Qualitative Analysis of Transience in Population Dynamics

by

Mojdeh Mohtashemi

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

April 2000
[June 2000]

Author .....
Department of Electrical Engineering and Computer Science
April 28, 2000

Certified by .....
Peter Szolovits
Professor
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

Qualitative Analysis of Transience in Population Dynamics

by

Mojdeh Mohtashemi

Submitted to the Department of Electrical Engineering and Computer Science
on April 28, 2000, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

To date, mathematical models of population dynamics have consistently focused on understanding the long term behavior of the interacting components, where the steady state solutions are paramount.

For most acute infections, however, the long run behavior of the pathogen population is of little importance to the host and population health. We introduce the notion of *transient pathology*, where the short term dynamics of within-host interaction between the immune system and pathogens is the principal focus. We identify the amplifying effect of the absence of a fully operative immune system on the pathogenesis of the initial inoculum, and its implication for the acute severity of the infection. We then formalize the underlying dynamics, and derive two measures of transient pathogenicity: the *peak of infection* (maximum pathogenic load) and the *time to peak of infection*, both crucial to understanding the early dynamics of infection and its consequences for early intervention.

Moreover, the differential importance of the windows of opportunity for different pathologies in individual patients, is related to the time structure of the lives of people in different vulnerability classes. Today, despite decades of studies and the relative abundance of mortality data, our understanding of the distribution of vulnerability to disease and death remains inadequate. We introduce a model of mortality selection partially offset by social mobility, to simulate the dynamics of vulnerability of a population cohort that is heterogeneous in health and death. We then propose a methodology for transforming mortality data from the "age-domain" to the "time-domain". Using our model of vulnerability, applied to mortality data in the time-domain, we identify the significance of selection and mobility for the dynamics of mortality of different vulnerability classes and the population as a whole. Finally, we compare the mortality experience of different populations; focusing on the transient phenomenon of "mortality crossover" between different pairs of populations, we identify the potential factors producing the phenomenon, and make diagnostic use of its underlying processes.

Thesis Supervisor: Peter Szolovits
Title: Professor

Acknowledgments

For four months short of three years, I have had the honor of working with two remarkable men. It never occurred to me that in such a short time I would become so fond of them.

Peter Szolovits, my advisor and the leader of MEDG gave me the support and freedom to develop an unconventional dissertation topic. Pete is open to possibilities; promotes independence in one's intellectual development; and encourages his students to be fully imaginative. His broad intelligence coupled with strong leadership have made him the scientific authority and the exceptionally effective advisor that he is. I have come to depend on his generosity, good advice, and wonderful stories. He has created a very comfortable, yet professional and stimulating, research environment at MEDG.

Richard Levins, my mentor from Harvard, with whom I worked closely during the past two years, is a visionary and a scholar in the classical tradition, who has devoted his life to "making the obscure obvious" in all matters of life. His questions are intriguing; his insights are shocking; and his kindness is overwhelming. His noble existence and passion for change have brought me back the hope that I had lost years ago. Lest I wax rhapsodic I shall conclude by saying that my love and admiration for this man may only be described effectively by resorting to Rumi, the 13th century Persian mystic, and his poems for his mentor Shams of Tabriz.

Other associates at MEDG whom I have been fortunate to know include: William Long, who gracefully offered to proof-read a substantial portion of my thesis and has taught me so much about the "comma" with his familiar gentle and calm manner. Jon Doyle, with whom I have enjoyed occasional mythological and philosophical chats; Jon has helped make the work environment at MEDG so pleasant and so invigorating for research. And Lik Mui who read the entire thesis enthusiastically and pointed out the typos.

Marilyn Pierce, at the MIT EECS Graduate Office, has been a dependable source of comfort and speedy assistance through various stages of my graduate studies. I am proud to know her.

Outside of academia, I have two favorite people to acknowledge. My sister Mojgan, for her selflessness when it comes to me; and Ali, without whom I would have been long gone. I adore them both.

To A. Jazayeri and many others we lost with the death of the social transience of 1979-1982, in Iran

And to Ali

Contents

List of Figures

Chapter 1

Introduction

1.1 Why Qualitative Modeling

Our world is increasingly complex. When we attempt to model problems of the real world, we tend to simplify these problems so we can study the models instead of the real systems. The question is: how much simplification is not too much so that the underlying model is still a reasonable representation of the system under study? When modeling, there are two methodologies to consider: *qualitative* and *quantitative*. With the computational advances in the past decade or so, quantitative modeling has gained so much popularity, yet the shortcomings of this methodology persist. Let us examine this more closely through a brief comparative study of the two approaches.

1.1.1 Generalizability of Results

Consider, as an example, a typical predator-prey system as follows

$$\begin{aligned} S(0) &= s_0, \quad I(0) = i_0 \\ \frac{dS}{dt} &= B - \mu_s S - \lambda SI \\ \frac{dI}{dt} &= \lambda SI - \mu_i I \end{aligned}$$

where S and I are the populations of *susceptible* and *infected* with respect to an infectious agent; s_0 and i_0 are initial values of S and I respectively; λ is the effective infection rate when individuals of type S and I come into contact; μ_s and μ_i are the death rates of susceptible and infected people respectively; and B is the size of the birth cohort. We assume new people are born into the class of susceptible.

If the constants of the above system are known, numerical methods can give numbers for S and I over time that later can be graphed to examine temporal behavior of the variables, relation between S and I, the impact of different intervention strategies, etc. [1].

But suppose that one or more of the parameters are no longer constant, in which case we may ask the question: how does variation in the parameters of one variable

affect the other variable and the dynamics of the system as a whole? With numerical methods this question can be tested for different values of the changing parameter, but a set of numerical answers are only valid answers for particular situations, and therefore cannot be generalized to unexamined situations. Hence, our understanding of the underlying processes will not be enhanced and will be limited to special cases.

In contrast, qualitative analysis does not provide us with exact solutions or precise answers to the behavior of systems, but qualitative methodologies such as loop analysis [7, 8, 19], do allow us to study the effect of variation in the parameters of systems in different directions and in great depth.

1.1.2 Measurability of Parameters

Now suppose that our numerical modeler learns, due to discrepancies between the numerical results and observation or other inconsistencies, that in our epi-system the rate of contagion λ, cannot be treated as constant and is dependent on people's behavior as well as the prevalence of the disease which itself is time-variant. So we get the following dynamics

$$\begin{aligned} S(0) &= s_0, \quad I(0) = i_0, \quad \lambda(0) = \lambda_0 \\ \frac{dS}{dt} &= B - \mu_s S - \lambda SI \\ \frac{dI}{dt} &= \lambda SI - \mu_i I \\ \frac{d\lambda}{dt} &= \alpha\left(\frac{\lambda_0}{1+p(t)} - \lambda\right) \end{aligned}$$

where λ_0 is the initial infection rate; α is the responsiveness of behavior; and $p(t)$ is the prevalence of the disease at time t. The new equation has an interesting interpretation. On the one hand, it is a decreasing function of the prevalence $p(t)$, implying that as the proportion of infected individuals approaches 100%, people panic and modify their behavior so that the infection rate goes up at a smaller rate with time. On the other hand, an increase in the prevalence itself is a result of unsafe social behavior and hazardous contacts. At the same time, public health education and awareness campaigns positively impact people's preventive behavior which reduces the infection rate independent of prevalence; thus the term $-\alpha\lambda$.

Adding one more variable to the system complicates the analysis to a great extent. In the new equation, the parameter α may not be readily measurable, and collecting any amount of quantitative information on α may be quite a costly and tedious task. In fact, most social parameters are not measurable. In such situations, our numerical modeler is likely to omit the new variable and its corresponding equation from the dynamic system. This as we know, is routinely practiced in quantitative modeling. But if the effect of the new variable is indeed critical in explaining the general dynamics and possible anomalies, our understanding of the dynamics of S will be flawed, and therefore the mismatch between the state of knowledge and the numerical results will not be resolved. Any claims made on the basis of such "linear"

results are hazardous and at best misleading.

With qualitative modeling, however, we may include as many variables as we find necessary to best describe the dynamics of the system under study despite the lack of data [7, 8, 19, 20]. In other words, "Instead of accumulating all the relevant information, we see how much we can avoid measuring and still understand the system." [19]

1.2 Why Modeling Transience

Daily transitory changes in the physical and social environment can make our bodies undergo a wide range of fluctuations. Changes in the temperature; time and content of food intake; patterns of physical and emotional stress; social policies; etc, can cause rapid variation in the blood pressure, heart rate, induction of immune elements, neural capacity, emotional stability, and other aspects of our physiology. Hence, we ask: *When are fluctuations of particular durations relevant to disease processes, and when can they be averaged out?* Mathematics of dynamical systems has routinely explored the long term, or average, behavior of dynamical components. But if some fluctuations are important in defining the health of a population, then we need to identify the underlying processes producing them.

When an invading pathogen finds its way through the epithelial surfaces, and the normal physiology of the body is perturbed by infection, components of cellular immunity are able to respond quickly, but it takes several days for the humoral immunity to be activated. But the immune system may also be inoperative for an initial period, for many different reasons. For instance, "a pulse of 500 grams of sugar can inhibit immune activity for up to 5 hours; an immunization can tie it up for a week or two; a major emotional trauma can reduce immune activity for months." (Richard Levins, unpublished); malnutrition can delay the immune response indefinitely. Therefore, if the immune system is unable to respond for some initial period, the pathogen will reproduce freely until the host's defenses are activated. The newly activated immunity is then to face an exponentially "amplified initial inoculum". Therefore we ask: *How does the size of the initial inoculum and the initial delay period affect the outcome of infection?*

This can be approached empirically: for what diseases does the initial inoculum matter? It can also be approached theoretically: after an initial infection, the pathogen's reproduction races against the removal by the immune system. If the immune system is responsive, it can exceed the reproduction of the pathogen from the start, so that the pathogen population decreases. But if the immune response has to be induced, the pathogen may increase for a time before the immune system is activated and overwhelmed. During this period, the pathogen may produce sufficient damage to kill the host.

In Chapter 2, we develop a mathematical model for simulating the transient dynamics of within-host interaction between the immune system and a pathogen. We will then derive two measures of transience pertaining to the early dynamics of an acute infection namely, the "peak of infection" (h_{peak}), and the "time to peak of

infection" (t_{peak}). We relate these transient outcomes of the model to the initial inoculum, the period during which the immune system is not fully effective, and various infection-specific parameters of reproductive rate, induction rate of the immune system, and efficacy of the immune elements. Finally, we make inferences about the diagnostic use of such inter-dependencies, as "how to intervene" and "when to intervene".

In Chapter 3, we relate the differential importance of the windows of opportunity for different pathologies to the time structure of the lives of people in different vulnerability classes of a population. A population is composed of individuals who are heterogeneous in their vulnerability to disease and death. In public health and demography, the distribution of vulnerability is only understood in terms of age and the underlying genetic basis. Although age and the genetic predisposition are important causes of variation in the distribution of vulnerability, they are not the only such causes. A substantial part of such variation is due to heterogeneity in social conditions of individuals. Therefore, it is essential to understand how certain factors such as mortality selectivity and differential mobility in a population affect morbidity and mortality of different vulnerability classes and the population as a whole.

To model the dynamics of vulnerability, we develop a mathematical model of selection partially offset by mobility, to simulate the dynamics of mortality of a population heterogeneous in health. In essence, we divide a population cohort into two vulnerability classes with respect to disease processes. Over time, people may move from one vulnerability class to another due to aging or factors affecting social mobility. They also die. The relation between these two processes will determine whether people die mostly in conditions of chronic good health or of high risk and poor health. Therefore, we have two outcomes of concern: total mortality and the proportion dying in different vulnerability classes. Our goal is to identify the domain of conditions under which total mortality (and mortality from classes with poor health) would decline, while death from classes with good health would rise or stay steady. This means that people will live relatively healthy lives up to the time of death. For the society as a whole, this results in lowering the terminal investment in health.

When comparing populations, interesting "anomalies" in the mortality curves of two populations may arise. One such anomaly is the occurrence of a mortality crossover between the Black and White populations of the United States: after about age 80, Blacks seem to have a lower mortality rate than Whites. Today, despite the relative richness of mortality data and nearly 80 years after the discovery of the phenomenon, our understanding of mortality crossover remains inadequate. Therefore we ask, *When will a population, initially exposed to greater force of mortality, have a lower death rate*? Section 3.4 is an attempt to explain the phenomenon in light of our mathematical model of vulnerability. We first propose a methodology for transforming mortality data from the age-domain to the time-domain. This is necessary because our model simulates the mortality experience of a non-aging cohort over time. We will then identify the conditions under which the mortality curves of two populations cross as well as identifying the processes that govern the dynamics of the mortality crossover. We will then extend the results to other pairs of populations. Finally, in section 3.5, we generalize the results of the 2-vulnerability-class to $n \geq 2$.

Chapter 2

Transience in Infectious Disease

2.1 Motivation and Background

Microorganisms or *pathogens* are often characterized by their small size and their short generation time during which they replicate at very high rates. The short infection period, relative to the life span of the host, is another important feature of microorganisms that causes *acute* pathology in their hosts. Once recovered from infection, the host acquires life-long or transient immunity against re-infection. Such a characterization of pathogens broadly includes viruses, bacteria, fungi, and protozoa.

2.1.1 Immunology of Infectious Disease

Our bodies are constantly exposed to infectious agents, yet we seldom acquire an infectious disease. The epithelial surfaces of the body form an efficient barrier to most foreign substances (antigens). However, those that bypass this barrier and successfully establish a site of infection, will face the body's immune defenses. The immune system performs two functions: it distinguishes between the organism's own cells and any intruding pathogens, and it fights the invaders by means *innate* and *adaptive* immunity.

The first line of body's defense mechanisms, are the phagocytes of the innate immunity. These cells pre-exist in all vertebrates and are able to react to infection almost immediately. Phagocytic cells, recognize, engulf and digest a wide range of common bacterial infections. If the innate defenses of the host do not recognize the invasive pathogens and are evaded, the next line of defense, *i.e.* the lymphocytes of the *induced* or adaptive immunity will be required to combat the infection.

Adaptive immunity is characterized by two responses: *cellular* and *humoral.* The cellular immunity consists of two major components: the $CD4^+$ T lymphocytes (a.k.a. *helper T cells*), and the $CD8^+$ T lymphocytes (a.k.a. *killer T cells*). The helper T cells reproduce to form a command center which in turn stimulates the production of the killer T cells and the B cells in the humoral response. The killer T cells then seek out the site of infection, and destroy the infected cells. In the humoral (or *antibody*) response, helper T cells activate the B lymphocytes, which are blood cells. The B cells

can then produce antigen-specific antibody molecules which can target and destroy the specific pathogen at hand.

Components of cellular immunity are able to respond to infection within 4-96 hours, but there is usually a delay of 4-7 days before humoral immunity makes its initial response. On re-exposure to same microbial antigens later in life there is an accelerated response in which larger amounts of specific antibodies are formed after only one or two days. The capacity to respond in this manner often persistes for life, and depends on the presence of "memory cells".

2.1.2 Transient Pathology – Window of Vulnerability

Every infection is a race between the ability of the invading microorganism to multiply and cause disease, and the ability of the host to mobilize defenses. When an invading pathogen finds its way through the epithelial surfaces and establishes a site of infection, components of cellular immunity are able to respond to infection rather quickly. However, it takes several days for the humoral immunity to be induced. When an infection is naturally acquired, the infecting dose generally consists of a small number of microorganisms. This, on its own, is quite insufficient to cause significant pathology or even stimulate an immune response. But the pathogen then multiplies, and if the immune system has to be induced, and this may indeed take several days, then such an "initial delay period" on the part of the host can be critical. Hence, the invading pathogen is most successful if it requires for its neutralization antibodies that have not previously been stimulated. For most acute infections, the pathogen replicates rapidly and at very high rates and thus can manage to stay ahead of the host's humoral defenses and take advantage of such a delay period.

If the immune system is readily activated, it may surpass the reproduction of the pathogen from the start, so that the pathogenic load declines. But if the immune response has to be induced, during the window of opportunity for the pathogen, there can be extensive increase in antigenic mass which may lead to the death of the host. It is during this initial delay period that the pathogen replicates freely until the humoral defenses are activated. Hence, the initial inoculum that only consisted of a few microorganisms at the time of inoculation, in the absence of a fully effective immunity grows almost without bound. The newly activated immunity is now to face an exponentially "amplified initial inoculum". Therefore we ask: *How does the size of the initial inoculum and the initial delay period affect the outcome of infection*?

It is reported that for *Shigella dysteriae* as little as 10 bacteria are sufficient to cause oral infection [13]. In malaria, Marsh reports [12], that for *Plasmodium falciparum* "in a small child a sporoziote dose in the hundreds could result in a clinically significant patent parasitaemia on the first round of blood-stage cycle". It is claimed by the same author that death occurs within a few days of inoculation with $> 10^5$ sporozoites. In measles it is reported by Hope-Simon [6], that the second case in a household is always more severe than the first, presumably because more intense contact results in a greater inoculum. For bacterial meningitis, the pathogen is present in some 10-20% of the population without causing symptoms and then occasionally escapes from control (Richard Levins, "Mini Essays on Health", unpublished).

Hence, it is of great importance to have a mathematical framework for modeling the early dynamics of the immuno-patho system, and for making inferences about the transient outcomes of infection namely, the "time of crisis" and "extent of crisis", based on the size of the initial inoculum, the period during which the immune system is not fully effective, and the infection-specific parameters.

Finally, understanding the early dynamics of infection has direct consequences for clinical medicine and public health intervention. Anticipating the maximum pathogenic load and the time of its occurrence has immediate implication for the choice of plausible intervention scheme(s), and assessing the duration of the critical period for intervention beyond which any intervention strategy may prove ineffective.

2.2 A Framework for Analysis of Transience

2.2.1 The Interaction Model

Our model of the dynamical interaction between the immune system and a pathogen is a time-dependent, two-variable, nonlinear system of ordinary differential equations. The variables are $I(t)$, the immune level of an infected host at time t; and $P(t)$, the pathogenic load at time t. The following system of ordinary differential equations models the dynamics of the immuno-patho system

$$\frac{dI}{dt} = a_0 - \mu I(t) + kP(t) \tag{2.1}$$

$$\frac{dP}{dt} = rP(t) - mI(t)P(t) \tag{2.2}$$

where a_0 represents the innate immunity; μ is the rate of decay of the immune system; k is the rate of induction of the immune system; r is the reproductive rate of the pathogen; and m is the rate of removal of the pathogen by the immune system. It is interesting to note the dual effects of some of the parameters: a_0 and μ are both properties of the immune system, yet μ favors the pathogen; m and k, on the other hand, are properties of interaction between the immune system and the pathogen, yet they both favor the immune system.

Now suppose that humoral immunity is not able to respond for some initial period, θ. During this *initial delay period*, the pathogen will replicate more freely since the only control element is due to the innate immunity a_0, partially offset by μ: $\frac{a_0}{\mu}$. Hence, in the absence of fully responsive immune system the dynamics of the patho-system is as follows

$$P(0) = p_0 \tag{2.3}$$

$$\frac{dP}{dt} = P(t)(r - m\frac{a_0}{\mu}) \qquad \text{for} \quad 0 \leq t \leq \theta \tag{2.4}$$

where p_0 is the arbitrary initial pathogenic inoculum. Equations 2.3-2.4 have the simple solution: $P(t) = p_0 e^{(r - m\frac{a_0}{\mu})t}$. This initial dynamics at $t = \theta$, when the immune

system becomes fully responsive, sets the initial conditions for the system of equations 2.1-2.2 so that $P(0) = p_0 e^{(r - m\frac{a_0}{\mu})\theta}$ becomes the initial pathogenic load for the new dynamics. We will refer to this phenomenon as the *amplified initial inoculum*, $P(0)$ below. Therefore, we have

$$I(0) = \frac{a_0}{\mu}, \quad P(0) = p_0 e^{(r - m\frac{a_0}{\mu})\theta} \tag{2.5}$$

$$\frac{dI}{dt} = a_0 - \mu I(t) + kP(t) \tag{2.6}$$

$$\frac{dP}{dt} = rP(t) - mI(t)P(t) \tag{2.7}$$

2.2.2 Regions of Pathogenicity

Our dynamic model of interaction between the immune elements and a pathogen consists of three qualitatively distinct regions of pathogenicity depending on the steady state values and the relative magnitude of the parameters of the system of equations 2.5-2.7. The equilibrium values (I^*, P^*) of the system are the solutions of the algebraic equations

$$a_0 - \mu I + kP = 0 \quad \text{and} \quad rP - mIP = 0$$

namely, the points $(I^*, P^*) = (\frac{a_0}{\mu}, 0)$ and $(I^*, P^*) = (\frac{r}{m}, \frac{r\mu - ma_0}{mk})$. The qualitative behavior of the solutions of the system of equations 2.5-2.7 and the conditions for local stability in each region can be determined by linearizing the system in the neighborhood of each critical point.

1. Linearizing the system in the neighborhood of $(\frac{a_0}{\mu}, 0)$, we have

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix} I \\ P \end{pmatrix} &= \begin{pmatrix} \frac{\partial}{\partial I}(\frac{dI}{dt}) & \frac{\partial}{\partial P}(\frac{dI}{dt}) \\ \frac{\partial}{\partial I}(\frac{dP}{dt}) & \frac{\partial}{\partial P}(\frac{dP}{dt}) \end{pmatrix}\begin{pmatrix} I \\ P \end{pmatrix} \\ &= \begin{pmatrix} -\mu & k \\ 0 & \frac{\mu r - ma_0}{\mu} \end{pmatrix}\begin{pmatrix} I \\ P \end{pmatrix} \end{aligned} \tag{2.8}$$

The characteristic polynomial, $p(\lambda)$ of the above system is

$$p(\lambda) = \lambda^2 + \lambda\frac{\mu^2 + ma_0 - r\mu}{\mu} + (ma_0 - r\mu)$$

Setting $p(\lambda) = 0$ and solving for the eigenvalues, we get: $\lambda_1 = -\mu$ and $\lambda_2 = \frac{\mu r - ma_0}{\mu}$.

For the system to be locally stable, both eigenvalues must be negative real numbers. Hence, we must have the following conditions on the system parameters

1-(a) $\mu > 0$

1-(b) $ma_0 > \mu r$

2. Linearizing the system in the neighborhood of $(\frac{r}{m}, \frac{r\mu - ma_0}{mk})$, we have

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix} I \\ P \end{pmatrix} &= \begin{pmatrix} \frac{\partial}{\partial I}(\frac{dI}{dt}) & \frac{\partial}{\partial P}(\frac{dI}{dt}) \\ \frac{\partial}{\partial I}(\frac{dP}{dt}) & \frac{\partial}{\partial P}(\frac{dP}{dt}) \end{pmatrix}\begin{pmatrix} I \\ P \end{pmatrix} \\ &= \begin{pmatrix} -\mu & k \\ \frac{ma_0 - \mu r}{k} & 0 \end{pmatrix}\begin{pmatrix} I \\ P \end{pmatrix} \end{aligned} \tag{2.9}$$

The characteristic polynomial of the above system is: $p(\lambda) = \lambda^2 + \mu\lambda + (\mu r - ma_0)$, which can be solved to get the eigenvalues: $\lambda_3 = \frac{-\mu + \sqrt{\mu^2 - 4(r\mu - ma_0)}}{2}$ and $\lambda_4 = \frac{-\mu - \sqrt{\mu^2 - 4(r\mu - ma_0)}}{2}$. For the system to be locally stable around $(\frac{r}{m}, \frac{r\mu - ma_0}{mk})$ we must have $\mathrm{Re}(\lambda_i) < 0$, $i = 3, 4$. We derive the stability conditions using the Routh-Hurwitz algorithm.

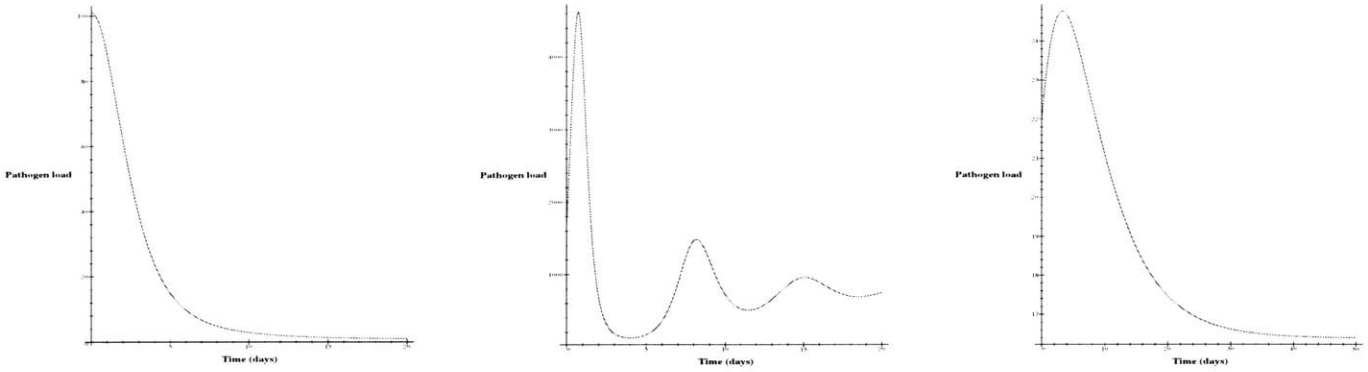
$$\begin{array}{c|cc} \lambda^2 & 1 & r\mu - ma_0 \\ \lambda^1 & \mu & 0 \\ \lambda^0 & \frac{\mu(r\mu - ma_0)}{\mu} & 0 \end{array}$$

The stability criterion requires that all the terms in the left-hand column of the above table have the same sign. That is, the following conditions must be satisfied for the system to be locally stable.

2-(a) $\mu > 0$

2-(b) $\mu r > ma_0$

If the parameters of the system are such that condition 1-(b) is satisfied, then clearly condition 2-(b) is violated, in which case the eigenvalues λ_3, λ_4 will be real and of opposite signs since $\sqrt{\mu^2 + 4(ma_0 + r\mu)} > \mu$. Consequently, the critical point $(\frac{r}{m}, \frac{r\mu - ma_0}{mk})$ will be an unstable saddle point. On the other hand, if condition 2-(b) is satisfied, then λ_1, λ_2 are real and of opposite signs, and the critical point $(\frac{a_0}{\mu}, 0)$ will be an unstable saddle point. We distinguish two different regions under this



(a) (b) (c)

Figure 2-1: Regions of pathogenesis. (a) Region I: Pathogen population growth cannot pick up from the start; immune elements are strong. (b) Region II: Pathogen produces sufficient damage before the immune elements are able to regulate its population growth. (c) Region III: Pathogen may exhibit small growth, but is not capable of producing extensive damage due to its relatively small reproduction rate; immune elements regulate its growth to an equilibrium value quite rapidly.

condition. If $\mu^2 < 4(r\mu - ma_0)$, then λ_3, λ_4 are complex conjugates and the dynamics will be a damped oscillation to a steady state. Otherwise, if $\mu^2 > 4(r\mu - ma_0)$, then λ_3, λ_4 are negative real, in which case, due to the relatively small reproduction rate, r, the pathogen is not capable of producing extensive damage and is quickly regulated by the immune elements. In this case, a small peak may be observed before the steady decline to a steady state. Putting it all together, we distinguish three regions of pathogenicity as follows:

Region I: If $(\frac{a_0}{\mu}, 0)$ is a stable steady state, then by condition 1-(b), it must be that the offensive parameters of the immune system, i.e., a_0 and m are much larger compared to r and μ. Hence, the immune system can exceed the reproduction of the pathogen from the start and the pathogenic load declines. Figure 2-1(a) illustrates the dynamics of this region.

Region II: If $(\frac{r}{m}, \frac{r\mu - ma_0}{mk})$ is a stable steady state and $\mu^2 < 4(r\mu - ma_0)$, then the pathogenic load will take off until the immune system is able to react and reverse the pathogen population growth and regulate it to some equilibrium value through damped oscillations. Figure 2-1(b) illustrates the dynamics of region II.

Region III: If $(\frac{r}{m}, \frac{r\mu - ma_0}{mk})$ is a stable steady state and $\mu^2 > 4(r\mu - ma_0)$, then it must be that the pathogen reproduction rate, r is relatively small but that condition 2-(b) is still satisfied. In this case, a small peak may be observed, but infection is promptly regulated by the immune elements. Figure 2-1(c) illustrates the dynamics of this region.

2.2.3 Measures of Pathogenicity: t_{peak} and h_{peak}

Clearly, we are interested in the analytic properties of region II and III since these are the regions in which the dynamics of infection may exhibit interesting transient properties. In the remainder of this paper, we will focus on the analytical properties of region II, since region III can be characterized as a trivial case of region II. We define two measures of transient pathogenicity corresponding to the dynamics of region II as follows:

Definition: Region II, wherein the pathogenic overload can cause severe disease or death in a short time period, best characterizes the dynamics of acute infections. We define two measures of pathogenesis pertaining to the transient dynamics of this region:

1. Maximum pathogenic load, or h_{peak} (for height of the peak), defined as the pathogen population at the peak of infection. If this is large enough, the host may not recover.

2. Time to peak of infection, or t_{peak} (for time-to-peak), defined as the time from the full activation of the immune system to the peak of infection. If the intervention is applied past this point, the host may not recover.

In later sections, we develop the analytic tools needed to explore the dynamics of these measures of pathogenicity.

2.3 A Quadratic Model of Interaction

To date, we do not know of any exact analytic solutions for the nonlinear system of equations 2.5-2.7. Using the method of time averaging [19], however, we do know how to approximate these solutions. We are interested in functional solutions of $I(t)$ and $P(t)$ in terms of the initial conditions and various parameters of the system. To obtain such solutions, we first integrate equation 2.6

$$\int_0^t \frac{dI}{d\tau} d\tau = \int_0^t [a_0 - \mu I(\tau) + kP(\tau)] d\tau$$

which implies that

$$I(t) - I(0) = a_0 t - \mu \int_0^t I(\tau) d\tau + k \int_0^t P(\tau) d\tau$$

or equivalently

$$\begin{aligned} I(t) &= \frac{a_0}{\mu} + a_0 t - \mu \int_0^t I(\tau) d\tau + k \int_0^t P(\tau) d\tau \qquad \text{since } I(0) = \tfrac{a_0}{\mu} \\ &\stackrel{Def}{=} \frac{a_0}{\mu} + a_0 t - \mu t \mathrm{E}[I] + kt \mathrm{E}[P] \end{aligned} \tag{2.10}$$

where $\mathrm{E}[x] \stackrel{Def}{=} \frac{1}{t}\int_0^t x(\tau)d\tau$ is the expected or average value of a variable $x(t)$. Adapting the notation $\overline{x}(t)$ for $\mathrm{E}[x]$, we rewrite equation 2.10 as follows

$$I(t) = \frac{a_0}{\mu} + a_0 t - \mu t\overline{I}(t) + kt\overline{P}(t) \tag{2.11}$$

where $\overline{I}(t)$ and $\overline{P}(t)$ are the average values of the immune level and the pathogen population respectively, for the period $[0, t]$. Clearly, $\overline{I}(t)$ and $\overline{P}(t)$ are time variant entities; furthermore, all our derivations have been exact thus far. To approximate the "transient" solutions of $I(t)$ and $P(t)$, we need to define a period during which these solutions would closely model the behavior of the immuno-patho system at the peak. This period is of particular importance for the upcoming transient analysis, and will be discussed in detail in section 2.3.1. Therefore, if we further approximate $\overline{I}(t)$ and $\overline{P}(t)$ with constant terms $\overline{I}$ and $\overline{P}$, derived for the period of interest, then $I(t)$ can be approximated as follows

$$\tilde{I}(t) = \frac{a_0}{\mu} + a_0 t - \mu\overline{I}t + k\overline{P}t \tag{2.12}$$

where $\tilde{I}(t)$ represents an approximate solution of $I(t)$ for the period of interest. Similarly, we can find an approximate functional form for $P(t)$ by integrating equation 2.7 for the period of interest

$$\begin{aligned}\int_0^t \frac{dP}{d\tau}d\tau &= r\int_0^t P(\tau)d\tau - m\int_0^t I(\tau)P(\tau)d\tau \\ &\stackrel{Def}{=} r\overline{P}(t) - m\int_0^t I(\tau)P(\tau)d\tau\end{aligned} \tag{2.13}$$

In equation 2.13, if we replace $I(t)$ with the value of $\tilde{I}(t)$ from equation 2.12, and $\overline{P}(t)$ with the constant term $\overline{P}$, we will then have

$$\begin{aligned}\tilde{P}_q(t) - P(0) &= r\overline{P}\,t - m\int_0^t (\frac{a_0}{\mu} + a_0\tau - \mu\overline{I}\tau + k\overline{P}\tau)P(\tau)d\tau \\ &= r\overline{P}\,t - m\frac{a_0}{\mu}\int_0^t P(\tau)d\tau - m(a_0 - \mu\overline{I} + k\overline{P})\int_0^t \tau P(\tau)d\tau \\ &= r\overline{P}\,t - m\frac{a_0}{\mu}\overline{P}\,t - m(a_0 - \mu\overline{I} + k\overline{P})\overline{P}\,\frac{t^2}{2} \qquad \text{using integration by part}\end{aligned}$$

or, equivalently

$$\tilde{P}_q(t) = P(0) + \overline{P}t(r - m\frac{a_0}{\mu}) - m\overline{P}\frac{t^2}{2}(a_0 - \mu\overline{I} + k\overline{P}) \tag{2.14}$$

where the quadratic function, $\tilde{P}_q(t)$ ("q" for quadratic), represents an approximation of $P(t)$. As will be demonstrated in later sections, having approximated $P(t)$ and $I(t)$ in terms of the initial inoculum, various parameters of the system, as well as

short term average measures of immunity and pathogenesis, namely $\overline{I}$ and $\overline{P}$, we are now equipped with a powerful analytic tool for studying the transient behavior of infection, as well as predicting the severity and time of the peak of infection.

2.3.1 Measures of Transient Pathogenicity: t_{peak} and h_{peak}

Equipped with the qualitative solutions of section 2.3, we derive the two transient measures of the patho-system t_{peak} and h_{peak} pertaining to region II. Differentiating $\tilde{P}_q(t)$ from equation 2.14 with respect to t we get

$$\dot{\tilde{P}}_q = \overline{P}(r - m\frac{a_0}{\mu}) - m\overline{P}t(a_0 - \mu\overline{I} + k\overline{P}) \tag{2.15}$$

since $\overline{I}$ and $\overline{P}$ are constant. Equation 2.15 can be equated to zero and solved for t for which the pathogenic load is maximum. That is

$$\begin{aligned} t_{peak} &= \frac{r - m\frac{a_0}{\mu}}{m(a_0 - \mu\overline{I} + k\overline{P})} \\ &= \frac{r\mu - ma_0}{\mu m(a_0 - \mu\overline{I} + k\overline{P})} \end{aligned} \tag{2.16}$$

For $t_{peak} > 0$, both the numerator and denominator must have the same sign. In section 2.2.2, we showed that in order for the interaction model of the system of equations 2.5-2.7 to be locally stable in region II, we must have the condition

$$r\mu - ma_0 > 0 \tag{2.17}$$

Hence for $t_{peak} > 0$, since its numerator is always positive, we must also have

$$a_0 - \mu\overline{I} + k\overline{P} > 0 \tag{2.18}$$

since $\mu > 0$ and $m > 0$. Therefore both the numerator and denominator of t_{peak} are positive. To derive h_{peak}, we evaluate $\tilde{P}_q(t)$ at t_{peak} as follows

$$\begin{aligned} h_{peak} &= \tilde{P}_q(t_{peak}) \\ &= P(0) + \overline{P}\frac{(r - m\frac{a_0}{\mu})^2}{2m(a_0 - \mu\overline{I} + k\overline{P})} \\ &= p_0 e^{(r - m\frac{a_0}{\mu})\theta} + \overline{P}\frac{(r\mu - ma_0)^2}{2\mu^2 m(a_0 - \mu\overline{I} + k\overline{P})} \end{aligned} \tag{2.19}$$

What are the consequences of such transient outcomes of an acute infection for the host? Can these measures be utilized to device effective early intervention strategies? How does the duration of inactivity of the humoral immunity θ, and the size of the amplified initial inoculum $P(0)$, affect these measures, the critical timing for intervention, and the choice of intervention scheme? θ not only influences the early dynamics of infection, it may also be influenced by factors such as the state of nutrition or the

stress level of the host. This is unlike some of the other parameters of the system, such as r or μ, which may be inherent physiological characteristics of a pathogen and the immune system, and that we may or may not be able to influence them. This has significant consequences for intervention. A boost to the immune system, by way of good nutrition or through intervention, may shorten θ. θ manifests its effect on the transient dynamics through $P(0)$; an increase in θ causes an exponential increase in $P(0)$ (see equation 2.5 and condition 2.17). In this paper, we will examine the effect of θ on the early dynamics of infection by way of examining that of $P(0)$. Hence, we ask: *How does the size of the amplified initial inoculum affect the early dynamics of infection*? Our transient outcomes, t_{peak} and h_{peak}, also depend on the average measures, $\overline{I}$ and $\overline{P}$. Therefore, in order to determine the impact of $P(0)$ on t_{peak} and h_{peak}, we must first examine the dependency of $\overline{I}$ and $\overline{P}$ on $P(0)$.

Approximating $\overline{I}$ and $\overline{P}$

In section 2.3, we derived the approximations $\tilde{I}(t)$ and $\tilde{P}_q(t)$ describing the transient dynamics of the immuno-patho system, based on the assumption that there exists a period during which $\tilde{P}_q(t)$ closely models the early dynamics of the pathogen population growth, in turn implying that this should lead to close approximations of t_{peak} and h_{peak}. Let T denote this period. We further define T to be the time at which $P(t)$ regains its initial value, i.e. when $P(T) = P(0)$. The constant terms $\overline{I}$ and $\overline{P}$, are then derived in a way that they would approximate the average values of $\overline{I}(t)$ and $\overline{P}(t)$ for the period of interest T. This reasonable assumption that T is defined in a way that $P(T) = P(0)$, not only greatly simplifies the analysis involved in the derivations of constant terms, it also results in close approximation of the transient measures t_{peak} and h_{peak} since it enforces symmetry in the shape of $\tilde{P}_q(t)$. This can be realized from the dynamics of Figure 2-2, where $\tilde{P}_q(t)$ from equation 2.14 is plotted against the numerical solution of $P(t)$ obtained by solving the system of equations 2.5-2.7 using a fourth order Runge-Kutta method.

The constant terms $\overline{I}$ and $\overline{P}$, are then derived in a way that they would approximate the average values of $\overline{I}(t)$ and $\overline{P}(t)$ for the period of interest T. To find $\overline{I}$, divide both sides of equation 2.7 by P and take expected value from both sides, for $0 \leq t \leq T$, to get

$$\begin{aligned} \mathrm{E}\left(\frac{1}{P}\frac{dP}{dt}\right) &\overset{Def}{=} \frac{1}{T}\int_0^T \frac{1}{P}\frac{dP}{dt}dt \overset{Def}{=} \frac{1}{T}(\ln P(T) - \ln P(0)) \\ &= \mathrm{E}(r - mI(t)) \qquad \text{by equation 2.7} \\ &= r - m\overline{I} \end{aligned} \tag{2.20}$$

Equation 2.20 together with the assumption that $P(T) = P(0)$ implies that

$$\overline{I} = \frac{r}{m} \tag{2.21}$$

which implies that for the duration T, $\overline{I}$ does not depend on the initial conditions

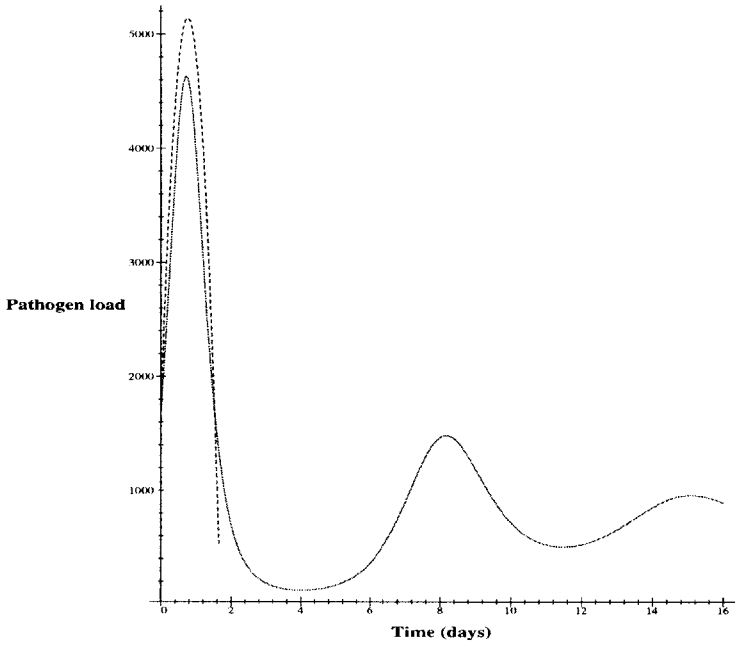


Figure 2-2: Pathogenic load in the first few days of a hypothetical infection after the full activation of the immune system. The dashed parabola is the plot of the quadratic function, $\tilde{P}_q(t)$; the solid plot is the graph of $P(t)$ based on the numerical solution of the system of equations 2.5-2.7 using a fourth order Runge-Kutta method. The constants of the system are: $a_0 = 0.66; \mu = 0.35; k = 0.005; r = 2.86; m = 0.22$; the initial immunity $I(0) = \frac{a_0}{\mu} = 1.886$; the initial inoculum $p_0 = 10$; the amplified initial inoculum $P(0) = 1,698$; and $\theta = 2.1$ days. At $T = 1.5296$, when $P(t) \approx P(0)$, then $\overline{I} = \frac{r}{m} = 13$, $\overline{P} \approx 3684$, $t_{peak} \approx 0.765$ days, and $h_{peak} \approx 5143$.

$I(0)$ and $P(0)$. To find $\overline{P}$, note that at $t = T$, equation 2.14 can be written as

$$\begin{aligned} \tilde{P}_q(T) &= P(0) + \overline{P}T(r - m\frac{a_0}{\mu}) - m\overline{P}\frac{T^2}{2}(a_0 - \mu\overline{I} + k\overline{P}) \\ &= P(0) \qquad \text{since } P(T) = P(0) \end{aligned}$$

which can then be solved for $\overline{P}$

$$\overline{P} = \frac{\mu r - m a_0}{mk} + \frac{2(r\mu - m a_0)}{\mu m k T} \qquad \text{for } 0 \leq t \leq T \tag{2.22}$$

which implies that given T, $\overline{P}$ is constant for the duration $t \in [0, T]$. Equations 2.21 and 2.22 are the constant approximations of the average values $\overline{I}(t)$ and $\overline{P}(t)$ respectively, for the duration $t \in [0, T]$. Substituting the right hand side of equation 2.22

for $\overline{P}$ in equation 2.16, we get

$$t_{peak} = \frac{T}{2}$$

as should be expected from the symmetry assumption. Figure 2-2 is illustrative of how closely our quadratic model approximates t_{peak} and h_{peak}.

$\overline{P}$ and T are inversely associated. As T gets smaller, $\overline{P}$ increases, which in turn causes t_{peak} to decrease. Clearly, T is an important measure; it provides a bound on the critical period for intervention beyond which any intervention scheme may prove ineffective. Such a critical period must be less than t_{peak}. The relationships between T, $\overline{P}$, t_{peak}, and h_{peak} are clear. But how do we determine T and what is its relationship with the initial inoculum $P(0)$? In other words, it is not clear how $\overline{P}$ depends on $P(0)$ since such a dependency is masked by T.

2.3.2 A Shortcoming of the Quadratic Model

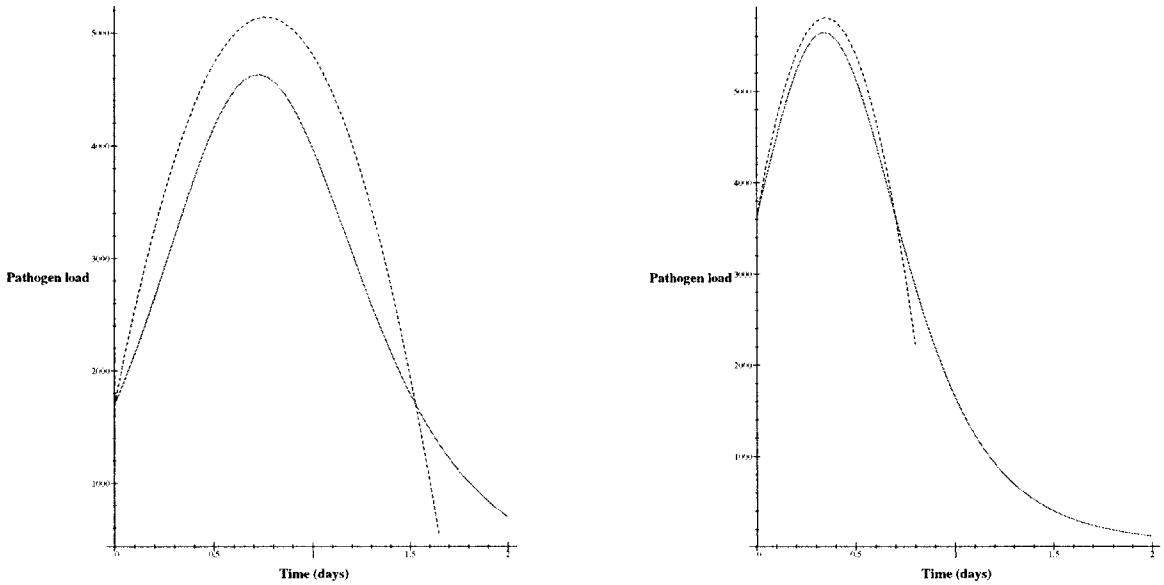
Our interest in studying the transient behavior of the immuno-patho system is motivated by the observation that while the humoral immunity has not yet been activated, there is potential for exponential increase in the antigenic mass. This means that upon activation, the humoral immunity is to encounter a massive pathogenic load. Although our quadratic model of pathogenesis closely models the transient measures t_{peak} and h_{peak}, the effect of $P(0)$ the amplified initial inoculum on these measures remains obscure since it is not clear how $P(0)$ is affecting $\overline{P}$ or T. To understand why our quadratic model lacks an explicit dependency on $P(0)$ in the equation for $\overline{P}$, note that $\tilde{P}_q$ is concave down if the initial slope is increasing and concave up otherwise. Assuming that in the absence of a fully effective immunity the initial slope is always increasing, and therefore $\tilde{P}_q$ is concave down, then it must be that the second derivative of the quadratic function is negative for the period of interest, that is, $\ddot{\tilde{P}}_q(t) < 0$ for $0 \leq t \leq T$. If $\tilde{P}_q(t)$ is indeed a good model of $P(t)$ at the end points 0 and T, then it must be that $\ddot{P} < 0$ as well, for $0 \leq t \leq T$. Consider the second derivative of the original equation at $t = 0$

$$\begin{aligned} \ddot{P}(0) &= \dot{P}(0)(r - mI(0)) - mP(0)\dot{I}(0) && \text{by equation 2.7} \\ &= P(0)(r - m\frac{a_0}{\mu})^2 - mk[P(0)]^2 && \text{by equations 2.5-2.7} \\ &< 0 \end{aligned}$$

which implies

$$P(0) > \frac{(r - ma_0)^2}{\mu^2 mk} \tag{2.23}$$

This means that our model is a good approximation to the actual curve at $t = 0$ if condition 2.23 is satisfied, that is, if $P(0)$ is large enough. This is because as $P(0)$ takes on larger values and becomes closer to the threshold of condition 2.23, $P(t)$

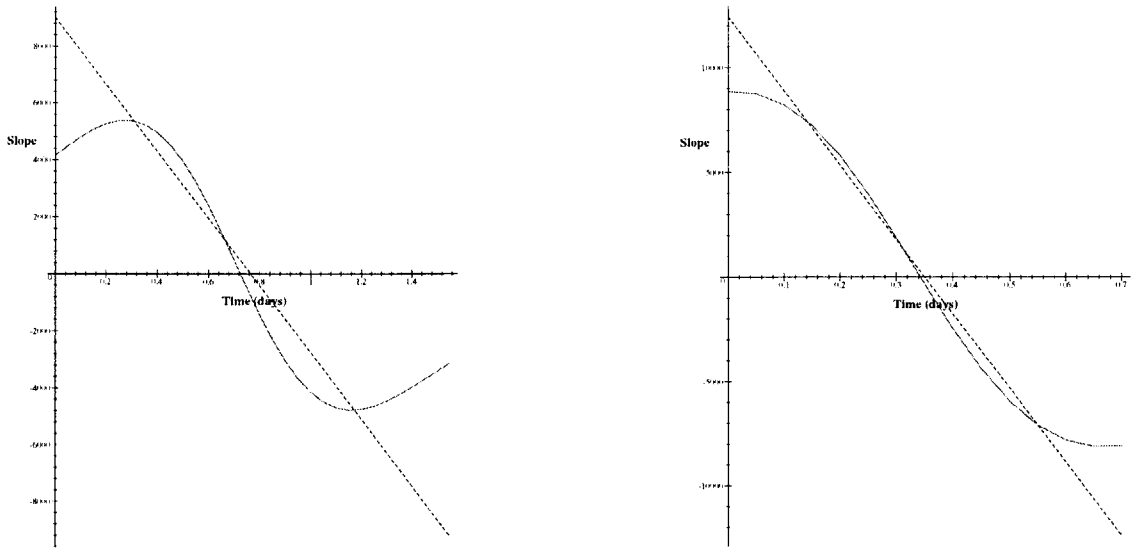


(a) (b)

Figure 2-3: Pathogenic load in the first two days of hypothetical infections for different values of $P(0)$ after the full activation of the immune system. (a) As in the dynamics of Figure 2-2, with $P(0) = 1,698$; and the threshold ≈ 5435 (b) A hypothetical infection with $k = 0.007$; $\theta = 2.41$ days; $T = 0.70078$; $P(0) = 3624$; the threshold ≈ 3882; the peak measures are estimated as $t_{peak} \approx 0.35$; and $h_{peak} \approx 5803$; all other parameters are as in figure 2-2. Notice the closer approximation of the peak measures as $P(0)$ approaches its threshold.

becomes more parabolic in shape, which is the dynamics of the quadratic function $\tilde{P}_q(t)$. In fact, this phenomenon can be observed from the dynamics of Figure 2-3. In Figure 2-3(a), where $P(0) = 1,698$ and is well below its threshold (≈ 5435), notice how the curve of $P(t)$ is "sigmoid" near the end points 0 and T. Compare that to the dynamics of Figure 2-3(b), where $P(0) = 3624$ and is much closer to its threshold (≈ 3882); notice how tight the peak approximations become as $P(0)$ approaches its threshold and $P(t)$ becomes more parabolic. This indicates that our quadratic model is not as good for modeling the end points as it is for modeling the peak measures, unless $P(0)$ is greater than some threshold. Figure 2-4 demonstrates the threshold effect for the dynamics of Figure 2-3 by the behavior of the first derivatives. Note, in Figure 2-4(b), as $P(0)$ approaches the threshold of condition 2.23, the approximation to the first derivative at the end points becomes tighter.

Our goal is to model the behavior of the pathogen population growth at the peak in a way that the role of the amplified initial inoculum in the dynamics is explicitly clear. Under the quadratic model, if $\overline{P}$ is as in equation 2.22, and if $P(0)$ is larger than the threshold of condition 2.23, then $\tilde{P}_q(t)$ closely follows the behavior of $P(t)$



(a) (b)

Figure 2-4: The slope of $P(t)$ versus the slope of $\tilde{P}_q(t)$ for the dynamics of Figure 2-3. The dashed line is the slope of the quadratic function $\tilde{P}_q(t)$; the solid curve is the slope of $P(t)$. (a) For the dynamics of Figure 2-3(a); (b) For the dynamics of Figure 2-3(b). Notice the tighter approximation of $\dot{P}$ by $\dot{\tilde{P}}_q$ as $P(0)$ approaches its threshold.

both at the peak and at the end points. However, in equation for $\overline{P}$, the nature of the dependency on $P(0)$ is not clear. Furthermore, there is no reason to assume that $P(0)$ is almost always larger than the threshold of condition 2.23. Hence, we need to expand upon our quadratic model of interaction.

2.4 Toward a *Hybrid* Model of Interaction

Our quadratic model of the immuno-patho system nicely models the peak measures, t_{peak} and h_{peak}, but it fails to express theses measures solely in terms of the amplified initial inoculum and various parameters of the system, unless $P(0)$ is greater than some threshold. Although this condition may well be realistic for many acute infections, it narrows the domain of possibilities. By now, it should be clear why the quadratic model fails to closely model the end points. Looking at the dynamics of $\dot{P}(t)$ from Figure 2-4, it is easy to see that a good model of $P(t)$ must be a fourth order polynomial and not a second order. Hence, it is no surprise that the peak measures can be traced so closely but not the end points, resulting in an obscure relationship between these measures and $P(0)$, which is masked by $\overline{P}$ and therefore T.

To arrive at the quadratic model, we doubly approximated the equation for $P(t)$

by replacing $I(t)$ with $\tilde{I}(t)$ and assuming the average measures as constants. In what follows, we will use only the approximation for $I(t)$ and will end up with an exponential function. We will then combine the results from the two models, thus making a "hybrid" to derive a new value of $\overline{P}$ with explicit dependency on $P(0)$. This means that the peak measures, t_{peak} and h_{peak}, can then be explicitly expressed in terms of $P(0)$ as well.

2.4.1 An Exponential Model of Interaction

In section 2.3, we defined $\tilde{P}_q$ to denote the quadratic model. By the same analogy, let $\tilde{P}_e$ denote the exponential model to be derived below. The first and second derivatives are defined accordingly. To derive $\tilde{P}_e$, divide both sides of equation 2.7 by P and integrate both sides to get

$$\begin{aligned}
\int_0^t \frac{1}{P}\frac{dP}{d\tau}d\tau &\stackrel{Def}{=} \ln P(t) - \ln P(0) \\
&= \int_0^t (r - mI(t)) \qquad \text{by equation 2.7} \\
&\approx \int_0^t (r - m\tilde{I}(t)) \\
&= rt - m\frac{a_0}{\mu}t - ma_0\frac{t^2}{2} + m\mu\overline{I}\frac{t^2}{2} - mk\overline{P}\frac{t^2}{2} \qquad \text{by equation 2.11}
\end{aligned}$$

Let $\Sigma(t) = rt - m\frac{a_0}{\mu}t - ma_0\frac{t^2}{2} + m\mu\overline{I}\frac{t^2}{2} - mk\overline{P}\frac{t^2}{2}$. Then we have

$$\tilde{P}_e = P(0)e^{\Sigma(t)}$$

Note that in the new model, the approximation comes from replacing $I(t)$ with $\tilde{I}(t)$ from section 2.3. This results in a higher order function, or an exponential one. Figure 2-5 demonstrates how closely $\tilde{P}_e$ models the end points. The approximations to t_{peak} and h_{peak} are not too shabby either; the new model provides a tight upper bound on t_{peak}, and a reasonable lower bound on h_{peak}. Compute the first derivative of the new model to get

$$\dot{\tilde{P}}_e = P(0)(r - m\frac{a_0}{\mu} - ma_0t + m\mu\overline{I}t - mk\overline{P}t)e^{\Sigma(t)}$$

Figure 2-6 illustrates the same effect of capturing the behavior of the pathogen at the end point 0 and T, by comparing the first derivatives.

Notice the goodness of the approximation by $\dot{\tilde{P}}_e$ at and close to the end points $t = 0$ and $t = 1.5296$ (compare to Figure 2-4). As for the peak measures, in fact it can be shown that both the quadratic and the exponential functions result in the same approximation for t_{peak}, if fed by the same value of $\overline{P}$. To demonstrate this, evaluate

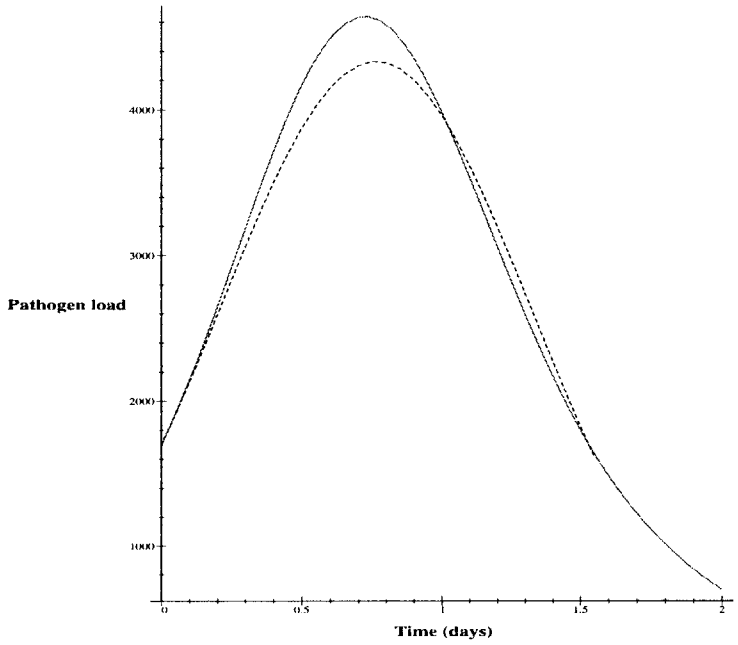


Figure 2-5: Pathogen population in the first two days of a hypothetical infection after the full activation of the immune system. The dashed curve is the plot of the exponential function, $\tilde{P}_e(t)$; the solid curve is the graph of $P(t)$ based on the numerical solution of the system of equations 2.5-2.7 using a fourth order Runge-Kutta method. All parameters are as in Figure 2-2. Notice how closely $\tilde{P}_e(t)$ approximates the end points, $P(0)$ and $P(T) = P(1.5296)$

$\dot{\tilde{P}}_e$ at t_{peak} from equation 2.16 as follows

$$\begin{aligned}
\dot{\tilde{P}}_e(t_{peak}) &= P(0)(r - m\frac{a_0}{\mu} - ma_0 t_{peak} + m\mu\overline{I}t_{peak} - mk\overline{P}t_{peak})e^{\Sigma(t_{peak})} \\
&= P(0)\left[(r - m\frac{a_0}{\mu}) - mt_{peak}(a_0 - \mu\overline{I} + k\overline{P})\right]e^{\Sigma(t_{peak})} \\
&= P(0)\left[(r - m\frac{a_0}{\mu}) - \frac{r - m\frac{a_0}{\mu}}{m(a_0 - \mu\overline{I} + k\overline{P})}m(a_0 - \mu\overline{I} + k\overline{P})\right]e^{\Sigma(t_{peak})} \\
&= 0
\end{aligned}$$

This proves the claim that both functions approximate t_{peak} the same way, since the equation for t_{peak} was derived from the quadratic model.

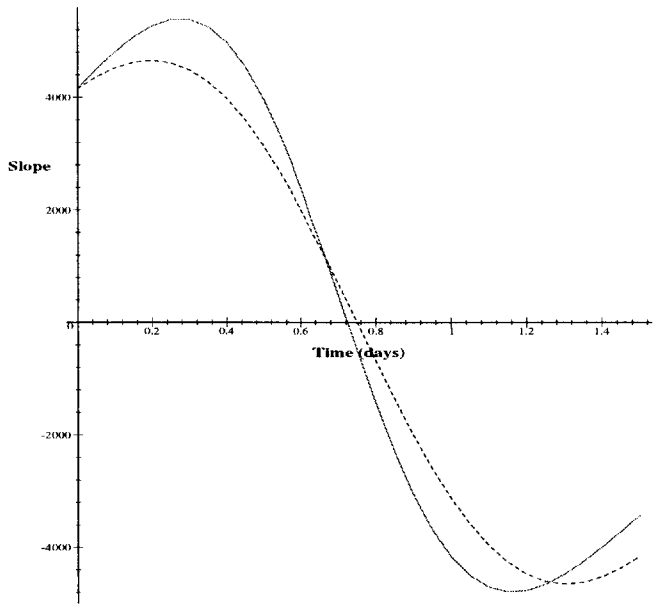


Figure 2-6: The slope of $P(t)$ versus the slope of $\tilde{P}_e(t)$ for the dynamics of Figure 2-2. The dashed curve is the slope of the exponential function $\tilde{P}_e(t)$; the solid curve is the slope of $P(t)$. Notice the goodness of the approximation at and around the end points $t = 0$ and $t = 1.5296$ (compare to Figure 2-4(a)).

2.4.2 Predicting Transience from Initial Inoculum

To derive a relationship between $\overline{P}$ and $P(0)$, we will combine our quadratic and exponential models in such a way that we will be able to capture the strengths of each model namely, close approximation of the peak measures by the quadratic model, and close approximation of the end points by the exponential model. Define $\overline{P}_{new}$ to be the new average measure of the pathogen population resulting from combining the quadratic and exponential models of interaction. Now, compute the second derivative of the exponential model at t_{peak}, to get

$$\begin{aligned}
\ddot{\tilde{P}}_e(t_{peak}) &= P(0)[(-ma_0 + m\mu\overline{I} - mk\overline{P}_{new})e^{\Sigma(t_{peak})} + \\
&\quad (r - m\frac{a_0}{\mu} - ma_0 t_{peak} + m\mu\overline{I}t_{peak} - mk\overline{P}_{new}t_{peak})^2 e^{\Sigma(t_{peak})}] \\
&= -mP(0)(a_0 - \mu\overline{I} + k\overline{P}_{new}) \exp\left(\frac{(r\mu - ma_0)^2}{2\mu^2 m(a_0 - \mu\overline{I} + k\overline{P}_{new})}\right) \quad (2.24)
\end{aligned}$$

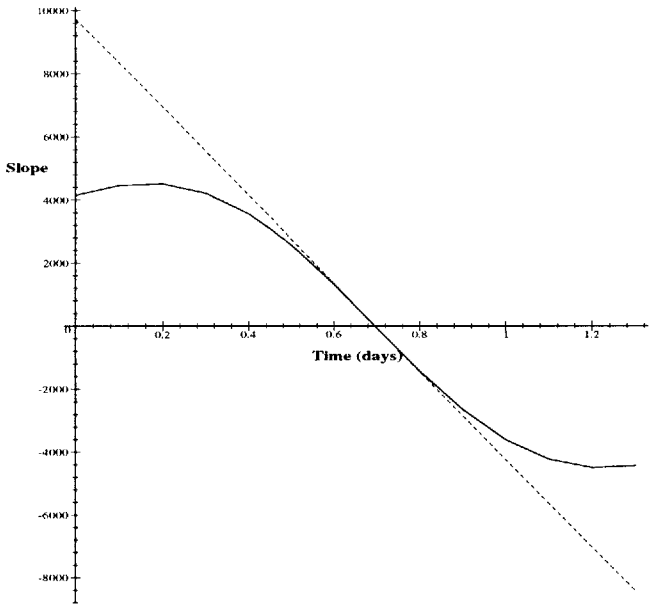


Figure 2-7: The slope of $\tilde{P}_q(t)$ versus the slope of $\tilde{P}_e(t)$ for the dynamics of Figure 2-2. The dashed line is the slope of the quadratic function $\tilde{P}_q(t)$; the solid curve is the slope of the exponential function $\tilde{P}_e(t)$. Notice that the two models coincide exactly at and around the peak and away from the end points.

since $t_{peak} = \frac{r\mu - ma_0}{\mu m(a_0 - \mu\bar{I} + k\bar{P}_{new})}$ by equation 2.16, and $\Sigma(t_{peak}) = \frac{(r\mu - ma_0)^2}{2\mu^2 m(a_0 - \mu\bar{I} + k\bar{P}_{new})}$.
On the other hand, $\ddot{\tilde{P}}_q$ at t_{peak} is

$$\ddot{\tilde{P}}_q(t_{peak}) = -m\bar{P}_{new}(a_0 - \mu\bar{I} + k\bar{P}_{new}) \qquad \text{by equation 2.15} \qquad (2.25)$$

Both $\tilde{P}_e$ and $\tilde{P}_q$ model the behavior of the system at the peak reasonably well, and we have shown that in fact both models approximate t_{peak} exactly the same. Figure 2-7 demonstrates this effect by comparing the slope of the quadratic function $\tilde{P}_q$, versus the slope of the exponential function $\tilde{P}_e$; the two models coincide exactly at and around the peak. Therefore it would be reasonable to assume that equations 2.24 and 2.25 must be roughly the same at t_{peak}. Equating equations 2.24 and 2.25, we get

$$P(0) = \bar{P}_{new} \exp\left(-\frac{(r\mu - ma_0)^2}{2\mu^2 m(a_0 - \mu\bar{I} + k\bar{P}_{new})}\right) \qquad (2.26)$$

Equation 2.26 is indeed the key to predicting the early behavior of infection from

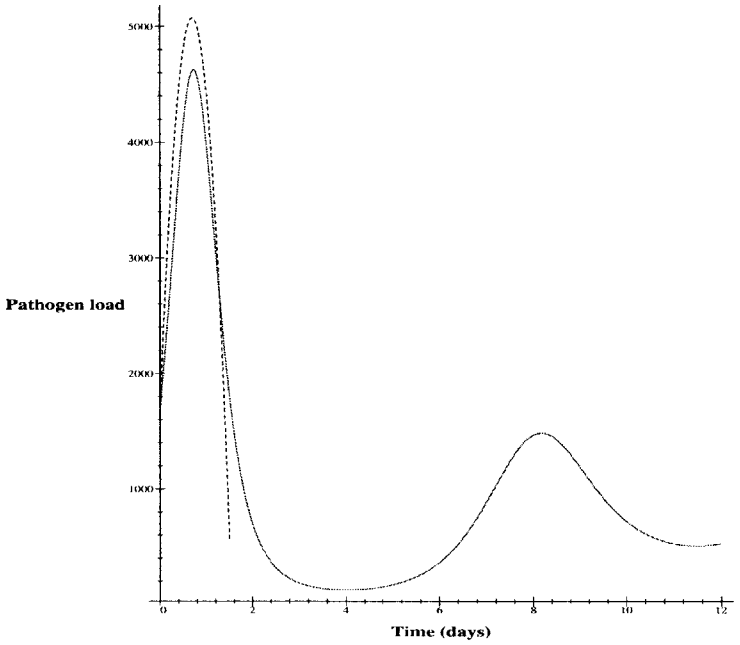


Figure 2-8: Pathogen population in the first few days of a hypothetical infection after the full activation of the immune system. The dashed parabola is the plot of the combined method of approximation using the quadratic and the exponential functions, where $\tilde{P}_q(t)$ is evaluated at $\overline{P}_{new}$; the solid plot is the graph of $P(t)$ based on the numerical solution of the system of equations 2.5-2.7 using a fourth order Runge-Kutta method. All parameters are as in Figure 2-2. $\overline{P}_{new} \approx 3974$; $t_{peak} \approx 0.695$ days; and $h_{peak} \approx 5077$.

$P(0)$ and the parameters of the system. By combining the quadratic and exponential models of interaction we were able to derive $P(0)$ as a function of $\overline{P}_{new}$. Although it may not be easy to solve for $\overline{P}_{new}$ analytically, given $P(0)$, we can always solve for it numerically. Notice that equation 2.26 defines an explicit relationship between $P(0)$ and $\overline{P}_{new}$, and there is no longer any dependency on T. Substituting the numerical value of $\overline{P}_{new}$, obtained from equation 2.26, into the quadratic function $\tilde{P}_q$ from equation 2.14, or equivalently into t_{peak} and h_{peak} from equations 2.16 and 2.19, we can derive numerical estimates of the peak measures solely in terms of the initial inoculum and parameters of the system. Figure 2-8 illustrates the result of the combined method of approximation. Compare that to the dynamics of Figure 2-2: the hybrid method provides a closer approximation of h_{peak} and a very tight lower bound on t_{peak}. By employing the combined method of approximation, not only were we able to keep the best features of each model, we were also able to estimate the peak measures explicitly in terms of the amplified initial inoculum $P(0)$.

Putting it together, equation 2.26 can be used to solve numerically for $\overline{P}_{new}$.

This in turn can be used to derive numerical estimations of t_{peak} and h_{peak} from equations 2.16 and 2.19. Furthermore, equating the numerical value of $\overline{P}_{new}$ from equation 2.26 with the right hand side of equation 2.22, we can solve for T. Alternatively, we can solve for t_{peak} in terms of $\overline{P}_{new}$ and use the derivation $t_{peak} = \frac{T}{2}$, from section 2.3.1 to evaluate T. Hence, the peak measure, as well as all intermediate measures can be numerically estimated based on $P(0)$ and the parameters of the system. Clearly, any information on the numerical values of t_{peak} and h_{peak} is of great importance. They can be used to assess the duration of the critical period for intervention, which is always $< t_{peak}$, and to decide on plausible intervention schemes in order to interfere with the early dynamics of infection.

2.5 Qualitative Effects of Initial Inoculum on Transient Measures

Before proceeding to the peak analysis, we need to clarify how a change in $P(0)$ impacts the average measure $\overline{P}_{new}$ since both of the transient measures, t_{peak} and h_{peak}, depend on this term. As established before, we assume that a change in the amplified initial inoculum is the result of a change in θ only. Furthermore, an increase in θ causes $P(0)$ to increase exponentially since the exponent of $P(0)$ in equation 2.5 is always positive by condition 2.17.

2.5.1 Effect of Initial Inoculum on Average Pathogenicity

By equation 2.26, it is clear that an increase in $\overline{P}_{new}$ causes an increase in $P(0)$ since both the numerator and the denominator in the exponent of equation 2.26 are positive (see conditions 2.17 and 2.18). On the other hand, although it is not easy to obtain an analytic solution for $\overline{P}_{new}$ in terms of $P(0)$ from equation 2.26, we can still show that an increase in $P(0)$ should cause an increase in $\overline{P}_{new}$ as well. To see this, note that if $P(0)$ is increasing, it can only be associated with a change in $\overline{P}_{new}$ since the increase in $P(0)$ is assumed to be associated with an increase in θ and all other parameters in equation 2.26 are assumed to be constant. Suppose for contradiction that an increase in θ causes $\overline{P}_{new}$ to decrease. This would imply that the exponent in equation 2.26 is increasing or the exponential term is decreasing. But if both terms in the product are decreasing, then $P(0)$ must be decreasing as well, which contradicts the original assumption. Therefore, if $P(0)$ is increasing, $\overline{P}_{new}$ must be increasing as well, implying that $P(0)$ and $\overline{P}_{new}$ are positively correlated.

2.5.2 Effect of Initial Inoculum on t_{peak}

An increase in $P(0)$ causes $\overline{P}_{new}$ to increase, which in turn causes t_{peak} to decrease (see equation 2.16). Hence, as the amplified initial inoculum takes on larger values as θ increases, the time of "crisis" will occur earlier, requiring faster intervention.

2.5.3 Effect of Initial Inoculum on h_{peak}

The effect of $P(0)$ on h_{peak} is less obvious since $\overline{P}_{new}$ appears in both the numerator and the denominator of the second term in the equation for h_{peak} (see equation 2.19). Differentiating h_{peak} with respect to $P(0)$ or $\overline{P}_{new}$ will not help. Instead, we will infer the direction of change in h_{peak} by deriving an equation for h_{peak} in terms of the second derivative of $P(t)$ from the original equations. Evaluation of equation 2.6 at t_{peak} produces

$$\begin{aligned} \dot{I}(t_{peak}) &= a_0 - \mu I(t_{peak}) + kP(t_{peak}) \\ &= a_0 - \mu\frac{r}{m} + kh_{peak} \end{aligned} \tag{2.27}$$

since $P(t_{peak}) = h_{peak}$, and since

$$\begin{aligned} I(t_{peak}) &= -\frac{\dot{P}(t_{peak})}{mP(t_{peak})} + \frac{rP(t_{peak})}{mP(t_{peak})} \qquad \text{by equation 2.7} \\ &= \frac{r}{m} \qquad \text{since } \dot{P}(t_{peak}) = 0 \end{aligned}$$

Differentiation of equation 2.7 at t_{peak} to get an explicit relationship between $\dot{I}(t_{peak})$, h_{peak}, and $\ddot{P}(t_{peak})$ produces

$$\begin{aligned} \ddot{P}(t_{peak}) &= r\dot{P}(t_{peak}) - m\left[I(t_{peak})\dot{P}(t_{peak}) + \dot{I}(t_{peak})P(t_{peak})\right] \\ &= -m\dot{I}(t_{peak})h_{peak} \qquad \text{since } \dot{P}(t_{peak}) = 0 \\ &= -m(a_0 - \mu\frac{r}{m} + kh_{peak})h_{peak} \qquad \text{by equation 2.27} \\ &= (\mu r - ma_0)h_{peak} - mk(h_{peak})^2 \end{aligned}$$

Solving the above quadratic equation for h_{peak} we get

$$h_{peak} = \frac{1}{2km}\left[(r\mu - ma_0) \pm \sqrt{(r\mu - ma_0)^2 - 4mk\ddot{P}(t_{peak})}\right] \tag{2.28}$$

Clearly, $h_{peak} > 0$, so we will only consider the positive solution of equation 2.28. Replacing $\ddot{P}(t_{peak})$ in equation 2.28 with either its quadratic or its exponential approximation is the final step needed in order to infer the direction of change in h_{peak} with respect to $P(0)$. We will do both. If we replace $\ddot{P}(t_{peak})$ with its quadratic approximation from equation 2.25 we get

$$h_{peak} = \frac{1}{2km}\left[(r\mu - ma_0) + \sqrt{(r\mu - ma_0)^2 + 4m^2k\overline{P}_{new}(a_0 - \mu\overline{I} + k\overline{P}_{new})}\right]$$

Now, an increase in $P(0)$ causes $\overline{P}_{new}$ to increase, which in turn causes h_{peak} to increase. Let $\beta = a_0 - \mu\overline{I} + k\overline{P}_{new}$. Then replacing $\ddot{P}(t_{peak})$ with its exponential

approximation from equation 2.24 gets

$$\begin{aligned} h_{peak} &= \frac{1}{2km}\left[(r\mu - ma_0) + \sqrt{(r\mu - ma_0)^2 + 4m^2kP(0)\beta \exp\left(\frac{(r\mu - ma_0)^2}{2\mu^2 m\beta}\right)}\right] \\ &> \frac{1}{2km}\left[(r\mu - ma_0) + \sqrt{(r\mu - ma_0)^2 + 4m^2kP(0)\beta}\right] \end{aligned} \tag{2.29}$$

since $0 < \frac{(r\mu - ma_0)^2}{2\mu^2 m\beta} < \infty$ by conditions 2.17 and 2.18 from section 2.3.1. An increase in $P(0)$ also causes an increase in $\overline{P}_{new}$, which together cause inequality 2.29 and h_{peak} to increase.

The intuition behind the argument that h_{peak} and $P(0)$ change in the same direction is much more straightforward than the preceding analysis. Note that an increase in $P(0)$ causes t_{peak}, and similarly T, to decrease. On the other hand, an increase in $P(0)$ causes $\overline{P}_{new}$ to increase. The only way to compensate for the effect of an increase in $\overline{P}_{new}$ in a shorter time period is for h_{peak} to increase, resulting in a "thinner" and "taller" parabolic shape.

2.6 Summary

We defined the notion of "transient pathogenicity" to encompass the early dynamics of the within-host pathogenesis of acute infectious disease. We recognized the amplifying effect of the initial period of inactivity of humoral immunity, θ, on the pathogenic growth of the initial inoculum. We further demonstrated that in the absence of a fully effective immunity, the initial pathogenic population consisting of a few microorganisms will grow freely and without bound until humoral immunity is induced. This "amplified initial inoculum", $P(0)$, is therefore the new initial pathogenic load at the time of full activation of the immune system. Using the method of time averaging, we developed a quadratic model and an exponential model of the transient behavior of the pathogen after the full activation of the immune system when the newly activated immunity is to face extensive antigenic mass. We further derived the transient measures, t_{peak} (the time from the full activation of the immune system to the peak of infection) and h_{peak} (maximum pathogenic load of infection), under the quadratic model. The two models were then combined in order to express the outcomes of early dynamics of infection in terms of $P(0)$ and the parameters of the system only. The qualitative analysis of this combined method of approximation enabled us to derive explicit relationships between $P(0)$ and and the peak measures, t_{peak} and h_{peak}. Therefore, we have established qualitatively that

1. As the duration of inactivity of the humoral immunity θ increases, the initial antigenic mass increases exponentially with θ. Upon activation, the humoral immunity must therefore face and combat a massive pathogenic load.

2. As the duration of inactivity of the humoral immunity θ increases, t_{peak} decreases, causing the "crisis" to occur earlier, making the critical window for intervention smaller.

3. As the duration of inactivity of the humoral immunity θ increases, h_{peak} increases, making the intensity of "crisis" greater.

4. Putting 2 and 3 together, as the duration of inactivity of the humoral immunity θ increases, the damage to the host will be much more extensive in a shorter time period.

Although some of the parameters of the immuno-patho dynamics may be inherent physiological properties of the interacting components that we may not be able to influence, θ may very well be influenced by the nutrition state or the stress level of the host. A boost to the immune system may be equivalent to reducing θ; thus reducing the maximum pathogenic load and delaying the time of its occurrence. This has significant consequences for diagnostics and intervention. Having derived the peak measures t_{peak} and h_{peak} explicitly in terms of $P(0)$ and the parameters of the system, we established quantitatively that

1. Given θ, $P(0)$, and various parameters of the immuno-patho dynamics, we can numerically estimate the time of "crisis" t_{peak}, and therefore assess the duration of the critical period for intervention beyond which any intervention scheme may prove ineffective.

2. Given θ, $P(0)$, and various parameters of the immuno-patho dynamics, we can numerically estimate the extent of "crisis" h_{peak}, and therefore decide on the domain of plausible intervention strategies needed to interfere with the dynamics.

If it is indeed possible to predict the within-host early behavior of acute infections from the initial inoculum and a few related immunological and pathological parameters, and if there is sufficient data on average parameter values for a cohort, then we can classify the transient dynamics of infections based on their peak measures, critical period to intervene, and plausible intervention schemes.

Chapter 3

Transience in Population Health

3.1 Motivation and Background

Our interest in studying the dynamics of vulnerability in a population is doubly motivated by the subtle existence of unidentified "nonlinearity" embedded in the mathematics of "linear" systems; and the obscure, yet significant, notion of "mortality" in public health and demography.

On the one hand, although linear dynamic systems have been well studied and in a sense are considered trivial to analyze, but in fact in an intuitive sense, they are neither understood nor trivial. The mystery lies in the observation that even if the dynamics of a multi-variable system is linear, its characteristic equation is nonlinear. This means that the relationship between the eigenvalues and the parameters of the system is not necessarily linear and may not be at all obvious. Although the solutions to linear systems are relatively easy to obtain, they involve the eigenvalues and therefore such explicit solutions do not necessarily reveal much about the underlying dynamics, and how perturbation in a system, acting through the system parameters, affect each variable and the dynamics of the system as a whole. Hence, our goal is to make intuitive sense of the underlying formality and gain insight into the nonlinear and obscure relationships between the parameters and different outcomes of the system.

On the other hand, despite decades of studies and the relative abundance of mortality data, our understanding of the dynamics of vulnerability to disease processes and mortality has remained inadequate. A population is defined by its constituent individuals who are heterogeneous in their vulnerability to disease and death. Today, the distribution of vulnerability is recognized and understood in terms of the underlying genetic basis and age. Although both the genetic predisposition and age are important causes of variation in the distribution of vulnerability, a substantial part of such variation is due to heterogeneity in the social conditions of individuals. Therefore, a sound theory of population vulnerability is to take account of both the inherent physiological and social heterogeneity in the population. If we divide a population into different vulnerability classes, then over time, people may move from one vulnerability class to another due to aging, disease processes, health improvement policies, acquiring knowledge or wealth; they also die. The relationship between these

processes namely, mortality selectivity and mobility, will determine whether people in different vulnerability classes die mostly in conditions of chronic good health or of high risk and poor health. This is particularly important when we compare the age-specific mortality rates of different populations and encounter irregularities that may defy common sense.

One of the "anomalies" of mortality data is that after about age 80, the death rate for the White population appears to be higher than that of the Black population, a phenomenon referred to as "mortality crossover" in demography and public health literature. For nearly 80 years the crossover phenomenon has intrigued scholars. In more recent years, in particular since the inclusion of "Black" in the ethnic and racial vocabulary of the Vital Statistics of the United States, the Black-White mortality crossover has puzzled demographers and social scientists in the United States, and has divided them into two schools of thought. Some argue that the crossover phenomenon is not real and is the result of inaccuracies in data due to the tendency of age over-reporting among the Black elderly. Others insist that it is real and hypothesize that it is due to selective processes. Today, despite the relative wealth of mortality data and a long history of debates over the existence of mortality crossovers and characterization of the intersecting populations, the mystery remains unraveled. Therefore we ask: *When will a population, initially exposed to a greater force of mortality, have a lower death rate?*

This chapter is an attempt to educate the common sense about the hidden nonlinearity in the dynamics of linear systems by developing the mathematics of the dynamics of vulnerability in a population; identify the factors influencing the mortality of people residing in different vulnerability classes in a population and the mortality of the population as a whole; to compare the mortality experiences of different pairs of populations whose mortality curves cross; identify the underlying processes governing the existence and dynamics of mortality crossover; and finally interpret the consequential outcomes of the analysis as different public health intervention strategies.

3.2 A Model of 2-Vulnerability-Class

Suppose that a population cohort is distributed among two vulnerability classes V_1 and V_2, each associated with its own mortality rate μ_1 and μ_2. Here, "vulnerability" is defined with respect to disease processes and the general health of a population; thus a high vulnerability class implies higher morbidity and mortality as well. Hence, we assume V_2 is more vulnerable than V_1, and therefore $\mu_2 > \mu_1$. We define the differential mortality $\mu_2 - \mu_1$, as the rate of selection which is a measure of inequality due to physiological or social conditions. Furthermore, individuals in one class may move to another due to disease processes, natural processes such as aging, loss of fortune, or health-detrimental policies, in which case the flow a_{21}, is from V_1 to V_2; or due to factors affecting viability such as acquiring knowledge or wealth, or through health-improvement policies, in which case the flow a_{12}, is from V_2 to V_1. We further assume that $a_{21} > a_{12}$; that is, for an individual at any point in time it is more likely

to move to the class with poor health than to the class with good health. Since there is no birth in this model, then $\forall t > 0$, we have the condition that $\sum_{i=1}^{2} V_i(t) < 1$, with the base case $V_1(0) + V_2(0) = 1$. Further, all parameters of the system are assumed to be constant. The following linear system of ordinary differential equations captures the dynamics of vulnerability of a population that is heterogeneous in health and death, as outlined above and depicted in Figure 3-1.

$$\frac{dV_1}{dt} = -V_1(a_{21} + \mu_1) + a_{12}V_2 \tag{3.1}$$

$$\frac{dV_2}{dt} = a_{21}V_1 - V_2(a_{12} + \mu_2) \tag{3.2}$$

The system of equations 3.1-3.2 has one critical point $(0, 0)$, which is asymptotically stable. The characteristic polynomial $p(\lambda)$, of the coefficient matrix of the above system is:

$$p(\lambda) = \begin{vmatrix} -\mu_1 - a_{21} - \lambda & a_{12} \\ a_{21} & -\mu_2 - a_{12} - \lambda \end{vmatrix}$$

Expanding the determinant and collecting the terms in λ, we get:

$$p(\lambda) = \lambda^2 + (a_{12} + \mu_2 + a_{21} + \mu_1)\lambda + (\mu_1\mu_2 + \mu_1 a_{12} + a_{21}\mu_2) \tag{3.3}$$

Solving for the roots of the characteristic polynomial, we get

$$\lambda_1 = \frac{1}{2}\left(-\mu_2 - a_{12} - a_{21} - \mu_1 - \sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}\right) \tag{3.4}$$

$$\lambda_2 = \frac{1}{2}\left(-\mu_2 - a_{12} - a_{21} - \mu_1 + \sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}\right) \tag{3.5}$$

Both eigenvalues are real, since $\sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}} > 0$. Clearly $\lambda_1 < 0$; for $\lambda_2 < 0$, we must show

$$|-\mu_2 - a_{12} - a_{21} - \mu_1| > \sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}$$

Subtract the square of the right-hand-side from the square of the left-hand-side of the above inequality to get

$$\begin{aligned} & (\mu_2 + a_{12} + a_{21} + \mu_1)^2 - (a_{12} + \mu_2 - a_{21} - \mu_1)^2 - 4a_{12}a_{21} \\ = \; & (\mu_2 + a_{12})^2 + (a_{21} + \mu_1)^2 + 2(\mu_2 + a_{12})(a_{21} + \mu_1) - (\mu_2 + a_{12})^2 + (a_{21} + \mu_1)^2 \\ & +2(\mu_2 + a_{12})(a_{21} + \mu_1) - 4a_{12}a_{21} \\ = \; & 2(a_{21} + \mu_1)^2 + 4(\mu_1\mu_2 + \mu_1 a_{12} + \mu_2 a_{21}) \\ > \; & 0 \end{aligned}$$

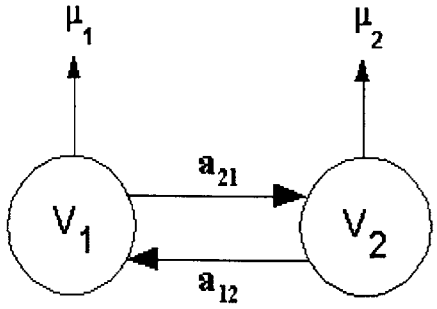


Figure 3-1: A population cohort consisting of two vulnerability classes. People in one class may move to another according to the transition rates a_{12} and a_{21}; or they may die according to the class-specific mortality rates μ_1 and μ_2.

Hence $\lambda_2 < 0$ and the system is asymptotically stable. Throughout this chapter we will assume that λ_2 is the dominant eigenvalue since $|\lambda_2| < |\lambda_1|$ by equations 3.4 and 3.5.

3.2.1 From Frequency to Ratio: A Transformation

In the absence of birth or migration in the model, the population in each class, and therefore the population as a whole, will eventually become extinct. In principle, the vulnerability-ratio $\frac{V_1}{V_2}$ reaches a steady state value before either V_1 or V_2 reach their final null values. We will prove the existence of a steady state solution for the vulnerability-ratio in Theorem 1. An important implication of the theorem is that at time t, when $\frac{V_1}{V_2}$ is in steady state, the total population $V_1(t) + V_2(t)$, is still in transience. Ideally, for the steady state solution to be worthy of analysis, we would like: (1) a "reasonable" proportion of the population to be alive when $\frac{V_1}{V_2}$ is in or near the steady state; (2) the steady state solution of the vulnerability-ratio can be utilized to infer information about the transient state of the population mortality and the proportion dying in each vulnerability class. In later sections, we will demonstrate how the steady state solutions, to be derived in this section, can be used to make inferences about the transient state of population mortality.

Hence, we transform our two dimensional linear system describing the rate of change in the frequency of each vulnerability class, into one quadratic equation describing the rate of change in the ratio of the frequency of the two vulnerability classes. First, multiply equation 3.1 by V_2 and equation 3.2 by V_1 to get

$$\frac{dV_1}{dt}V_2 = -V_1V_2(a_{21} + \mu_1) + a_{12}{V_2}^2 \tag{3.6}$$

$$\frac{dV_2}{dt}V_1 = a_{21}{V_1}^2 - V_1V_2(a_{12} + \mu_2) \tag{3.7}$$

Subtract equation 3.7 from equation 3.6 and divide both sides of the new equation

by ${V_2}^2$ to get a single, quadratic, differential equation as follows

$$\begin{aligned}\frac{\frac{dV_1}{dt}V_2 - \frac{dV_2}{dt}V_1}{{V_2}^2} &\stackrel{def}{=} \frac{d}{dt}\left(\frac{V_1}{V_2}\right) \\ &= -a_{21}\left(\frac{V_1}{V_2}\right)^2 + (a_{12} + \mu_2 - a_{21} - \mu_1)\frac{V_1}{V_2} + a_{12} \qquad (3.8)\end{aligned}$$

The resulting equation describes the rate of change in the population vulnerability-ratio. In the following theorem, we will prove the existence of the steady state solution for the population vulnerability-ratio.

Theorem 1 *The population vulnerability-ratio* $\frac{V_1}{V_2}$*, corresponding to the system of ordinary differential equations 3.1-3.2, is asymptotically stable.*

Proof: Let C_1, C_2, C_3, C_4 be constants. Then $\frac{V_1}{V_2}$ can be written in its general form as follows

$$\begin{aligned}\frac{V_1}{V_2} &= \frac{C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}}{C_3 e^{\lambda_1 t} + C_4 e^{\lambda_2 t}} \\ &= \frac{C_1 + C_2 e^{(\lambda_2 - \lambda_1)t}}{C_3 + C_4 e^{(\lambda_2 - \lambda_1)t}} \\ &= \frac{C_1}{C_3 + C_4 e^{(\lambda_2 - \lambda_1)t}} + \frac{C_2}{C_3 e^{(\lambda_1 - \lambda_2)t} + C_4} \\ &\to \frac{C_2}{C_4} \qquad \text{as } t \to \infty, \text{ and if } \lambda_2 \text{ is the dominant eigenvalue} \\ &\to \frac{C_1}{C_3} \qquad \text{as } t \to \infty, \text{ and if } \lambda_1 \text{ is the dominant eigenvalue}\end{aligned}$$

which proves the existence of a steady state solution. ■

Therefore, by Theorem 1, and since by equation 3.8 the result of the transformation is a polynomial of degree two, the approach to the steady state must be monotone. Note That The validity of Theorem 1 stands still even if the system is unidirectional. Without loss of generality, suppose that $a_{12} = 0$. Then we have

$$\begin{aligned}\frac{V_1}{V_2} &= \frac{C_1 e^{\lambda_1 t}}{C_2 e^{\lambda_1 t} + C_3 e^{\lambda_2 t}} \\ &= \frac{C_1}{C_2 + C_3 e^{(\lambda_2 - \lambda_1)t}} \\ &\to \frac{C_1}{C_2} \qquad \text{as } t \to \infty, \text{ and if } \lambda_1 \text{ is the dominant eigenvalue} \\ &\to 0 \qquad \text{as } t \to \infty, \text{ and if } \lambda_2 \text{ is the dominant eigenvalue}\end{aligned}$$

which again proves the existence of a steady state solution. However, if there is no transitional flow between the vulnerability classes, then there cannot be any steady

state solution. In the absence of any inter-class flow, the eigenvalues are simply the class-specific mortality rates μ_1 and μ_2; and since there is no flow to partially regulate the population, the ratio will not be stable. To see this, note that

$$\begin{aligned}
\frac{V_1}{V_2} &= \frac{C_1 e^{\lambda_1 t}}{C_2 e^{\lambda_2 t}} \\
&= \frac{C_1 e^{(\lambda_1 - \lambda_2)t}}{C_2} \\
&\to 0 \qquad \text{as } t \to \infty, \text{ and if } \lambda_2 \text{ is the dominant eigenvalue} \\
&\to \infty \qquad \text{as } t \to \infty, \text{ and if } \lambda_1 \text{ is the dominant eigenvalue}
\end{aligned}$$

We will explore these special cases in detail in section 3.3.1.

3.2.2 Steady State Solutions

At the steady state, when $\frac{d(\frac{V_1}{V_2})}{dt} = 0$, we can solve equation 3.8 for $\frac{V_1}{V_2}$ to get

$$\begin{aligned}
\left(\frac{V_1}{V_2}\right)^* &= \frac{a_{12} + \mu_2 - a_{21} - \mu_1 + \sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}}{2a_{21}} \\
&= \frac{1}{2}\left(\frac{a_{12}}{a_{21}} + \frac{\mu_2}{a_{21}} - 1 - \frac{\mu_1}{a_{21}} + \sqrt{(\frac{a_{12}}{a_{21}} + \frac{\mu_2}{a_{21}} - 1 - \frac{\mu_1}{a_{21}})^2 + 4\frac{a_{12}}{a_{21}}}\right) \qquad (3.9)
\end{aligned}$$

where $\left(\frac{V_1}{V_2}\right)^*$ represent the steady state solution of the population vulnerability-ratio $\frac{V_1}{V_2}$. Let μ represent the total population mortality rate. Then by definition,

$$\begin{aligned}
\mu &= \frac{\mu_1 V_1 + \mu_2 V_2}{V_1 + V_2} \\
&= \frac{\mu_1 \frac{V_1}{V_2} + \mu_2}{\frac{V_1}{V_2} + 1} \qquad (3.10)
\end{aligned}$$

As equation 3.10 reveals, since μ_1 and μ_2 are constant, μ must also have a steady state solution. Furthermore, μ should reach its steady state value at the same time when $\frac{V_1}{V_2}$ does. However, μ's dynamics is opposite that of $\frac{V_1}{V_2}$. To see this, differentiate equation 3.10 with respect to $\frac{V_1}{V_2}$ to get

$$\begin{aligned}
\frac{\partial \mu}{\partial \frac{V_1}{V_2}} &= \frac{\mu_1(1 + \frac{V_1}{V_2}) - \mu_1 \frac{V_1}{V_2} - \mu_2}{(1 + \frac{V_1}{V_2})^2} \\
&= \frac{\mu_1 - \mu_2}{(1 + \frac{V_1}{V_2})^2} \\
&< 0 \qquad \text{since } \mu_1 < \mu_2 \text{ by definition} \qquad (3.11)
\end{aligned}$$

This implies a negative correlation between μ and $\frac{V_1}{V_2}$, which in turn implies that μ also monotonically approaches its steady state. In Theorems 2 and 3, we will further establish that the vulnerability-ratio $\frac{V_1}{V_2}$ is indeed a monotonically "increasing" process, and therefore the population mortality μ must be a monotonically "decreasing" process. Let μ^* represent the steady state population mortality. Then by equation 3.10 we have

$$\mu^* = \frac{\mu_1 \left(\frac{V_1}{V_2}\right)^* + \mu_2}{\left(\frac{V_1}{V_2}\right)^* + 1} \tag{3.12}$$

Similarly, we define μ_1^* and μ_2^* as the steady state proportion of people dying in V_1 and V_2 respectively. Then by definition we have

$$\mu_1^* = \frac{\mu_1 \left(\frac{V_1}{V_2}\right)^*}{\left(\frac{V_1}{V_2}\right)^* + 1} \tag{3.13}$$

and

$$\mu_2^* = \frac{\mu_2}{\left(\frac{V_1}{V_2}\right)^* + 1} \tag{3.14}$$

Our interest in the investigation of the dynamics of vulnerability is twofold. On the one hand we want to reduce the total population mortality; on the other hand, we want a larger proportion of deaths in the population to come from the class with good health. This means that people will live relatively healthy up to the time of death. For the society as a whole, this results in lowering the terminal investment in health. Hence, we need to identify the underlying processes that govern the dynamics of mortality; this will be done in the next section. But before we proceed to the next section, first we need to establish that mortality, as produced by the vulnerability model, is a monotonically decreasing process to its steady state μ^*. To prove this property, we require another important concept in the vocabulary of the dynamics of vulnerability namely, "the time to steady state". Let t_p denote the time it takes for $\frac{V_1}{V_2}$ to reach $(p \times 100)\%$ of its steady state value. Then we have the following theorems.

Theorem 2 *The population vulnerability-ratio* $\frac{V_1}{V_2}$ *is a monotonically increasing process.*

Proof: Recall the definition of t_p as the time it takes for $\frac{V_1}{V_2}$ to reach $(p \times 100)\%$ of its steady state value $\frac{C_2}{C_4}$. Rewrite $\frac{V_1}{V_2}$ in equation 3.8 in terms of its steady state value, i.e. as $p\frac{C_2}{C_4}$, to get

$$\frac{d}{dt}\left(\frac{V_1}{V_2}\right) = \frac{C_2}{C_4}\frac{dp}{dt} = -a_{21}\left(p\frac{C_2}{C_4}\right)^2 + (a_{12} + \mu_2 - a_{21} - \mu_1)\left(p\frac{C_2}{C_4}\right) + a_{12}$$

which implies

$$\begin{aligned}\frac{dp}{dt} &= -a_{21}\frac{C_2}{C_4}p^2 + (a_{12}+\mu_2-a_{21}-\mu_1)p + a_{12}\frac{C_4}{C_2} \\ &= -a_{21}\frac{C_2}{C_4}\left(p^2 - \frac{C_4}{C_2}\frac{\Sigma}{a_{21}}p - \frac{a_{12}}{a_{21}}\left(\frac{C_4}{C_2}\right)^2\right) \\ &= a_{21}\frac{C_2}{C_4}(p-a)(b-p)\end{aligned}$$

where $\Sigma = \mu_2 + a_{12} - a_{21} - \mu_1$; and a and b are the roots of the new polynomial, i.e when $\frac{dp}{dt} = 0$. Since p can attain a maximum value of 1, we already know that one of the roots of the polynomial must be 1. Let $\Delta = \sqrt{(\mu_2 + a_{12} - a_{21} - \mu_1)^2 + 4a_{12}a_{21}} = \sqrt{\Sigma^2 + 4a_{12}a_{21}}$; then solve explicitly for a and b to get

$$\begin{aligned}a &= \frac{\Sigma - \Delta}{2a_{21}\frac{C_2}{C_4}} \\ &= \frac{C_4}{C_2}\frac{\Sigma - \Delta}{2a_{21}} \\ &= \left(\frac{2a_{21}}{\Sigma + \Delta}\right)\left(\frac{\Sigma - \Delta}{2a_{21}}\right) \qquad \text{by equation 3.9} \\ &= \frac{\Sigma - \Delta}{\Sigma + \Delta} < 0 \qquad \text{since } \Delta > |\Sigma| \end{aligned} \tag{3.15}$$

and

$$\begin{aligned}b &= \frac{\Sigma + \Delta}{2a_{21}\frac{C_2}{C_4}} \\ &= \frac{C_4}{C_2}\frac{\Sigma + \Delta}{2a_{21}} \\ &= \left(\frac{2a_{21}}{\Sigma + \Delta}\right)\left(\frac{\Sigma + \Delta}{2a_{21}}\right) \qquad \text{by equation 3.9} \\ &= \frac{\Sigma + \Delta}{\Sigma + \Delta} \\ &= 1\end{aligned}$$

Next using linear fractional transformation, we linearize the differential equation by introducing a new variable u as follows

$$u = \frac{p-a}{b-p} = \frac{p-a}{1-p} \qquad \text{since } b = 1 \tag{3.16}$$

Solve for p in terms of u to get

$$p = \frac{a+bu}{1+u} = \frac{a+u}{1+u} \qquad \text{since } b = 1$$

Differentiate the new variable u with respect to t to get

$$\begin{aligned}
\frac{du}{dt} &= \frac{du}{dp}\frac{dp}{dt} = \frac{(1-p)+(p-a)}{(1-p)^2} a_{21}\frac{C_2}{C_4}(p-a)(1-p) \\
&= \left(\frac{p-a}{1-p}\right) a_{21}\frac{C_2}{C_4}(1-a) \\
&= \left(\frac{p-a}{1-p}\right)\Delta \qquad \text{since } 1-a = b-a = \tfrac{C_4}{C_2}\tfrac{2\Delta}{2a_{21}} \\
&= \Delta u
\end{aligned}$$

Solve the above ODE to get

$$u(t) = u_0 e^{\Delta t} \tag{3.17}$$

where u_0 is the initial value of u at time $t = 0$. Replace t with t_p in equation 3.17 and solve the equation for t_p to get

$$t_p = \frac{\ln\left(\frac{u(t_p)}{u_0}\right)}{\Delta} \tag{3.18}$$

where $u(t_p)$ is as in equation 3.16, and $u_0 = \frac{p_0-a}{b-p_0} = \frac{p_0-a}{1-p_0}$ since $b = 1$. For large values of p, $u(t_p) \approx \frac{1-a}{1-p}$. Hence, $u(t_p)$ can be approximated as follows

$$\begin{aligned}
u(t_p) &\approx \frac{1-a}{1-p} \\
&= \frac{1}{1-p}\left(1 - \frac{\Sigma-\Delta}{\Sigma+\Delta}\right) \qquad \text{by equation 3.15} \\
&= \frac{1}{1-p}\left(\frac{2\Delta}{\Sigma+\Delta}\right)
\end{aligned} \tag{3.19}$$

To rewrite u_0 in terms of Σ and Δ, let $V_0 = \frac{V_1(0)}{V_2(0)}$. Note that $p_0\left(\frac{C_2}{C_4}\right) = V_0$, which implies that $p_0 = V_0\left(\frac{C_4}{C_2}\right) = \frac{2V_0a_{21}}{\Sigma+\Delta}$ (see equation 3.9). Hence we have

$$\begin{aligned}
u_0 &= \frac{p_0-a}{1-p_0} \\
&= \frac{\frac{2V_0a_{21}}{\Sigma+\Delta} - \frac{\Sigma-\Delta}{\Sigma+\Delta}}{1-\frac{2V_0a_{21}}{\Sigma+\Delta}} \\
&= \frac{\Delta-\Sigma+2V_0a_{21}}{\Delta+\Sigma-2V_0a_{21}}
\end{aligned} \tag{3.20}$$

For t_p to have a positive definite value, $\ln\left(\frac{u(t_p)}{u_0}\right) > 0$. This implies that both $u(t_p) > 0$ and $u_0 > 0$. Clearly, $u(t_p) > 0$ since $0 \le p < 1$ and $a < 0$. What about u_0? The numerator in equation 3.20 is always positive since $\Delta > |\Sigma|$. But u_0 has a singularity at $\Delta + \Sigma - 2V_0a_{21} = 0$. Clearly the denominator must also always be positive or else

$u_0 < 0$. For the denominator to be positive we must have

$$\Delta + \Sigma - 2V_0 a_{21} \quad > \quad 0$$

which implies

$$\frac{\Delta + \Sigma}{2a_{21}} \quad > \quad V_0$$

or equivalently

$$\left(\frac{V_1}{V_2}\right)^* \quad > \quad V_0 \qquad \text{by equation 3.9} \tag{3.21}$$

This means that for $t_p > 0$, $\Delta + \Sigma - 2V_0 a_{21} > 0$, which in turn implies that the steady state vulnerability-ratio must be larger than the initial vulnerability-ratio. Theorem 1 established that the vulnerability-ratio $\frac{V_1}{V_2}$, is a monotone process. Inequality 3.21 further establishes that it must be monotonically "increasing" since its final value is larger than its initial value. ■

Theorem 3 *The population mortality μ, as defined by equation 3.10, is a monotonically decreasing process.*

Proof: Immediate by Theorem 2 and inequality 3.11. ■

The following theorem is of particular importance to the upcoming analysis.

Theorem 4 *$\mu^* = -\lambda_d$, where μ^* is the steady state population mortality and λ_d is the dominant eigenvalue of the system of equations 3.1-3.2.*

Proof: First we show $-\mu^*$ is an eigenvalue of the system. Then we argue that it must be the dominant one. If $-\mu^*$ is indeed an eigenvalue, then it must have the same qualitative properties of the eigenvalues derived from the characteristic polynomial in equation 3.3. Let's examine this. Let δ be such eigenvalue. Then we have:

$$\begin{aligned} \mu^* &= \frac{\mu_1 \left(\frac{V_1}{V_2}\right)^* + \mu_2}{\left(\frac{V_1}{V_2}\right)^* + 1} \qquad \text{by equation 3.12} \\ &= -\delta \end{aligned}$$

or equivalently

$$\begin{aligned} \mu_1 \left(\frac{V_1}{V_2}\right)^* + \mu_2 &= -\delta \left(\left(\frac{V_1}{V_2}\right)^* + 1\right) \\ \equiv \quad \delta + \mu_2 &= \left(\frac{V_1}{V_2}\right)^* (-\delta - \mu_1) \end{aligned}$$

which implies

$$\left(\frac{V_1}{V_2}\right)^* = \frac{\delta + \mu_2}{-\delta - \mu_1} \tag{3.22}$$

Replace the new steady-state vulnerability-ratio from equation 3.22 into equation 3.8 when $\frac{d(\frac{V_1}{V_2})}{dt} = 0$, we get

$$-a_{21}\left(\frac{\delta + \mu_2}{-\delta - \mu_1}\right)^2 + (a_{12} + \mu_2 - a_{21} - \mu_1)\left(\frac{\delta + \mu_2}{-\delta - \mu_1}\right) + a_{12} = 0$$

Expand the squared term and set the numerator to zero to get

$$\begin{aligned} -a_{21}(\delta^2 + 2\mu_2\delta + {\mu_2}^2) - (a_{12} + \mu_2 - a_{21} - \mu_1)(\delta^2 + \mu_1\mu_2 + \mu_1\delta + \mu_2\delta) \\ +a_{12}(\delta^2 + 2\mu_1\delta + {\mu_1}^2) = 0 \end{aligned}$$

Collect the terms in δ as follows

$$\begin{aligned} \delta^2(-a_{21} - a_{12} - \mu_2 + a_{21} + \mu_1 + a_{12}) \\ +\delta(-2\mu_2 a_{21} - \mu_1 a_{12} - \mu_1\mu_2 + \mu_1 a_{21} + {\mu_1}^2 - \mu_2 a_{12} - {\mu_2}^2 + \mu_2 a_{21} + \mu_1\mu_2 + 2\mu_1 a_{12}) \\ +(-a_{21}{\mu_2}^2 - \mu_1\mu_2 a_{12} - \mu_1{\mu_2}^2 + \mu_1\mu_2 a_{21} + {\mu_1}^2\mu_2 + a_{12}{\mu_1}^2) = 0 \end{aligned}$$

Simplify further to get

$$\begin{aligned} \delta^2(\mu_1 - \mu_2) + \delta[a_{21}(\mu_1 - \mu_2) + a_{12}(\mu_1 - \mu_2) + (\mu_1 + \mu_2)(\mu_1 - \mu_2)] \\ +(\mu_1\mu_2(\mu_1 - \mu_2) + a_{21}\mu_2(\mu_1 - \mu_2) + a_{12}\mu_1(\mu_1 - \mu_2)) = 0 \end{aligned}$$

Factor out the common terms and simplify further to finally get

$$\delta^2 + \delta(a_{21} + a_{12} + \mu_1 + \mu_2) + (\mu_1\mu_2 + a_{21}\mu_2 + a_{12}\mu_1) = 0 \tag{3.23}$$

Note that the new polynomial in δ, is the same as the characteristic polynomial of the coefficient matrix of the system of equations 3.1-3.2 (see equation 3.3). Therefore, it must be that $\mu^* = -\lambda_i,\ i = 1, 2$. But since at the steady state, as we have shown in Theorem 1, only the effect of the dominant eigenvalue remains, it must be that $-\mu^*$ is the dominant eigenvalue. ■

3.3 Qualitative Effects of Parameter Change on Steady State Solutions

3.3.1 Special Cases of Interest

No Transition

Suppose there is no transition flow between the vulnerability classes. In this case, since each class V_i has its own decay rate μ_i, $i = 1, 2$, and is independent of the other class, then the eigenvalues of the system will simply correspond to the decay rates μ_1 and μ_2. If $\mu_2 > \mu_1$, then simple selection would shift the population toward the low vulnerable class as it declines. Hence, the mortality for the whole population will approach that of V_1. Furthermore, if there is no transition flow between the vulnerability classes, then the vulnerability-ratio does not have a steady state. To see this, write V_1 and V_2 in their general forms with $\lambda_1 = -\mu_1$ and $\lambda_2 = -\mu_2$. That is, let $V_1 = C_1 e^{-\mu_1 t}$ and $V_2 = C_2 e^{-\mu_2 t}$, where C_1, C_2 are constants. Then

$$\begin{aligned} \frac{V_1}{V_2} &= \frac{C_1 e^{-\mu_1 t}}{C_2 e^{-\mu_2 t}} \\ &= \frac{C_1 e^{(\mu_2 - \mu_1)t}}{C_2} \\ &\to \infty \qquad \text{as } t \to \infty; \text{ if } \mu_2 > \mu_1 \end{aligned}$$

This also implies that $\mu_2^* \to 0$, as $t \to \infty$; $\mu_1^* = \frac{\mu_1}{\left(\frac{V_2}{V_1}\right)^* + 1} \to \mu_1$, as $t \to \infty$; and therefore $\mu^* = \mu_1^* + \mu_2^* \to \mu_1$ as $t \to \infty$, as we suspected. Figure 3-2 illustrates an example of the dynamics of mortality in the absence of flow.

Unidirectional Transition

Now suppose that the system in unidirectional in transition between the vulnerability classes. Without loss of generality suppose that $a_{12} = 0$. Due to the effect of aging, which is incorporated in the flow to the high-vulnerable class, as well as the reality of the nature of societal mobility, it would only be sensible to assume that $a_{21} \neq 0$. Write V_1 and V_2 in their general forms, i.e., $V_1 = C_1 e^{\lambda_1 t}$ and $V_2 = C_2 e^{\lambda_1 t} + C_3 e^{\lambda_2 t}$, where C_1, C_2, C_3 are constants. Then

$$\begin{aligned} \frac{V_1}{V_2} &= \frac{C_1 e^{\lambda_1 t}}{C_2 e^{\lambda_1 t} + C_3 e^{\lambda_2 t}} \\ &= \frac{C_1}{C_2 + C_3 e^{(\lambda_2 - \lambda_1)t}} \\ &\to \frac{C_1}{C_2} \qquad \text{as } t \to \infty, \text{ and if } \lambda_1 \text{ is the dominant eigenvalue} \\ &\to 0 \qquad \text{as } t \to \infty, \text{ and if } \lambda_2 \text{ is the dominant eigenvalue} \end{aligned}$$

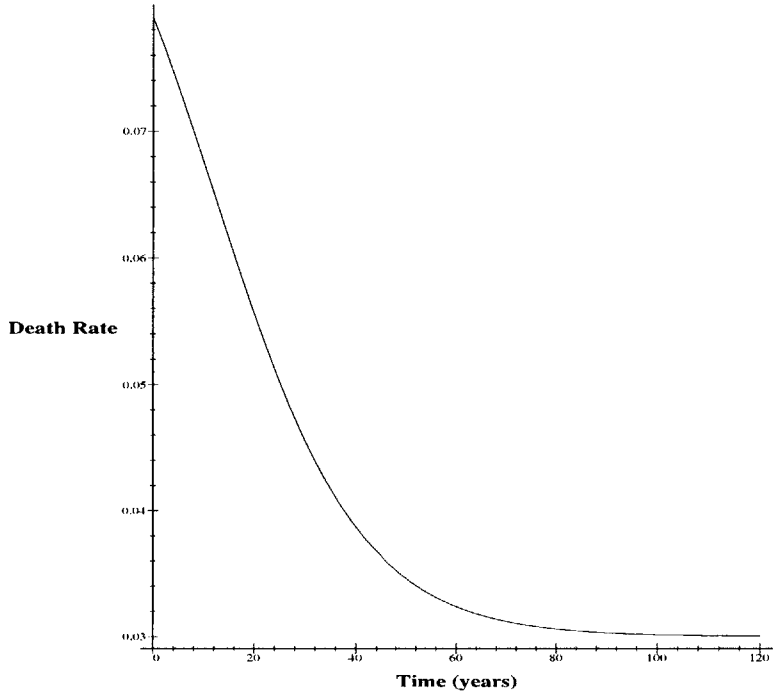


Figure 3-2: Dynamics of mortality when there is no transition flow between the vulnerability classes and the process is driven only by the force of selection $\mu_2 - \mu_1$. Parameter values are $\mu_1 = 0.03; \mu_2 = 0.1; a_{12} = a_{21} = 0$. The initial conditions are: $V_1(0) = 0.3$; and $V_2(0) = 0.7$. Note that total mortality approaches the mortality of the low-vulnerable class μ_1, as time progresses.

Solve for the eigenvalues to get: $\lambda_1 = -\mu_1 - a_{21}$ and $\lambda_2 = -\mu_2$. Then, for the steady state solution to exit, we must have $\lambda_2 - \lambda_1 < 0$, which is equivalent to the condition that $\boxed{\mu_2 > \mu_1 + a_{21}}$. Now let's derive the steady state solutions of the vulnerability-ratio, total mortality, and the proportion dying in each class.

$$\left(\frac{V_1}{V_2}\right)^* = \frac{1}{a_{21}}(\mu_2 - a_{21} - \mu_1) \tag{3.24}$$

$$\begin{aligned} \mu_1^* &= \frac{\mu_1(\mu_2 - \mu_1 - a_{21})}{\mu_2 - \mu_1} \\ &= \mu_1\left(1 - \frac{a_{21}}{\mu_2 - \mu_1}\right) \end{aligned} \tag{3.25}$$

$$\mu_2^* = \frac{a_{21}\mu_2}{\mu_2 - \mu_1} \tag{3.26}$$

$$\begin{aligned}
\mu^* &= \mu_1^* + \mu_2^* \\
&= \frac{\mu_1(\mu_2 - \mu_1 - a_{21}) + a_{21}\mu_2}{\mu_2 - \mu_1} \\
&= \frac{\mu_1(\mu_2 - \mu_1) + a_{21}(\mu_2 - \mu_1)}{\mu_2 - \mu_1} \\
&= \mu_1 a_{21}
\end{aligned} \tag{3.27}$$

Figure 3-3 demonstrates the dynamics of population mortality under unidirectional transition from V_1 to V_2. Notice that the dynamics is the opposite of that of Figure 3-2, where mortality is a monotonically decreasing process. In the absence of any flow from V_2 to V_1 to partially regulate the population, mortality will increase monotonically to a steady state. We are now ready to investigate the effect of change in the system parameters μ_1, μ_2, and a_{21}, on the steady state solutions.

Effect of Change in μ_1, μ_2, and a_{21} on Steady State Solutions

To study the effect of change, we first differentiate the steady state solutions 3.24,3.25,3.26, and 3.27 with respect to all three parameters μ_1, μ_2, a_{21}, as follows

$$\begin{aligned}
\frac{\partial}{\partial \mu_1}\left(\frac{V_1}{V_2}\right)^* &= \frac{-1}{a_{21}} \\
&< 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \mu_2}\left(\frac{V_1}{V_2}\right)^* &= \frac{1}{a_{21}} \\
&> 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial a_{21}}\left(\frac{V_1}{V_2}\right)^* &= \frac{\mu_1 - \mu_2}{(a_{21})^2} \\
&< 0 \qquad \text{since } \mu_1 < \mu_2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mu_1^*}{\partial \mu_1} &= \frac{(\mu_2 - \mu_1 - a_{21} - \mu_1)(\mu_2 - \mu_1) + \mu_1(\mu_2 - \mu_1 - a_{21})}{(\mu_2 - \mu_1)^2} \\
&= \frac{(\mu_2 - \mu_1 - a_{21})(\mu_2 - \mu_1 + \mu_1) - \mu_1(\mu_2 - \mu_1)}{(\mu_2 - \mu_1)^2} \\
&= \frac{\mu_2^2 - 2\mu_1\mu_2 + \mu_1^2 - a_{21}\mu_2}{(\mu_2 - \mu_1)^2} \\
&= \frac{(\mu_2 - \mu_1)^2 - a_{21}\mu_2}{(\mu_2 - \mu_1)^2}
\end{aligned}$$

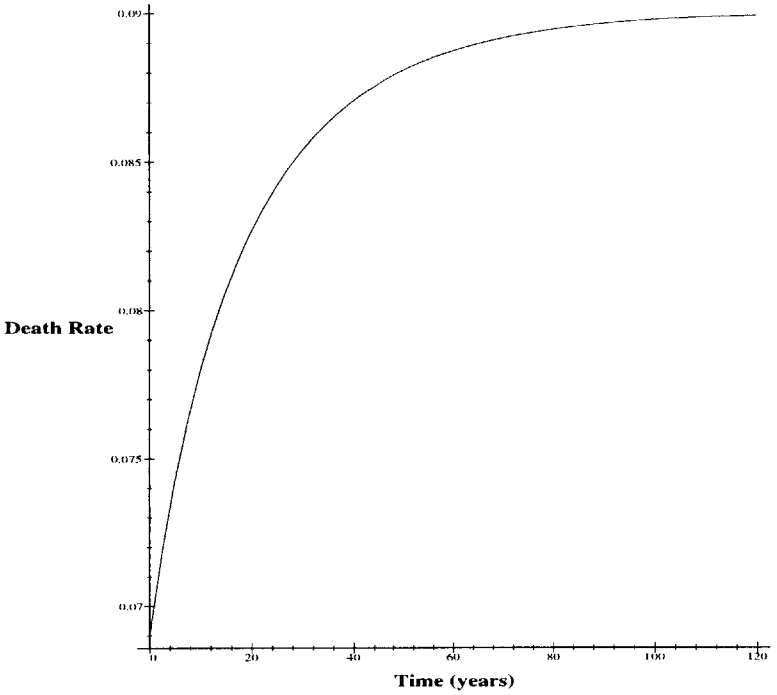
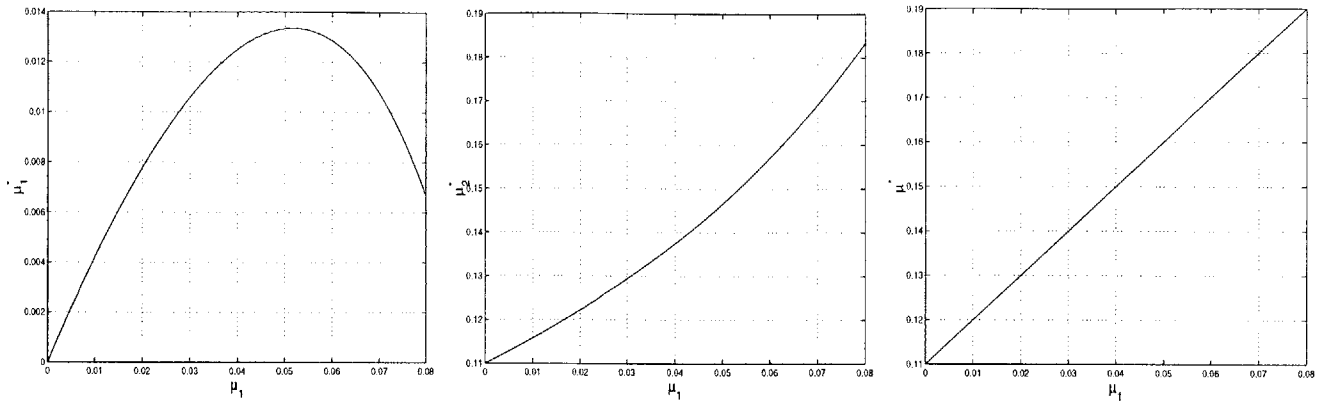


Figure 3-3: Dynamics of mortality when the transition flow is unidirectional. The process is driven by the force of selection $\mu_2 - \mu_1$, partially offset by mobility from V_1 to V_2. Parameter values are $\mu_1 = 0.02; \mu_2 = 0.09; a_{12} = 0$; and $a_{21} = 0.11$. The initial conditions are: $V_1(0) = 0.3$; and $V_2(0) = 0.7$.

$$\begin{array}{lll} > & 0 & \text{if } (\mu_2 - \mu_1)^2 > a_{21}\mu_2 \\ < & 0 & \text{otherwise} \end{array} \tag{3.28}$$

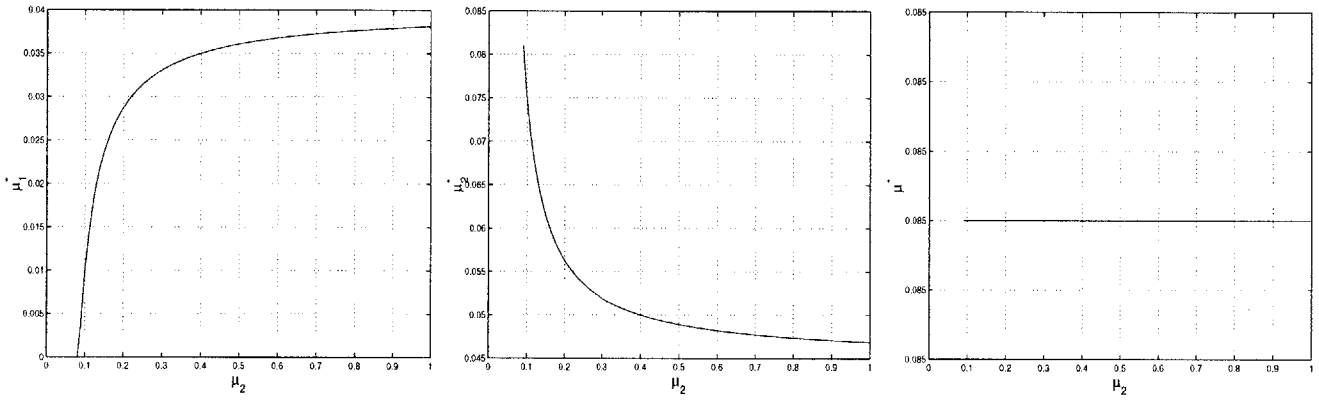
This dual effect of change in μ_1 can be better understood simply by analyzing the effect of each term in the product in the right hand side of equation 3.25. The first term is simply μ_1, which therefore implies that μ_1^* increases with the death rate in V_1. The second term is the fraction of the population in V_1, which decreases with μ_1 and in turn will cause the proportion dying in V_1, i.e. μ_1^*, to decrease. These opposing processes, i.e. the mortality and vitality of V_1, result in an intermediate peak, which can be observed in Figure 3-4(a).

$$\begin{aligned} \frac{\partial \mu_1^*}{\partial \mu_2} &= \frac{\mu_1 - \mu_1(\mu_2 - \mu_1 - a_{21})}{(\mu_2 - \mu_1)^2} \\ &= \frac{\mu_1(1 - \mu_2 + \mu_1 + a_{21})}{(\mu_2 - \mu_1)^2} \end{aligned}$$



(a) (b) (c)

Figure 3-4: Effect of change in μ_1 on the unidirectional steady state mortality (a) on μ_1^* (b) on μ_2^* (c) on μ^*. Parameter values are $\mu_2 = 0.2; a_{21} = 0.11$; and $0 \leq \mu_1 \leq 0.08$.



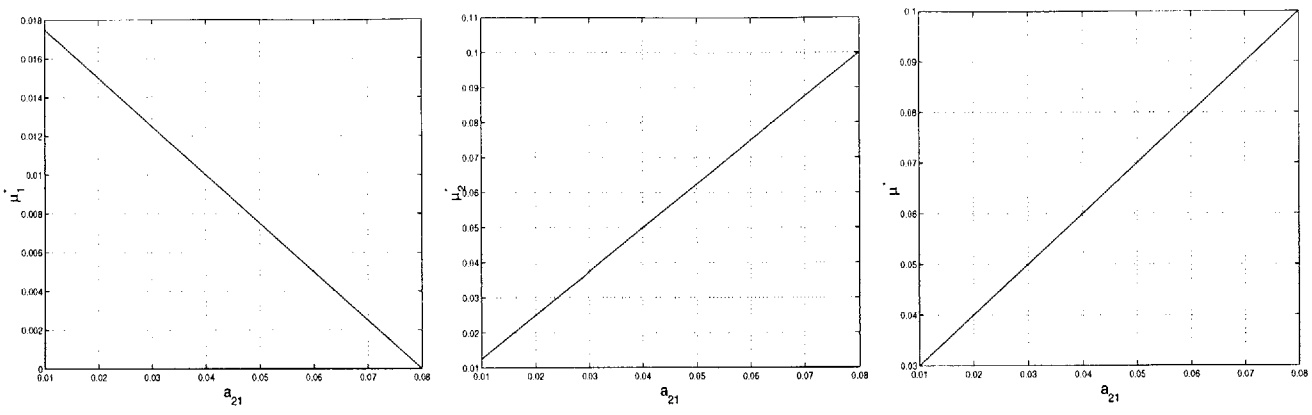
(a) (b) (c)

Figure 3-5: Effect of change in μ_2 on the unidirectional steady state mortality: (a) on μ_1^* (b) on μ_2^* (c) on μ^*. Parameter values are $\mu_1 = 0.04; a_{21} = 0.045$; and $0.09 \leq \mu_2 \leq 1$.

$$\geq \ 0 \qquad \text{since } 0 < \mu_1, \mu_2, a_{21} \leq 1 \text{ and } |\mu_2 - \mu_1 - a_{21}| \leq 1$$

Here, an increase in μ_2 further increases the rate of selection against V_2, killing off more people as they arrive in V_2 from V_1. This reduces the fraction alive in V_2, thereby decreasing the ratio $\frac{V_2}{V_1}$, which in turn causes $\mu_1^* = \frac{\mu_1}{1+\frac{V_2}{V_1}}$ to increase. This effect can be seen in Figure 3-5(a).

$$\begin{aligned} \frac{\partial \mu_1^*}{\partial a_{21}} &= \frac{-\mu_1}{\mu_2 - \mu_1} \\ &< 0 \end{aligned}$$



(a) (b) (c)

Figure 3-6: Effect of change in a_{21} on the unidirectional steady state mortality: (a) on μ_1^* (b) on μ_2^* (c) on μ^*. Parameter values are $\mu_1 = 0.02; \mu_2 = 0.1$; and $0.01 \leq a_{21} \leq 1$.

An increase in the out-flow of V_1 causes more people to move to V_2 to die. This reduces the fraction residing in V_1, which presumably opens more room in V_1 and increases the vitality of V_1 due to the availability of resources, which in turn causes μ_1^* to decrease monotonically with a_{21}. This effect can be seen in Figure 3-6(a).

$$\begin{aligned} \frac{\partial \mu_2^*}{\partial \mu_1} &= \frac{a_{21}\mu_2}{(\mu_2 - \mu_1)^2} \\ &> 0 \end{aligned}$$

An increase in μ_1 decreases the fraction of people alive in V_1, thereby decreasing the ratio $\frac{V_1}{V_2}$, which in turn causes $\mu_2^* = \frac{\mu_2}{1+\frac{V_1}{V_2}}$ to increase. This effect can be seen in Figure 3-4(b).

$$\begin{aligned} \frac{\partial \mu_2^*}{\partial \mu_2} &= \frac{a_{21}(\mu_2 - \mu_1) - \mu_2 a_{21}}{(\mu_2 - \mu_1)^2} \\ &= \frac{-\mu_1 a_{21}}{(\mu_2 - \mu_1)^2} \\ &< 0 \end{aligned}$$

An increase in μ_2 further increases the rate of selection against V_2; thus decreases the fraction alive in V_2. Therefore the proportion dying in V_2, i.e. μ_2^*, decrease monotonically with μ_2. This effect can be seen in Figure 3-5(b).

$$\begin{aligned} \frac{\partial \mu_2^*}{\partial a_{21}} &= \frac{\mu_2}{\mu_2 - \mu_1} \\ &> 0 \end{aligned}$$

Increase of flow to V_2 exposes the new residents to a greater force of mortality μ_2, which in turn causes μ_2^* to increase. Figure 3-6(b) is illustrative of this effects.

$$\frac{\partial \mu^*}{\partial \mu_1} = 1$$

This is the result of combined effects of μ_1 on μ_1^* and μ_2^*. Figure 3-4(c) illustrates this effect.

$$\frac{\partial \mu^*}{\partial \mu_2} = 0$$

This is the result of the opposing effects of μ_2 on μ_1^* and μ_2^*. Figure 3-5(c) illustrates this effect.

$$\frac{\partial \mu^*}{\partial a_{21}} = 1$$

Increased flow to V_2 decreases the frequency of V_1, but it also decreases that of V_2 since more people will arrive in V_2 to be exposed to a greater force of mortality μ_2. The contribution to the total mortality is therefore positive since an increasingly greater proportion of the population is exposed to a higher death rate. This effect can be seen in Figure 3-6(c).

Consequences for Intervention: Putting It All Together

Assuming that the condition for the existence of the steady state solution in a system with unidirectional flow is satisfied, i.e. $\mu_2 - \mu_1 - a_{21} > 0$, for μ_1^* to change in the direction of μ_1, we further require the condition that $(\mu_2 - \mu_1)^2 > a_{21}\mu_2$ (see inequality 3.28). This means that if μ_1 is decreasing, and as long as $0 < (\mu_2 - \mu_1)^2 < a_{21}\mu_2$, μ_1^* will increase, while μ_2^* and μ^* decrease. Although, this effect produces the desired outcomes, i.e. lowers the general mortality; lowers the proportion dying in the class with poor health; and increases the proportion who die in the class with good health, it can be misleading in terms of its implication for intervention. This would imply that in order to achieve the desired objectives, μ_1 must be further reduced! On the other hand, increasing μ_2 also produces a similar effect: μ^* does not change as we increase μ_2; μ_2^* decreases; and μ_1^* increases. This would imply that in order to achieve the desired effect, μ_2 must be further increased!

Whether the desired effect is achieved by way of a reduction in μ_1 or an increase in μ_2, the rate of selection $\mu_2 - \mu_1$, will be further increased as an immediate result. This as we know, is which is a measure of inequality either due to physiological or social conditions. Hence, a reduction in the total mortality of a population cannot necessarily be interpreted as improvement in the state of health of the population. It could be due to either a rise in the mortality of the least healthy, or a fall in the mortality of those with good health, and overall due to a stronger selection.

The same effect is produced by way of a reduction in the flow rate a_{21}. Put in terms of public health intervention, in the absence of transition flow from the least healthy

class to the one with good health, further reduction in a_{21} would result in a "rigid" population, divided in health and death, resembling the dynamics of a population driven by simple selection without mobility.

3.3.2 General Case

In this section, we investigate the effect of change in the system parameters μ_1, μ_2, a_{12}, and a_{21}, on the steady state solutions: $\left(\frac{V_1}{V_2}\right)^*$, μ^*, μ_1^*, and μ_2^*. As indicated in section 3.2.2, we have two objectives of interest: lowering the death rate for the population; and increasing the proportion dying in the class with good health. Therefore, we need to capture those events that cause μ^* and μ_2^* to decrease, while at the same time cause μ_1^* to increase or stay steady. Hence we address two fundamental questions: (1) "*What type of processes cause μ^* and μ_1^* to behave differently?*", and (2) "*What are the consequences of such effects for intervention?*"

Effect of Change in μ_1

A change in μ_1 causes $\left(\frac{V_1}{V_2}\right)^*$ to change in the opposite direction. To see this effect, we differentiate equation 3.9 with respect to μ_1 to get

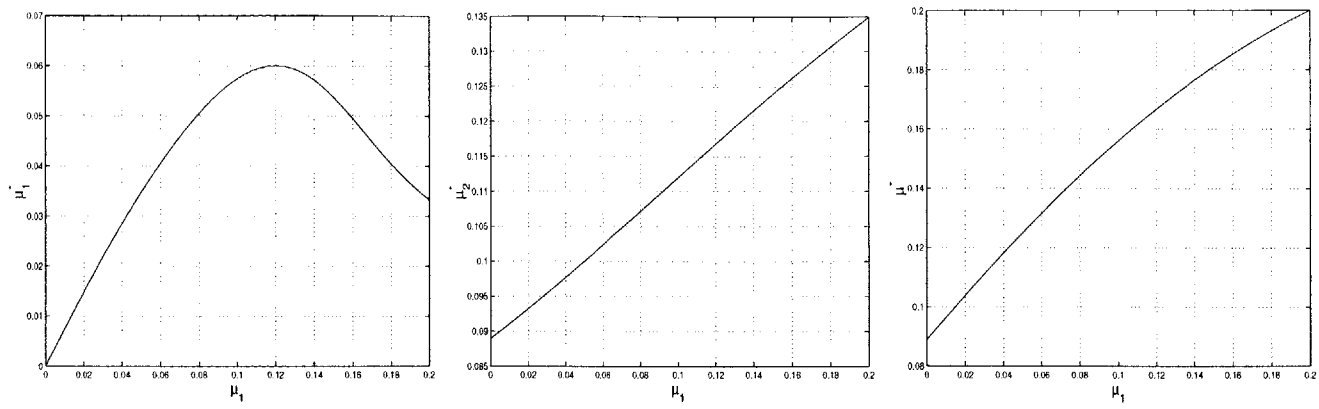
$$\begin{aligned}\frac{\partial}{\partial \mu_1}\left(\frac{V_1}{V_2}\right)^* &= \frac{-1}{2a_{21}}\left(1+\frac{2(a_{12}+\mu_2-a_{21}-\mu_1)}{2\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right)\\ &< 0 \qquad \text{since } \left|\frac{a_{12}+\mu_2-a_{21}-\mu_1}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right| < 1\end{aligned}$$

This in turn causes μ_2^* to change in the direction of μ_1 since $\left(\frac{V_1}{V_2}\right)^*$ appears in the denominator of μ_2^* (see equation 3.14). Figure 3-7(b) illustrates this effect.

The effect on μ^* can be realized by differentiating $-\lambda_2$ with respect to μ_1 since by Theorem 4 we know μ^* is the magnitude of the dominant eigenvalue. Therefore we have

$$\begin{aligned}\frac{\partial \mu^*}{\partial \mu_1} = \frac{\partial(-\lambda_2)}{\partial \mu_1} &= \frac{\partial}{\partial \mu_1}\left\{\frac{1}{2}\left(\mu_2+a_{12}+a_{21}+\mu_1-\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}\right)\right\}\\ &= \frac{1}{2}\left(1+\frac{2(a_{12}+\mu_2-a_{21}-\mu_1)}{2\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right)\\ &> 0 \qquad \text{since } \left|\frac{a_{12}+\mu_2-a_{21}-\mu_1}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right| < 1\end{aligned}$$

which implies that μ^* changes in the direction of μ_1. Figure 3-7(c) demonstrates the dynamics of μ^* with respect to μ_1. The effect on μ_1^* is less obvious since both the numerator and the denominator of equations 3.12 and 3.13 contain the ratio $\left(\frac{V_1}{V_2}\right)^*$. Let $\left(\frac{V_1}{V_2}\right)^* = X$. Note that $\frac{\partial X}{\partial \mu_1} = \frac{-X}{\Delta}$, where $\Delta = \sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}$.



(a) (b) (c)

Figure 3-7: Effect of change in the value of μ_1 on steady state solutions: (a) on μ_1^* (b) on μ_2^* (c) on μ^*. Parameter values are $\mu_2 = 0.1; a_{12} = 0.07; a_{21} = 0.145$; and $0 \leq \mu_1 \leq \mu_2$.

Then differentiating μ_1^* with respect to μ_1 we get

$$\begin{aligned}
\frac{\partial \mu_1^*}{\partial \mu_1} &= \frac{(X + \mu_1 \dot{X})(1+X) - \mu_1 X \dot{X}}{(1+X)^2} \\
&= \frac{X + X^2 + \mu_1 \dot{X}}{(1+X)^2} \\
&= \frac{X + X^2 + \frac{\mu_1 X}{\Delta}}{(1+X)^2} \\
&= \frac{X(1 + X - \frac{\mu_1}{\Delta})}{(1+X)^2} \\
&> 0 \qquad \text{if } 1 + X > \tfrac{\mu_1}{\Delta} \\
&< 0 \qquad \text{otherwise}
\end{aligned} \tag{3.29}$$

Note that when $1 + X < \frac{\mu_1}{\Delta}$, we have $\mu_1 \gg \mu_2$, which is in contrast with the original assumption that $\mu_2 > \mu_1$. Hence, we can exclude this situation from the analysis and conclude that μ_1^* always increases with μ_1. Figure 3-7(a) illustrates this effect; notice that μ_1^* increases to a peak at $\mu_1 \approx \mu_2$, after which it declines.

Effect of Change in μ_2

$\left(\frac{V_1}{V_2}\right)^*$ changes in the direction of μ_2. To see this, differentiate equation 3.9 with respect to μ_2 to get

$$\begin{aligned}
\frac{\partial}{\partial \mu_2}\left(\frac{V_1}{V_2}\right)^* &= \frac{1}{2a_{21}}\left(1 + \frac{2(a_{12} + \mu_2 - a_{21} - \mu_1)}{2\sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}}\right) \\
&> 0 \qquad \text{since } \left|\frac{a_{12}+\mu_2-a_{21}-\mu_1}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right| < 1
\end{aligned}$$

The effect of change in μ_2 on μ_1^* can be simply realized if we divide both the numerator and denominator of equation 3.13 by $\left(\frac{V_1}{V_2}\right)^*$ and rewrite $\mu_1^* = \frac{\mu_1}{\left(\frac{V_2}{V_1}\right)^* + 1}$. It is now easy to see that μ_1^* changes in the direction of μ_2 as well. Figure 3-8(a) demonstrates this effect.

The effect on μ^* can be realized by differentiating $-\lambda_2$ with respect to μ_2 as follows

$$\begin{aligned}
\frac{\partial \mu^*}{\partial \mu_2} = \frac{\partial(-\lambda_2)}{\partial \mu_2} &= \frac{\partial}{\partial \mu_2}\left\{\frac{1}{2}\left(\mu_2 + a_{12} + a_{21} + \mu_1 - \sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}\right)\right\} \\
&= \frac{1}{2}\left(1 - \frac{2(a_{12} + \mu_2 - a_{21} - \mu_1)}{2\sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}}\right) \\
&> 0 \qquad \text{since } \left|\frac{a_{12}+\mu_2-a_{21}-\mu_1}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right| < 1
\end{aligned}$$

which implies that μ^* also changes in the direction of μ_2. Figure 3-8(c) demonstrates the dynamics of μ^* with respect to μ_2. The effect on μ_2^* is less obvious since both the numerator and denominator of equations 3.14 will change with μ_2. Again let $\left(\frac{V_1}{V_2}\right)^* = X$, and note that $\frac{\partial X}{\partial \mu_2} = \frac{X}{\Delta}$, where $\Delta = \sqrt{(a_{12} + \mu_2 - a_{21} - \mu_1)^2 + 4a_{12}a_{21}}$. Then differentiate μ_2^* with respect to μ_2 to get

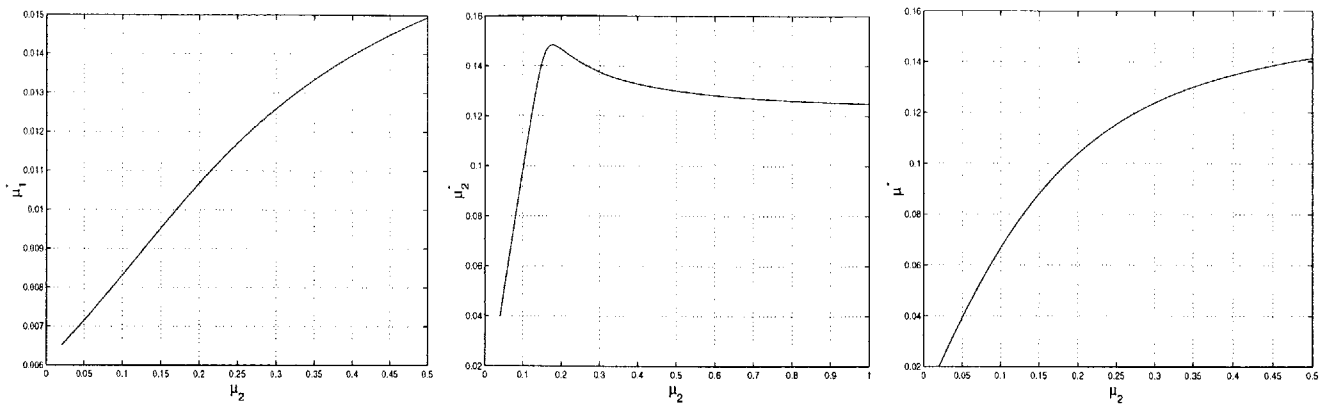
$$\begin{aligned}
\frac{\partial \mu_2^*}{\partial \mu_2} &= \frac{1 + X - \mu_2 \dot{X}}{(1+X)^2} \\
&= \frac{1 + X - \mu_2 \frac{X}{\Delta}}{(1+X)^2} \\
&> 0 \qquad \text{if } 1 + X > \mu_2 \frac{X}{\Delta} \\
&< 0 \qquad \text{otherwise}
\end{aligned} \tag{3.30}$$

This dual effect of change in μ_2 on the proportion dying in V_2 is due to two opposing processes in the dynamics. On the one hand, an increase in μ_2 results in the death of more people as they arrive from V_1; this causes the proportion dying in V_2 to increase. On the other hand, as μ_2 takes on larger values, the strong selection against V_2 causes the fraction alive in V_2 to decrease; this causes the proportion dying in V_2 to decrease. These opposing processes, i.e. the mortality and vitality of V_2, result in an intermediate peak, which can be observed in Figure 3-8(b).

Effect of Change in a_{12} and a_{21}

If μ_1, μ_2 are fixed and changes in the system are the result of change in the flow rates, then the relationship between the direction of change in $\left(\frac{V_1}{V_2}\right)^*$ and those of the steady state mortality can be uniquely defined. In other words, using the following corollaries, we can infer the direction of change in μ_1^*, μ_2^*, and μ^*, from the direction of change in $\left(\frac{V_1}{V_2}\right)^*$.

Corollary 1 *If μ_1, μ_2 are fixed and changes in the system are only due to changes*



(a) (b) (c)

Figure 3-8: Effect of change in the value of μ_2 on steady state solutions: (a) on μ_1^* (b) on μ_2^* (c) on μ^*. Parameter values are $\mu_1 = 0.02; a_{12} = 0.07; a_{21} = 0.145$; and $\mu_1 \leq \mu_2 \leq 0.5$.

in the flow rates, then μ^* *and* $\left(\frac{V_1}{V_2}\right)^*$ *change in opposite directions.*

Proof: Let $\left(\frac{V_1}{V_2}\right)^* = X$ and a be the parameter that is going under change (either a_{12} or a_{21}). Then, differentiating equation 3.12 with respect to a, we get:

$$\begin{aligned}
\frac{\partial \mu^*}{\partial a} &= \frac{\partial}{\partial a}\left(\frac{\mu_1 X + \mu_2}{X+1}\right) \\
&= \frac{\mu_1 \dot{X}(1+X) - \dot{X}(\mu_1 X + \mu_2)}{(1+X)^2} \\
&= \frac{\mu_1 \dot{X} + \mu_1 \dot{X} X - \mu_1 \dot{X} X - \mu_2 \dot{X}}{(1+X)^2} \\
&= \frac{\dot{X}(\mu_1 - \mu_2)}{(1+X)^2} \\
&< 0 \qquad \text{if } X \text{ is increasing and } \dot{X} > 0, \text{ since } \mu_1 - \mu_2 < 0 \\
&> 0 \qquad \text{if } X \text{ is decreasing and } \dot{X} < 0, \text{ since } \mu_1 - \mu_2 < 0
\end{aligned}$$

Hence, if the change in the system is due to changes in the flow rates, then μ^* changes in the opposite direction of $\left(\frac{V_1}{V_2}\right)^*$. ■

Corollary 2 *If* μ_1, μ_2 *are fixed and changes in the system are only due to changes in the flow rates, then* μ_1^* *and* $\left(\frac{V_1}{V_2}\right)^*$ *change in the same direction.*

Proof: Again, let $\left(\frac{V_1}{V_2}\right)^* = X$ and a be as in Corollary 1. Then, differentiate equation 3.13 with respect to a to get

$$\frac{\partial \mu_1^*}{\partial a} = \frac{\partial}{\partial a}\left(\frac{\mu_1 X}{X+1}\right)$$

$$
\begin{aligned}
&= \frac{\mu_1\dot{X}(1+X) - \mu_1 X\dot{X}}{(1+X)^2} \\
&= \frac{\mu_1\dot{X} + \mu_1\dot{X}X - \mu_1\dot{X}X}{(1+X)^2} \\
&= \frac{\mu_1\dot{X}}{(1+X)^2} \\
&> 0 \qquad \text{if } X \text{ is increasing and } \dot{X} > 0 \\
&< 0 \qquad \text{if } X \text{ is decreasing and } \dot{X} < 0
\end{aligned}
$$

Hence, if the change in the system is due to change in the flow rates, then μ_1^* and $\left(\frac{V_1}{V_2}\right)^*$ change in the same direction. □

Corollary 3 *If μ_1, μ_2 are fixed and changes in the system are only due to changes in the flow rates, then μ_2^* and $\left(\frac{V_1}{V_2}\right)^*$ change in opposite directions.*

Proof: Immediate from equation 3.14. ■

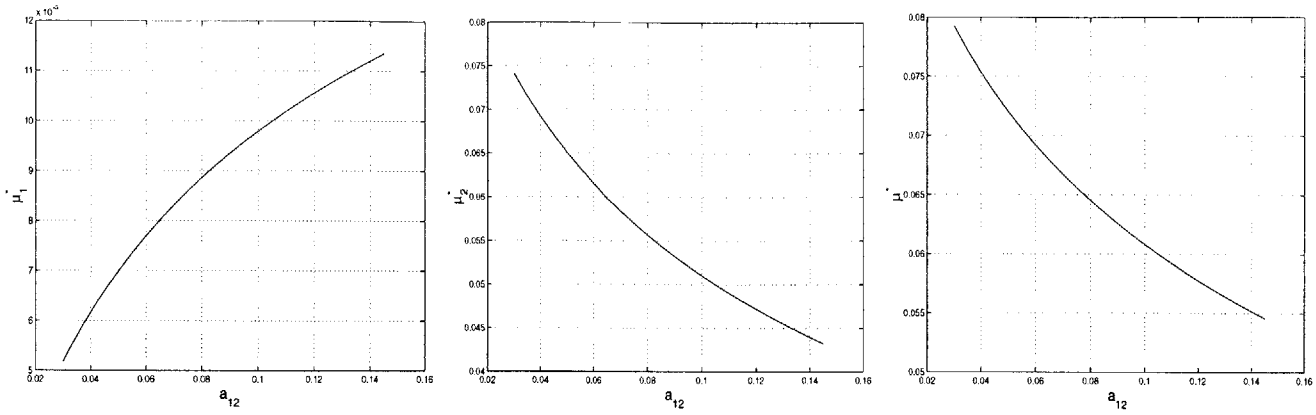
Equipped with these results, we are now ready to investigate the effect of change in the flow rates on the steady state solutions. A change in a_{12} changes $\left(\frac{V_1}{V_2}\right)^*$ in the same direction. To see this effect, we differentiate equation 3.9 with respect to a_{12} as follows

$$
\begin{aligned}
\frac{\partial}{\partial(a_{12})}\left(\frac{V_1}{V_2}\right)^* &= \frac{1}{2a_{21}}\left(1 + \frac{2(a_{12}+\mu_2-a_{21}-\mu_1)+4a_{21}}{2\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right) \\
&= \frac{1}{2a_{21}}\left(1 + \frac{(a_{12}+\mu_2-a_{21}-\mu_1)+2a_{21}}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right) \\
&> 0 \qquad \text{since } \left|\frac{a_{12}+\mu_2-a_{21}-\mu_1}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right| < 1 \\
&\qquad\qquad \text{and } \frac{2a_{21}}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}} > 0
\end{aligned}
$$

Thus by Corollaries 1,2, and 3, a change in the flow rate a_{12} causes μ_1^* to change in the same direction, while μ_2^* and μ^* will change in the opposite direction. Figure 3-9 demonstrates these effects.

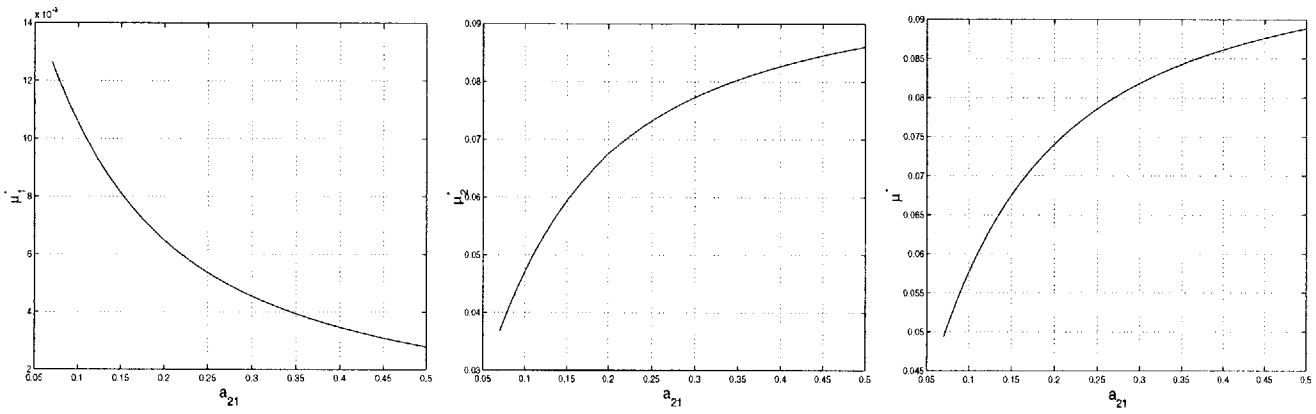
A change in a_{21} causes $\left(\frac{V_1}{V_2}\right)^*$ to change in the opposite direction. To see this effect, we differentiate equation 3.9 with respect to a_{21} to get

$$
\begin{aligned}
\frac{\partial}{\partial(a_{21})}\left(\frac{V_1}{V_2}\right)^* &= \frac{-1}{2(a_{21})^2}\left[a_{21}\left(1 + \frac{2(a_{12}+\mu_2-a_{21}-\mu_1)+4a_{12}}{2\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right)\right] \\
&\quad -\frac{1}{2(a_{21})^2}\left(a_{12}+\mu_2-a_{21}-\mu_1+\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}\right)
\end{aligned}
$$



(a) (b) (c)

Figure 3-9: Effect of change in the value of a_{12} on steady state solutions: (a) on μ_1^* (b) on μ_2^* (c) on μ^*. Parameter values are $\mu_1 = 0.02; \mu_2 = 0.1; a_{21} = 0.145$; and $0.02 \leq a_{12} \leq a_{21}$.



(a) (b) (c)

Figure 3-10: Effect of change in the value of a_{21} on steady state solutions: (a) on μ_1^* (b) on μ_2^* (c) on μ^*. Parameter values are $\mu_1 = 0.02; \mu_2 = 0.1; a_{12} = 0.07$; and $a_{12} \leq a_{21} \leq 0.5$.

$$
\begin{aligned}
&= \frac{-1}{2a_{21}}\left(1 + \frac{(a_{12}+\mu_2-a_{21}-\mu_1)+2a_{12}}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right) - \frac{1}{a_{21}}\left(\frac{V_1}{V_2}\right)^* \\
&< 0 \qquad \text{since } \left|\frac{a_{12}+\mu_2-a_{21}-\mu_1}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}}\right| < 1 \\
&\qquad\qquad \text{and } \frac{2a_{21}}{\sqrt{(a_{12}+\mu_2-a_{21}-\mu_1)^2+4a_{12}a_{21}}} > 0 \text{ and } \left(\frac{V_1}{V_2}\right)^* > 0
\end{aligned}
$$

Hence, by Corollaries 1,2, and 3, a change in the flow rate a_{21} causes μ_1^* to change in the opposite direction, while μ_2^* and μ^* will change in the direction of a_{21}. Figure 3-10 demonstrate these effects. Notice the symmetry in the dynamics of Figure 3-9 and that of Figure 3-10.

Consequences for Intervention: Putting It All Together

A decrease in μ_1 causes all three steady state mortality outcomes μ_1^*, μ_2^*, and μ^* to decline. Although, this effect produces two of the three desired outcomes, i.e. lowers the general mortality μ^*, and lowers the proportion dying in the class with poor health μ_2^*, yet it can be misleading in terms of its implication for intervention. This would imply that in order to achieve the desired objectives, μ_1 must be further reduced! Therefore a reduction in the population mortality and the mortality of those with poor health, is not necessarily conclusive of improvement in health; it may very well be due to further improvement in the health of those already having good health. On the other hand, as μ_2 decreases below a threshold, i.e. as long as $1 + X > \mu_2 \frac{X}{\Delta}$ (see inequality 3.30), all three outcomes μ^*, μ_1^*, and μ_2^* decrease. Although the proportion dying in the class with good health also declines, between a decrease in μ_1 and a decrease in μ_2, the latter is more sensible in terms of public health intervention and produces a more desirable outcome: it lowers the general mortality μ^* and the proportion dying the class with poor health μ_2^*, at a much faster rate. A reduction in μ_2 can be interpreted as medical interventions such as improved emergency room care or increased ambulance services for the poor and elderly.

It is only through the changes in the flow rates a_{12} and a_{21} that all three outcomes of interest can be achieved in a sensible manner for intervention. In other words, an increase in a_{12} or a decrease in a_{21}, generates opposite dynamics for μ^* and μ_1^*. That is, while the general mortality μ^*, and the proportion dying in the class with poor health decline, at the same time the proportion dying in the class with good health rises. This means that people will live generally healthy until the time of death.

What is the domain of different intervention strategies pertaining to such changes in the flow rates? The answer may vary with the age group of the population under study. For instance, if there is a push to move people from the high-vulnerable class to the low, through social programs or educational efforts or health improvement policies such as cleaning up the minority neighborhoods or however an incremental change in a_{12} (or a decremental change in a_{21}) may come about, the bottom line is that it will change the dynamics of vulnerability as we live it today.

3.4 Comparing Populations: Mortality Crossover

One of the "anomalies" of mortality data is that after about age 80, the death rate for the White population appears to be higher than that of the Black population, a phenomenon referred to as "mortality crossover" in demography and public health literature. Figure 3-11 demonstrates the crossover phenomenon for the all-cause Black and White mortality, for the male and female populations of the United States in 1996. The same phenomenon of gradually converging and intersecting the mortality curves has been observed with respect to disease processes. Figure 3-12 and Figure 3-13 illustrate the crossover phenomenon for the Black and White mortality due to diseases of heart, for the respective male and female populations of the United States during the period 1960-1996. Note that the age-at-crossover is slightly higher for

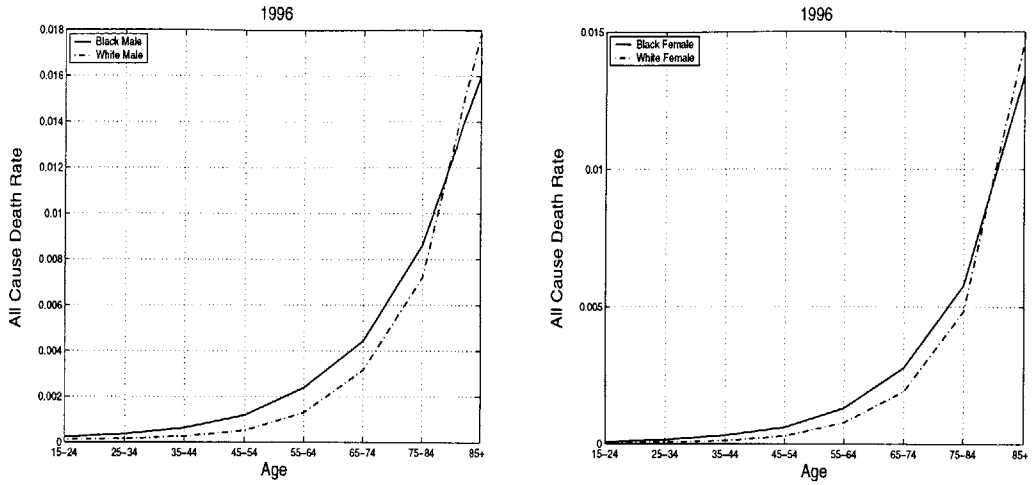


Figure 3-11: Age-specific all-cause death rates for the Black and White populations of the United States. (a) Male; (b) Female. Source: Vital Statistics of the United States, 1996.

the female population than the male population, for both all-cause and cause-specific mortality. Further the cause-specific age-at-crossover is slightly lower than the all-cause age-at-crossover for both the male and female populations. But this is a rather sensible outcome since the all-cause mortality is the weighted average of all cause-specific mortality.

For nearly 80 years, the crossover phenomenon has intrigued scholars, and has divided them into two schools of thought. Some argue that the crossover phenomenon is not real and is the result of inaccuracies in data due to age over-reporting at older ages. Others insist that it is real and hypothesize that it is due to selective processes. In the early 1900's, Raymond Pearl, when comparing the mortality experiences of different populations, noticed the phenomenon for the first time. At first, he attributed it to erroneous data and found it inconsistent with common sense. Why should a population with lower life expectancy which is exposed to a harsher environment and greater force of mortality, exhibit better survival in the long run (relative to the life span)? However, in 1922, as he was developing his theory of lifetime chances of death [17], he rediscovered the phenomenon and found it entirely consistent with the process of mortality selectivity.

Since Pearl, there has been substantial debate over the existence of the phenomenon and its meaning. Sam Preston and colleagues have argued extensively against its realness [3, 4, 5, 18]. They have demonstrated by matching the death certificates of persons 65 years and older to records of the same individuals in U.S. censuses that there are serious errors and inconsistencies in the mortality data, although they were not been able to eliminate the Black-White mortality crossover; the age-at-crossover merely increased to an older age as data was corrected. On the other hand, Nam et al. [15, 16] have examined the existence of crossovers for a large number of paired populations at different points in time and reported that about one-third of all possible pairs of age-specific curves exhibit the crossover phenomenon. Vaupel et

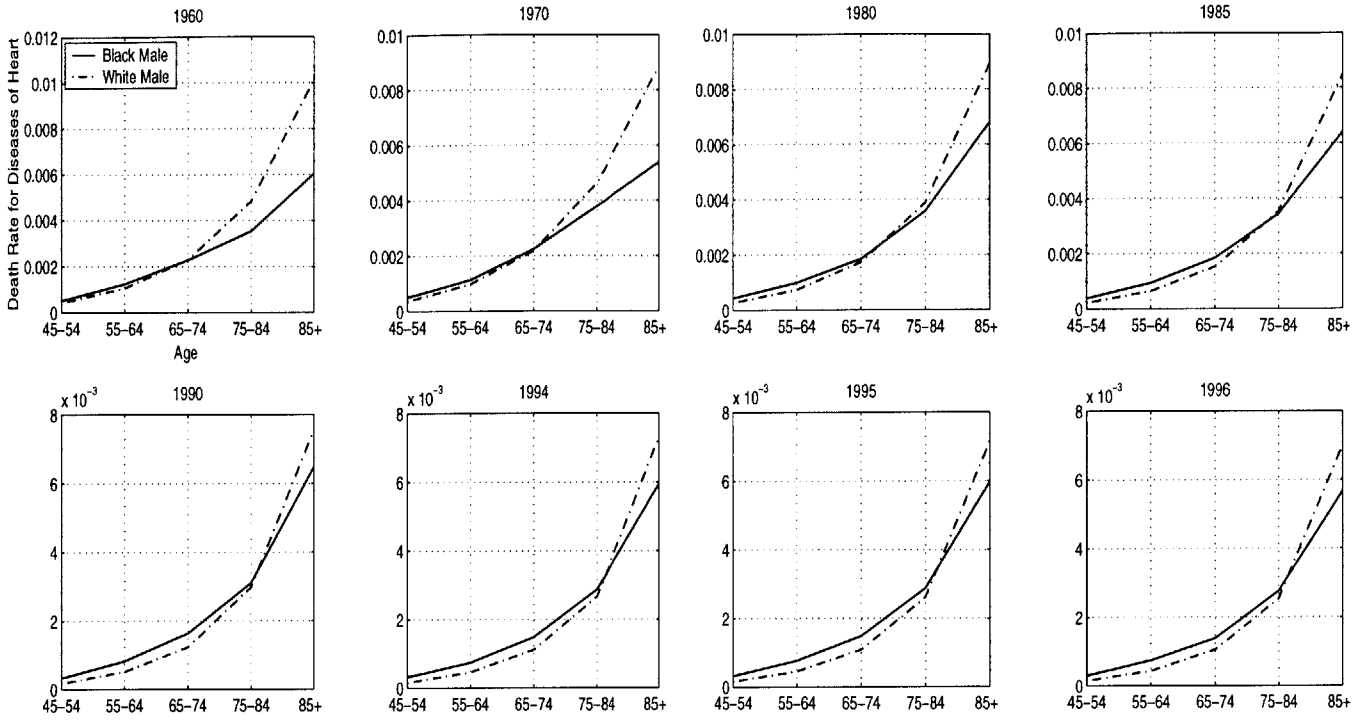


Figure 3-12: Age-specific death rate for diseases of heart; Black male versus White male; 1960-96. Source: Vital Statistics of the United States, 1960-96.

al. [22], and Manton and Stellard [9, 10, 11], have further elaborated on the selectivity hypothesis using a selection model of frailty. Nam [14] provides an overview of the history of the debate over the crossover phenomenon.

More recently, a group of scientists at the North Carolina site of the Established Populations for Epidemiologic Studies of the Elderly (EPESE), founded by the National Institute on Aging, examined the phenomenon through a prospective cohort study [2]. They devised a cohort study of 4136 men and women, consisting of 1875 Whites and 2261 Blacks, 65 years and older, living in North Carolina. They were interviewed in 1986 and were followed up until 1994. All-cause and cardiovascular disease mortality rates were calculated, with adjustment made for sociodemographic and coronary heart disease (CHD) risk factors. They recorded the age and date of birth for Black and White participants in the same manner up to 8 years before death, so that their ages were more likely to be accurate than ages on death certificates that would be normally reported by relatives. They further hypothesized that if age over-reporting among Blacks is to explain the appearance of mortality crossover, then the mortality crossover must be observed across all causes of death. However, in their study the mortality crossover was only observed for CHD and not for any other diseases the participants died from. Therefore they concluded that the mortality crossover is indeed a real phenomenon between the Black and White populations of North Carolina, and they attributed it to selective survival of the healthiest oldest Blacks or other biomedical factors affecting longevity after age 80.

Today, despite the relative wealth of mortality data and the long history of debates

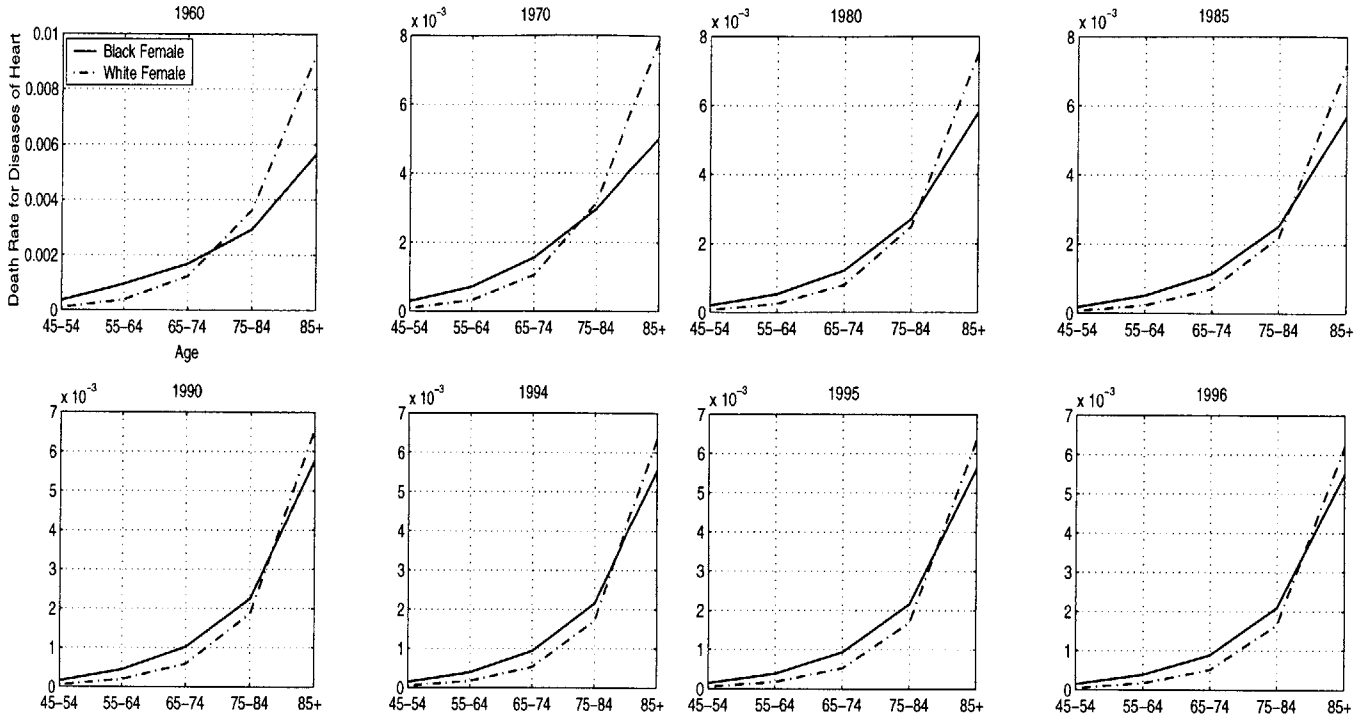


Figure 3-13: Age-specific death rate for diseases of heart; Black female versus White female; 1960-96. Source: Vital Statistics of the United States, 1960-96.

over the existence of mortality crossovers and characterization of the intersecting populations, our understanding of the crossover phenomenon remains inadequate. This chapter is an attempt to explain the phenomenon in light of our vulnerability model, driven by the processes of selection and mobility. The model was developed in section 3.2. Here, we ask: *When will a population, initially exposed to greater force of mortality, have a lower death rate*? To shed light on this matter, we first argue that the crossover phenomenon is indeed a dynamic one and propose a method for transforming data from the "age-domain" to the "time-domain". This step is necessary in order for our model to mimic the transient dynamics of mortality experience of a population in a realistic manner. We will then identify the conditions under which the mortality curves of two populations cross as well as identifying the processes that govern the dynamics of the mortality crossover. Finally, we extend the results to other populations.

3.4.1 Dynamics of Crossover

Is the mortality crossover between the Black and White populations a dynamic phenomenon? Or, has the age-at-crossover stayed relatively the same over time? In fact, it turns out that the age-at-crossover for the Black and White populations has been a time varying phenomenon. Prior to 1960, the national vital registration system divided the population of the United States into "White" and "Others", and therefore we are left with no clue for investigating the existence of crossover phenomenon

before 1960. In 1960, "Black" was recognized as a distinct race or color among the "non-White" population. Starting in 1980, other racial and ethnic groups were added into the vital statistics vocabulary. Today, these groups include "American Indian or Alaskan Native", "Asian or Pacific Islander", "Hispanic", and "White, non-Hispanic". Despite the lack of racially distinct and reliable mortality data prior to 1970[1], looking at the dynamics of the age-at-crossover during the seemingly short period of 1970-1996, can still be enlightening. Figure 3-14 demonstrates the crossover phenomenon between the Black and White male populations in the United States, for the period 1970-1996. In 1970, the age-at-crossover was at the brink of entering the age group 75-84; after 1980, the age-at-crossover moved up into the upper half of the age group 75-84; in 1996 it is on the brink of entering into the age group 85+. One could deduce from all this that prior to 1970, the-age-at-crossover must have been in the younger age groups, although we cannot say with certainty what the initial age-at-crossover might have been. The only evidence toward the validity of this claim would be the comparison of the mortality data from the years 1960 and 1965. In those years the vital registration system included records of deaths for non-residents of the United States. But if on average, the effect of such inclusion is the same for both the Black and White populations, then studying these records can still help put matters in perspective. Figure 3-15 demonstrates the dynamics of mortality crossover for the years 1960 and 1965. Compare them to the dynamics of Figure 3-14. In 1960 and 1965, the age-at-crossover is well within the age group 65-74, whereas starting in 1970 it is at the brink of entering the next age group 75-84.

If the mortality crossover is indeed a dynamic phenomenon, what are the underlying processes that govern its dynamics? To identify and analyze the effect of such processes using our model of vulnerability, we need to link the outcomes of our model and the implications derived from mortality data. Our abstract model of transition among different vulnerability classes models the mortality experience of a population cohort over time and does not take account of the effect of aging explicitly. In section 3.2.2 we established that mortality, as produced by our vulnerability model, is a monotonically decreasing process, as should be the mortality experience of a non-aging cohort over time. For our abstract model to simulate the transient behavior of mortality at the crossover, first we need to make a transformation from mortality data in the "age-domain" to mortality data in the "time-domain". In other words, given a set of age-specific mortality data at different points in time, as in Figure 3-14, we will transform them to a set of time-dependent mortality data for different age groups. That is, we still work with the same data, only we change our perspective of it.

3.4.2 From "age-domain" to "time-domain"

Let $A_\times$ represent the age-at-crossover, and $Y_\times$ be the year-at-crossover. Then the following transformation is necessary to map the mortality data from the age-domain

[1]For the period 1960-1969 the vital registration system included deaths of non-residents of the United States.

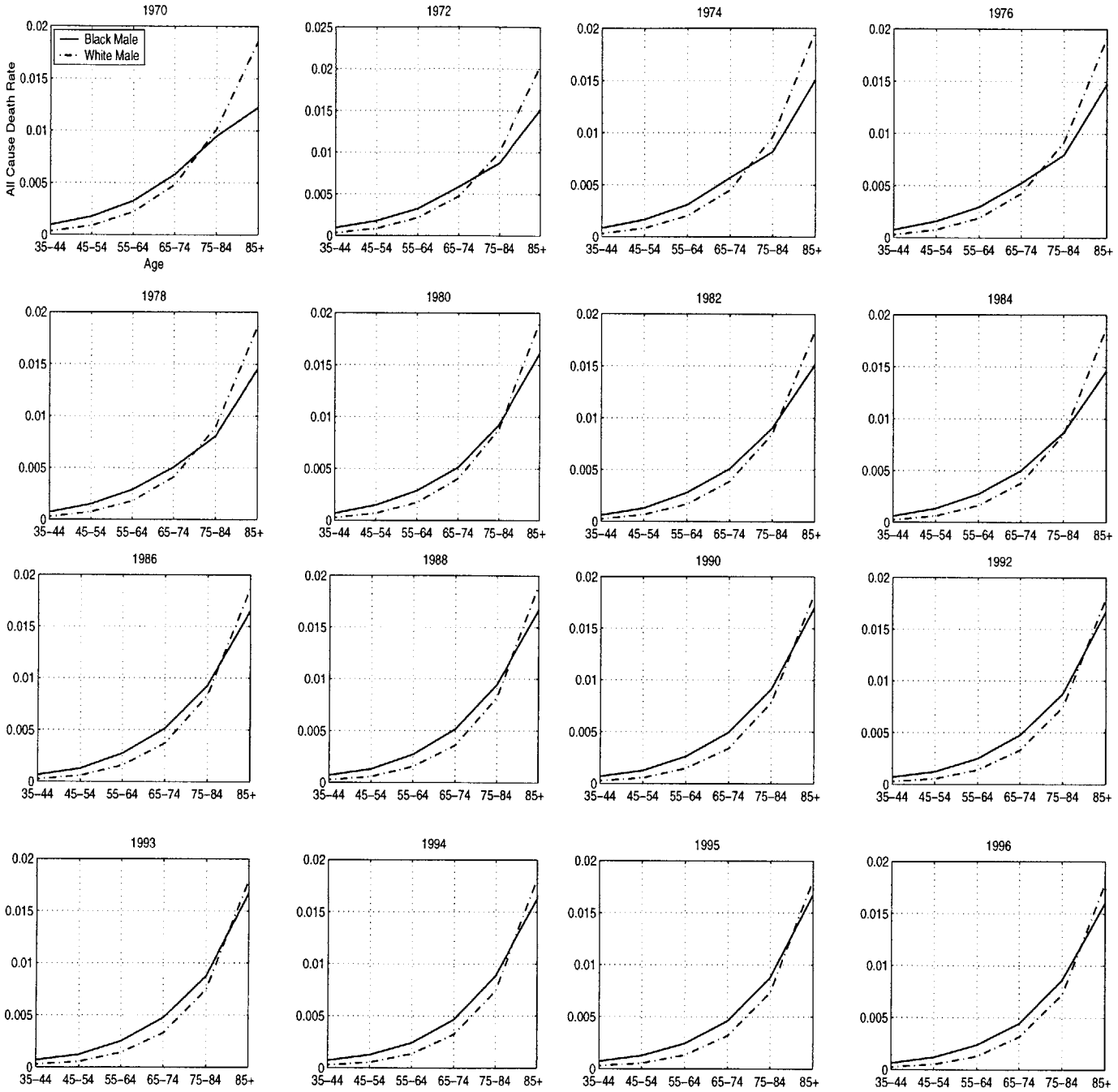


Figure 3-14: Age-specific all-cause death rates; Black male versus White male; 1970-96. Source: Vital Statistics of the United States, 1970-96.

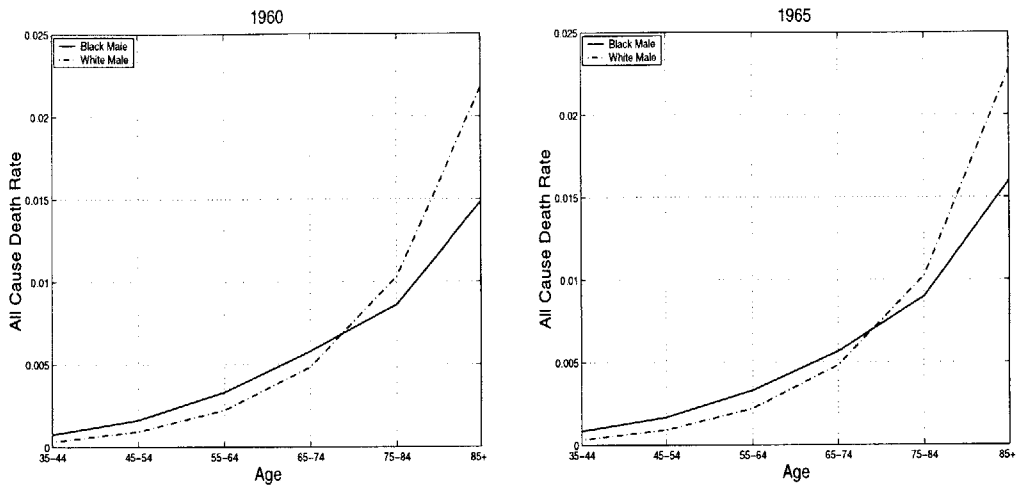


Figure 3-15: Age-specific all-cause death rates; Black male versus White male for the years 1960 and 1965. (a) 1960; (b) 1965. Source: Vital Statistics of the United States, 1960, 1965.

to the time-domain.

1. For every age group, plot the mortality curves of Black and White populations of that age group over time. We will end up with a collection of cohort-specific Black-White mortality in the time-domain. Figure 3-16 demonstrates the transformation process for the age groups 65-74, 75-84, and 85+.

2. To determine the correspondence of $A_\times$ to the transformed data from step 1, for every age group in the time-domain, one of the following must be true:

   (a) If the Black curve is consistently dominating the White curve consistently over a time period, it means that $A_\times$ must be in an older age group for that time window. Figure 3-16(a) demonstrates an example of the case, where the mortality of the White population is consistently below that of the Black population, indicating that $A_\times$ must be in one of the age groups older than 65-74.

   (b) If the Black curve is consistently dominated by the White curve over a time period, it means that $A_\times$ must be in a younger age group for that time window. Figure 3-16(c) illustrates an example of the case, where the mortality curve of the White population is consistently above that of the Black population for the period 1970-96. This means that $A_\times$ must be in one of the age groups younger than 85+.

   (c) If the two curves are crossing at a particular year $Y_\times$, then it is indicative of a change in $A_\times$. Assuming that the $A_\times$ is only increasing over time, then the Black curve is dominated by the White curve prior to $Y_\times$, and dominating the White curve thereafter. In this case, the interpretation of change is the following: $A_\times$ passed the median age of the age group at year

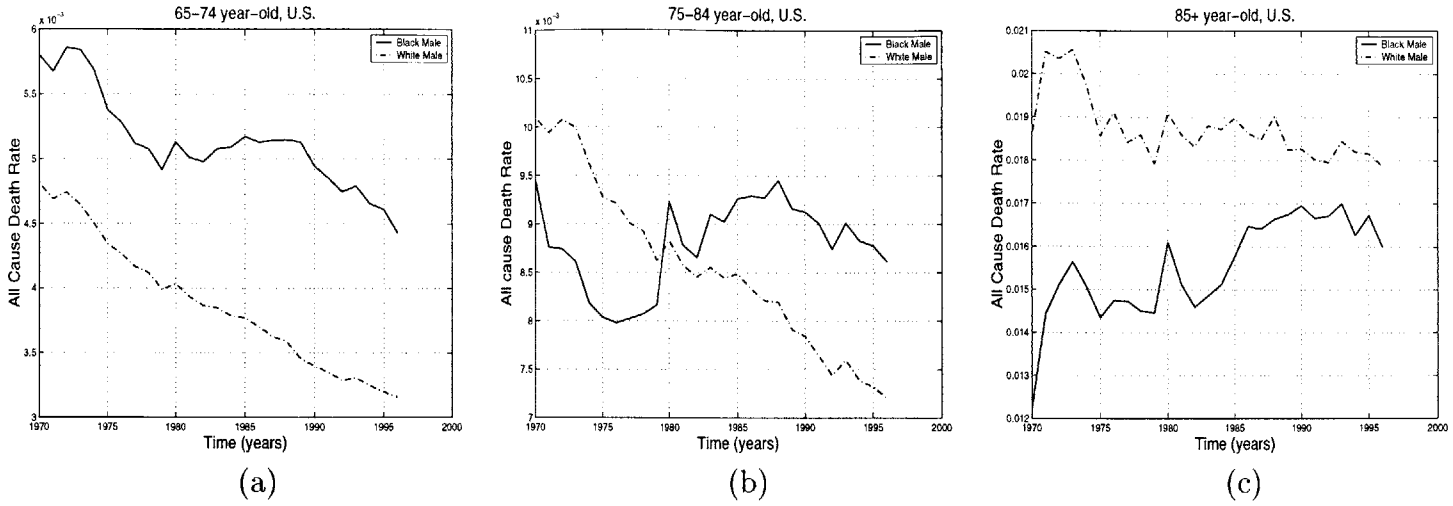


Figure 3-16: All cause longitudinal death rates; Black male versus White male; 1970-96. (a) cohort: 65-74 year-old; (b) cohort: 75-84 year-old; (c) cohort: 85+ year-old. Source: Vital Statistics of the United States, 1970-96.

$Y_\times$. For example, the crossover in Figure 3-16(b) indicates that in 1980, $A_\times$ has "roughly"[2] moved up to the upper half of the age group 75-84.

Clearly, the above mapping is a one-to-one correspondence . A collection of instances in the age-domain can be uniquely transformed into a collection of instances in the time-domain and vice versa. Although $Y_\times$ in the time-domain has a different interpretation than $A_\times$ in the age-domain, the above mapping is sound in the sense that the dynamics of crossover is preserved and can therefore be analyzed using our abstract model of vulnerability. But more importantly, the transformation of mortality data obviates the need for further modification of our model to directly incorporate the effect of aging. we took a short cut of transforming the data to match the specifications of the model rather than modifying the model to directly accommodate for the aging effect in the data.

To identify the processes governing the existence of a mortality crossover as well as its dynamics, we will make use of the steady state solutions of the vulnerability-ratio and population mortality, derived in section 3.2.2, to make inferences about the transient dynamics of mortality of a population at the crossover. That is, despite the lack of an explicit solution for the transient vulnerability of a population, we will be able to identify the conditions for the existence and change in the age-at-crossover, and make diagnostic use of them.

[2]Needless to say that the mortality rate of an age group, taken from the Vital Statistics of the United States, is a weighted average for that age group. Under uniform distribution, however, if for an age group in the time domain, two populations cross in the above manner, it must be that the majority of the strata of that age group have higher death rates in one population. Thus the median age of an age group can be used to represent the threshold for change in $A_\times$. This rough interpretation of $A_\times$ would not have been necessary, had we access to single age group mortality data.

3.4.3 A Modified Model of 2-Vulnerability-Class

Although the dynamics of population mortality, as produced by our model of 2-Vulnerability-Class, matches that of a non-aging cohort, the dynamics of population "vitality"[3] does not match that of a non-aging cohort. Recall from section 3.2.1 that in the absence of birth or migration in the model, the population in each vulnerability class and therefore the total population will eventually become extinct. But this is descriptive of the behavior of an "aging" cohort, where a fixed number of individuals who are born at the same time, will age simultaneously and die as time progresses. Hence in this case, the notion of progression in time matches that of age. However, the vitality of a non-aging population cohort should grow over time, as it has been, due to the continuous increase in infant survivability and therefore the increase in the life expectancy of man.

Hence, we need to modify our original system so that it can accommodate for the increase in the vitality of a non-aging cohort, while the dynamics of mortality stays the same. In other words, we want the population to grow without bound, as opposed to becoming extinct, while the vulnerability-ratio $\frac{V_1}{V_2}$, maintains its original dynamics. To achieve these objectives, we modify the original system of ODE's 3.1-3.2 as follows

$$\frac{dV_1}{dt} = -V_1(a_{21} + \mu_1 - c) + a_{12}V_2 \tag{3.31}$$

$$\frac{dV_2}{dt} = a_{21}V_1 - V_2(a_{12} + \mu_2 - c) \tag{3.32}$$

where c is a positive constant that in principle should be larger than the rate of selection $\mu_2 - \mu_1$. This new addition makes the population vitality to increase unbounded. The new system is now unstable; implying that at least one of its eigenvalues is positive. Let δ_1, δ_2 be the eigenvalues of the new system. The characteristic polynomial $p(\delta)$ of the coefficient matrix of the new system is

$$p(\delta) = \begin{vmatrix} -\mu_1 - a_{21} + c - \delta & a_{12} \\ a_{21} & -\mu_2 - a_{12} + c - \delta \end{vmatrix}$$

Expanding the determinant and collecting the terms in δ, we get

$$\begin{aligned} p(\delta) &= \delta^2 + (a_{12} + \mu_2 + a_{21} + \mu_1 - 2c)\delta \\ &\quad +(\mu_1\mu_2 + \mu_1 a_{12} + a_{21}\mu_2) - c(a_{12} + \mu_2 + a_{21} + \mu_1) + c^2 \end{aligned}$$

Solving for the roots of the new characteristic polynomial, we get

$$\begin{aligned} \delta_1 &= \lambda_1 + c \\ \delta_2 &= \lambda_2 + c \end{aligned} \tag{3.33}$$

where λ_1, λ_2 are the eigenvalues of the original system from section 3.2. Furthermore,

[3]The term vitality refers to the proportion of population alive at a point in time.

if we assume that λ_2 is the dominant eigenvalue in the original system, then δ_2 is the dominant eigenvalue in the new system, i.e. δ_2 is the positive eigenvalue. To transform the new two dimensional linear system describing the rate of change in the frequency of each vulnerability class, into one quadratic equation describing the rate of change in the ratio of the frequency of the two vulnerability classes, we follow the same procedure as in section 3.2.1. First multiply equation 3.31 by V_2 and equation 3.32 by V_1 to get

$$\frac{dV_1}{dt}V_2 = -V_1V_2(a_{21} + \mu_1 - c) + a_{12}{V_2}^2 \tag{3.34}$$

$$\frac{dV_2}{dt}V_1 = a_{21}{V_1}^2 - V_1V_2(a_{12} + \mu_2 - c) \tag{3.35}$$

Next, subtract equation 3.35 from equation 3.34 and divide both sides of the new equation by ${V_2}^2$ to get a single quadratic differential equation as follows

$$\begin{aligned}
\frac{\frac{dV_1}{dt}V_2 - \frac{dV_2}{dt}V_1}{{V_2}^2} &\stackrel{def}{=} \frac{d}{dt}\left(\frac{V_1}{V_2}\right) \\
&= -\frac{V_1}{V_2}(a_{21} + \mu_1) + c\frac{V_1}{V_2} - c\frac{V_1}{V_2} + \frac{V_1}{V_2}(a_{12} + \mu_2) \\
&\quad +a_{12} - a_{21}\left(\frac{V_1}{V_2}\right)^2 \\
&= -a_{21}\left(\frac{V_1}{V_2}\right)^2 + (a_{12} + \mu_2 - a_{21} - \mu_1)\frac{V_1}{V_2} + a_{12}
\end{aligned}$$

which is exactly the same as equation 3.8 that was derived from the original system. This means that all the results pertaining to the dynamics of vulnerability-ratio, including the existence of a steady state solution, stand still. What about population mortality? Recall the definitions of total mortality μ and its steady state value μ^* from section 3.2.2, as functions of μ_1, μ_2, and the ratio $\frac{V_1}{V_2}$. If the vulnerability-ratio in the new system is exactly the same as the one from the original system, then there is no reason to assume that mortality should be any different. In other words, the population mortality in the new system is a monotonically decreasing process as well. Furthermore, the final mortality μ^* is the same as the dominant eigenvalue in the original system, or put in terms of the dominant eigenvalue of the new system, $\mu^* = c - \delta_2 = \lambda_2$ (see equation 3.33). Therefore, we conclude that all major results pertaining to the dynamics of the vulnerability-ratio and population mortality, derived for the 2-Vulnerability-Class, will strongly hold for the Modified-2-Vulnerability-Class. In subsequent sections, we will make use of these results to infer knowledge about the dynamics of mortality of two non-aging cohorts at the crossover.

3.4.4 From Steady State to Transience: Conditions for Existence of Crossover

Suppose we have two populations B and W, each modeled after the dynamics of the system of equations 3.1-3.2 and that of Figure 3-1. Let B_1, B_2 represent the low and high vulnerability classes of B respectively; W_1, W_2 be those of W; μ_B and μ_W be the respective population mortality for B and W; and μ_B^*, μ_W^* represent the steady state population mortality for B and W respectively.

In the age-domain, when B and W cross, B initially has a higher mortality rate, but then the mortality rates of the two populations gradually converge, cross at some age, after which W attains a higher death rate. This crossover phenomenon, once transformed into the time-domain, has the following correspondence: for all age groups older than the age group containing $A_\times$, B must have a lower death rate at all times over the period of investigation, and for all age groups younger than the age group containing $A_\times$, B must have a higher death rate at all times over the period of investigation. Figure 3-16 illustrates this mapping for three adjacent cohorts in the Black and White populations. Hence, the relationship between the mortality curves in the oldest age group of two populations provides sufficient clues into the existence of crossover. Recall from the transformation recipe in section 3.4.2 that given an age group, if the B curve is consistently dominated by the W curve, it indicates that $A_\times$ for B and W must be within a younger age group. Whatever that younger age group may be, the implication of such an observation is straightforward: the age-specific mortality curves of B and W do cross. Therefore, to identify the conditions under which the age-specific mortality curves of B and W cross, we need to identify the conditions under which the longitudinal mortality curve of B is consistently below W in the oldest age group. Ultimately, the implications suggested by such analysis should expand to all intermediate age groups older than the age group containing $A_\times$. By symmetry, the inverse of the results is implied for all age groups younger than the age group containing $A_\times$.

For our abstract model of vulnerability to simulate the dynamics of the mortality experience of the oldest age group of two populations in the time-domain, the following conditions must be satisfied:

1. The initial mortality rate for B must be lower than the initial mortality rate for W. In other words, $\mu_B(0) < \mu_W(0)$.

2. The final mortality rate for B must also be lower than that for W. In other words, $\mu_B^* < \mu_W^*$.

The existence of these two conditions is sufficient to deduce that the mortality curve of the oldest cohort in B is consistently below that of W at all times, and therefore deduce the existence of crossover in a younger age group. This is because the only other way for B to have lower initial and final mortality rates than W, is for B to have higher mortality rates at all other intermediate points, which would imply a double-crossing in the two curves. In the following lemma we show this cannot happen.

Lemma 1 *There cannot exist a double-crossing between the mortality curves of two populations.*

Proof: Suppose for contradiction that the mortality curves of B and W cross twice. Then this means that W, which initially is above B, must "curve" twice at the points of crossing. In the vulnerability model, however, mortality is a monotone process to a steady state. This requires for a mortality curve to "bend" only once. Hence, if the initial and final mortality of B are below those of W, then B must have lower mortality at all intermediate points in time. ■

Therefore by Lemma 1, the existence of the above two conditions in the oldest cohort of B and W necessarily implies the existence of $A_\times$ in one of the younger age groups. This result has the remarkable implication that the underlying processes governing the existence of crossover can be inferred from the steady state solutions of the vulnerability-ratio and population mortality. In subsequent sections, we map the above two conditions into conditions in terms of the system parameters and the initial values, under two different assumptions: (1) when class-specific mortality rates μ_1, μ_2 are the same for both populations but the flow rates are different; (2) when the flow rates a_{12}, a_{21} are the same for both populations but the class-specific mortality rates are different.

Different Flow Rates; Same Mortality Rates

Suppose B and W have the same class-specific mortality rates μ_1 and μ_2. Let $B_1(0)$, $B_2(0)$ be the initial conditions for B; $W_1(0)$, $W_2(0)$ be the initial conditions for W; and $\frac{B_1}{B_2}$, $\frac{W_1}{W_2}$ represent the population vulnerability-ratio for B and W respectively. If μ_1 and μ_2 are the same for both populations, then the first condition on the existence of crossover implies

$$\begin{aligned}
& \mu_B(0) < \mu_W(0) \\
\equiv\ & \mu_1 B_1(0) + \mu_2 B_2(0) < \mu_1 W_1(0) + \mu_2 W_2(0) \\
\equiv\ & \mu_1(1 - B_2(0)) + \mu_2 B_2(0) < \mu_1(1 - W_2(0)) + \mu_2 W_2(0) \\
\equiv\ & \mu_1(W_2(0) - B_2(0)) < \mu_2(W_2(0) - B_2(0)) \\
\equiv\ & (W_2(0) - B_2(0))(\mu_2 - \mu_1) > 0 \\
\equiv\ & W_2(0) > B_2(0) \quad \text{since by definition } \mu_2 > \mu_1
\end{aligned} \tag{3.36}$$

which in turn implies

$$B_1(0) > W_1(0) \tag{3.37}$$

Putting inequalities 3.36 and 3.37 together, we get

$$\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)} \tag{3.38}$$

This means that the oldest cohort in B, initially had a higher vulnerability-ratio

than that of W. Here, "initially" refers to the initial year in the scope of the study. In this work, the initial year is 1970 as depicted in Figure 3-16. Inequality 3.38 is a rather sensible outcome: lower initial death rate for the oldest cohort in B is due to a higher initial vulnerability-ratio in B; the class with good health in B contained a higher proportion of old-aged residents compared to the class with good health in W. By symmetry, the class with poor health in B had a smaller residency by old-aged people than the one in W.

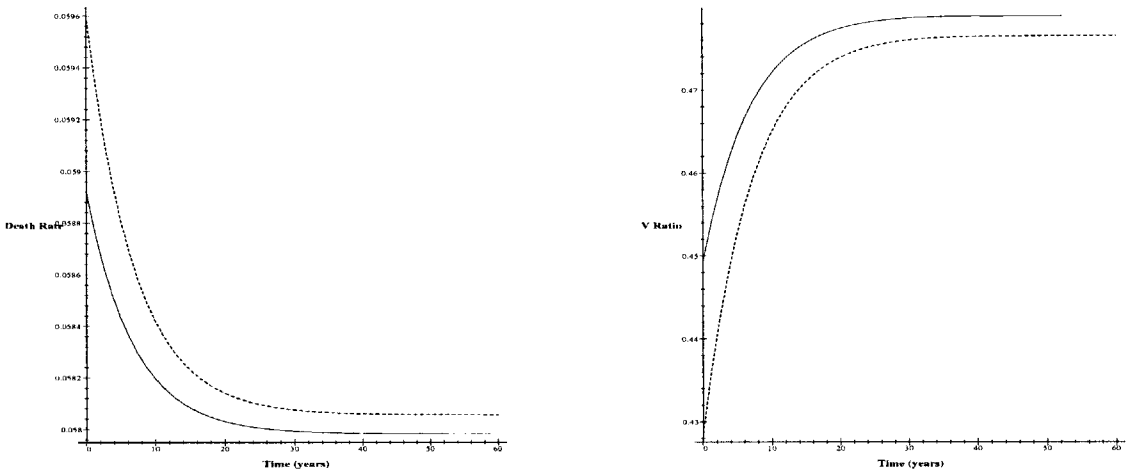
To express condition 2 for the existence of crossover in terms of the system parameters, let $\left(\frac{B_1}{B_2}\right)^*$ be the steady state vulnerability-ratio for B, and $\left(\frac{W_1}{W_2}\right)^*$ be the steady state vulnerability-ratio for W. Then condition 2 for the existence of crossover implies

$$\begin{aligned}
\mu_B^* &< \mu_W^* \\
\equiv \quad \frac{\mu_1\left(\frac{B_1}{B_2}\right)^* + \mu_2}{\left(\frac{B_1}{B_2}\right)^* + 1} &< \frac{\mu_1\left(\frac{W_1}{W_2}\right)^* + \mu_2}{\left(\frac{W_1}{W_2}\right)^* + 1} \\
\equiv \quad \frac{\mu_1\left(\frac{B_1}{B_2}\right)^*}{\left(\frac{B_1}{B_2}\right)^* + 1} - \frac{\mu_1\left(\frac{W_1}{W_2}\right)^*}{\left(\frac{W_1}{W_2}\right)^* + 1} &< \frac{\mu_2}{\left(\frac{W_1}{W_2}\right)^* + 1} - \frac{\mu_2}{\left(\frac{B_1}{B_2}\right)^* + 1} \\
\equiv \quad \mu_1\left(\left(\frac{B_1}{B_2}\right)^* - \left(\frac{W_1}{W_2}\right)^*\right) &< \mu_2\left(\left(\frac{B_1}{B_2}\right)^* - \left(\frac{W_1}{W_2}\right)^*\right) \\
\equiv \left(\left(\frac{B_1}{B_2}\right)^* - \left(\frac{W_1}{W_2}\right)^*\right)(\mu_2 - \mu_1) &> 0 \\
\equiv \quad \left(\frac{B_1}{B_2}\right)^* &> \left(\frac{W_1}{W_2}\right)^* \quad \text{since } \mu_2 > \mu_1
\end{aligned} \tag{3.39}$$

This means that for population B to have a lower final death rate, its final vulnerability-ratio must be higher than that of W. In other words, the conditions for the existence of mortality crossover can be inferred from the conditions for the existence of the vulnerability-ratio crossover. Recall that the existence of mortality crossover in the age-domain implies non-intersecting mortality curves of the oldest cohort in the time-domain. Inequality 3.39 further asserts that the question of whether two mortality curves in the time-domain intersect, can be confirmed from the status of the corresponding vulnerability-ratio curves. Figure 3-17 portrays an example of this situation, where the dynamics of mortality (Figure 3-17(a)) can be inferred from the dynamics of vulnerability-ratio (Figure 3-17(b)). Notice the symmetry in the relative dominance of the curves between Figure 3-17(a) and Figure 3-17(b).

Let the flow rates for B be as before, and the flow rates for population W be represented by A_{12} and A_{21}. Then by equation 3.9 we have

$$\left(\frac{B_1}{B_2}\right)^* = \frac{1}{2}\left(\frac{a_{12}}{a_{21}} + \frac{\mu_2}{a_{21}} - 1 - \frac{\mu_1}{a_{21}} + \sqrt{\left(\frac{a_{12}}{a_{21}} + \frac{\mu_2}{a_{21}} - 1 - \frac{\mu_1}{a_{21}}\right)^2 + 4\frac{a_{12}}{a_{21}}}\right) \tag{3.40}$$



(a) (b)

Figure 3-17: Inferring the transient dynamics of mortality of two hypothetical populations in the time-domain from the dynamics of their vulnerability-ratio. Both populations have similar class-specific mortality rates but they differ in their mobility rates. The initial conditions are: $B_1(0) = 0.31$; $B_2(0) = 0.69$; $W_1(0) = 0.3$; and $W_2(0) = 0.7$. The system parameters are: $a_{12} = 0.041$; $a_{21} = 0.1316$; $A_{12} = 0.04$; $A_{21} = 0.13$; $\mu_1 = 0.012$; $\mu_2 = 0.08$. The mobility ratios are: $\frac{a_{12}}{a_{21}} \approx 0.316$ and $\frac{A_{12}}{A_{21}} \approx 0.308$. (a) Comparing the mortality rates; (b) Comparing the vulnerability ratios.

and

$$\left(\frac{W_1}{W_2}\right)^* = \frac{1}{2}\left(\frac{A_{12}}{A_{21}} + \frac{\mu_2}{A_{21}} - 1 - \frac{\mu_1}{A_{21}} + \sqrt{\left(\frac{A_{12}}{A_{21}} + \frac{\mu_2}{A_{21}} - 1 - \frac{\mu_1}{A_{21}}\right)^2 + 4\frac{A_{12}}{A_{21}}}\right) \quad (3.41)$$

The second condition for the existence of crossover, i.e. whether $\mu_B^* < \mu_W^*$, can now be investigated by asking whether $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$. We will answer this question by examining all possible relationships between the flow rates a_{ij} and A_{ij}, $i = 1, 2$. Let $\frac{a_{12}}{a_{21}}$ and $\frac{A_{12}}{A_{21}}$ denote the "mobility-ratio" for B and W respectively, where the mobility-ratio of a population is defined as a measure of the relative ease of mobility to the class with good health. Then there are five cases to consider:

1. $a_{21} = A_{21}$.

   Then if $a_{12} > A_{12}$, clearly $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$, and therefore $\mu_B^* < \mu_W^*$. Hence, under this assumption, the only way for population B to have a smaller final mortality rate, is to have a stronger mobility from the high vulnerability class to the low vulnerability class, compared to that of W. This condition translated in terms of the mobility ratios, clearly implies that $\frac{a_{12}}{a_{21}} > \frac{A_{12}}{A_{21}}$.

2. $a_{21} > A_{21}$.

Then if $(a_{12} > A_{12})$, then the only way for $\mu_B^* < \mu_W^*$, is to have $(\frac{a_{12}}{a_{21}} \gg \frac{A_{12}}{A_{21}})$. This increases the possibility that $(\frac{a_{12}+\mu_2-\mu_1}{a_{21}} > \frac{A_{12}+\mu_2-\mu_1}{A_{21}})$, and therefore $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$. Hence under stronger mobility from B_1 to B_2 compared to that of W, the only way for population B to have a smaller final mortality rate, is to also have a stronger mobility from B_2 to B_1 compared to that of W. Note that once again this condition reduces to a condition in terms of the mobility ratios, i.e. $\frac{a_{12}}{a_{21}} \gg \frac{A_{12}}{A_{21}}$.

3. $a_{21} < A_{21}$. Then we have 2 cases to consider:

   3-(a) $(a_{12} \geq A_{12})$. Then since $(\frac{a_{12}}{a_{21}} > \frac{A_{12}}{A_{21}})$ and $(\frac{\mu_2-\mu_1}{a_{21}} > \frac{\mu_2-\mu_1}{A_{21}})$, it must be that $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$, and therefore $\mu_B^* < \mu_W^*$. Hence, under this assumption, for population B to have a lower final mortality rate, is to have the rate of mobility from B_2 to B_1 to be at least as strong as mobility rate from W_2 to W_1. This condition is also equivalent to having $\frac{a_{12}}{a_{21}} > \frac{A_{12}}{A_{21}}$.

   3-(b) $(a_{12} < A_{12})$. Then there are 2 cases to consider:

      3-(b)(a) $\frac{a_{12}}{a_{21}} \geq \frac{A_{12}}{A_{21}}$. Then since $(a_{21} < A_{21})$, we have $(\frac{\mu_2-\mu_1}{a_{21}} > \frac{\mu_2-\mu_1}{A_{21}})$, which implies $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$, and therefore $\mu_B^* < \mu_W^*$. Hence, under the assumption that the rate of mobility from B_1 to B_2 is smaller than that from W_1 to W_2, one way for population B to have a lower final mortality rate, is to also have a smaller mobility rate from B_2 to B_1, to the extent that $\frac{a_{12}}{a_{21}} \geq \frac{A_{12}}{A_{21}}$. This condition portrays B as an "almost" rigid population with low rates of social mobility subject to B_2 being more mobile than B_1.

      3-(b)(b) $\frac{a_{12}}{a_{21}} < \frac{A_{12}}{A_{21}}$. Under this condition, the only way for $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$ would be if $(a_{21} \ll A_{21})$; this is to guarantee that $\frac{\mu_2-\mu_1+a_{12}}{a_{21}} > \frac{\mu_2-\mu_1+A_{12}}{A_{21}}$. Hence, under the assumptions that the mobility rate from B_2 to B_1 and B's mobility-ratio are weaker than their counterparts in W, the only way for B to have a smaller final mortality rate, is if the mobility rate from B_1 to B_2 is "much" weaker than that in W. This portrays B as a "very" rigid population with almost no social mobility (compare this to condition 3-(b)(a)).

Note that all sensible conditions for the existence of crossover, expressed in terms of the flow rates, reduce to a single condition in terms of the mobility ratios, that is $\frac{a_{12}}{a_{21}} > \frac{A_{12}}{A_{21}}$.

Putting it all together, under the assumption that B and W have the same class-specific mortality rates and only differ in their mobility rates and the initial conditions,

for $A_\times$ to exist, or for the oldest cohort in B to have consistently lower death rate than W, the following must hold:

1. $\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)}$ for the oldest cohort. This implies that initially B had a higher proportion of its old-aged individuals residing in the class with good health, compared to its counterpart in W.

2. The mobility-ratio for B must be higher than that of W. In other words, we must have $\frac{a_{12}}{a_{21}} > \frac{A_{12}}{A_{21}}$, which indicates that the relative rate of mobility to the class with good health is more easily facilitated for the oldest cohort in B than the oldest cohort in W.

3. Conditions 1 and 2 apply to all cohorts older than the cohort containing $A_\times$.

4. By symmetry, the converse of conditions 1 and 2 must hold for all cohorts younger than the cohort containing $A_\times$.

Different Mortality Rates; Same Flow Rates

Now suppose that the flow rates are the same for B and W, but they differ in their class-specific mortality rates. Let μ_1, μ_2 be the class-specific mortality rates for B, and m_1, m_2 be those of W. Further, define $S_B = \mu_2 - \mu_1$ and $S_W = m_2 - m_1$ to be the respective selection rates for B and W, where the selection rate is a measure of inequality due to physiological or social conditions.

For the first condition on the existence of crossover to hold, we must have

$$\begin{aligned}
& \mu_B(0) < \mu_W(0) \\
\equiv\ & \mu_1 B_1(0) + \mu_2 B_2(0) < m_1 W_1(0) + m_2 W_2(0) \\
\equiv\ & \mu_1(1 - B_2(0)) + \mu_2 B_2(0) < m_1(1 - W_2(0)) + m_2 W_2(0) \\
\equiv\ & B_2(0)(\mu_2 - \mu_1) + \mu_1 < W_2(0)(m_2 - m_1) + m_1 \\
\equiv\ & \frac{B_2(0)}{W_2(0)} < \frac{m_2 - m_1}{\mu_2 - \mu_1} + \frac{m_1 - \mu_1}{W_2(0)(\mu_2 - \mu_1)}
\end{aligned} \tag{3.42}$$

Given the system parameters and the initial conditions, inequality 3.42 can be used to numerically verify whether the ratio of the initial conditions is within the bound determined by the selection ratio and the relative difference of mortality rates of the classes with good health. Whether $B_2(0) > W_2(0)$ or $B_2(0) < W_2(0)$, will be determined when we qualitatively analyze the second condition on the existence of crossover. For the second condition to hold, i.e. for $\mu_B^* < \mu_W^*$, we need the following lemma in order to establish a bound in terms of the differential selection and the system parameters.

Lemma 2 *If $x > 0$ and $y > 0$, then $x - y < \frac{x^2 - y^2}{2y}$.*

Proof: There are two cases to consider:

1. If $x > y > 0$, then we have

$$2y \;<\; x+y \qquad\qquad \text{since } x > y$$

Multiply both sides of the inequality by $(x-y)$ and divide by $2y$ to get

$$x-y \;<\; \frac{x^2-y^2}{2y} \quad \text{since } x > y > 0$$

2. If $x < y < 0$, then we have

$$1 \;>\; \frac{y+x}{2y} \qquad \text{since } y > x$$

Multiply both sides of the inequality by $(y-x)$ to get

$$y-x \;>\; \frac{y^2-x^2}{2y} \quad \text{since } y > x > 0$$

Multiply both sides of the inequality by -1 to get

$$x-y \;<\; \frac{x^2-y^2}{2y}$$

■

Let λ_B and λ_W denote the dominant eigenvalues of the dynamical systems represented by B and W respectively; $f = a_{21} - a_{12}$ be the differential flow rate; and $\Delta_B = \sqrt{(S_B - f)^2 + 4a_{12}a_{21}}$ and $\Delta_W = \sqrt{(S_W - f)^2 + 4a_{12}a_{21}}$. Then, for $\mu_B^* < \mu_W^*$, by Theorem 4 we have

$$-\lambda_B \;<\; -\lambda_W$$

which implies

$$\mu_2 + \mu_1 - \sqrt{(S_B - f)^2 + 4a_{12}a_{21}} \;<\; m_2 + m_1 - \sqrt{(S_W - f)^2 + 4a_{12}a_{21}}$$

or equivalently

$$\begin{aligned}
\mu_2 + \mu_1 - m_2 - m_1 &< \Delta_B - \Delta_W \\
&< \frac{\Delta_B^2 - \Delta_W^2}{2\Delta_W} \qquad \text{by Lemma 2} \\
&= \frac{1}{2\Delta_W}\left((S_B - f)^2 + 4a_{12}a_{21} - (S_W - f)^2 - 4a_{12}a_{21}\right) \\
&= \frac{1}{2\Delta_W}\left(S_B^2 - S_W^2 - 2f(S_B - S_W)\right)
\end{aligned}$$

$$= \frac{1}{2\Delta_W}((S_B - S_W)(S_B + S_W) - 2f(S_B - S_W))$$
$$= \frac{1}{2\Delta_W}(S_B - S_W)(S_B + S_W - 2f) \quad (3.43)$$

Given the system parameters, inequality 3.43 can be used to numerically verify the relationship between the selection terms and the system paramcters. However, to interpret the second condition for the existence of crossover in terms of simple relationships between the class-specific mortality rates of the two populations, we need to qualitatively analyze all possible cases. This task can be carried out by analyzing the differential selection. Hence, there are two cases to consider:

1. If $S_B = \mu_2 - \mu_1 > S_W = m_2 - m_1$, that is, if the rate of selection for B is higher than the rate of selection for W, then it must be that $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$, since the flow rates are the same (see equations 3.40 and 3.41). Under stronger selection in B, there are two possible cases that can occur:

   (a) $\mu_2 > m_2$ but $\mu_1 < m_1$. The opposite cannot happen since it would contradict the assumption that $S_B > S_W$.

   (b) $\mu_i < m_i$ for $i = 1, 2$. The opposite cannot happen since then $\mu_B^* > \mu_W^*$ and this would contradict the assumption that $\mu_B^* < \mu_W^*$.

   We are now prepared to determine whether $B_2(0) > W_2(0)$ or $B_2(0) < W_2(0)$. The following lemma establishes that if condition 1(a) is satisfied, then $B_2(0) < W_2(0)$. For condition 1(b), one must resort to inequality 3.42 to verify the relationship numerically, since either $B_2(0) > W_2(0)$ or $B_2(0) < W_2(0)$ can hold.

Lemma 3 *If $S_B > S_W$; $\mu_1 < m_1$; and $\mu_2 > m_2$, then $B_2(0) < W_2(0)$.*

Proof: Suppose for contradiction that $B_2(0) > W_2(0)$. Then this would imply that $B_1(0) < W_1(0)$. Therefore we have the following inequalities

$$\begin{cases} \mu_2 B_2(0) > m_2 W_2(0) \\ \mu_1 B_1(0) < m_1 W_1(0) \end{cases}$$

since $\mu_2 > m_2$ and $\mu_1 < m_1$ by the assumptions in the lemma. Multiply the second inequality by -1 to get

$$\begin{cases} \mu_2 B_2(0) > m_2 W_2(0) \\ -\mu_1 B_1(0) > -m_1 W_1(0) \end{cases}$$

Subtract the second inequality from the first to get

$$\mu_1 B_1(0) + \mu_2 B_2(0) > m_1 W_1(0) + m_2 W_2(0)$$

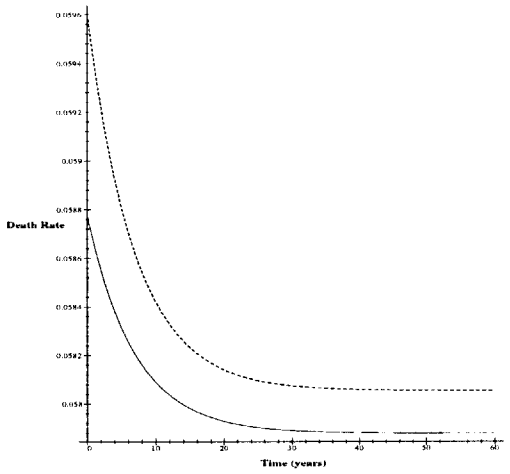


Figure 3-18: Comparing the mortality experience of two hypothetical populations in the time-domain, where one population (solid curve) has a higher selection rate. Both populations have similar mobility rates, but they differ in their class-specific mortality rates. The initial conditions are: $B_1(0) = 0.315$; $B_2(0) = 0.685$; $W_1(0) = 0.3$; and $W_2(0) = 0.7$. The system parameters are: $a_{12} = 0.04$; $a_{21} = 0.13$; $\mu_1 = 0.01$; $\mu_2 = 0.0812$; $m_1 = 0.012$; and $m_2 = 0.08$.

or equivalently

$$\mu_B(0) > \mu_W(0)$$

But this contradicts the assumption that B has a lower initial death rate than W, i.e. $\mu_B(0) < \mu_W(0)$. Therefore it must be that $B_2(0) < W_2(0)$. ■

Rewriting the result of Lemma 3 in terms of the ratio of the initial conditions, we get

$$\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)}$$

Figure 3-18 demonstrates an example of the case when $\mu_2 > m_2$ and $\mu_1 < m_1$.

2. If $S_B = \mu_2 - \mu_1 < S_W = m_2 - m_1$, that is, if the rate of selection for B is lower than the rate of selection for W, then it must be that $\left(\frac{B_1}{B_2}\right)^* < \left(\frac{W_1}{W_2}\right)^*$, since the flow rates are the same (see equations 3.40 and 3.41). Under a weaker selection in B, there are two possible cases that can occur:

   (a) $\mu_1 > m_1$ but $\mu_2 < m_2$. The opposite cannot happen since it would contradict the assumption that $S_W > S_B$.

(b) $\mu_i < m_i$ for $i = 1, 2$. The opposite cannot happen since then $\mu_B^* > \mu_W^*$ and this would contradict the assumption that $\mu_B^* < \mu_W^*$.

The following lemma establishes that if condition 2(a) is satisfied, then $B_2(0) > W_2(0)$. For condition 2(b), one must resort to inequality 3.42 to verify the relationship numerically, since either $B_2(0) > W_2(0)$ or $B_2(0) < W_2(0)$ can hold.

Lemma 4 *If $S_B < S_W$; $\mu_1 > m_1$; and $\mu_2 < m_2$, then $B_2(0) > W_2(0)$.*

Proof: Suppose for contradiction that $B_2(0) < W_2(0)$. Then this would imply that $B_1(0) > W_1(0)$. Therefore we have the following inequalities

$$\begin{cases} \mu_1 B_1(0) > m_1 W_1(0) \\ \mu_2 B_2(0) < m_2 W_2(0) \end{cases}$$

since $\mu_1 > m_1$ and $\mu_2 < m_2$ by the assumptions in the lemma. Multiply the second inequality by -1 to get

$$\begin{cases} \mu_1 B_1(0) > m_1 W_1(0) \\ -\mu_2 B_2(0) > -m_2 W_2(0) \end{cases}$$

Subtract the second inequality from the first to get

$$\mu_1 B_1(0) + \mu_2 B_2(0) > m_1 W_1(0) + m_2 W_2(0)$$

or equivalently

$$\mu_B(0) \quad > \quad \mu_W(0)$$

But this contradicts the assumption that B has a lower initial mortality rate than W, i.e. $\mu_B(0) < \mu_W(0)$. Therefore it must be that $B_2(0) > W_2(0)$. ■

Rewriting the result of Lemma 4 in terms of the ratio of the initial conditions, we get

$$\frac{B_1(0)}{B_2(0)} \quad < \quad \frac{W_1(0)}{W_2(0)}$$

Figure 3-19 demonstrates an example of the case when $\mu_2 < m_2$ and $\mu_1 > m_1$.

Putting it all together, under the assumption that B and W only differ in their class-specific mortality rates, for $A_\times$ to exist, or for the oldest cohort in B to have consistently lower death rate than W over time, one of the following must hold:

1. Either both classes in B have lower death rates than the classes in W, irrespective of differential selection or differential initial conditions, as long as inequality 3.42 is satisfied.

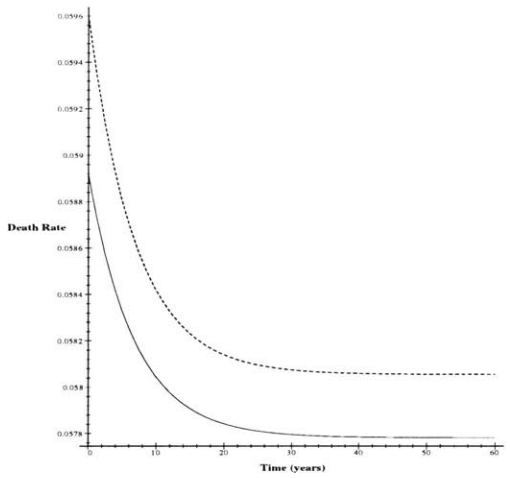


Figure 3-19: Comparing the mortality experience of two hypothetical populations in the time-domain, where one population (dashed curve) has a higher selection rate. Both populations have similar mobility rates, but they differ in their class-specific mortality rates. The initial conditions are: $B_1(0) = 0.29$; $B_2(0) = 0.71$; $W_1(0) = 0.3$; and $W_2(0) = 0.7$. The system parameters are: $a_{12} = 0.04$; $a_{21} = 0.13$; $\mu_1 = 0.0171$; $\mu_2 = 0.076$; $m_1 = 0.012$; and $m_2 = 0.08$.

2. Or the outcome depends on the differential selection:

   (a) If $S_B > S_W$, i.e. B has a higher selection rate than W, then $\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)}$, and the class with poor health in B is experiencing higher mortality than its counterpart in W. In other words, under stronger selection in B, although people in the class with poor health are dying at a faster rate than their counterparts in W, but since a higher proportion of people initially resided in the class with good health in B, and enjoyed a lower rate of mortality than their counterparts in W, then on average in the long run, B attains a lower death rate. This portrays a situation, where the mortality selectivity is high in B so that B is a population heterogeneous in health and death.

   (b) If $S_B < S_W$, i.e. B has a lower selection rate than W, then $\frac{B_1(0)}{B_2(0)} < \frac{W_1(0)}{W_2(0)}$, and the class with good health in B is experiencing higher mortality than its counterpart in W. In other words, under a weaker selection in B, since people in the class with poor health in B are not as badly off as their counterparts in W, and since people in the class with good health in B are dying faster than their counterparts in W, then in the long run B exhibits a lower death rate. This portrays a situation, where B is a population more homogeneous in health and death.

3.4.5 From Steady State to Transience: Conditions for Change in Crossover

In the time-domain, a change in the age-at-crossover to an older age can be detected as follows: for any age group, if the mortality curves of B and W cross conditioned upon that B has lower death rates prior to the year at intersection $Y_\times$, and higher death rates thereafter, it would be indicative of a change in $A_\times$. For the age group exhibiting this phenomenon, this would mean that $A_\times$ was in the lower half of that age group prior to $Y_\times$; and $A_\times$ increased to the upper half of the age group after $Y_\times$ (see Figure 3-16(b)). The opposite would imply a decline in $A_\times$. In this section, we investigate the conditions for change under the assumption that $A_\times$ for B and W can only increase over time, as it has been evident from the Black and White mortality data thus far. Such a correspondence enforces the following conditions on the initial and final mortality of the cohort exhibiting the crossover:

1. The initial mortality rate for B must be lower than the initial mortality rate for W. In other words, $\mu_B(0) < \mu_W(0)$.

2. The final mortality rate for B, however, must be higher than that for W. In other words, $\mu_B^* > \mu_W^*$.

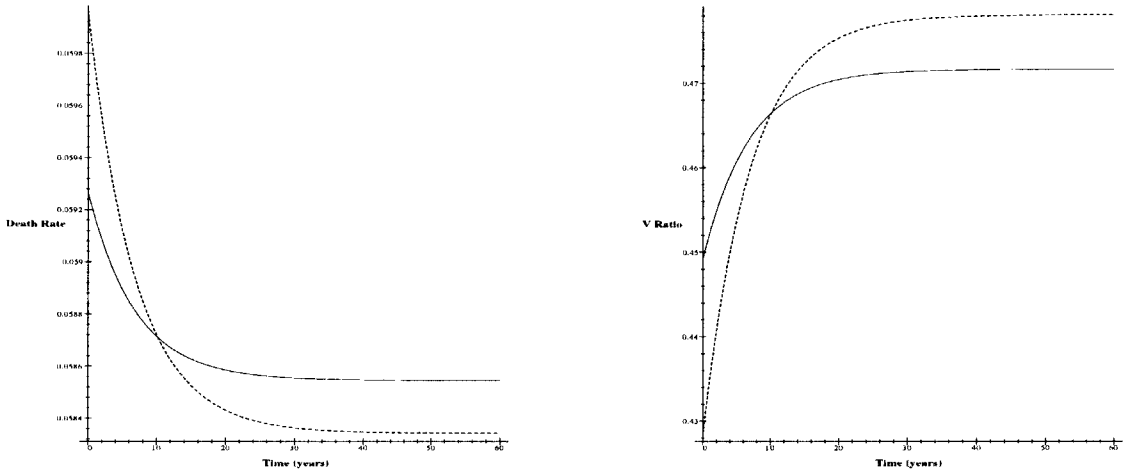
Clearly, by Lemma 1 there cannot be a double-crossing in the mortality curves of B and W. Therefore, the above two conditions are necessary and sufficient to imply a crossover at time $Y_\times$. One remarkable implication of conditions 1 and 2 is that the underlying processes governing the dynamics of crossover can be inferred from the steady state solution of the population mortality. In subsequent sections, we investigate the conditions for change in $A_\times$ under two different assumptions: (1) when class-specific mortality rates μ_1, μ_2 are the same for both populations but the flow rates are different; (2) when the flow rates a_{12}, a_{21} are the same for both populations but the class-specific mortality rates are different.

Different Flow Rates; Same Mortality Rates

Suppose B and W have the same class-specific mortality rates μ_1 and μ_2, but they differ in their mobility rates. As before, let a_{12}, a_{21} be the flow rates in B, and A_{12}, A_{21} be those in W. If μ_1 and μ_2 are the same for both populations B and W, then necessarily μ_B and μ_W cross at the same time as when $\frac{B_1}{B_2}$ and $\frac{W_1}{W_2}$ cross (see equation 3.10 for μ). Therefore, it is possible to infer the conditions for the crossing of the mortality curves, from the conditions for the crossing of the vulnerability-ratio curves. Figure 3-20 demonstrates an example of this situation, where the dynamics of mortality (Figure 3-20(a)) can be inferred from the dynamics of vulnerability-ratio (Figure 3-20(b)). Note that $Y_\times$ is the same in both figures.

If μ_1 and μ_2 are the same for both populations, then the first condition for change in $A_\times$ will be similar to condition 3.38. Therefore we have

$$\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)}$$



(a) (b)

Figure 3-20: Inferring the conditions for change in the age-at-crossover for two hypothetical populations in the time-domain, from the conditions for crossing of their vulnerability-ratio curves. Both populations have similar class-specific mortality rates but they differ in their mobility rates. The initial conditions are: $B_1(0) = 0.31$; $B_2(0) = 0.69$; $W_1(0) = 0.3$; and $W_2(0) = 0.7$. The system parameters are: $a_{12} = 0.0396$; $a_{21} = 0.1305$; $A_{12} = 0.04$; $A_{21} = 0.13$; $\mu_1 = 0.012$; $\mu_2 = 0.0805$. The mobility ratios are: $\frac{a_{12}}{a_{21}} \approx 0.303$ and $\frac{A_{12}}{A_{21}} \approx 0.308$. (a) Comparing the dynamics of mortality; (b) Comparing the dynamics of the vulnerability-ratio. Note that $Y_\times = 10$ years in both figures.

This means that the initial vulnerability-ratio for the cohort containing $A_\times$ is higher for B than for W. This is a rather intuitive outcome: lower initial death rate for the cohort containing $A_\times$ in B, is due to having a higher proportion of people of that age group residing in the low-vulnerability class of B compared to that of W.

Note that under the assumption that only the flow rates differ across the two populations, the second condition for change in $A_\times$ is the opposite of the second condition for the existence of $A_\times$ (see section 3.4.4). Therefore to express the second condition for change in $A_\times$, in terms of the flow rates and the initial conditions, we need to reverse the conclusive remark pertaining to the final mortality and redirect it to the age group containing $A_\times$ instead of the oldest cohort. Hence, we conclude that under the assumption that the flow rates are different across the two populations, for B and W to cross, or for $A_\times$ to be a dynamic phenomenon, the following must hold:

1. $\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)}$ for the cohort containing $A_\times$. This implies that initially B had a higher proportion of individuals of that age group residing in the class with good health, compared to that in W.

2. The mobility-ratio for the cohort containing $A_\times$ in B, must be lower than that in W. In other words, $\frac{a_{12}}{a_{21}} < \frac{A_{12}}{A_{21}}$ indicates that the relative rate of mobility to

the class with good health is better facilitated for the individuals of that age group in W than those in B.

Putting it together, for the cohort exhibiting the crossover phenomenon, initially B must have had a higher proportion living in the class with better health relative to W; thus B enjoyed a lower initial death rate. But then as time progressed, due to a lower mobility-ratio in B compared to W, the dynamics of vulnerability changed so that B has attained a higher death rate in the long run.

Different Mortality Rates; Same Flow Rates

Now suppose that the flow rates are the same for both populations, but they differ in their class-specific mortality rates. As before, let μ_1, μ_2 be the class-specific mortality rates for B, and m_1, m_2 be those of W. Under this assumption, when μ_B and μ_W cross, $\frac{B_1}{B_2}$ and $\frac{W_1}{W_2}$ may cross at a different point in time or may not cross at all. Hence, we can no longer infer the dynamics of mortality from the dynamics of vulnerability-ratio. Figure 3-21 demonstrates an example of a situation in which the mortality curves of two populations cross but their vulnerability-ratio curves do not. In contrast, Figure 3-22 exemplifies a situation when both the mortality and the vulnerability-ratio curves of two populations cross but the time at intersection is different for each.

For the first condition on change in $A_\times$ to hold, inequality 3.42 must be satisfied. Whether $B_2(0) > W_2(0)$ or $B_2(0) < W_2(0)$, will be determined when we qualitatively analyze the second condition for change in $A_\times$. For the second condition to hold, i.e. for $\mu_B^* > \mu_W^*$, exactly the opposite of inequality 3.43 must be satisfied. That is, we have

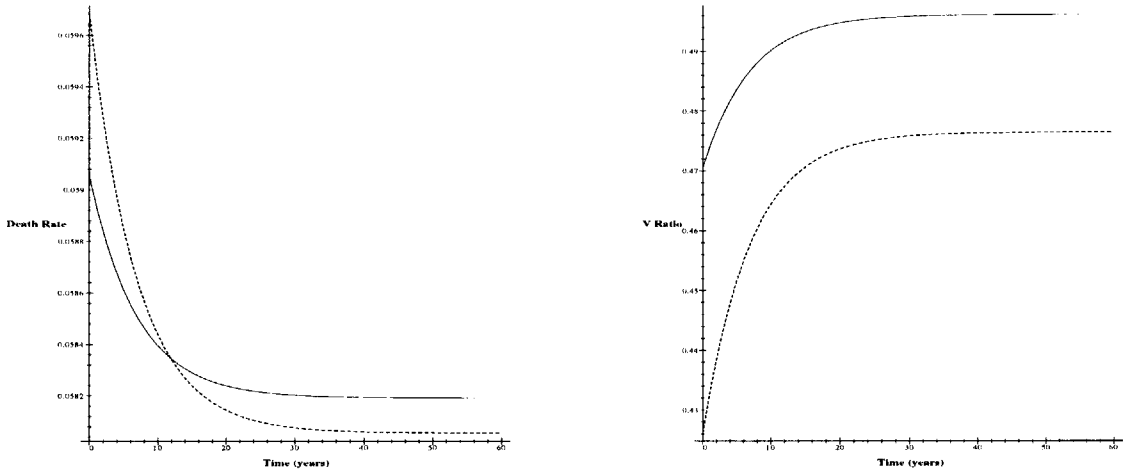
$$\mu_2 + \mu_1 - m_2 - m_1 \;>\; \frac{1}{2\Delta_W}(S_B - S_W)(S_B + S_W - 2f) \tag{3.44}$$

Given the system parameters, inequality 3.44 can be used to numerically verify the relationship between the selection terms and the system parameters. However, to interpret the second condition for change in the age-at-crossover in terms of simple relationships between the class-specific mortality rates of the two populations, we need to qualitatively analyze all possible cases. This can be carried out by analyzing the differential selection. Hence, there are two cases to consider:

1. If $S_B = \mu_2 - \mu_1 > S_W = m_2 - m_1$, that is, if the rate of selection for B is higher than the rate of selection for W, then it must be that $\left(\frac{B_1}{B_2}\right)^* > \left(\frac{W_1}{W_2}\right)^*$, since the flow rates are the same (see equations 3.40 and 3.41). Under stronger selection in B, there are two possible cases that can occur:

   (a) $\mu_2 > m_2$ but $\mu_1 < m_1$. The opposite cannot happen since it would contradict the assumption that $S_B > S_W$.

   (b) $\mu_i > m_i$ for $i = 1, 2$. The opposite cannot happen since then $\mu_B^* < \mu_W^*$ and this would contradict the assumption that $\mu_B^* > \mu_W^*$.



(a) (b)

Figure 3-21: (a) Detecting change in $A_\times$ by comparing the mortality experience of two hypothetical populations in the time-domain; (b) Comparing the dynamics of their vulnerability-ratio. Here, one population (solid curve) has a higher selection rate. Both populations have similar mobility rates, but they differ in their class-specific mortality rates. The initial conditions are: $B_1(0) = 0.32$; $B_2(0) = 0.68$; $W_1(0) = 0.29882$; and $W_2(0) = 0.70118$. The system parameters are: $a_{12} = A_{12} = 0.042$; $a_{21} = A_{21} = 0.13$; $a_{12} = A_{12} = 0.04$; $\mu_1 = 0.0088$; $\mu_2 = 0.0827$; $m_1 = 0.012$; and $m_2 = 0.08$.

By Lemma 3, condition 1(a) further requires that $B_2(0) < W_2(0)$. For condition 1(b), one must resort to inequality 3.42 to verify the relationship numerically, since either $B_2(0) > W_2(0)$ or $B_2(0) < W_2(0)$ can hold. Putting it in terms of the ratio of the initial conditions, the result of Lemma 3 can also be stated as follows

$$\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)}$$

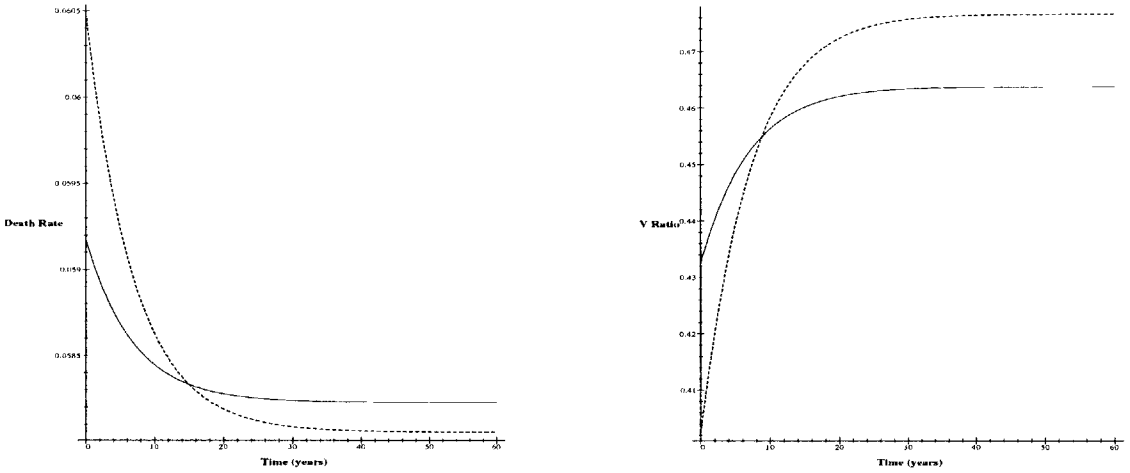
Figure 3-21 demonstrates an example of the case when $\mu_2 > m_2$ and $\mu_1 < m_1$.

2. If $S_B = \mu_2 - \mu_1 < S_W = m_2 - m_1$, that is, if the rate of selection for B is lower than the rate of selection for W, then it must be that $\left(\frac{B_1}{B_2}\right)^* < \left(\frac{W_1}{W_2}\right)^*$. Under a weaker selection in B, there are two possible cases that can occur:

   (a) $\mu_1 > m_1$ but $\mu_2 < m_2$. The opposite cannot happen since it would contradict the assumption that $S_W > S_B$.

   (b) $\mu_i > m_i$ for $i = 1, 2$. The opposite cannot happen since then $\mu_B^* < \mu_W^*$ and this would contradict the assumption that $\mu_B^* > \mu_W^*$.

   By Lemma 4, condition 1(a) further requires that $B_2(0) > W_2(0)$. For condition



(a) (b)

Figure 3-22: (a) Detecting change in $A_\times$ by comparing the mortality experience of two hypothetical populations in the time-domain; (b) Comparing the dynamics of their vulnerability-ratio. Here, one population (solid curve) has a lower selection rate. Both populations have similar mobility rates, but they differ in their class-specific mortality rates. The initial conditions are: $B_1(0) = 0.302$; $B_2(0) = 0.698$; $W_1(0) = 0.287$; and $W_2(0) = 0.713$. The system parameters are: $a_{12} = 0.042$; $a_{21} = A_{21} = 0.13$; $a_{12} = A_{12} = 0.04$; $\mu_1 = 0.0145$; $\mu_2 = 0.0785$; $m_1 = 0.012$; and $m_2 = 0.08$. Note that $Y_\times$ is different in each figure, thus $Y_\times$ in the mortality curves cannot be inferred from $Y_\times$ in the vulnerability-ratio curves.

1(b), one must resort to inequality 3.42 to verify the relationship numerically, since either $B_2(0) > W_2(0)$ or $B_2(0) < W_2(0)$ can hold. Putting it in terms of the ratio of the initial conditions, the result of Lemma 4 can also be stated as follows

$$\frac{B_1(0)}{B_2(0)} < \frac{W_1(0)}{W_2(0)}$$

Figure 3-22 demonstrates an example of the case when $\mu_2 < m_2$ and $\mu_1 > m_1$.

Putting it all together, under the assumption that B and W only differ in their class-specific mortality rates, for $A_\times$ to be a dynamic phenomenon (or for B to have a higher final death rate, while having a lower initial death rate), one of the following must hold:

1. Either both classes in B have higher death rates than the classes in W, irrespective of differential selection or differential initial conditions, as long as inequality 3.42 is satisfied.

2. Or the outcome depends on the differential selection:

(a) If $S_B > S_W$, i.e. B has a higher selection rate than W, then $\frac{B_1(0)}{B_2(0)} > \frac{W_1(0)}{W_2(0)}$, and the class with poor health in B is experiencing a higher mortality than its counterpart in W. In other words, under a stronger selection in B, although a higher proportion of people initially resided in the class with good health in B, and enjoyed a lower rate of mortality than their counterparts in W, but since people in the class with poor health are dying at a faster rate, relative to W, in the long run B attains a higher death rate. This portrays the B population in the age group containing $A_\times$, as a population heterogeneous in health.

(b) If $S_B < S_W$, i.e. B has a lower selection rate than W, then $\frac{B_1(0)}{B_2(0)} < \frac{W_1(0)}{W_2(0)}$, and the class with good health in B is experiencing a higher mortality than its counterpart in W. In other words, under a weaker selection in B, since people in the class with good health in B are worse off than their counterparts in W, even though people in the class with poor health in B are dying at a slower rate than their counterparts in W, in the long run B exhibits a higher death rate. This portrays the B population in the age group containing $A_\times$, as a population "almost" homogeneous in health.

3.4.6 The Black and White Finale: Putting It All Together

If the Black and White age-at-crossover has indeed been steadily rising over time, and if the age-at-crossover would increase at a faster rate than the average life span, then one could only deduce that the crossover phenomenon must eventually disappear. Therefore, we ask: *What are the implications of such diminishing effects*?

Using our abstract model of vulnerability, we investigated the conditions for the existence, as well as the conditions governing the dynamics, of the mortality crossover for the Black and White populations, in sections 3.4.4 and 3.4.5. Assuming that the class-specific mortality rates were the same for Blacks and Whites, we were able to examine the effect of social processes, acting through the flow rates, as potential underlying causes of the crossover and its dynamics. On the other hand, assuming that the mobility rates were the same for Blacks and Whites, we were able to examine the effect of differential selection as the underlying cause of the crossover and its dynamics. The class of all mathematically plausible conditions, so derived, is now selectively reduced to a sub-class of conditions that are also socially sensible, given the real-life Black and White dynamics. Therefore, we conclude that:

1. If class-specific mortality rates μ_1, μ_2 are the same for both populations, then:

   (a) For the crossover phenomenon to exist within an age group, in all cohorts older than the one containing $A_\times$, the mobility-ratio for Blacks must be greater than that for Whites. This means that among the older Black population, movement to the class with good health must be better facilitated, compared to the White elderly. In other words, the social mobility gap is not as wide among the Black elderly and therefore there is a higher degree of homogeneity in the distribution of vulnerability to disease and

death among them; the polarization into distinct vulnerability classes has not yet extended to, or is not as well defined for, all age groups in B. Older aged population in B, is still enjoying a somewhat homogeneous social conditions of life, while the younger population is more heterogeneous in vulnerability to disease and death, and is exposed to a high rate of selection.

(b) For $A_\times$ to increase to an older age, the cohort containing $A_\times$ must exhibit a smaller mobility-ratio for Blacks than for Whites. This means that for all age groups younger than, and including the lower half of the age-group containing $A_\times$, it is harder for Blacks to move to the class with good heath than it is for Whites. In other words, as the entire Black population is aging, the polarization into distinct vulnerability classes extends to the age group containing $A_\times$. This means that a larger proportion of the Black population is now experiencing heterogeneity in the conditions of life, and is exposed to a greater force of selection.

2. If the mobility rates a_{12}, a_{21} are the same for both populations, then:

   (a) For the crossover phenomenon to exist within an age group, all cohorts older than the one containing $A_\times$, must either exhibit lower mortality rates in both classes of B than those in W, irrespective of differential selectivity; or the selection rate is weaker among the Blacks, with the class with good health experiencing a greater force of mortality in B compared to its counterpart in W. This means that vulnerability to death is more homogeneous among the Black elderly than the Whites of the same generation.

   (b) For $A_\times$ to increase to an older age, the cohort containing $A_\times$ must either exhibit a higher mortality in both classes in B than in W, irrespective of differential selectivity; or the selection rate is higher for Blacks than for Whites, with the class with poor health in B experiencing a greater force of mortality. This means that the younger Black population is heterogeneous in vulnerability to death, and class formation is extending to the older population.

To the extent that the Black population is not yet entirely divided into distinct classes of vulnerability, and the population is partially homogeneous in their vulnerability to disease and death, the age-at-crossover exists and is representative of inequality in the conditions of life between Blacks and Whites. As the Black population is further polarizing into different vulnerability classes by: further increase in differential mortality, or further reduction in social mobility, or both, as the population is aging, inequality in the conditions of health extends to older age groups, thereby increasing the age-at-crossover with the White population. Hence, we attribute the dynamics of mortality crossover, or in the case of Black versus White, the increase in the age-at-crossover over time, to the Black-White conditions of life becoming more similar in the sense that the gap between different vulnerability classes in B is increasing and such class-based vulnerability structure is extending to the older Black population.

As distinct vulnerability classes are forming among the younger Black population and extending to older generations, as the population ages as a whole, social and selective processes further modify the dynamics of inequality. These processes, although in the modern sense are new for Blacks, for the White population of the United States are perhaps as old as the Civil War.

3.4.7 Other Populations

Is the crossover phenomenon unique to the mortality dynamics of the Black and White populations? Or, are there other pairs of populations that exhibit a similar transient behavior in their mortality curves? From time to time, demographers have reported the discovery of the crossover phenomenon for different pairs of populations in different parts of the world. Spiegelman [21] reported an incident of the phenomenon when studying the age-specific mortality rates of Jews and non-Jews in Canada, and made a reference to similar findings for Jews and non-Jews in New York City and Berlin. For him Jews were the Whites and non-Jews were the Blacks. The age-at-crossover was observed to be at young adult ages for these populations. He consequently attributed the existence of the crossover to detrimental life style of Jews, particularly dietary. Nam [14] compared the age-specific mortality rates of several pairs of provinces in China for which the quality of data was regarded as exceptionally sound. He found convergence of mortality curves in all pairs; and the appearance of crossover in older age groups in a few. For an overview of discovery of mortality crossovers for different pairs of populations in different parts of the world, see [15, 16, 14].

In fact, if we have reasons to believe that the crossover phenomenon is merely a side effect of inequality in the mortality selectivity, as well as inequality in the socio-economic conditions of life between populations (as demonstrated above), one should expect to revisit the crossover phenomenon for different pairs of populations exhibiting dissimilar conditions of life. As the socio-economic gap increases in a population, as a population continues to lose homogeneity in conditions of health and death among its individuals, as the boundaries of class divisions in health and vulnerability are better defined and become more pronounced in a population, the age-at-crossover also continues to increase to older age groups. Therefore, we conjecture that the younger the process of polarization into different health classes in a population, the younger the age-at-crossover.

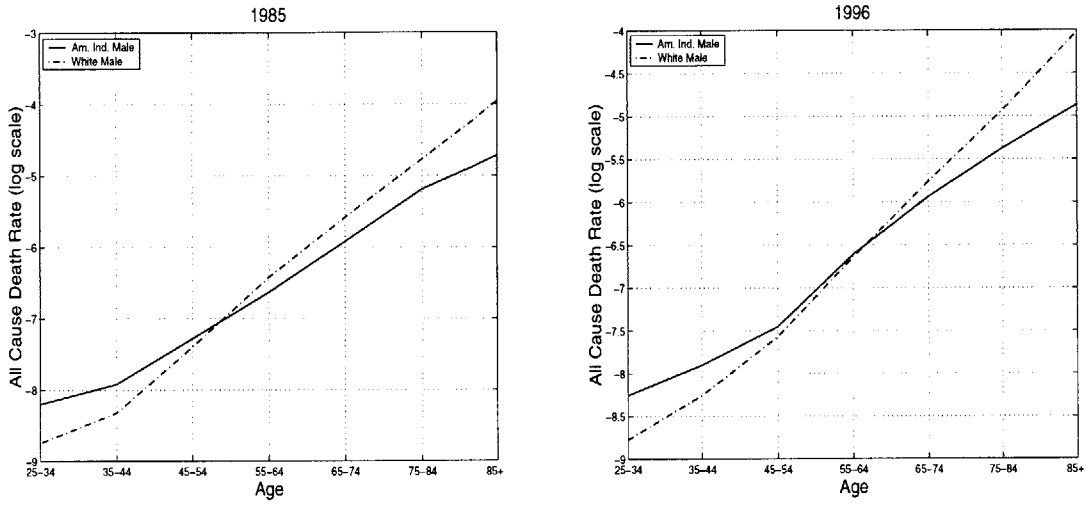
It is interesting to note that our interpretation of the direction of change in the age-at-crossover is even in contrast with the arguments made by one of the proponents of the selectivity process, as a potential cause of the crossover. Charles Nam, one of the prominent demographers arguing in favor of the existence of the crossover, writes: "The mortality convergence and crossover process is a dynamic one. As socio-economic characteristics of two populations become more alike, as frailty conditions of those populations become more alike, and as the quality of demographic data collected for the populations become more alike, mortality convergence and crossover will decline and eventually disappear. To the extent that those equalities are not obtained, mortality convergence will probably continue, and mortality crossover will probably occur at some age, however late." [14]. To date, we have not found any

evidence of prior examination of the dynamics of crossover in a non-speculative manner. Therefore, any proclamation that is not the least based on studying the trends in mortality data, should be considered at best speculative and open to investigation. Moreover, despite the controversial nature and significance of the Black-White crossover, the mortality trends for other pairs of populations in the United States, have never been studied. If in fact the crossover phenomenon can be observed for pairs of populations other than the Black-White pair, then our perceptions of the crossover phenomenon may have to be reexamined.

We have examined the above conjecture in light of limited data available for two other minority populations in the United States: the American-Indian and the Hispanic. Figure 3-23 demonstrates the dynamics of mortality crossover for the American-Indian and Alaskan Native male versus the White male over the years 1985 and 1996. In 1985, the age-at-crossover was in the upper half of the age group 45-54, whereas in 1996 it was well within the age group 55-64. Hence, in 11 years, $A_\times$ has increased by about 8 years (from ~ 53 to ~ 61). Compare that to the Black-White age-at-crossover for those years; in 1996, the Black-White age-at-crossover is about 24 years older than $A_\times$ for Native-White (see Figure 3-11). Figure 3-24 is the result of transforming the mortality data for American Indians from the age-domain to the time-domain in the manner presented in section 3.4.2. Figure 3-24(b) further illustrates that $A_\times$ passed the median age of 59.5 in 1995.

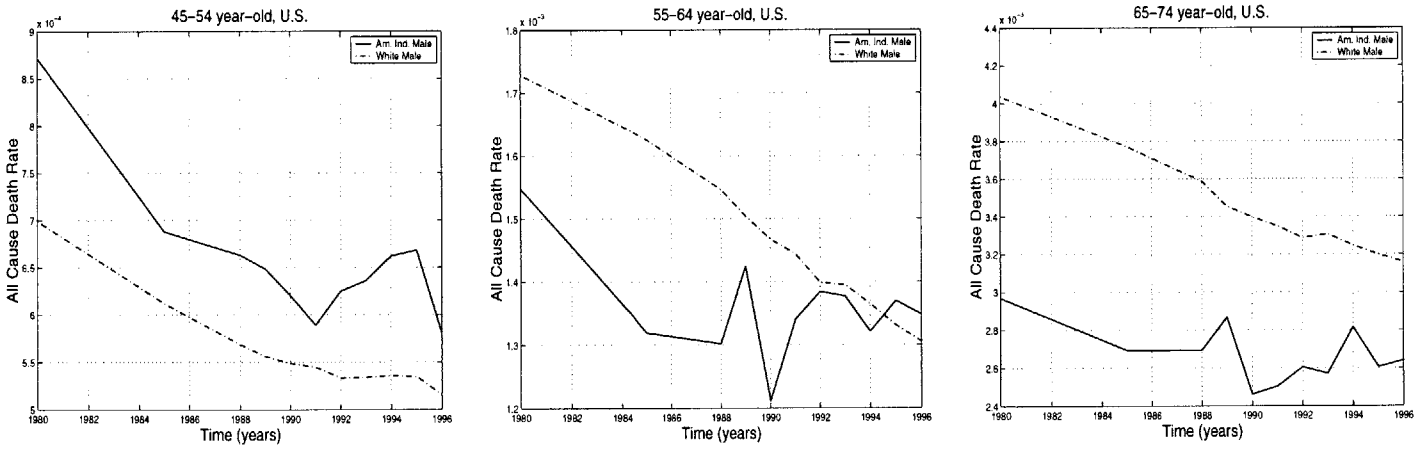
Similarly, examining the dynamics of mortality crossover in the age-domain for the Hispanic male versus the White male reveals that the phenomenon is still very young for this pair of populations. Figure 3-25 shows that in 1985, $A_\times \sim 45$ whereas in 1996, $A_\times$ seems to have passed the median age 49.5 of the age group 45-54. That is, in the period of 11 years, the age-at-crossover has increased only by about 5 years. The Hispanic-White age-at-crossover is not only younger than the age-at-crossover for Native-White by 10 years; it has also exhibited a much slower rate of growth during the same time period. An interesting dynamics is observed when we change the scope of the study to the female population. The age-at-crossover is even younger for the Hispanic-White female population. Figure 3-26 demonstrates that in 1985, the age-at-crossover was at the brink of entering the age group 5-14, whereas in 1996, it has entered the lower half of the next age group 15-24. Although the age-at-crossover is younger among the female population, but it seems to have had a faster rate of growth: in 11 years, it has increased by about 10 years, which is 2.5 times faster than the growth for the Hispanic-White male. However, compared to the dynamics of Black-White or Native-White, the Hispanic-White dynamics is rather new. Therefore, further investigation into the dynamics of the age-at-crossover for this pair of populations must await until more, and better quality, data is available.

Comparing the mortality trends of other pairs of populations is quite enlightening. In particular, the appearance of mortality crossovers for the Native-White and Hispanic-White should strengthen the doubt that mortality crossover is merely the result of erroneous data. If the Black-White mortality crossover is indeed hypothesized to be a side effect of age over-reporting among the Black elderly, then an alternative premise is needed to rationalize the appearance of mortality crossovers at very young ages for the Native-White and Hispanic-White.



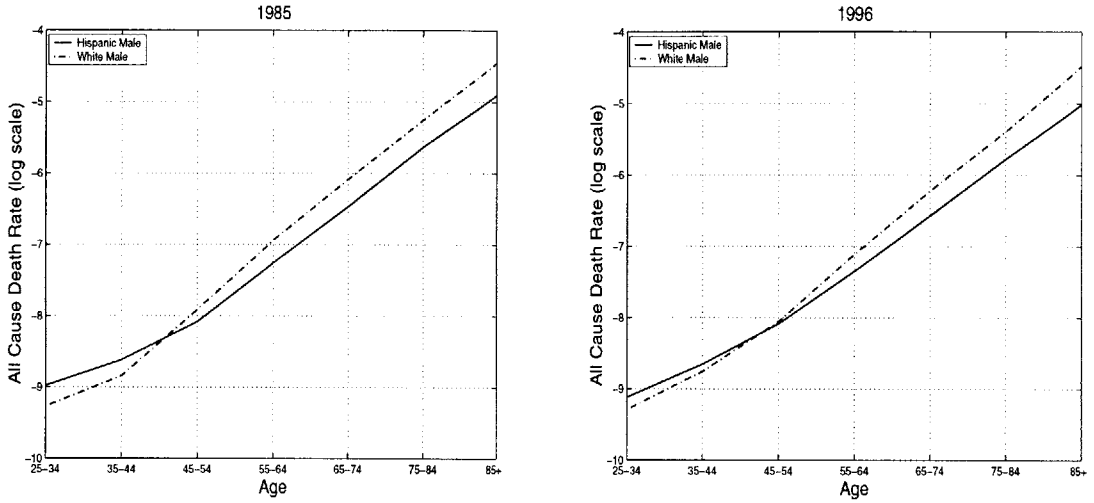
(a) (b)

Figure 3-23: Age-specific all-cause death rates; Native-American male versus White male, for the years 1985 and 1996. (a) 1985; (b) 1996. Source: Vital Statistics of the United States, 1985, 1996.



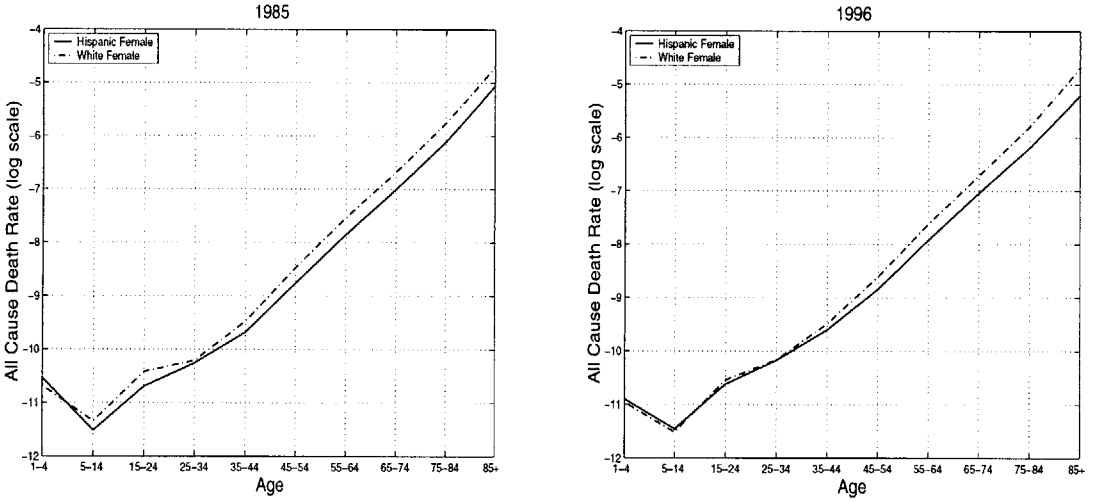
(a) (b) (c)

Figure 3-24: All cause longitudinal death rates; Native-American male versus White male. (a) cohort: 45-54 year-old; (b) cohort: 55-64 year-old; (c) cohort: 65-74 year-old. Source: Vital Statistics of the United States, 1980, 1985, 1988-96.



(a) (b)

Figure 3-25: Age-specific all-cause death rates; Hispanics male versus White male, for the years 1985 and 1996. (a) 1985; (b) 1996. Source: Vital Statistics of the United States, 1985, 1996.



(a) (b)

Figure 3-26: Age-specific all-cause death rates; Hispanics female versus White female, for the years 1985 and 1996. (a) 1985; (b) 1996. Source: Vital Statistics of the United States, 1985, 1996.

3.5 n-Vulnerability-Class: Generalization

Suppose that a population cohort is distributed among $n \geq 2$ vulnerability classes V_i, each associated with its own mortality rate μ_i, $i = 1 \ldots n$. The degree of vulnerability is assumed to increase with the subscript i; thus $\mu_n > \mu_{n-1} > \ldots > \mu_2 > \mu_1$. Furthermore, individuals in one class may move to another due to disease processes, aging, loss of fortune, or health-detrimental social policies, in which case the flow $a_{i+1,i}$, $i \leq n-1$, is to the adjacent higher vulnerability class; or due to factors affecting viability such as acquiring knowledge or wealth, or through health-improvement policies, in which case the flow $a_{i-1,i}$, $i \geq 2$ is to the adjacent lower vulnerability class. Since there is no birth in this model, at any point in time, we must have the condition that $\sum_{i=1}^{i} V_i \leq 1$. Further, all parameters of the system are assumed to be constant. The following linear system of n ordinary differential equations captures the dynamics of vulnerability of a population, heterogeneous in health and death.

$$\begin{aligned}
\frac{dV_1}{dt} &= -V_1(a_{21} + \mu_1) + a_{12}V_2 \\
\frac{dV_2}{dt} &= a_{21}V_1 - V_2(a_{12} + \mu_2 + a_{32}) + a_{23}V_3 \\
&\vdots \qquad\qquad \vdots \\
\frac{dV_n}{dt} &= a_{n,n-1}V_{n-1} - V_n(a_{n-1,n} + \mu_n)
\end{aligned} \tag{3.45}$$

Let A be the $n \times n$ coefficient matrix corresponding to system 3.45

$$A = \begin{pmatrix}
a_{11} & a_{12} & 0 & 0 & \ldots & 0 & 0 \\
a_{21} & a_{22} & a_{23} & 0 & \ldots & 0 & 0 \\
0 & a_{32} & a_{33} & a_{34} & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & 0 & a_{n-1,n-2} & a_{n-1,n-1} & a_{n-1,n} \\
0 & 0 & \ldots & 0 & 0 & a_{n,n-1} & a_{nn}
\end{pmatrix}$$

where the main diagonal entries a_{ii}, $i = 1 \cdots n$, are the self-damping terms: for V_1, $a_{11} = -a_{21} - \mu_1$; for V_n, $a_{nn} = -a_{n-1,n} - \mu_n$; and for all intermediate V_i, $i = 2 \cdots n-1$, $a_{ii} = -a_{i-1,i} - a_{i+1,i} - \mu_i$.

3.5.1 Major Results

Matrix A is a tridiagonal matrix. A tridiagonal matrix is defined as a matrix whose only nonzero entries occur on the main diagonal or on the two diagonals which are immediate neighbors of the main diagonal. The diagonals immediately below and above the main diagonal are referred to as the sub-diagonal and super-diagonal respectively. Tridiagonal matrices and their properties have been well studied. Based on such known properties, here we will generalize two of the major results for the

2-dimensional system, from section 3.2, to the n-dimensional system 3.45.

Lemma 5 *The eigenvalues of the $n \times n$ coefficient matrix A, are all real.*

Proof: Recall from elementary linear algebra, that a tridiagonal matrix is irreducible if and only if every sub-diagonal and super-diagonal element is nonzero. Furthermore, if a (real) matrix is tridiagonal and irreducible, then its eigenvalues are real and distinct. Hence, matrix A is irreducible and all its eigenvalues are real. ■

Theorem 5 *$\mu^* = -\lambda_d$, where μ^* is the steady state population mortality and λ_d is the dominant eigenvalue of the $n \times n$ coefficient matrix A.*

Proof: Recall from elementary linear algebra that if the eigenvalues of a matrix are distinct, then the matrix can be diagonalized. In the proof of Lemma 5, we already established that matrix A is irreducible and therefore all its eigenvalues are distinct. Hence, matrix A must be diagonalizable. This means that A can be written as

$$A = U \Lambda U^{-1} \tag{3.46}$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

is an $n \times n$ matrix of n distinct eigenvalues of A; U is an $n \times n$ matrix whose columns are the eigenvectors of A; and U^{-1} is its inverse. Expressed in vector notation, the n-dimensional dynamical system 3.45 can be written as follows

$$A\,V = \dot{V}$$

where

$$V = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix}$$

is the $n \times 1$ column vector of n vulnerability classes; and

$$\dot{V} = \begin{pmatrix} \dot{V}_1 \\ \dot{V}_2 \\ \vdots \\ \dot{V}_n \end{pmatrix}$$

is the $n \times 1$ column vector of the first derivatives of n vulnerability classes. The solution of the system can also be written in vector notation as follows

$$V = e^{At} V(0) \tag{3.47}$$

where

$$V(0) = \begin{pmatrix} V_1(0) \\ V_2(0) \\ \vdots \\ V_n(0) \end{pmatrix}$$

is an $n \times 1$ column vector of the initial conditions; and

$$e^{At} = U \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n t} \end{pmatrix} U^{-1} \qquad \text{since } A \text{ is diagonalizable} \tag{3.48}$$

Without loss of generality, suppose $\lambda_1 = \lambda_d$ is the dominant eigenvalue for the coefficient matrix A. Then, at the steady state, when $e^{\lambda_1 t} \gg e^{\lambda_i t}$, $\forall i \geq 2$, equation 3.47 can be rewritten as follows

$$V = U \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} U^{-1} V(0) \tag{3.49}$$

Now, note that the sum of n ODE's of the dynamical system 3.45 is the negative of the weighted average of all V_i's, $i = 1, \cdots n$. That is

$$\sum_{i=1}^{n} \frac{dV_i}{dt} = -\sum_{i=1}^{n} \mu_i V_i \tag{3.50}$$

To write equation 3.50 in matrix notation, let

$$\Pi = [\mu_1, \mu_2, \cdots, \mu_n]$$

be an $1 \times n$ row vector of class-specific mortality rates. Then equation 3.50 can be rewritten in vector notation as follows

$$-\Pi V = \mathbf{1} A V$$

where $\mathbf{1}=\overbrace{[1,1,\cdots,1]}^{n}$ is an $1 \times n$ row vector of 1's. Expanding A and V as in equations 3.46 and 3.47 we get

$$-\Pi\, V \;=\; \mathbf{1}\, U\, \Lambda\, \underbrace{U^{-1}\, U}_{\mathbf{I}} \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} U^{-1}\, V(0)$$

where $\mathbf{I}$ is an $n \times n$ identity matrix. Write Λ in its explicit matrix form to get

$$\begin{aligned} -\Pi\, V &= \mathbf{1}\, U \underbrace{\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}}_{\Lambda} \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} U^{-1}\, V(0) \\ &= \mathbf{1}\, U \begin{pmatrix} \lambda_1 e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} U^{-1}\, V(0) \\ &= \lambda_1\, \mathbf{1}\, \underbrace{U \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} U^{-1}\, V(0)}_{V} \qquad \text{since } \lambda_1 \text{ is a scalar} \\ &= \lambda_1\, \mathbf{1}\, V \end{aligned} \tag{3.51}$$

In equation 3.51, λ_1 is a scalar; $\mathbf{1}=\overbrace{[1,1,\cdots,1]}^{n}$; V is an $n \times 1$ column vector; and Π is an $1 \times n$ row vector. Multiply through on each side of the equation to reduce each side to its scalar form, and reset $\lambda_1 = \lambda_d$, to get

$$\lambda_d\, [1,1,\cdots,1] \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix} \;=\; -[\mu_1, \mu_2, \cdots, \mu_n] \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix} \tag{3.52}$$

or equivalently

$$\lambda_d \sum_{i=1}^{n} V_i \;=\; -\sum_{i=1}^{n} \mu_i V_i$$

which in turn implies

$$\lambda_d = -\frac{\sum_{i=1}^{n} \mu_i V_i}{\sum_{i=1}^{n} V_i}$$

Divide both the numerator and denominator of the right-hand-side of the above equation by V_1 to get the steady state vulnerability-ratios. Then we have

$$\begin{aligned} &= -\frac{\sum_{i=1}^{n} \mu_i \left(\frac{V_i}{V_1}\right)^*}{\sum_{i=1}^{n} \left(\frac{V_i}{V_1}\right)^*} \\ &= -\mu^* \qquad \text{by definition} \end{aligned}$$ ■

3.5.2 Qualitative Complexity of the General Case

Recall from section 3.2.1 that we transformed a system of two ODE's into one quadratic ODE describing the rate of change in the vulnerability-ratio $\frac{V_1}{V_2}$. Utilizing the ratio and its steady state value $\left(\frac{V_1}{V_2}\right)^*$, we then made inferences about the dynamics of population mortality μ and its steady state value μ^*. For the 3-dimensional dynamical system, we would need to know two such ratios (e.g. $\left(\frac{V_1}{V_2}\right)^*$ and $\left(\frac{V_3}{V_2}\right)^*$), the solution of each being dependent on the other. To demonstrate the complexity of the analysis involved, consider the 3×3 case. To arrive at an ODE for the ratio $\frac{V_1}{V_2}$, multiply $\dot{V}_1$ by V_2, and $\dot{V}_2$ by V_1, to get

$$\begin{aligned} \dot{V}_1 V_2 &= -V_1 V_2 (a_{21} + \mu_1) + a_{12} V_2^2 \\ \dot{V}_2 V_1 &= a_{21} V_1^2 - (a_{12} + \mu_2 + a_{32}) V_1 V_2 + a_{23} V_1 V_3 \end{aligned}$$

Subtract the second equation from the first and divide by ${V_2}^2$ to get

$$\frac{\dot{V}_1 V_2 - \dot{V}_2 V_1}{{V_2}^2} \stackrel{def}{=} \frac{d}{dt}\left(\frac{V_1}{V_2}\right) =$$
$$-a_{21}\left(\frac{V_1}{V_2}\right)^2 + (a_{12} + \mu_2 + a_{32} - a_{21} - \mu_1)\frac{V_1}{V_2} - a_{23}\left(\frac{V_1}{V_2}\right)\left(\frac{V_3}{V_2}\right) + a_{12}$$

Similarly, to arrive at an ODE for the ratio $\frac{V_3}{V_2}$, multiply $\dot{V}_2$ by V_3, and $\dot{V}_3$ by V_2, to get

$$\begin{aligned} \dot{V}_2 V_3 &= a_{21} V_1 V_3 - (a_{12} + \mu_2 + a_{32}) V_2 V_3 + a_{23} V_3^2 \\ \dot{V}_3 V_2 &= a_{32} V_2^2 - (a_{23} + \mu_3) V_2 V_3 \end{aligned}$$

Subtract the second equation from the first and divide by ${V_2}^2$ to get

$$\frac{\dot{V}_3 V_2 - \dot{V}_2 V_3}{{V_2}^2} \stackrel{def}{=} \frac{d}{dt}\left(\frac{V_3}{V_2}\right) =$$

$$-a_{23}\left(\frac{V_3}{V_2}\right)^2+(a_{12}+\mu_2+a_{32}-a_{23}-\mu_3)\frac{V_3}{V_2}-a_{21}\left(\frac{V_1}{V_2}\right)\left(\frac{V_3}{V_2}\right)+a_{32}$$

Define the population mortality μ as follows

$$\begin{aligned}\mu &= \frac{\mu_1 V_1+\mu_2 V_2+\mu_3 V_3}{V_1+V_2+V_3}\\ &= \frac{\mu_1\left(\frac{V_1}{V_2}\right)+\mu_2+\mu_3\left(\frac{V_3}{V_2}\right)}{\left(\frac{V_1}{V_2}\right)+1+\left(\frac{V_3}{V_2}\right)}\end{aligned}$$

and similarly the steady state population mortality μ^* as follows

$$\mu^* = \frac{\mu_1\left(\frac{V_1}{V_2}\right)^*+\mu_2+\mu_3\left(\frac{V_3}{V_2}\right)^*}{\left(\frac{V_1}{V_2}\right)^*+1+\left(\frac{V_3}{V_2}\right)^*}$$

Since by Theorem 5 $\mu^*=-\lambda_d$, then we must have

$$-\lambda_d = \frac{\mu_1\left(\frac{V_1}{V_2}\right)^*+\mu_2+\mu_3\left(\frac{V_3}{V_2}\right)^*}{\left(\frac{V_1}{V_2}\right)^*+1+\left(\frac{V_3}{V_2}\right)^*}$$

which implies

$$\mu_1\left(\frac{V_1}{V_2}\right)^*+\mu_2+\mu_3\left(\frac{V_3}{V_2}\right)^* = -\lambda_d\left(\left(\frac{V_1}{V_2}\right)^*+1+\left(\frac{V_3}{V_2}\right)^*\right)$$

or equivalently

$$\lambda_d+\mu_2 = -\left(\frac{V_1}{V_2}\right)^*(\lambda_d+\mu_1)-\left(\frac{V_3}{V_2}\right)^*(\lambda_d+\mu_3)$$

Solve for the steady state vulnerability ratios to get

$$\left(\frac{V_1}{V_2}\right)^* = \frac{(\lambda_d+\mu_2)+\left(\frac{V_3}{V_2}\right)^*(\lambda_d+\mu_3)}{-\lambda_d-\mu_1}$$

and

$$\left(\frac{V_3}{V_2}\right)^* = \frac{(\lambda_d+\mu_2)+\left(\frac{V_1}{V_2}\right)^*(\lambda_d+\mu_1)}{-\lambda_d-\mu_3}$$

Clearly, without making further assumptions on the relationship between the two ratios, we would be left with one equation and two unknowns. Similarly, for the n-dimensional case, we can transform n ODE's into $n-1$ ODE's for vulnerability-ratios, each dependent on the dominant eigenvalue and the remaining $n-2$ ratios. Although in the context of the population dynamics of vulnerability, it is reasonable to assume $n=2$ or $n=3$, the qualitative analysis carried out for $n=2$ in the previous sections,

cannot be repeated with ease even for $n = 3$. Therefore, the validity of all the results other than the ones proved in this section, i.e. Lemma 5 and Theorem 5, would have to be examined numerically for $n = 3$. To further demonstrate the difficulty of the qualitative analysis involved, consider the characteristic polynomial for $n = 3$

$$\begin{aligned} p(\lambda) &= \lambda^3 + \lambda^2(\mu_1 + \mu_2 + \mu_3 + a_{12} + a_{21} + a_{23} + a_{32}) \\ &\quad +\lambda(\mu_1\mu_2 + \mu_1\mu_3 + \mu_2\mu_3 + a_{12}a_{23} + a_{12}\mu_1 + a_{12}\mu_3 + a_{21}a_{32} + a_{21}\mu_2 + a_{21}\mu_3 \\ &\quad +a_{23}\mu_1 + a_{23}\mu_2 + a_{32}\mu_1 + a_{32}\mu_3 + a_{21}a_{23}) + (a_{21}\mu_2\mu_3 + a_{21}\mu_2a_{23} + a_{12}\mu_1\mu_3 \\ &\quad +\mu_1\mu_2\mu_3 + \mu_1\mu_2a_{23} + \mu_1\mu_3a_{32} + \mu_1a_{12}a_{23} + \mu_3a_{21}a_{32}) \end{aligned} \tag{3.53}$$

Although our dynamical system is linear and therefore in principle we are able to solve explicitly for V_i's, but in fact the characteristic polynomial is significantly "nonlinear". Compare the characteristic polynomial 3.3 for the 2-dimensional case, with the 3-dimensional characteristic polynomial 3.53. In polynomial 3.3, there are 7 possible combinations of 4 system parameters, 3 of which are of order 2; in polynomial 3.53, however, there are 29 possible combinations of 7 system parameters, 8 of which are of order 3 and 14 are of order 2. This means that as the number of vulnerability classes n grows, the number of system parameters grow, and therefore the roots of the characteristic polynomial not only become highly nonlinear, but the number of nonlinear terms constituting the eigenvalues become enormous. In the 2-dimensional case, the eigenvalues' nonlinearity is of order 2; the eigenvalues contain all possible parameter combinations of orders 1 and 2 (see equations 3.4 and 3.5). In the 3-dimensional case, the eigenvalues' nonlinearity is of order 6; the eigenvalues contain all possible parameter combinations of orders $1 \ldots 6$; thus the effect of a change, acting through any of the 7 system parameters, on the mortality-related outcomes may not be at all a straightforward task. Let m denote the number of parameters in a system of n vulnerability classes. Then, we hypothesize that the qualitative complexity of the analysis is $O(m^n)$.

3.5.3 Quantitative Results

We have numerically examined the validity of the results of the previous sections, for $n = 3, 4, 5$. All major results pertaining to the qualitative analysis of the 2-dimensional case, maintained their validity for $n \leq 5$. Focusing on the principal outcome of this chapter, namely the mortality crossover and its dynamics, if the results are indeed valid for $n \geq 2$, then this means that irrespective of the number of vulnerability classes, the existence of a mortality crossover for two populations is always an indication of inequality in the conditions of life, where one of the two intersecting populations, on average, is subjected to a harsher environment and a greater force of mortality. One such population experiences duality in the dynamics of vulnerability of the age structure of the population: the proportion of the population that is younger than the age-at-crossover is already divided into distinct vulnerability classes and is therefore "more" heterogeneous in health and death; the rest of the population that is older than the age-at-crossover, however, is "more" homogeneous

in vulnerability to disease and death. But then the population is aging, and with it, the age-at-crossover is changing. A change in the age-at-crossover to an older age is therefore an indication of formation and further extension of distinct vulnerability classes into the older generation of the population. This further increases the inequality in the conditions of life within one population; thus making the vulnerability structure of the two populations more similar.

3.6 Summary

We developed a dynamic model of selection partially offset by mobility, to mimic the inherent heterogeneity in the vulnerability structure of a population cohort. In the absence of birth or migration in the model, the population eventually becomes extinct, while the ratio of the frequency of vulnerability classes reaches a steady state. Therefore a transformation from a system of two linear ordinary differential equations, describing the rate of change in the frequency of each vulnerability class V_i, $i = 1, 2$, into a single quadratic ordinary differential equation, describing the rate of change in the vulnerability-ratio $\frac{V_1}{V_2}$, made it possible to infer the dynamics of the steady state mortality from the dynamics of the steady state vulnerability-ratio. We established qualitatively that:

1. As the class-specific mortality rates μ_1, μ_2 increase, the final population mortality μ^* also increases.

2. As the mortality of the class with good health μ_1 increases, the final population mortality μ^*, the final proportion dying in the class with good health μ_1^*, and the final proportion dying in the class with poor health μ_2^*, will all increase.

3. As the mortality for the class with poor health μ_2 increases,

   (a) the final population mortality μ^*, and the final proportion dying in the class with good health μ_1^*, also increase.

   (b) the final proportion dying in the class with poor health μ_2^*, however, increases to a peak at high values of μ_2, after which it slowly declines to a high steady state value. This is because at high values of μ_2, when selection is strong and against the high-vulnerability class V_2, almost everyone in the population dies early in the process. This causes the system to reach its steady state at a faster rate, thus causing a slight decline toward the steady state.

4. As the flow rate a_{21}, from the low-vulnerability class V_1 to the high-vulnerability class V_2 increases, the final population mortality μ^* and the final proportion dying in the high-vulnerability class μ_2^* increase, while the final proportion dying in the low-vulnerability class μ_1^* declines.

5. As the flow rate a_{12}, from the high-vulnerability class V_2 to the low-vulnerability class V_1 increases, the final population mortality μ^* and the final proportion dying in the high-vulnerability class μ_2^* decline, while the final proportion dying in the low-vulnerability class μ_1^* increases.

To make diagnostic use of the above qualitative analysis, we defined our goal in investigating the dynamics of mortality to be twofold: on the one hand, we want to reduce the population mortality; on the other hand, we want a larger proportion of deaths in the population to come from the low-vulnerability class. This means that people will live relatively healthy up to the time of death. For society as a whole, this results in lowering the terminal investment in health. As can be seen from the outcomes of the above analysis, only increasing a_{12} or decreasing a_{21} will facilitate these goals. This can be accomplished through health improvement policies, cleaning up the minority neighborhoods, educational campaigns, etc.

Motivated by the existing "anomaly" of a mortality crossover between the Black and White populations of the United States, we were then set to compare populations. To utilize our model of vulnerability that simulates the mortality experience of a non-aging cohort over time, we devised a transformation strategy to map mortality data from the age-domain to the time-domain. Equipped with the results of the qualitative analysis of the steady state solutions of mortality, as outlined above, we were able to make inferences about the transient behavior of population mortality at the crossover. Hence, by transforming the domain of data, as well as transforming the "explicit" solutions at the steady state to "implicit" derivations at the transient state, we established qualitatively that:

1. For the crossover to exist between the Black and White populations, the following must hold:

   (a) If both the Black and White populations have the same class-specific mortality rates μ_1, μ_2, and only differ in their mobility rates, then the relative rate of mobility to the class with better health must be higher for the Black cohorts older than the cohort containing the age-at-crossover, than their White counterparts. By symmetry, the relative rate of mobility to the class with better health must be lower for the Black cohorts younger than the cohort containing the age-at-crossover, than their White counterparts.

   (b) If both the Black and White populations have the same mobility rates a_{12}, a_{21}, and only differ in their class-specific mortality rates, then all Black cohorts older than the cohort containing the age-at-crossover must have:

      i. Either lower class-specific mortality rates than their White counterparts, irrespective of the rate of selection.

      ii. Or a lower selection rate (due to the class with good health having a higher death rate and the class with poor health having a lower death rate) compared to their White counterparts.

   By symmetry, all Black cohorts younger than the cohort containing the age-at-crossover must have:

i. Either higher class-specific mortality rates than their White counterparts, irrespective of the rate of selection.

ii. Or a higher selection rate (due to the class with good health having a lower death rate and the class with poor health having a higher death rate) compared to their White counterparts.

Hence, the existence of a mortality crossover between the Black and White populations is the result of a greater force of mortality selectivity or a smaller rate of relative mobility to the class with good health, among the younger Black population: the younger population is heterogeneous in the conditions of life, where distinct vulnerability classes have been formed. The older Black population, however, is more homogeneous in the conditions of life: higher rate of relative mobility and a weaker force of selection; the vulnerability classes are not so distinct in the older population. Therefore, the appearance of a mortality crossover is an indication of dissimilarity in the "vulnerability structure"of the two populations.

2. For the mortality crossover to be a dynamic phenomenon, or more specifically, for the Black-White age-at-crossover to increase to older age groups over time, the following must hold:

   (a) If both the Black and White populations have the same class-specific mortality rates μ_1, μ_2, and only differ in their mobility rates, then the relative rate of mobility to the class with better health must be lower for the Black cohort containing the age-at-crossover, than their White counterparts.

   (b) If both the Black and White populations have the same mobility rates a_{12}, a_{21}, and only differ in their class-specific mortality rates, then the Black population in the age group containing the age-at-crossover must have:

      i. Either higher class-specific mortality rates than their White counterparts, irrespective of the rate of selection;

      ii. Or a higher selection rate (due to the class with good health having a lower death rate and the class with poor health having a higher death rate) compared to their White counterparts.

   Hence, the Black-White age-at-crossover increases: either as the rate of mortality selectivity in the Black cohort containing the age-at-crossover increases; or as the relative rate of mobility to the class with good health slows down for the Black cohort containing the age-at-crossover, more so than it does for the White cohort. Therefore an increase in the age-at-crossover is an indication of further extension of the increasing gap between the high- and low-vulnerability classes, to the older Black population. As the older Black population becomes more heterogeneous in the conditions of life and the vulnerability classes form and extend to the entire population , the "vulnerability structure" of the Black and White populations becomes more similar.

Chapter 4

Conclusions

We introduced the notion of "transience in population dynamics" in the context of within-host dynamic interaction of the immune system with a pathogen; and population health. We demonstrated that qualitative analysis of transience can be quite enlightening and rich in information. We further reasoned that the information so obtained, will not only enhance our understanding of the underlying processes, but more importantly, it contains diagnostic values for making "timely" public health intervention.

Our interest in investigating the impacts of transitory changes in the physical and social environment on the dynamics of interacting components was spurred by the following observations:

1. Although the mathematics of linear systems is well studied and looked upon as trivial, it is in fact in an intuitive sense, neither understood nor trivial. This is because despite the dynamics of the system being linear, the characteristic polynomial is nonlinear; thus the equations governing the eigenvalues are nonlinear. Therefore, having explicit solutions that are in terms of the eigenvalues provides no comfort when trying to understand "why things are the way they are" and "why things become the way they become". Hence, the need for identifying the effect of each system parameter and the initial conditions, on the components of a system and the system as a whole.

2. Although mathematical models of population dynamics have almost always focused on the long term behavior of the interacting components, but in fact the steady state solutions are of little diagnostic value for the hosts invaded by pathogens or for the population health. This becomes clear when we realize that under transitory changes in the physical and social environment, our physiology is subject to a wide range of fluctuations, some of which are related to, or determinant of, disease processes and therefore detrimental to our health.

3. Although genetic predisposition and age are important causes of variation in the distribution of vulnerability of a population, they are not the only ones. While a substantial part of such variation is due to heterogeneity in the social conditions of individuals, in public health and demography, the distribution of vulnerability

is only understood in terms of age and the underlying genetics. Thus, despite the relative abundance of mortality data and decades of investigations, our understanding of the distribution of vulnerability of a population to disease and death, remains incomplete.

The results of the qualitative analysis of transience in infectious disease and population health, as developed in this thesis, can be summarized as follows:

1. We developed a dynamic model of within-host interaction between the immune system and a pathogen to simulate the early dynamics of acute infectious disease in the absence of a fully effective immune system. Consequently,

   (a) We identified the amplifying effect of the absence of a fully operative immune system on the pathogenesis of the initial inoculum at the start of infection.

   (b) We derived two measures of transience h_{peak} and t_{peak}, to respectively measure the "the peak of infection" (maximum pathogenic population) and the "time to peak of infection", in terms of the initial inoculum, initial immunity, and various infection-specific parameters of reproductive rate, induction rate of the immune system, and efficacy of the immune elements.

   (c) We made diagnostic use of such outcomes for devising effective, early intervention strategies, as "how to intervene" and "when to intervene".

2. Motivated by the occurrence of a mortality crossover between the Black and White populations of the United States, we developed a model of selection partially offset by mobility, to simulate the dynamics of vulnerability in a population heterogeneous in health. This abstract model mimics the mortality experience of a non-aging cohort over time, which requires mortality to be a monotonically decreasing process to a steady state. Therefore, to analyze the dynamics of mortality:

   (a) We transformed the linear model of two ordinary differential equations into a single quadratic differential equation, describing the rate of change in the ratio of the two vulnerability classes. This was based on the realization that in the absence of birth or migration in the model, the population of each vulnerability classes should eventually become extinct, while the ratio of the population of the two classes reaches a steady state.

   (b) We derived and analyzed the steady state population mortality based on the steady state population vulnerability-ratio.

   (c) We proposed and implemented a transformation strategy for mapping mortality data from the age-domain to the time-domain. This transformation was the key intermediate step to making inferences about the underlying processes governing the existence of crossover and its dynamics; the model was then used to analyze the trends in mortality data.

(d) We used the "explicit" steady state solutions for the vulnerability-ratio and population mortality, derived in earlier sections, to make inferences about the transient state of mortality at the crossover. We concluded that:

i. For the crossover to exist the Black population must be heterogeneous in health among its younger population, up to and including the age group containing the age-at-crossover, while the population older than the age-at-crossover is more homogeneous in health. This means that distinct vulnerability classes are formed in the Black population, but are not quite extended to the entire population. These effects are achieved by factors such as a higher selection rate among the younger population and a lower selection rate among the older population; or a lower relative rate of mobility to the class with good health among the younger population and a higher relative rate of mobility among the older population.

ii. For the age-at-crossover to change, or more specifically for the Black-White age-at-crossover to dynamically increase to older age groups, the Black population in the age group containing the age-at-crossover must become more heterogeneous in health. This means that vulnerability classes have been formed and are extending to the older Black population as the population ages as a whole. These effects are achieved by way of either a higher selection rate or a lower relative rate of mobility to the class with good health, for the Black cohort containing the age-at-crossover, compared to their White counterparts.

iii. We compared the mortality experiences of other pairs of populations in the United States. Specifically, we compared the mortality trends of the Native-American population with that of the White; and the mortality trends of the Hispanic population with that of the White. We learned that:

A. For both pairs of populations, their age-specific mortality curves cross.

B. The age-at-crossover appears to be "younger" for the Native-White pair, compared to Black-White; and it is "much younger" for the Hispanic-White pair.

Open problems are:

1. To apply the results pertaining to the within-host early behavior of acute infectious disease to infection-specific data. The goal is to ultimately classify infections based on their transient properties, the critical period to intervene, and plausible intervention schemes.

2. To extend the results pertaining to the within-host detection of the peak of acute infectious disease to the within-population detection of the peak of outbreaks of infectious disease, thereby facilitating the development of early surveillance techniques.

3. To extend the results of the Black-and-White mortality crossover to other populations, thereby enhancing our understanding of the dynamics of mortality crossover in its entirety. This will not only further refine our notion of the multi-faceted distribution of vulnerability in a population, but it can also educate the common sense as to how different people, under different social and physical transitory changes of different durations, age and die.

Bibliography

[1] Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans.* Oxford Science Publications, 1991.

[2] Maria-Chiara Corti, Jack M. Guralnik, Luigi Ferrucci, Grant Izmirlian, Suzanne G. Leveille, Marco Pahor, Harvey J. Cohen, Carl Pieper, and Richard J. Havlik. Evidence for a black-white crossover in all-cause and coronary heart disease mortality in an older population: The north carolina epese. *American Journal of Public Health*, 89(3):308–314, 1999.

[3] Irma T. Elo and Samuel H. Preston. The effect of early life conditions on adult mortality: A review. *Population Index*, 58:186–212, 1992.

[4] Irma T. Elo and Samuel H. Preston. Quality of data for estimating african-american mortality, 1930-1990. *Ms.*, 1993.

[5] Irma T. Elo and Samuel H. Preston. Estimating african-american mortality from inaccurate data. *Demography*, 31:427–58, 1994.

[6] R. E. Hope-Simon. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps. *The Lancet*, pages 549–554, 1952.

[7] Richard Levins. Evolution in communities near equilibrium. In Martin L. Cody and Jared M. Diamond, editors, *Ecology and Evolution of Communities*, pages 16–50. The Belknap Press of Harvard University Press, 1975.

[8] Richard Levins and Brian Schultz. Effects of density dependence, feedback and environmental sensitivity on correlations among predators, prey and plant resources: models and practical implications. *Journal of Animal Ecology*, 65:802–812, 1996.

[9] Kenneth G. Manton and Eric Stallard. Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: Explaining the black/white mortality crossover by a model of mortality selection. *Human Biology*, 53(1):47–67, 1981.

[10] Kenneth G. Manton, Eric Stallard, and James W. Vaupel. Methods for comparing the mortality experience of heterogeneous populations. *Demography*, 18(3):389–410, 1981.

[11] Kenneth G. Manton, Eric Stallard, and James W. Vaupel. *Recent Trends in Mortality Analysis.* Academic Press, Orlando, 1984.

[12] K. Marsh. Malaria – a neglected disease? *Parasitology*, pages S53–69, 1992.

[13] Cedric Mimms, Nigel Dimmock, Anthony Nash, and John Stephen. *Mims' Pathogenesis of Infectious Disease.* Academic Press, fourth edition, 1995.

[14] Charles B. Nam. Another look at mortality crossovers. *Social Biology*, pages 133–142, 1995.

[15] Charles B. Nam and Kathleen A. Ockay. Factors contributing to the mortality crossover pattern: Effects of development level, overall mortality level, and causes of death. International Union for the Scientific Study of Population, Mexico City, 1977.

[16] Charles B. Nam, Norman L. Weatherby, and Kathleen A. Ockay. Causes of death which contribute to the mortality crossover effect. *Social Biology*, 25(4):306–314, 1978.

[17] Raymond Pearl. *The Biology of Death.* J. B. Lippincott, Philadelphia, 1922.

[18] Samuel H. Preston, Irma T. Elo, Ira Rosenwaike, and Mark Hill. African-american mortality at older ages: Results of a matching study. *Demography*, 33(2):193–209, 1996.

[19] Charles J. Puccia and Richard Levins. *Qualitative Modeling of Complex Systems.* Harvard University Press, 1985.

[20] David Rapport, Robert Costanza, Paul R. Epstein, Connie Gaudet, and Richard Levins. *Ecosystem Health.* Blackwell Science, 1997.

[21] M. Spiegelman. The longevity of jews in canada, 1940-1942. *Pop. Stud.*, 2:292–304, 1948.

[22] James W. Vaupel, Kenneth G. Manton, and Eric Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–454, 1979.