration 4
Data 2010



Discussion

The key finding of this model is that if water treatment behavior is prevalence dependent, then the population can converge to a water treatment share that enables cholera to persist in the population. This result can be seen in the graphs presented in section A.2. High costs and low benefits of water treatment enable cholera to persist in the population. This is due to the fact that the water treatment strategy will only grow when the costs of water treatment are less than the costs of cholera infection. However, high costs of water treatment imply that high levels of cholera can occur before water treatment strategies will change.

Biological research on cholera has attempted to explain why cholera may persist in pop-

ulations. While there are undoubtedly biological reasons that enable cholera to persist in populations, section A.3 calibration 4 demonstrates that even if V. cholerae bacteria exists in the aquatic reservior, cholera can be eliminated with zero or near zero costs of water treatment. Therefore, the role of prevalence dependent water treatment behavior should not be ignored in research and policies that aim to eliminate cholera. This seems to fit with historical evidence on the elimination of water transmitted diseases in developed countries, as well. Historically, water transmitted diseases were eliminated through public and private water and sanitation infrastructure not through individual water treatment (Currie, 2000). Countries that developed water and sanitation infrastructures significantly reduced the time costs associated with water treatment thereby reducing overall costs of water treatment. This effectively forced the population to coordinate on the water treatment equilibrium which subsequently reduced water transmitted disease.

Additionally, the calibrations in section A.2 provide a good characterization of the initial cholera outbreak, but they fail to match the drop-off in cholera incidence that began around week 7 of the observed data. The free provision of clean water, water purification tablets, and sanitation information during the outbreak effectively reduced the cost of water treatment to zero in the middle of the outbreak. However, this change in the cost of water treatment is not capture by the model. Therefore, the outbreak magnitude presented in the calibrations in section A.2 should be thought of as the epidemic magnitude in absence of the WHO interventions. Unfortunately, these interventions were short-term policies aimed at outbreak elimination. Although they were successful in their aim, cancellation of free water provision, without corresponding changes to the costs and benefits associated with water treatment, will increase population susceptibility to cholera infection.

While this model provides a first pass at incorporating outbreak related behavior into a model of cholera outbreak, it does not provide a full biological analysis of a cholera epidemic. Future research should seek to incorporate more biological realness into the model of

cholera incidence to provide better analysis of potential epidemic magnitudes. Specifically, incorporating the highly infectious cholera state and the associated health behavior, post-defecation hand-washing, will provide a more thorough analysis of the impacts of behavior change on cholera epidemics. Additionally, it is not clear how the promotion of one health behavior affects engagement in other health behaviors. In this case, it is important to understand whether water treatment behavior and hand-washing behavior are substitutes or complements. Future research examining both health behaviors should aim to understand whether policies that promote one behavior will crowd-out the other behavior or whether those policies that promote one behavior will work to complement existing health behaviors.

B Appendix

Behavior Dynamic Derivation

Table B.7: Disease Prevention Investment Game

	Invest	Do Not Invest
Invest	b-c	b-c
Do Not Invest	b-c(I(t))	-c(I(t))

REPLICATOR DYNAMIC: I use an evolutionary game with the replicator dynamic to derive the changes in the share of the population not investing in disease prevention. The replicator dynamic can be derived from a variety of strategy revision protocols. However, for simplicity, I describe the dynamic as arising from a process where individuals are randomly paired in each time period. If an individual is paired with someone that has a different investment strategy, then she can compare the payoffs under the two strategies. If the alternative strategy provides a higher payoff, then she will adopt the alternative strategy with some probability.

GAME: I provide the game and payoffs used to derive the dynamic in table B.7. Following

convention for evolutionary games, I report only one set of payoffs in the table, the payoffs for the left hand side (LHS) player. The row headers indicate the two possible strategies available to the LHS player: invest in disease prevention or do not invest in disease prevention. The column headers give the two possible strategies available to the random player with whom the LHS player is periodically paired: invest in disease prevention or do not invest in disease prevention.

PAYOFFS: I use the payoffs given in table B.7 to derive the dynamic for changes in the share of the population not investing in disease prevention. In this game, b represents the benefit from avoiding infection, and c is the cost of investing in disease prevention. The cost of not investing in disease prevention is given by $c(I(t))$. For simplicity, I assume that the cost of not investing in disease prevention is the same regardless of whether an individual is paired with someone that invests in disease prevention. For individuals that choose not to invest in disease prevention and are paired with someone that invests in disease prevention, the cost of not investing in disease prevention represents the cost of increased disease monitoring and the cost of social ostracism for non investment. For individuals that choose not to invest in disease prevention and are paired with someone that does not invest in disease prevention, the cost of not investing in disease prevention represents the cost of becoming infected. The cost of becoming infected includes both the costs associated with treating the illness and the costs associated with missing work.

$$\frac{d\alpha(t)}{dt} = \alpha(t)\left(\Pi(NI) - \bar{\Pi}\right) \tag{B.6}$$

REPLICATOR EXPLANATION: Equation 5 uses the replicator dynamic to capture the evolution of the change in the share of the population not investing in disease prevention, $\frac{d\alpha(t)}{dt}$ (Sandholm (2010)). It is easy to see from this equation that the change in the share of

the population not investing in disease prevention is a function of the share of the population not investing in disease prevention, $\alpha(t)$; the payoff for not investing in disease prevention, $\Pi(NI)$; and the average payoff, $\bar{\Pi}$. This equation demonstrates that the share of the population not investing in disease prevention will grow if the payoff from not investing in disease prevention is larger than the average payoff. Conversely, the share of the population not investing in disease prevention will shrink if the average payoff is larger than the payoff from not investing in disease prevention. The payoff for not investing in disease prevention is:

$$\Pi(NI) = (1 - \alpha(t))\,[b - c(I(t))] + \alpha(t)\,[-c(I(t))] \tag{B.7}$$

The average payoff is:

$$\bar{\Pi} = (1-\alpha(t))\,\{(1 - \alpha(t))\,[b - c] + \alpha(t)\,[b - c]\} + \alpha(t)\,\{(1 - \alpha(t))\,[b - c(I(t))] + \alpha(t)\,[-c(I(t))]\} \tag{B.8}$$

With a little algebra, equation 7 reduces to:

$$\bar{\Pi} = (1 - \alpha(t))\,[b - c] + \alpha(t)\,\{[(1 - \alpha(t))b - c(I(t))]\} \tag{B.9}$$

Now substituting equations 6 and 8 into equation 5 and doing a little algebra, I find:

$$\frac{d\alpha(t)}{dt} = \alpha(t)(1-\alpha(t))\left\{c - c(I(t)) - \alpha(t)b\right\} \tag{B.10}$$

B.1 Non-Linear Model Specification

SPECIFICATION PROBLEMS. Determining the appropriate model specification for this analysis is problematic, because both linear and non-linear specifications potentially violate some specification assumptions. The linear probability model with both fixed and random effects reported in the main body of the paper requires that the dependent variable is continuous. However, in this paper, the dependent variable is the decision to invest in disease prevention, a binary variable. Violation of this assumption introduces heteroskedasticity into the standard errors, an issue that can be corrected, and violates the assumption that the errors are normally distributed, an issue that can not be corrected (Wooldridge (2012)). Both linear and non-linear model specifications require that the model error terms are uncorrelated with each other unless the correlation structure can be specified using fixed or random effects (Wooldridge (2012)). Since individuals make repeated decisions over time in the experiment, the error terms are likely correlated within subject. However, it is not clear whether the entire correlation structure can be accurately modeled using fixed effects or random effects specifications. Although violating this assumptions is an issue for both linear and non-linear models, violation of this assumption with a linear model only produces inefficient standard errors, an issue that can be corrected (Wooldridge (2012)). However, violating this assumption in non-linear models creates inconsistent parameter estimates, an issue that can not be corrected (Wooldridge (2012)). For this reason, I test the robustness of the reported results to both linear and non-linear specifications.

LINEAR PROBLEMS. Using a linear specification with a binary dependent variable introduces heteroskedasticity into the standard errors and violates the assumption that the standard errors are normally distributed. I address the issue of heteroskedasticity by reporting clustered, robust standard errors in the linear model specification. However, I can not control for violating the assumption that the standard errors are normally distributed. Since the asymptotic properties of the ols estimator are derived from the assumptions on the distribution of the standard errors, violating the normality assumption causes the typical tests for statistical significance to be invalid (Wooldridge (2012)). Therefore, a potential concern is that the significance of the coefficients reported in the main body of the paper may not hold when a non-linear model is specified. Although linear probability models create misspecification problems, the non-linear models are problematic for this analysis, as well.

NONLINEAR PROBLEMS. Although specifying a non-linear model can correct for the two issues with linear probability models, linear, logit, and probit models all require that the error terms be independent or uncorrelated with each other unless the correlation structure can be modeled explicitly. This experiment considers individual choices over time which likely causes correlation in individual errors over time. I approach this issue by explicitly modeling the correlation structure using fixed effects and random effects in both the linear and non-linear models. According to Greene (2012), fixed effect specifications control for any subject specific effects that do not vary over time, and they allow the unobserved effect to be correlated with the regressors. Random effect specifications control for an individual specific error term, but they requires the stronger assumption that the unobserved effect is uncorrelated with the regressors. The issue arises if the errors are still correlated with each other after explicitly modeling the correlation structure. If the correlation among errors is not correctly modeled using fixed or random effects in the linear model, then the coefficient estimates are still consistent and unbiased, but the standard errors are inefficient (Greene (2012)). I correct for this issue in the linear model by reporting clustered, robust standard errors. However, if the correlation among errors is not correctly modeled using fixed or

random effects in the non-linear model, then the coefficient estimates are inconsistent (Greene (2012)). Unfortunately, I can not correct this issue by clustering the standard errors, because the coefficient vector β is still an inconsistent estimator. For further discussion of this issue see Greene (2012). For these reasons and ease of interpretation, I chose to report the linear model in the main body of the paper. However, I demonstrate that the reported results are robust to a non-linear specification.

LOGIT. For the robustness analysis, I chose to report a logit model specification over a probit model specification. Theoretically, the logit model specification requires the assumption that the standard errors are logistically distributed while the probit model requires the assumption that the standard errors are normally distributed (Wooldridge (2012)). It is not clear which assumption is correct in this case. Furthermore, Wooldridge (2012) argues that explicitly modeling the error correlation structure by estimating the fixed effects produces an inconsistent and potentially biased estimation of the coefficients in both logit and probit models. This issue is known as incidental parameter bias. However, using the conditional logit estimator does not treat the unobserved effects as parameters to be estimated (Wooldridge (2012)). Instead, conditional logit finds a conditional density that depends only on observables and the vector of coefficients (Wooldridge (2012)). In this way, conditional logit controls for unobserved, non time-varying effects or fixed effects and produces a consistent estimation of the vector of coefficients. Although I could estimate a conditional probit model, estimation of the conditional probit model requires that an additional assumption be made about the distribution of the unobserved effects (Wooldridge (2012)). For further discussion of this issue see Wooldridge (2012). Since conditional logit allows me to control for unobserved individual effects without the incidental parameters problem, I chose to report robustness results from the non-linear logit specification. However, it should be noted that using a probit specification and estimating fixed or random effects (not reported) does not change the sign or significance of these results.

ODDS RATIO: When estimating non-linear models, it is customary to report the marginal effect to allow for interpretation of the regression coefficients. However, the magnitude of an interaction term in a non-linear model is not equal to the marginal effect of the interaction term (Ai and Norton (2003)). Since the estimated model contains an interaction term, the marginal effects will be incorrect for the regression coefficient squared incidence: 1 lag. To avoid this issue, I follow Buis (2010) and report the odds ratios for the estimated logit model. The odds ratios give the ratio by which the dependent variable changes for a one unit change in an independent variable. Unlike computing the marginal effects, the odds ratio provides the correct magnitude, sign, and significance for interaction terms.

Table B.8: Effect of Treatments on Decision to Invest in Disease Prevention

	Investment in Disease Prevention		
		Logit RE	
	(1)	(2)	(3)
High Cost	-2.74***	-2.84***	-2.87***
	(0.11)	(0.16)	(0.17)
Order Effects	-0.07	-0.20	-0.21
	(0.33)	(0.37)	(0.37)
High Cost * Order Effects		0.18	0.18
		(0.22)	(0.22)
Round Effects	No	No	Yes
Observations	4000	4000	4000

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TREATMENT EFFECTS: Table B.8 provides a robustness check of result 1 to a logit specification. This table demonstrates that being in the high cost treatment has a significant negative impact on the probability of investing in disease prevention. This effect holds when controlling for order and round effects. Additionally, table B.9 provides the odds ratios for the models reported in table B.8. The coefficients and significance levels are consistent with the linear model. This demonstrates that the odds of investing in disease prevention for subjects in the low cost treatment are approximately 17 times higher than the odds

Table B.9: Effect of Treatments on Decision to Invest in Disease Prevention

	Investment in Disease Prevention		
	Logit REs Odds Ratios		
	(1)	(2)	(3)
High Cost	0.06***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)
Order Effects	0.94	0.82	0.81
	(0.31)	(0.30)	(0.30)
High Cost * Order Effects		1.19	1.20
		(0.26)	(0.27)
Round Effects	No	No	Yes
Observations	4000	4000	4000

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

of investing in disease prevention for subjects in the high cost treatment. This effect is significant at the 1% level. This demonstrates that result 1 is robust to a non-linear model specification.

PREVALENCE ELASTICITY: Table B.10 equations 1-6 show that result 2 is robust to a logit model specification. Higher disease incidence in the prior round increases the probability of investing in disease prevention, an effect significant at the 1% level. Thus, I find evidence that subjects exhibit prevalence elastic demand for disease prevention. Table B.11 equation 5 demonstrates that the odds of investing in disease prevention are 1.52 times higher if there is a one person increase in the infection level than if there is no change in the infection level. This effect is significant at the 1% level. As in the linear specification, the effect of disease incidence on prevention investment is not constant as infection levels rise.

SQUARED INCIDENCE: Table B.10 equations 4-6 give the results of the robustness of result 4 to a logit specification. The impact of disease incidence on the probability of investing in disease prevention decreases as infection levels rise. This effect is significant at the 1% level. Table B.11 equation 4 shows that for each one person decrease in the infection level, the odds

Table B.10: Effect of Changes in Disease Incidence on Decision to Invest in Disease Prevention

	Investment in Disease Prevention					
	Logit					
	FE					RE
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	0.04	0.06	0.07*	0.29**	0.42***	0.42***
	(0.03)	(0.04)	(0.04)	(0.12)	(0.13)	(0.13)
Infected: 1 Lag		-0.53***	-0.49***	-0.54***	-0.49***	-0.59***
		(0.11)	(0.12)	(0.11)	(0.12)	(0.12)
Incidence: 2 Lags			0.08*		0.09**	0.09**
			(0.04)		(0.04)	(0.04)
Squared Incidence: 1 Lag				-0.001**	-0.01***	-0.01***
				(0.00)	(0.00)	(0.00)
Free Rider Factor						-0.43***
						(0.16)
High Cost Treatment	-3.30***	-3.27***	-4.40***	-4.41***	-6.33***	-6.27***
	(0.46)	(0.46)	(0.74)	(0.73)	(1.02)	(1.02)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3552	3552	3404	3552	3404	3358

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.11: Effect of Changes in Disease Incidence on Decision to Invest in Disease Prevention

	Investment in Disease Prevention Logit Odds Ratios					
			FE			RE
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	1.04	1.06	1.07*	1.33**	1.52***	1.52***
	(0.04)	(0.04)	(0.04)	(0.16)	(0.20)	(0.20)
Infected: 1 Lag		0.59***	0.61***	0.59***	0.61***	0.55***
		(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
Incidence: 2 Lags			1.08*		1.09**	1.09**
			(0.04)		(0.04)	(0.04)
Squared Incidence: 1 Lag				0.99**	0.99***	0.99***
				(0.00)	(0.00)	(0.00)
Free Rider Factor						0.65***
						(0.11)
High Cost Treatment	0.04***	0.04***	0.01***	0.01***	0.002***	0.002***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3552	3552	3404	3552	3404	3358

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

of investing in disease prevention are 1.01 times larger than the odds of investing in disease prevention if infection levels do not change. This effect is significant at the 1% level. Taken with the findings on prevalence elastic demand for disease prevention, these results capture the fact that the probability of investing in disease prevention increases at a decreasing rate as infection levels rise.

PRIOR INFECTION: Table B.10 equations 2-6 demonstrate that result 3 is robust to a logit model specification. Being infected in the previous period decreases the probability that a subject invests in disease prevention, an effect significant at the 1% level. Table B.11 equations 1 and 4 provide the interpretation that the odds of investing in disease prevention are 1.69 times larger for subjects that were not infected in the previous round than the odds of investing in disease prevention for subjects that were infected in the previous round. This effect is significant at the 1% level. This provides support for the finding that an infection in the previous period reduces the current probability of investment.

INCIDENCE HISTORY: Table B.10 equations 3, 5, and 6 show result 5 is robust to a logit model specification. Higher historical disease incidence has a persistent positive impact on the probability of investing in disease prevention, an effect significant at the 5% level. Table B.11 equations 5 and 6 give the interpretation that the odds of investing in disease prevention are 1.09 times higher if there is a one person increase in the infection level 2 rounds prior than if there is no change in the infection level. This effect is significant at the 5% level. This demonstrates that disease incidence has a persistent impact on the probability of investing in disease prevention.

FREE RIDER: Table B.10 equation 6 provides a robustness check of result 6 to a logit model specification. Subjects that report free-riding on the disease prevention investments of others in the real-world are less likely to invest in disease prevention in the experiment, an effect significant at the 1% level. Table B.11 equation 6 demonstrates that the odds of investing in disease prevention are 1.54 times higher for subjects that score lower on the free-rider factor

than the odds for subjects who score higher on the free-rider factor. This effect is significant at the 1% level. This finding suggests that subjects' behavior in the experiment is correlated with their real-world attitudes towards investment in disease prevention.

Self-Reported Disease Prevention Behaviors

Table B.12: Summary Statistics for Self-Reported Engagement in Disease Prevention Behaviors

	Observations	Median	Mean	Standard Deviation
Question 1	80	1	1.25	0.46
Question 2	80	2	2.44	1.04
Question 3	80	1	1.44	0.59

The questionnaire implemented at the end of the experiment asks subjects to rate the frequency with with they engaged in the following behaviors:

1. How often do you wash your hands after using the restroom?
2. How often do you wash your hands before you eat?
3. How often do you cover your mouth when coughing in public?

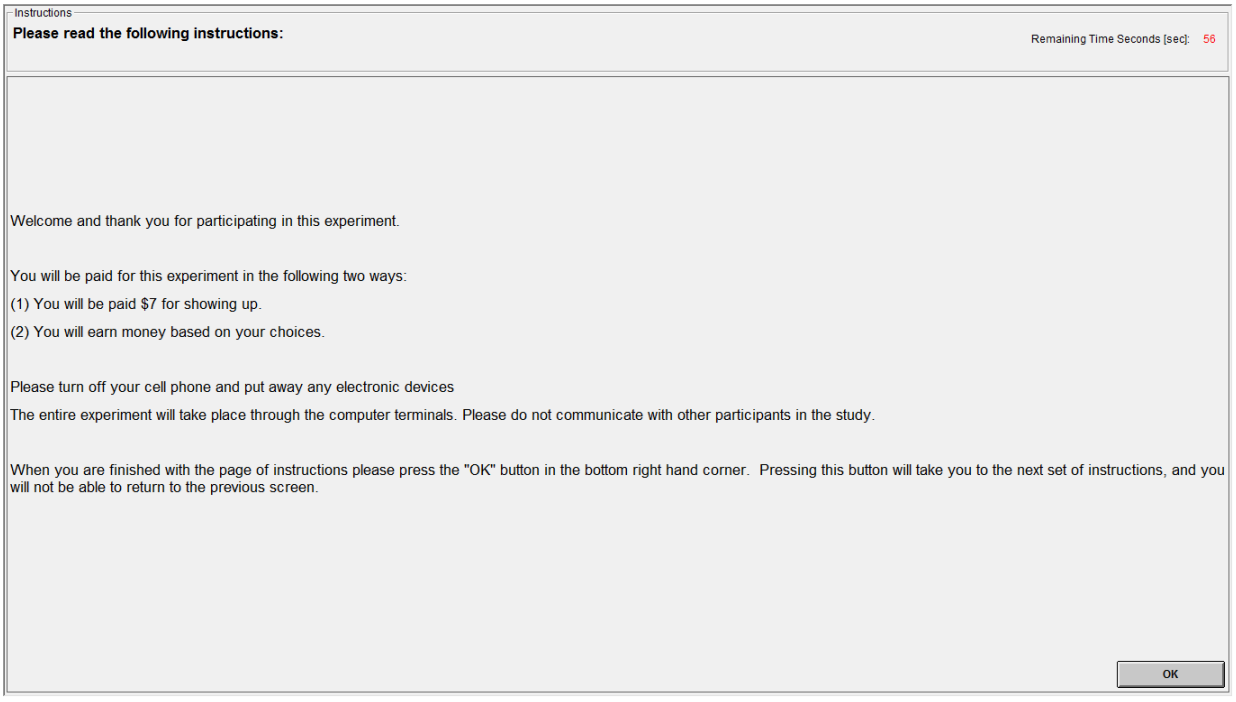
RESPONSE: Subjects were asked to rate the frequency with which they engage in these behaviors: Every time, Most of the time, Half of the time, Some of the Time, and Never. These responses were scaled from 1 to 5 with 1 representing Every time and 5 representing Never. Table B.12 provides the average numbered response for each of these questions. It is easy to see that a high number of subjects report frequent engagement in each of these behaviors meaning that there is very little variation in their responses.

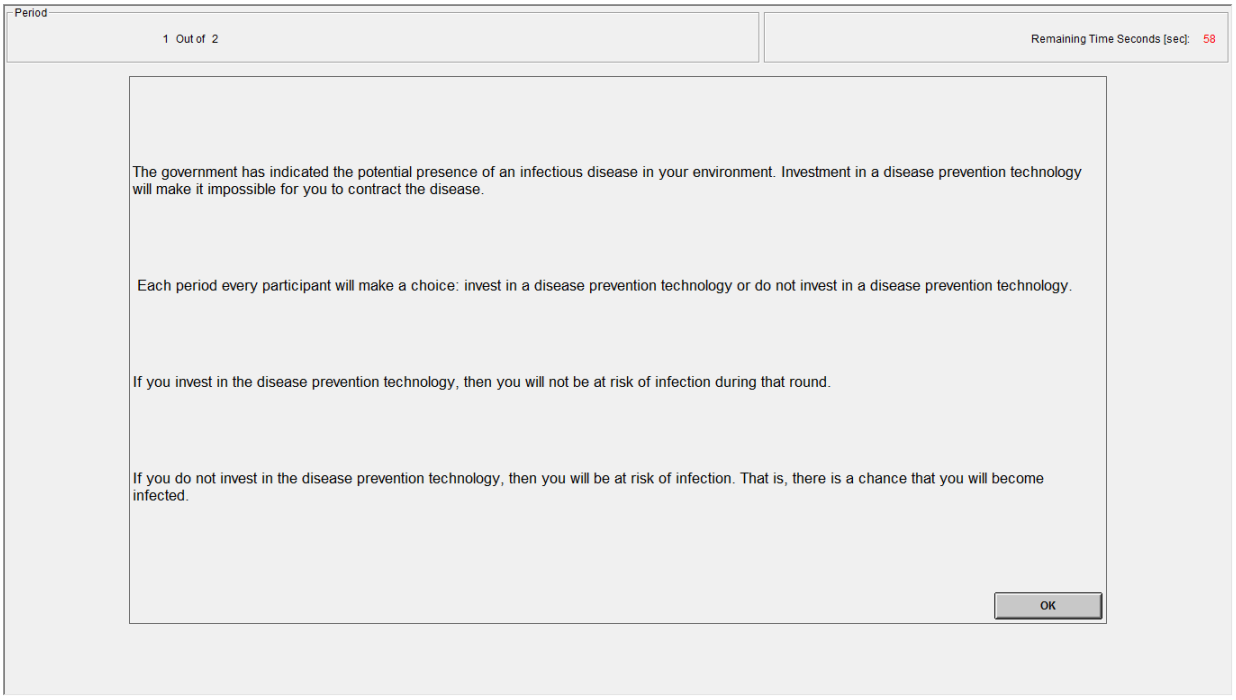
FUTURE: In the future, a better way to ask this question would be to ask subjects what percentage of the time they engage in each behavior. Asking the question in this way would

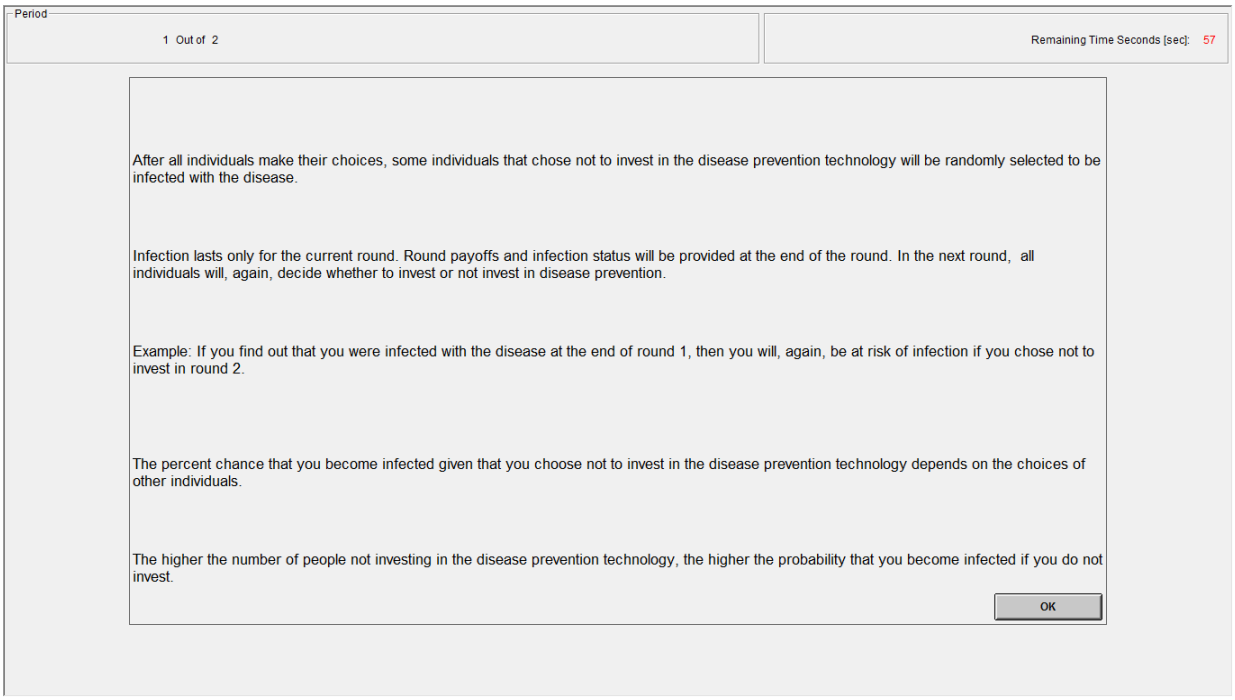
avoid truncating the distribution. Thus, as long as the over reporting only results in a shift of the distribution mean, the variables should still be correlated with behavior in the experiment.

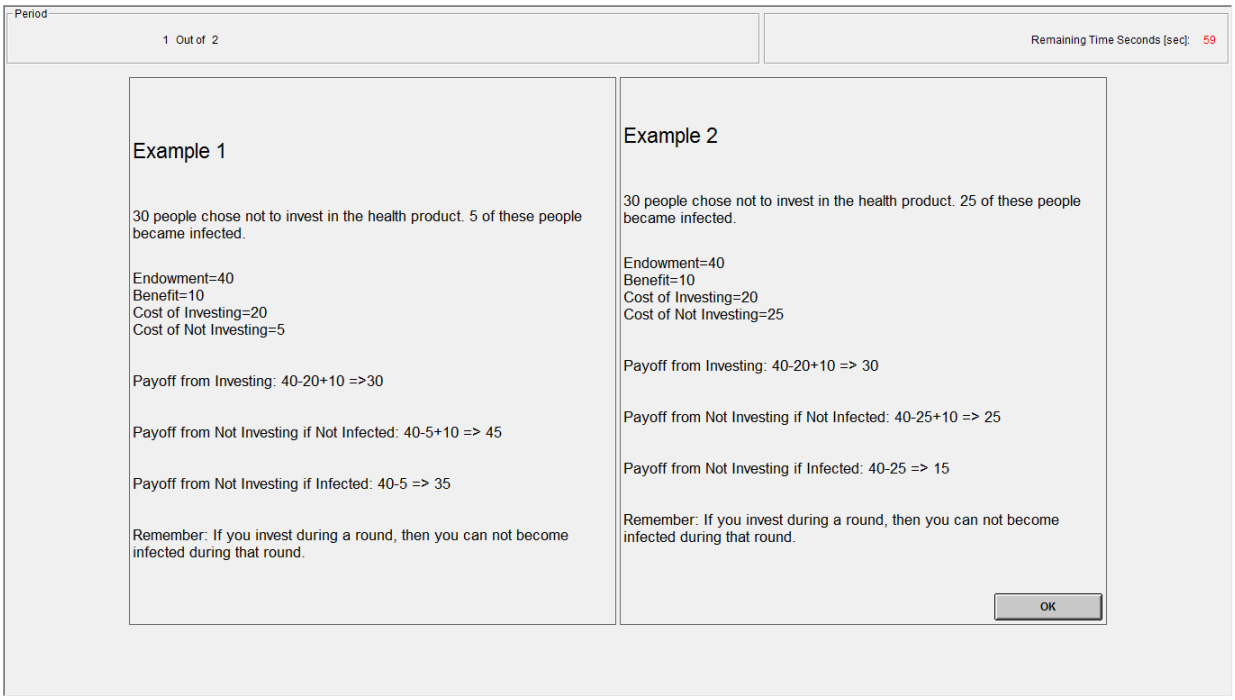
Experiment Screen Shots

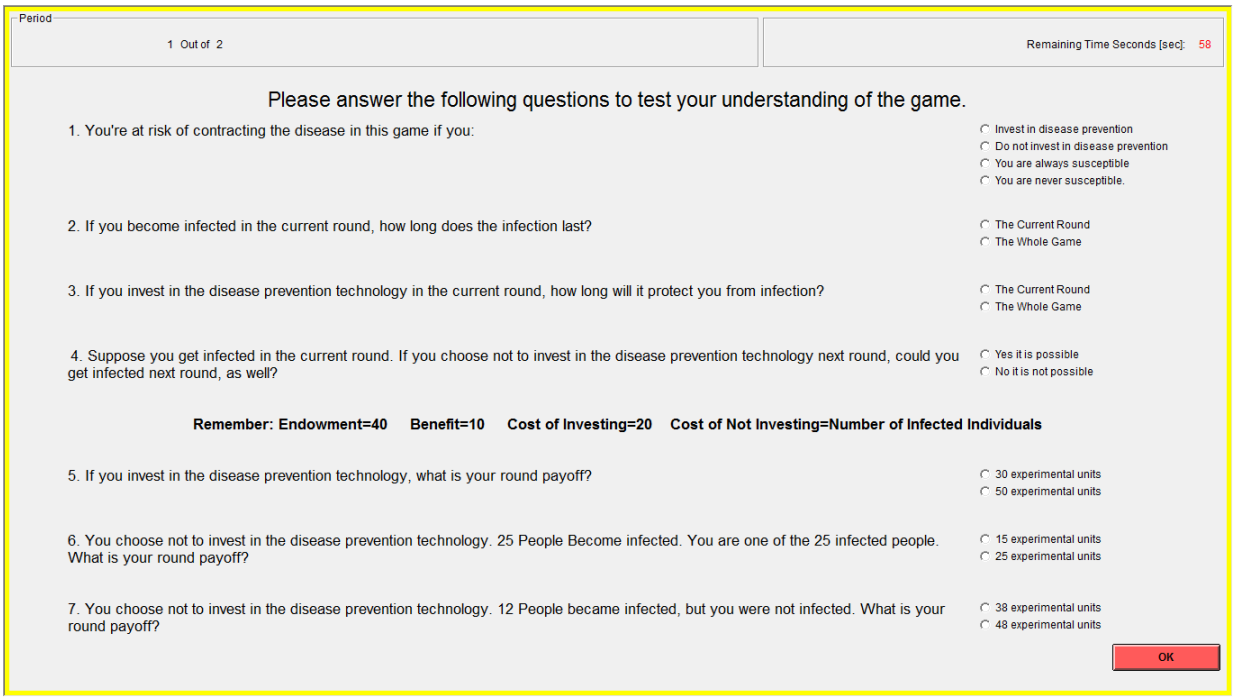
Part I

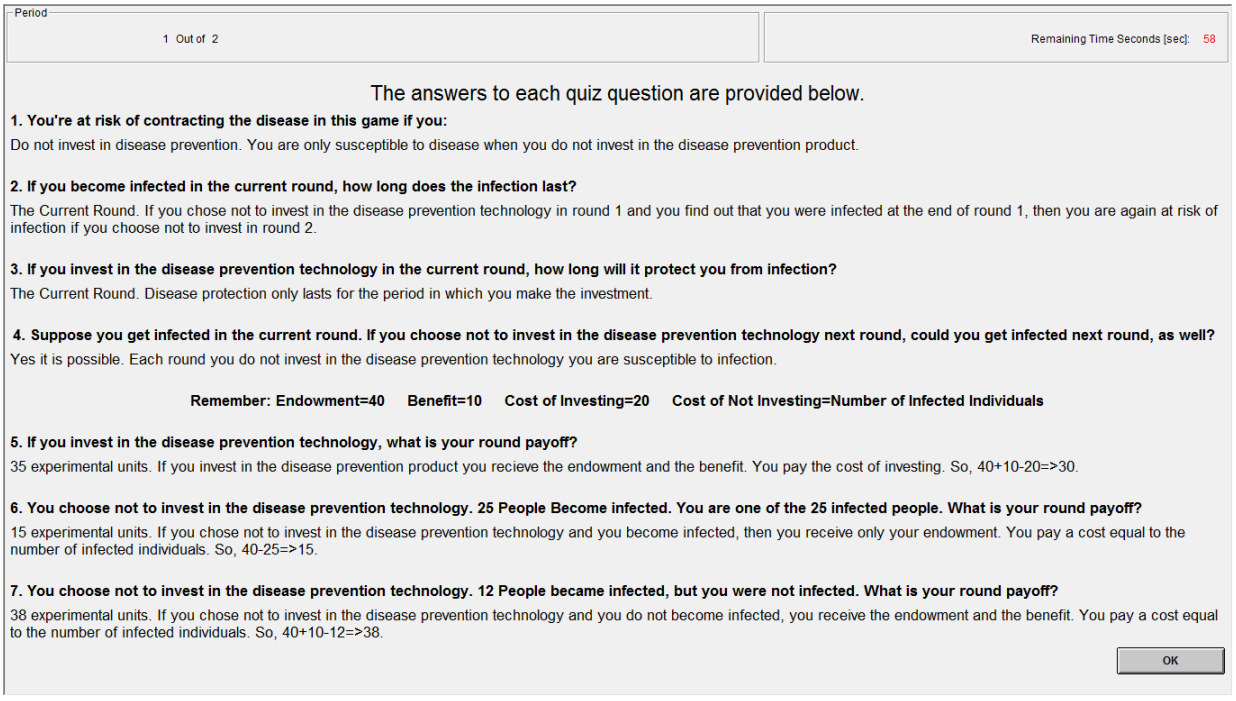


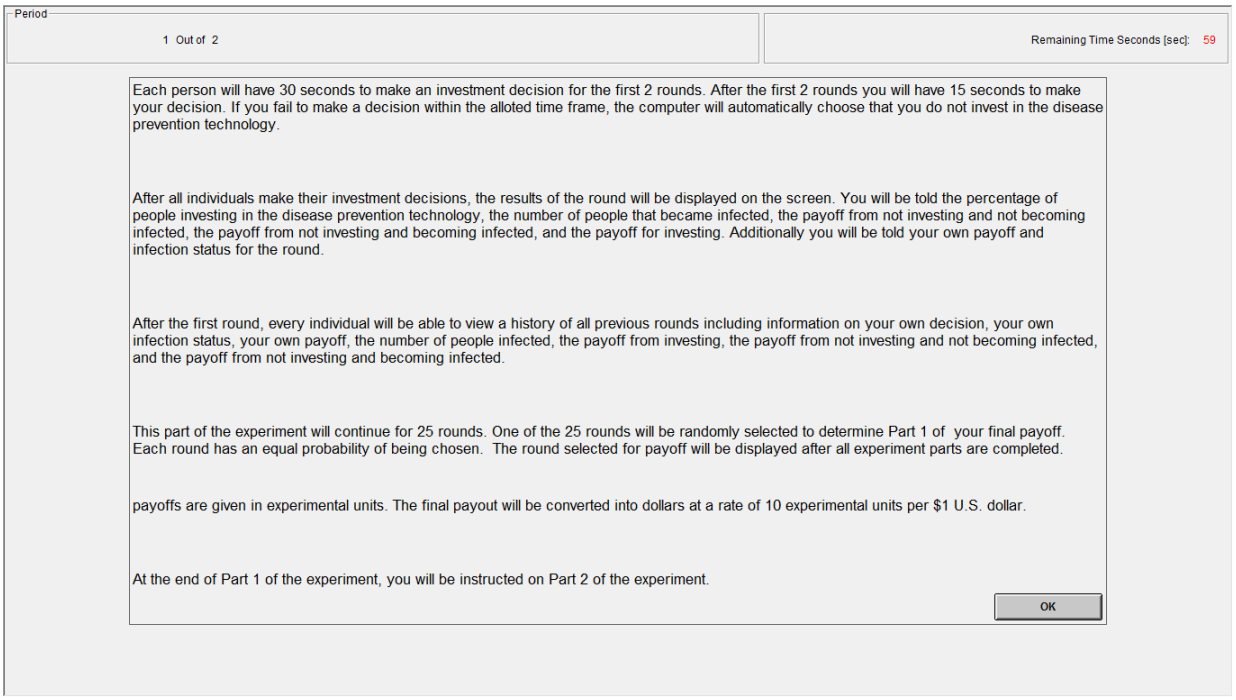


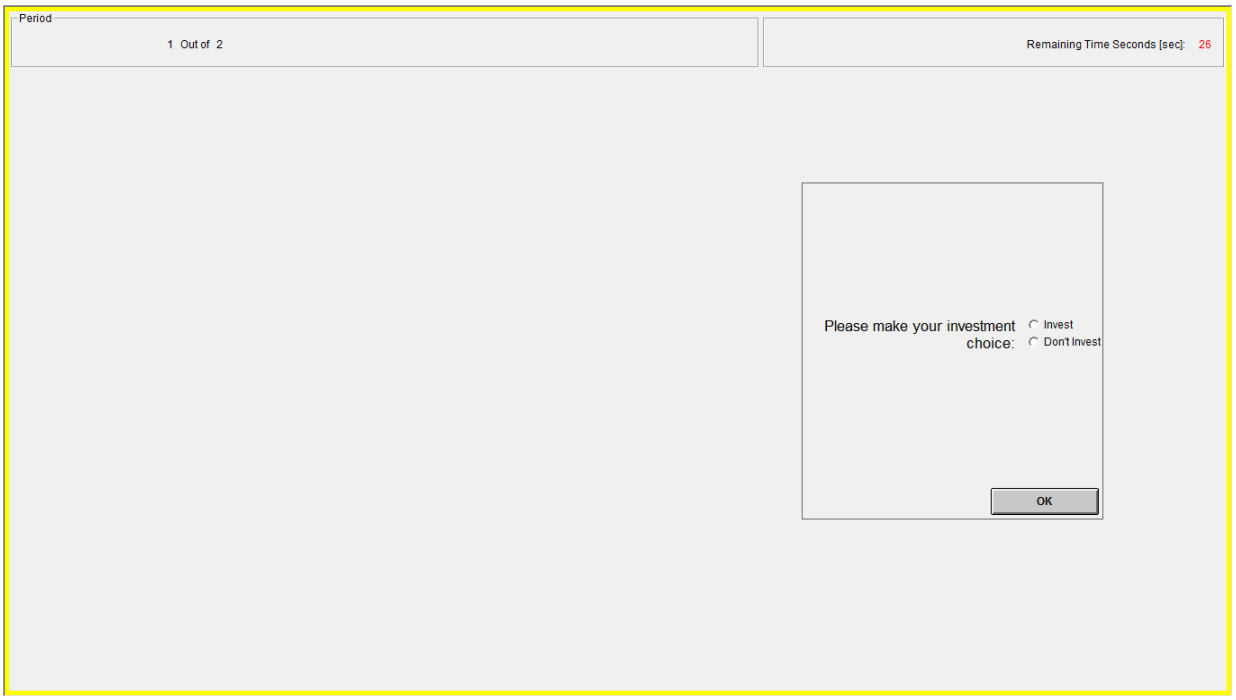

Period
1 Out of 2
Remaining Time Seconds [sec]: 57
After all individuals make their choices, some individuals that chose not to invest in the disease prevention technology will be randomly selected to be infected with the disease.
Infection lasts only for the current round. Round payoffs and infection status will be provided at the end of the round. In the next round, all individuals will, again, decide whether to invest or not invest in disease prevention.
Example: If you find out that you were infected with the disease at the end of round 1, then you will, again, be at risk of infection if you chose not to invest in round 2.
The percent chance that you become infected given that you choose not to invest in the disease prevention technology depends on the choices of other individuals.
The higher the number of people not investing in the disease prevention technology, the higher the probability that you become infected if you do not invest.
OK

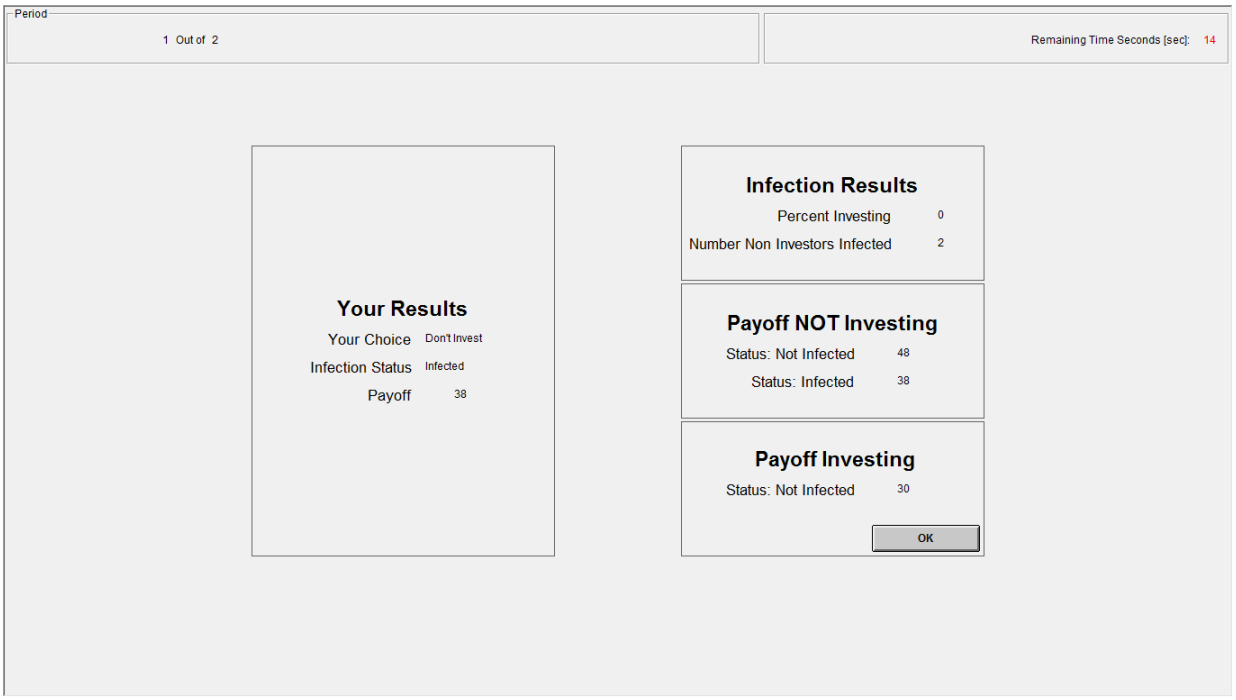

Period
1 Out of 2
Remaining Time Seconds [sec]: 59
Example 1
30 people chose not to invest in the health product. 5 of these people became infected.
Endowment=40
Benefit=10
Cost of Investing=20
Cost of Not Investing=5
Payoff from Investing: 40-20+10 =>30
Payoff from Not Investing if Not Infected: 40-5+10 => 45
Payoff from Not Investing if Infected: 40-5 => 35
Remember: If you invest during a round, then you can not become infected during that round.
Example 2
30 people chose not to invest in the health product. 25 of these people became infected.
Endowment=40
Benefit=10
Cost of Investing=20
Cost of Not Investing=25
Payoff from Investing: 40-20+10 => 30
Payoff from Not Investing if Not Infected: 40-25+10 => 25
Payoff from Not Investing if Infected: 40-25 => 15
Remember: If you invest during a round, then you can not become infected during that round.
OK

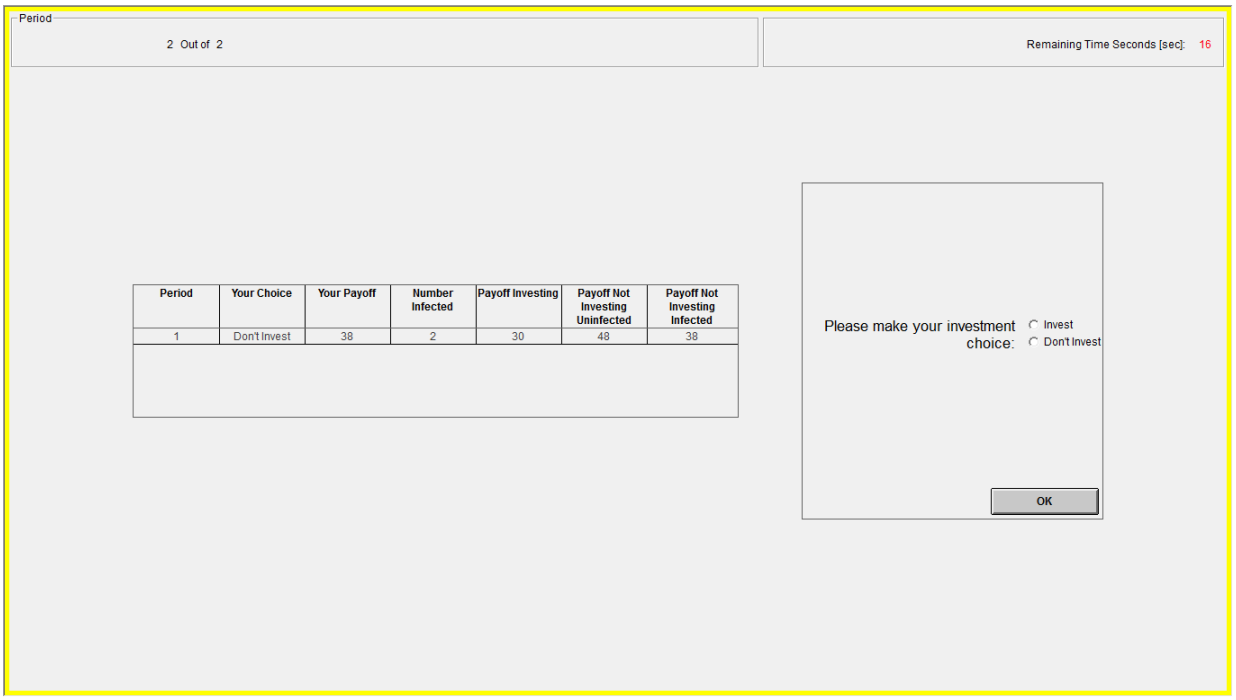

Period
1 Out of 2
Remaining Time Seconds [sec]: 58
Please answer the following questions to test your understanding of the game.
1. You're at risk of contracting the disease in this game if you:
Invest in disease prevention
Do not invest in disease prevention
You are always susceptible
You are never susceptible.
2. If you become infected in the current round, how long does the infection last?
The Current Round
The Whole Game
3. If you invest in the disease prevention technology in the current round, how long will it protect you from infection?
The Current Round
The Whole Game
4. Suppose you get infected in the current round. If you choose not to invest in the disease prevention technology next round, could you get infected next round, as well?
Yes it is possible
No it is not possible
Remember: Endowment=40 Benefit=10 Cost of Investing=20 Cost of Not Investing=Number of Infected Individuals
5. If you invest in the disease prevention technology, what is your round payoff?
30 experimental units
50 experimental units
6. You choose not to invest in the disease prevention technology. 25 People Become infected. You are one of the 25 infected people. What is your round payoff?
15 experimental units
25 experimental units
7. You choose not to invest in the disease prevention technology. 12 People became infected, but you were not infected. What is your round payoff?
38 experimental units
48 experimental units
OK


Period
1 Out of 2
Remaining Time Seconds [sec]: 58
The answers to each quiz question are provided below.
1. You're at risk of contracting the disease in this game if you:
Do not invest in disease prevention. You are only susceptible to disease when you do not invest in the disease prevention product.
2. If you become infected in the current round, how long does the infection last?
The Current Round. If you chose not to invest in the disease prevention technology in round 1 and you find out that you were infected at the end of round 1, then you are again at risk of infection if you choose not to invest in round 2.
3. If you invest in the disease prevention technology in the current round, how long will it protect you from infection?
The Current Round. Disease protection only lasts for the period in which you make the investment.
4. Suppose you get infected in the current round. If you choose not to invest in the disease prevention technology next round, could you get infected next round, as well?
Yes it is possible. Each round you do not invest in the disease prevention technology you are susceptible to infection.
Remember: Endowment=40 Benefit=10 Cost of Investing=20 Cost of Not Investing=Number of Infected Individuals
5. If you invest in the disease prevention technology, what is your round payoff?
35 experimental units. If you invest in the disease prevention product you recieve the endowment and the benefit. You pay the cost of investing. So, 40+10-20=>30.
6. You choose not to invest in the disease prevention technology. 25 People Become infected. You are one of the 25 infected people. What is your round payoff?
15 experimental units. If you chose not to invest in the disease prevention technology and you become infected, then you receive only your endowment. You pay a cost equal to the number of infected individuals. So, 40-25=>15.
7. You choose not to invest in the disease prevention technology. 12 People became infected, but you were not infected. What is your round payoff?
38 experimental units. If you chose not to invest in the disease prevention technology and you do not become infected, you receive the endowment and the benefit. You pay a cost equal to the number of infected individuals. So, 40+10-12=>38.
OK

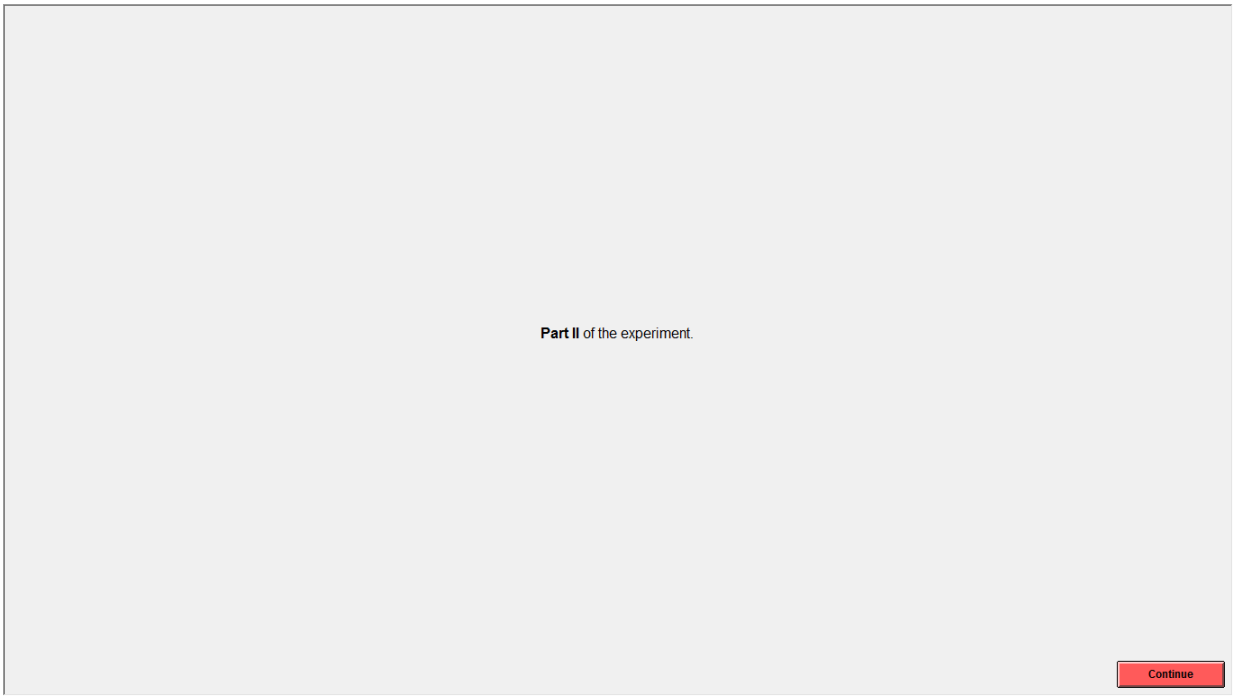

Period
1 Out of 2
Remaining Time Seconds [sec]: 59
Each person will have 30 seconds to make an investment decision for the first 2 rounds. After the first 2 rounds you will have 15 seconds to make your decision. If you fail to make a decision within the alloted time frame, the computer will automatically choose that you do not invest in the disease prevention technology.
After all individuals make their investment decisions, the results of the round will be displayed on the screen. You will be told the percentage of people investing in the disease prevention technology, the number of people that became infected, the payoff from not investing and not becoming infected, the payoff from not investing and becoming infected, and the payoff for investing. Additionally you will be told your own payoff and infection status for the round.
After the first round, every individual will be able to view a history of all previous rounds including information on your own decision, your own infection status, your own payoff, the number of people infected, the payoff from investing, the payoff from not investing and not becoming infected, and the payoff from not investing and becoming infected.
This part of the experiment will continue for 25 rounds. One of the 25 rounds will be randomly selected to determine Part 1 of your final payoff. Each round has an equal probability of being chosen. The round selected for payoff will be displayed after all experiment parts are completed.
payoffs are given in experimental units. The final payout will be converted into dollars at a rate of 10 experimental units per $1 U.S. dollar.
At the end of Part 1 of the experiment, you will be instructed on Part 2 of the experiment.
OK

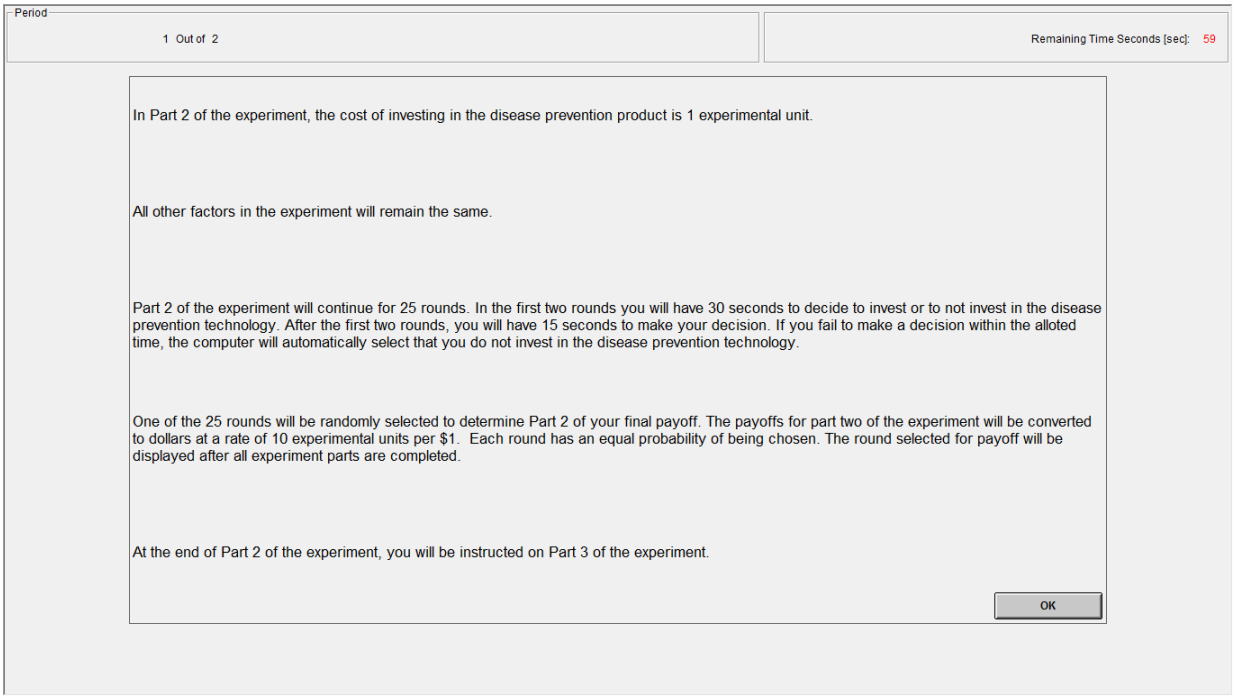

Period
1 Out of 2
Remaining Time Seconds [sec]: 26
Please make your investment choice:
Invest
Don't Invest
OK

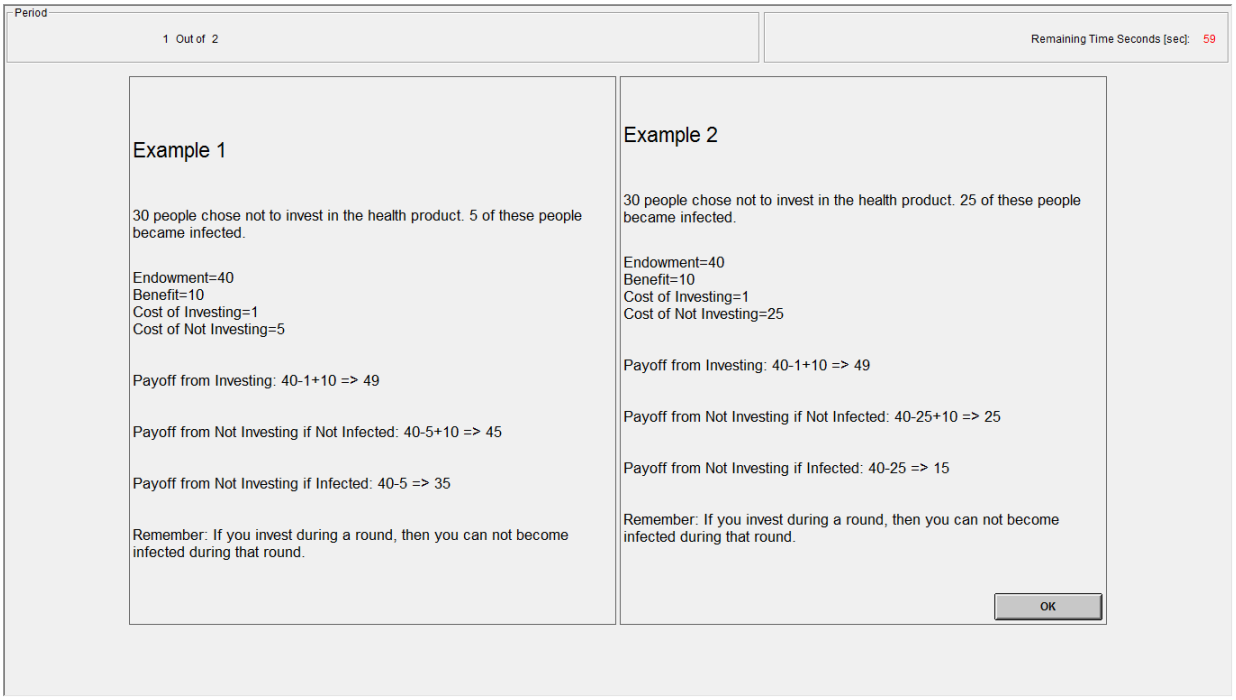

Period
1 Out of 2
Remaining Time Seconds [sec]: 14
Your Results
Your Choice Don't Invest
Infection Status Infected
Payoff 38
Infection Results
Percent Investing 0
Number Non Investors Infected 2
Payoff NOT Investing
Status: Not Infected 48
Status: Infected 38
Payoff Investing
Status: Not Infected 30
OK

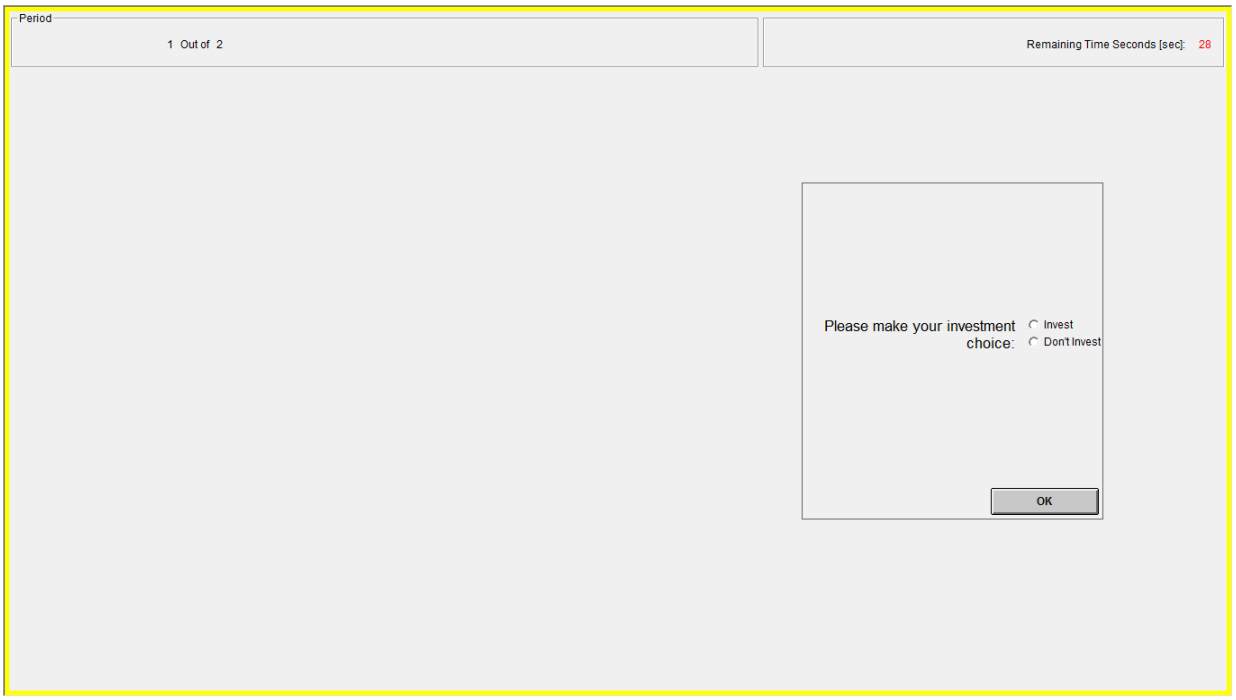

Period
2 Out of 2
Remaining Time Seconds [sec]: 16
Period
Your Choice
Your Payoff
Number Infected
Payoff Investing
Payoff Not Investing Uninfected
Payoff Not Investing Infected
1
Don't Invest
38
2
30
48
38
Please make your investment choice:
Invest
Don't Invest
OK

Part II


Part II of the experiment.
Continue


Period
1 Out of 2
Remaining Time Seconds [sec]: 59
In Part 2 of the experiment, the cost of investing in the disease prevention product is 1 experimental unit.
All other factors in the experiment will remain the same.
Part 2 of the experiment will continue for 25 rounds. In the first two rounds you will have 30 seconds to decide to invest or to not invest in the disease prevention technology. After the first two rounds, you will have 15 seconds to make your decision. If you fail to make a decision within the alloted time, the computer will automatically select that you do not invest in the disease prevention technology.
One of the 25 rounds will be randomly selected to determine Part 2 of your final payoff. The payoffs for part two of the experiment will be converted to dollars at a rate of 10 experimental units per $1. Each round has an equal probability of being chosen. The round selected for payoff will be displayed after all experiment parts are completed.
At the end of Part 2 of the experiment, you will be instructed on Part 3 of the experiment.
OK

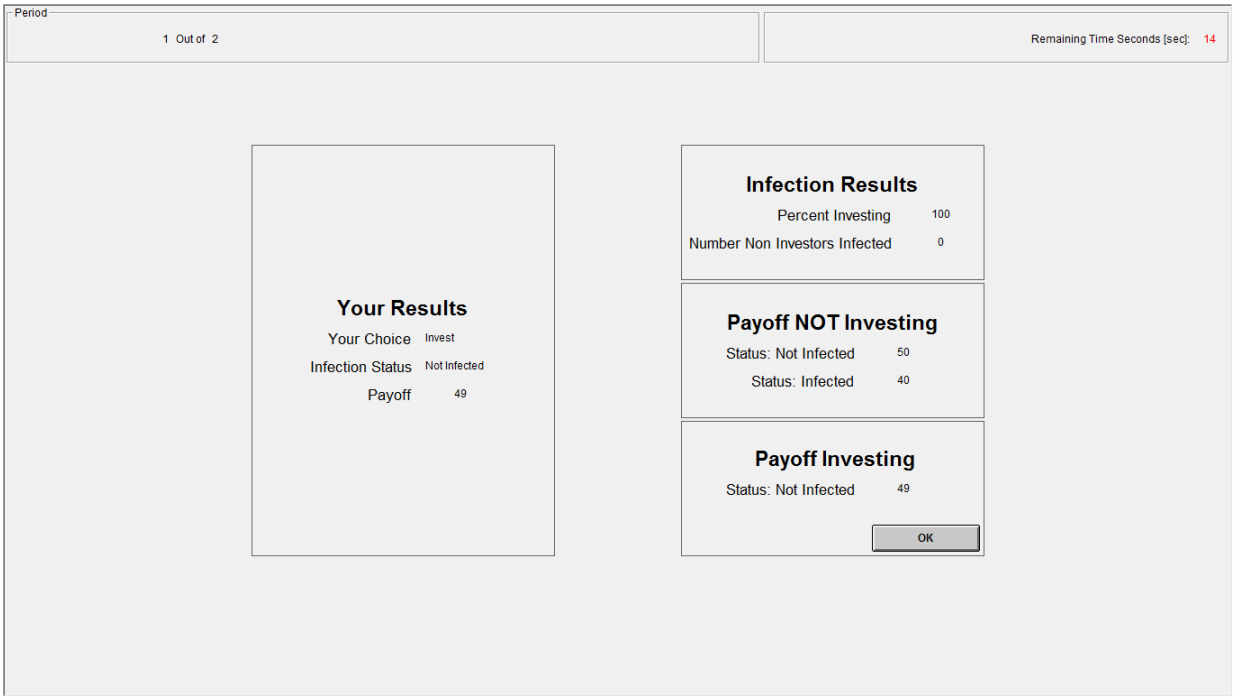


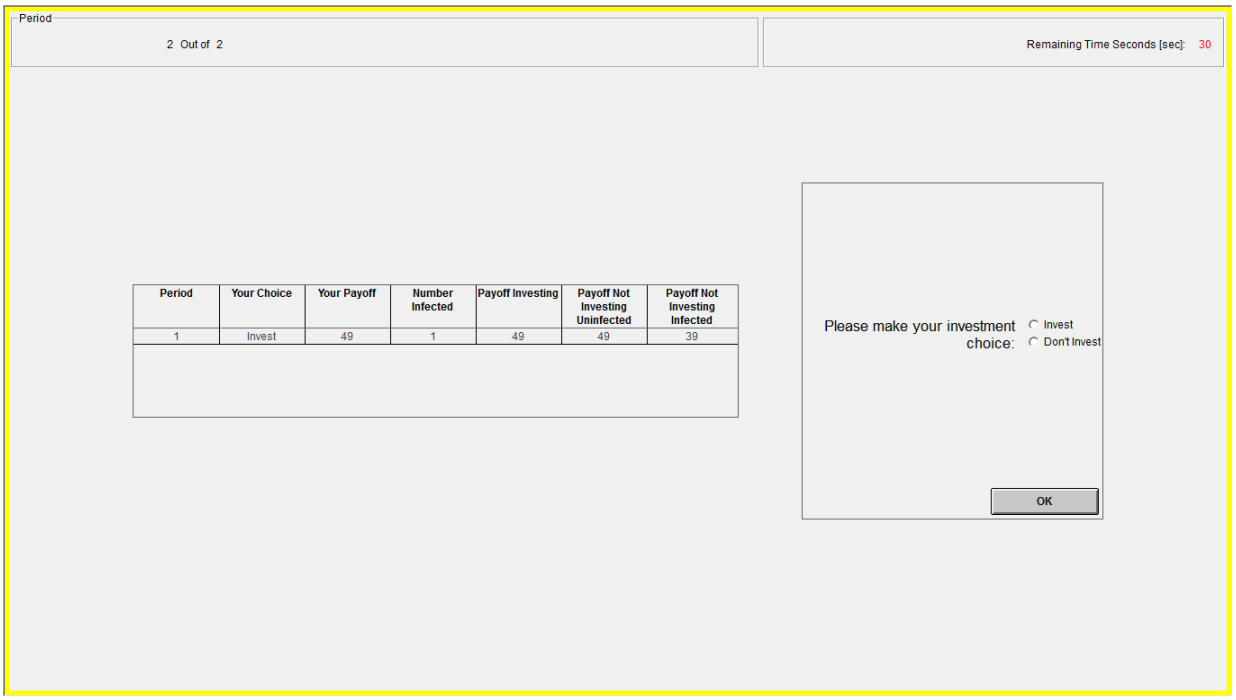


C Appendix

C.1 Behavior Dynamic Derivation

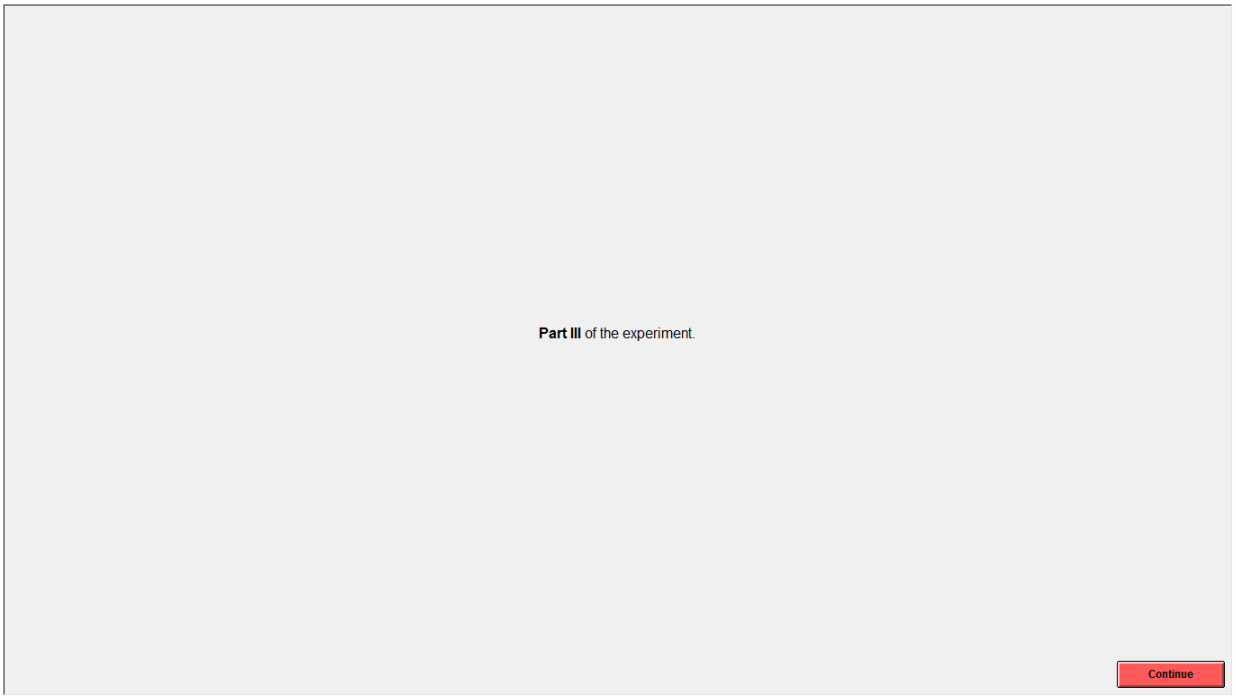
I use an evolutionary game with the replicator dynamic to derive the change in the share of the population not investing in disease prevention. Since the replicator dynamic can be

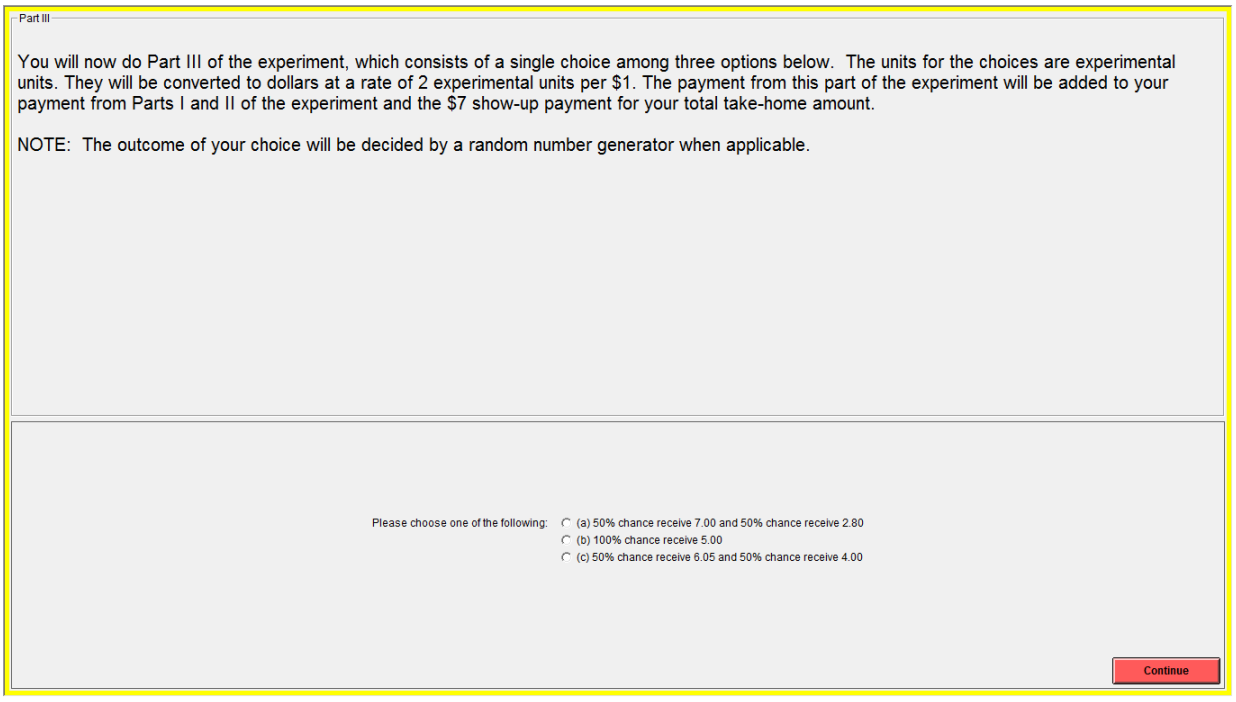




derived from different strategy revision protocols, I choose to describe the dynamic as arising from process where individuals are randomly paired each time period. When an individual is paired with someone with a different strategy, she can compare her payoff with the payoff under the other strategy. If the other strategy provides a higher payoff, she will adopt the

Part III





strategy with some probability.

The games underlying model 1 and model 2 are provided in table C.12 and table C.13, respectively. As is customary for evolutionary games, the payoffs are given for the player

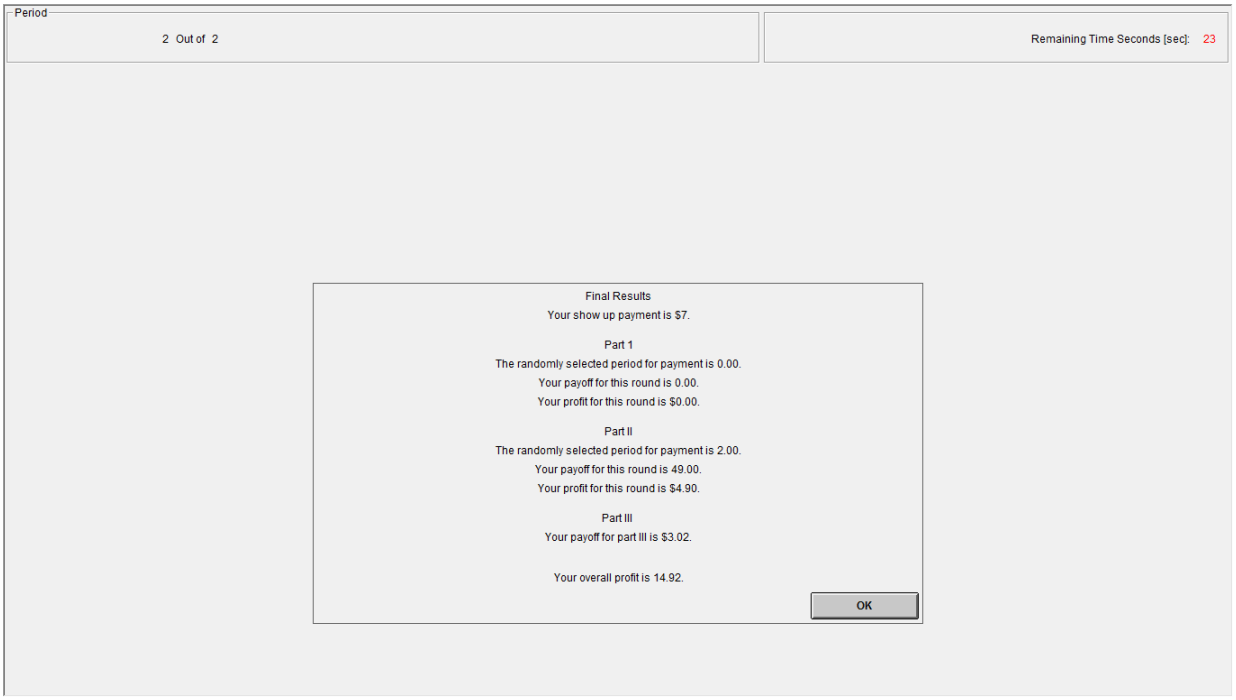


Table C.13: Prevention Investment Game: Model 1

	Invest	Do Not Invest
Invest	b-c	b-c
Do Not Invest	b-c(I(t))	-c(I(t))

on the left hand side (LHS). The row labels provide the two possible strategies for the LHS player: invest in disease prevention or do not invest in disease prevention. The column labels provide the two possible strategies for the random player with whom the LHS player is periodically paired: invest in disease prevention or do not invest in disease prevention.

Table C.14: Prevention Investment Game: Model 2

	Invest	Do Not Invest
Invest	b-c	b-c
Do Not Invest	b	-c(I(t))

Model 1 Derivation

Table C.12 provides the payoffs used to derive the changes in the share of the population investing in disease prevention for model 1. In table C.12, b is the benefit from avoiding infection, and c is the cost of investing in disease prevention. The cost of investing in disease prevention represents the monetary costs and the time costs associated with investing in disease prevention. The cost of not investing in disease prevention is given by $c(I(t))$. I assume that the cost of not investing in disease prevention is an increasing function of the infection rate. For simplicity, I assume that agents pay the same cost for not investing in disease prevention regardless of whether they are paired with an investor or a non-investor. For individuals that choose not to invest in disease prevention and are randomly paired with an investor, this cost represents the cost of social exclusion, including social ostracism for not investing in disease prevention. For individuals that choose not to invest in disease prevention and are randomly paired with a non-investor, this cost represents the cost of becoming infected.

$$\frac{d\alpha(t)}{dt} = \alpha(t)\left(\Pi(NI) - \bar{\Pi}\right) \tag{C.11}$$

Using the replicator dynamic, the change in the share of the population not investing in disease prevention, $\frac{d\alpha(t)}{dt}$, is a function of the share of the population not investing in prevention, $\alpha(t)$; the payoff from not investing in disease prevention, $\Pi(NI)$; and the average payoff, $\bar{\Pi}$. Equation 6 demonstrates that the change in the share of the population not investing in disease prevention evolves according to the underlying individual payoffs. If the payoff from not investing in disease prevention is larger than the average payoff, then the share of the population not investing in disease prevention will grow. Conversely, if the payoff from not investing in disease prevention is smaller than the average payoff, the the share of the

population not investing in disease prevention will shrink. The payoff for not investing in disease prevention is:

$$\Pi(NI) = (1 - \alpha(t)) [b - c(I(t))] + \alpha(t) [-c(I(t))] \tag{C.12}$$

The average payoff is:

$$\bar{\Pi} = (1-\alpha(t)) \{(1 - \alpha(t)) [b - c] + \alpha(t) [b - c]\} + \alpha(t) \{(1 - \alpha(t)) [b - c(I(t))] + \alpha(t) [-c(I(t))]\} \tag{C.13}$$

With a little algebra, equation 8 reduces to:

$$\bar{\Pi} = (1 - \alpha(t)) [b - c] + \alpha(t) \{[(1 - \alpha(t))b - c(I(t))]\} \tag{C.14}$$

Substituting equations 7 and 9 into equation 6 and doing some algebra, I find:

$$\frac{d\alpha(t)}{dt} = \alpha(t)(1 - \alpha(t)) \{c - c(I(t)) - \alpha(t)b\} \tag{C.15}$$

Model 2 Derivation

Table C.13 gives the payoffs used to derive the changes in the share of the population not investing in disease prevention used in model 2. As in Model 1 Derivation section, b is the benefit from avoiding infection, and c is the cost of investing in disease prevention. The cost of not investing in disease prevention is, $c(I(t))$. I assume that this cost is an increasing function of the infection level. Unlike model 1, an individual only pays a cost of not investing in disease prevention if she is randomly paired with another non-investor. In this model, the cost of not investing in disease prevention simply represents the cost of becoming infected.

$$\frac{d\alpha(t)}{dt} = \alpha(t)\left(\Pi(NI) - \bar{\Pi}\right) \tag{C.16}$$

Like the derivation for the behavior dynamic presented in the Model 1 Derivation section, I begin with the equation for the replicator dynamic. Equation 11 gives the change in the share of the population not investing in disease prevention, $\frac{d\alpha(t)}{dt}$. The change in the share of the population not investing in disease prevention is a function of the share of the population not investing in disease prevention, $\alpha(t)$; the payoff from not investing in disease prevention, $\Pi(NI)$; and the average payoff, $\bar{\Pi}$. From the replicator dynamic, it is easy to see that the share of the population not investing in disease prevention grows if the payoff from not investing in disease prevention is larger than the average payoff. Conversely, the share of the population not investing in disease prevention shrinks if the average payoff is larger than the payoff from not investing in disease prevention. The payoff for not investing in disease

prevention is:

$$\Pi(NI) = (1 - \alpha(t))\,[b] + \alpha(t)\,[-c(I(t))] \tag{C.17}$$

The average payoff is:

$$\bar{\Pi} = (1 - \alpha(t))\,\{(1 - \alpha(t))\,[b - c] + \alpha(t)\,[b - c]\} + \alpha(t)\,\{(1 - \alpha(t))\,[b] + \alpha(t)\,[-c(I(t))]\} \tag{C.18}$$

With a little algebra, equation 13 reduces to:

$$\bar{\Pi} = (1 - \alpha(t))\,[b - c] + \alpha(t)\,\{(1 - \alpha(t))\,[b] + \alpha(t)\,[-c(I(t))]\} \tag{C.19}$$

Substituting equations 12 and 14 into equation 11 and doing some algebra, I find:

$$\frac{d\alpha(t)}{dt} = \alpha(t)(1 - \alpha(t))\,\{c - \alpha(t)\,[b + c(I(t))]\} \tag{C.20}$$

C.2 Non Linear Model Specification

In this analysis, both linear and non-linear model specification assumptions are potentially violated. This makes it difficult to determine the most appropriate model for analysis. The linear probability model with random effects reported in the main body of the paper requires that the dependent variable is continuous. However, both dependent variables of interest, the probability of starting investment and the probability of stopping investment, are binary. Violating the assumption that the dependent variable is continuous creates two problems: heteroskedastic error terms and violation of the assumption that errors are normally distributed (Wooldridge (2012)). While the first problem, heterskedastic error terms, can be corrected, non-normality of error terms is not an issue that can be correct (Wooldridge (2012)). Furthermore, both linear and non-linear models require that the model error terms are uncorrelated with each other unless the correlation structure can be correctly modeled with fixed or random effects (Wooldridge (2012)). Because subjects make repeated decisions in this experiment, the error terms are likely correlated within subject. Since the treatments in this experiment were between subject, I can not explore the impact of the treatments on the dependent variables while controlling for subject fixed effects. So, I can only model the potential error term correlation structure using random effects. However, it is not clear whether the error term structure is correctly modeled using random effects. If correlation exists between error terms after controlling for random effects, the problem is worse for a non-linear specification than a linear specification. Violation of the assumption that error terms are uncorrelated with each other in a linear specification creates inefficient standard errors, an issue that can be corrected (Wooldridge (2012)). Violation of this assumption in a non-linear specification creates inconsistent parameter estimates, an issue that can not be corrected (Wooldridge (2012)). For this reason, I test the robustness of the reported results to both linear and non-linear model specifications.

Using a linear probability model creates heteroskedasticity and violates the assumption

that the error terms are normally distributed (Wooldridge (2012)). To correct for the heteroskedasticity created by the linear probability model, I report clustered, robust standard errors in the main body of the paper (Wooldridge (2012)). However, I am unable to correct for the effects of non-normally distributed error terms in the linear model. The asymptotic properties of the OLS estimator are derived from the distribution of the model's error terms, and violating the normality assumption causes the statistically tests of significance to be invalid (Wooldridge (2012)). Thus, the concern is that the reported significance of the results may not hold when a non-linear model is specified.

Both linear and non-linear models such as probit and logit require the assumption that error terms are uncorrelated with each other unless the correlation structure can be explicitly modeled (Greene (2012)). Because individuals make repeated decisions over time in the experiment, the error terms are likely correlated within subject. There are two ways to explicitly model the error correlation structure: fixed effects and random effects. Since the experimental treatments were varied between subjects, controlling for fixed effects would eliminate all treatment effects. So, I can not model the correlation structure using fixed effects. I attempt to model the error term correlation structure using random effects. While a random effect specification controls for an individual specific error term, it requires the strong assumption that the unobserved effect is uncorrelated with the independent variables (Greene (2012)). If the error correlation structure is not correctly modeled using random effects in the linear model, then coefficient estimates are still consistent and unbiased, but the standard errors will be inefficient (Wooldridge (2012)). If the error correlation structure is not correctly modeled using random effects in the non-linear model, then the coefficient estimates are inconsistent (Greene (2012)). Unlike the linear model, this problem is can not be corrected using clustered standard errors. For a more detailed explanation see Greene (2012). For these reasons and ease of interpretation, I chose to report the linear specification in the main body of the paper. However, I demonstrate below that the reported results are robust to a non-linear specification.

To conduct a robustness check of the linear model specification, I report a probit model specification rather than a logit model specification. Theoretically, a logit specification requires that the error terms are distributed logistically while a probit specification requires that the error terms are distributed normally(Greene (2012)) . It is not clear which assumption is correct in this analysis. I report the results of a random effects probit specification in this appendix, but both the sign and significance of coefficients are robust to a random effects logit specification (not shown).

Table C.15: Effect of Treatments on Changes in Disease Prevention Investment Decisions

	Decision: Start Investing			Decision: Stop Investing		
	RE					
	(1)	(2)	(3)	(4)	(5)	(6)
Outbreak Information	-0.06	-0.10	-0.10	-0.07	-0.09	-0.09
	(0.07)	(0.10)	(0.10)	(0.07)	(0.10)	(0.10)
Exclusion Cost	-0.05	-0.08	-0.08	-0.03	-0.05	-0.06
	(0.07)	(0.09)	(0.09)	(0.07)	(0.09)	(0.10)
Outbreak Information *		0.08	0.08		0.05	0.05
Exclusion Cost		(0.14)	(0.14)		(0.14)	(0.14)
Round Effects	No	No	Yes	No	No	Yes
Observations	3840	3840	3840	3840	3840	3840
ρ	0.08***	0.08***	0.08***	0.08***	0.08***	0.09***
Model χ^2	1.16	1.49	48.03	1.11	1.23	30.66

Clustered robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.14 provides the results of the effects of the treatments on the probability of starting and stopping investment in disease prevention. Additionally, table C.15 gives the results for the marginal effects of the treatments on the probability of starting and stopping investment in disease prevention. These tables demonstrate that neither the outbreak information treatment nor the exclusion costs treatment alone are significant in explaining the probability of starting or stopping investment in disease prevention. Furthermore, the treatment effects are not significant when controlling for interaction effects or round effects. As in the linear model specification, this shows that there are additional relevant factors for explaining the

Table C.16: Effect of Treatments on Changes in Disease Prevention Investment Decisions

	Decision: Start Investing			Decision: Stop Investing		
	Marginal Effects					
	(1)	(2)	(3)	(4)	(5)	(6)
Outbreak Information	-0.01	-0.03	-0.03	-0.02	-0.03	-0.03
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Exclusion Cost	-0.01	-0.02	-0.02	-0.01	-0.02	-0.01
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Outbreak Information *		0.03	0.03		-0.02	-0.02
Exclusion Cost		(0.03)	(0.03)		(0.03)	(0.03)
Round Effects	No	No	Yes	No	No	Yes
Observations	3840	3840	3840	3840	3840	3840
ρ	0.08***	0.08***	0.08***	0.08***	0.08***	0.09***
Model χ^2	1.16	1.49	48.03	1.11	1.23	30.66

Clustered robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

decisions to start and to stop investing in disease prevention.

In tables C.16 and C.18 equations 1-6, I test the robustness of result 1 to a random effects probit specification. I demonstrate that subjects who face costs of social exclusion for not investing in disease prevention are significantly more likely to start investing in disease prevention and significantly less likely to stop investing in disease prevention. These effects are significant at the 1% level. Tables C.17 equations 1-6 give the marginal effect of exclusion costs on the probability of starting investment in disease prevention. Table C.19 equations 1-6 provide the marginal effect of exclusion costs on the probability of stopping investment in disease prevention.

Tables C.16 and C.18 equations 1-6 provide a robustness check of result 2 to a random effects probit specification. I show that providing subjects with outbreak information reduces the probability that they start investing in disease prevention and increases the probability that they stop investing in disease prevention. These effects are significant at the 5% and 1% levels, respectively. As in the linear model, a two period lag in disease incidence is correlated

Table C.17: Effect of Changes in Disease Incidence on Decision to Start Investing in Prevention

	Decision: Start Investing					
	RE					
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	0.10***	0.09***	0.08***	0.07***	0.26***	0.24**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.10)	(0.10)
Infected: 1 Lag		1.52***	1.43***	1.45***	1.43***	1.45***
		(0.11)	(0.13)	(0.12)	(0.13)	(0.12)
# Decision Changes			0.20***	0.20***	0.20***	0.20***
			(0.01)	(0.01)	(0.01)	(0.01)
Incidence: 2 Lags				-0.01		-0.01
				(0.02)		(0.02)
Squared Incidence: 1 Lag					-0.01*	-0.01
					(0.00)	(0.00)
Outbreak Information	-0.17**	-0.20**	-0.10*	-0.10	-0.10*	-0.09
	(0.07)	(0.10)	(0.05)	(0.06)	(0.05)	(0.06)
Exclusion Cost	0.23***	0.28***	0.28***	0.24***	0.27***	0.23***
	(0.08)	(0.11)	(0.07)	(0.09)	(0.07)	(0.09)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3840	3840	3840	3680	3840	3680
ρ	0.09***	0.20***	0.00	0.00	0.00	0.00
Model χ^2	96.00	624.33	758.04	737.85	758.62	738.60

Cluster robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.18: Effect of Changes in Disease Incidence on Decision to Start Investing in Prevention

	Decision: Start Investing Marginal Effects					
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	0.02***	0.02***	0.01***	0.01***	0.01***	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Infected: 1 Lag		0.29***	0.26***	0.26***	0.26***	0.26***
		(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
# Decision Changes			0.04***	0.04***	0.03***	0.04***
			(0.00)	(0.00)	(0.00)	(0.00)
Incidence: 2 Lags				-0.002		-0.002
				(0.00)		(0.00)
Squared Incidence: 1 Lag						
Outbreak Information	-0.04**	-0.04**	-0.02*	-0.02	-0.01*	-0.02
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Exclusion Cost	0.05***	0.05***	0.05***	0.04***	0.05***	0.04***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3840	3840	3840	3680	3840	3680
ρ	0.09***	0.20***	0.00	0.00	0.00	0.00
Model χ^2	96.00	624.33	758.04	737.85	758.62	738.60

Cluster robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.19: Effect of Changes in Disease Incidence on Decision to Stop Investing in Prevention

	Decision: Stop Investing					
	RE					
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	-0.06***	-0.06***	-0.06***	-0.06***	-0.08	-0.05
	(0.01)	(0.01)	(0.01)	(0.01)	(0.08)	(0.09)
# Infections: 1 Lag		-0.05***	-0.04***	-0.03***	-0.04***	-0.03***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
# Decision Changes			0.16***	0.15***	0.16***	0.15***
			(0.01)	(0.01)	(0.01)	(0.01)
Incidence: 2 Lags				-0.03**		-0.03**
				(0.01)		(0.01)
Squared Incidence: 1 Lag					0.00	-0.00
					(0.00)	(0.00)
Outbreak Information	-0.00	0.01	0.07*	0.12***	0.07*	0.12***
	(0.07)	(0.08)	(0.04)	(0.04)	(0.04)	(0.04)
Exclusion Cost	-0.18**	-0.22***	-0.17***	-0.27***	-0.17***	-0.27***
	(0.08)	(0.08)	(0.05)	(0.06)	(0.05)	(0.06)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3840	3840	3840	3680	3840	3680
ρ	0.09***	0.10***	0.00	0.00	0.00	0.00
Model χ^2	46.36	63.02	298.13	286.78	298.12	286.82

Cluster robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.20: Effect of Changes in Disease Incidence on Decision to Stop Investing in Prevention

	Decision: Stop Investing Marginal Effects					
	(1)	(2)	(3)	(4)	(5)	(6)
Incidence: 1 Lag	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
# Infections: 1 Lag		-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
# Decision Changes			0.04***	0.03***	0.04***	0.03***
			(0.00)	(0.00)	(0.00)	(0.00)
Incidence: 2 Lags				-0.01**		-0.01**
				(0.00)		(0.00)
Squared Incidence: 1 Lag						
Outbreak Information	-0.000	0.003	0.02**	0.03***	0.02*	0.03***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Exclusion Cost	-0.04**	-0.05***	-0.04***	-0.06***	-0.04***	-0.06***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Round Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3840	3840	3840	3680	3840	3680
ρ	0.09***	0.10***	0.00	0.00	0.00	0.00
Model χ^2	46.36	63.02	298.13	286.78	298.12	286.82

Cluster robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

with outbreak information causing outbreak information to lose significant when I control for it. Tables C.17 equations 1-6 provide the marginal effect of providing subjects with outbreak information on the probability of starting investment in disease prevention. Table C.19 equations 1-6 shows the marginal effect of providing subjects with outbreak information on the probability of stopping investment in disease prevention.

A robustness check of result 3 to a random effects probit specification is provided in tables C.16 and C.18 equations 1-6. I find that higher disease incidence in the previous period make subjects significantly more likely to start investing in disease prevention and significantly less likely to stop investing in disease prevention. This effect is significant at the 1% level. Like the linear model specification, table C.16 equations 5 and 6 demonstrate that controlling for squared disease incidence, a variable found to be significant in prior work, creates a multicollinearity problem. This causes the coefficient for a one period lag in disease incidence to become insignificant. Table C.17 equations 1-6 provide the marginal effect for a change in disease incidence on the probability that a subject starts investing in disease prevention. Table C.19 equations 1-6 demonstrates the marginal effect for a change in disease incidence on the probability that a subject stops investing in disease prevention.

I show that result 4 is robust to a random effects probit specification in table C.16 equations 2-6. As in the linear specification, a subject is only classified as stopping investment if they invested in prevention in the previous period but not in the current period. However, if a subject invested in prevention in the previous period, then they were fully protected against infection. So, a prior period infection can not explain the decision to stop investing in disease prevention. I find that being infected in the previous period increases the probability that a subject starts investing in disease prevention. This is significant at the 1% level. Table C.17 equations 2-6 provide the marginal effect of being infected in the previous period on the probability that a subject starts investing in disease prevention.

A subject's infection history is likely relevant in their decision to stop investing in disease

prevention although a subjects that stops investing would not have been infected in the previous period. Table C.18 equations 2-6 demonstrate that result 5 is robust to a random effects probit specification. Having been infected more times in prior rounds decreases the probability that a subject will stop investing in disease prevention. This effect is significant at the 1% level. Table C.19 equations 2-6 gives the marginal effect of an increase in the number of prior infections on the probability that a subject stops investing in disease prevention.

Subjects may have different willingness to try out disease prevention investment strategies. I use the number of times a subject has changed their investment decision in prior rounds as a proxy for willingness to try different strategies. Tables C.16 and 18 give the results of a robustness check of result 6 to a random effects probit specification. I find that a willingness to try out different prevention investment strategies is associated with a significant increase in the probability that a subject starts investing in disease prevention and with a significant increase in the probability that a subject stops investing in disease prevention. These effects are significant at the 1% level. Tables C.17 and C.19 provide the marginal probability of the willingness to switch investment strategies on the probability of starting investment in prevention and stopping investment in prevention, respectively.

Tables C.16 and C.18 equations 4 and 6 demonstrate that result 7 is robust to a random effects probit model specification. In table C.16, I show that historical disease incidence is not significant in explaining the decision to start investing in disease prevention. In table C.18, I find higher historical disease incidence is associated with a significant reduction in the probability that a subject stops investing in disease prevention. This effect is significant at the 5% level. Tables C.17 and C.19 provide the results of the marginal effect of increased historical disease incidence on the probability of starting and stopping investment in disease prevention, respectively.

Carpenter (2016) previously found that there where non-linear effect in the impact of disease incidence on the probability of investing in disease prevention. Tables C.16 and 18

equations 5 and 6 provide the results of the random effects probit specification controlling for non-linearity in disease incidence (squared incidence: 1 lag). Equation 5 in table C.16 demonstrates the result found in Carpenter (2016) that the probability that a subject starts investing in disease prevention increases at a decreasing rate. However, this result is not robust to controlling for historical disease incidence, equation 6, or the linear specification reported in the main body of the paper. Additionally, as in the linear specification, I find that controlling for non-linearity in disease incidence in table C.18 equations 5 and 6 causes the coefficient for disease incidence to lose significance. While I think that disease incidence likely has a non-linear effect on the probability of starting or stopping investment in disease prevention, controlling for non-linearity in disease incidence creates a multicollinearity issue in the model.