Sexual volatility and the spread of HIV

by

Ethan Obie Romero-Severson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiologic Science)
in The University of Michigan
2011

Doctoral Committee:

Professor James S. Koopman, Chair
Professor Ana Diez-Roux
Associate Professor Joseph N.S. Eisenberg
Associate Professor Noah A. Rosenberg

To all of my teachers

ACKNOWLEDGEMENTS

Thank you to my parents, especially to my mother who went beyond the often thankless duty one has to their children by also advising me academically when no one else would. Thank you to Wade and the Radiant crew for the psycho-nautical adventures that inspired me to go back to school. Thanks to all my boys on team Science Gays featuring Mark the Architect and Ian the Straight (nobody's perfect). I'm the last to get a Ph.D., which, I think, makes me α. My first decree as α is that we all move to Hong Kong; a really nice elephant trainer that I met in a charming Western themed post-modern piano bar told me that its lovely this time of year. Thanks to the epidemiology G&L secret society. It truly was a sequence of events separated only by a sequence of non-events. To Emily and her long-time companion Marta: if this science thing does not work, let's expand our dominance of all quiches into an empire. Ian, put down that pint and IRUVADE! Meghan, you're cool, I guess. And, finally, to my BF Grant, you have been more supportive and understanding than I deserve. In time, the roles will switch: You be Octavian, and I'll be Agrippa. And finally, to everyone I have ever met: I am buying you...a pizza. End communication.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

CHAPTER I

Introduction

1.1 Motivation

HIV persists despite being weakly transmitted [28], well characterized, highly preventable [35], and having well-funded research and prevention programs. This persistence may stem from current prevention practices that are fundamentally sound but too small in magnitude, or because our theoretical understanding is lacking, leading to inefficient prevention recommendations. HIV prevention programs have been effective in reducing, but not eliminating, HIV transmission. For example, the sharp decline of HIV in Uganda in the 1990s was attributed to the implementation of effective, cooperatively designed prevention programs [29]. Since 2005 the prevalence of HIV remained stable at about 6.5% [44], indicating that even though these programs are clearly effective in reducing risk behavior, high-levels of transmission are still occurring. The question remains, what is generating the remaining force of infection?

The answer to this question is not obvious. HIV, unlike many other infectious diseases, has a very long infectious period leading to highly complex transmission dynamics. These dynamics are shaped not only by individual sexual behaviors, but also by large-scale forces that have no individual-level correspondence, such as the

shape of sexual networks [41] [73]. The building of a sound theoretical understanding of HIV transmission has been a decades-long project that is still underway. In this thesis I present work done in collaboration with many other scientists on a simple yet poorly understood aspect of HIV transmission: how short-term variation in sexual behavior affects HIV transmission dynamics.

My personal motivation for studying this topic came from a simple realization. While learning about mathematical models of infectious disease, I came to think of models less as systems of equations and computer simulations and more as collections of assumptions about how humans behave. I realized that the validity of a mathematical model is a two-fold question. First, do the model assumptions obtain in the real world? And, two, whether or not violation of those assumptions matters to the inference we can make from our models. One assumption that I kept coming back to was the stability of sexual behavior in general. The idea the someone who had risky sex once – maybe without a condom or under the influence of drugs – would continue that behavior pattern indefinitely seemed unlikely. In addition to seeming unlikely, the assumption of stable sexual behavior given the observed levels of sexual activity imply the unpleasant conclusion that a small subset of highly promiscuous individuals are responsible for the epidemic [52]. Good science is good science regardless of its conclusions; however, the question remains, is such a conclusion the result of unexamined assumptions? My own experience suggested that sexual behavior is more contextual in nature. Maybe a split from a long term partner would precipitate a brief period of sexual risk in an otherwise cautious person. Or, maybe, the new found freedom associated with coming out of the closet would be celebrated by a period of uncharacteristic promiscuity. In the much more elegant words of Ortega y Gassett, 'yo soy yo y mi circunstancia' [23].

In this thesis I formalize a representation of contextual sexual behavior that I call **contact rate volatility**. I will attempt to answer two questions: does the assumption of stable risk behavior obtain? And, does the stability or volatility of sexual behavior matter to our theoretical understanding of HIV transmission dynamics? In the second chapter I calculate the likelihood of a prospective survey of sexual behavior under a range of stable and volatile sexual behavior models. In the third chapter I use the theory of branching process to calculate a theoretical statistic that delineates the conditions under which an epidemic can occur assuming either stable or volatile sexual behavior. In the fourth chapter I use simulation to explore the mechanism and effects of volatile sexual behavior on the transmission dynamics of HIV.

1.2 Theoretical context

The intent of this thesis is to contribute to the broader theoretical understanding of HIV transmission by exploring the effects of an understudied aspect of human sexual behavior. In this section I will briefly review some of the relevant literature on the dynamics of HIV transmission to contextualize this work as the next logical step in a lineage of theory development. The HIV modeling literature is massive and its domain is well beyond the scope of this section. I have selected papers that I believe analyze key heterogeneities that must be considered as fundamental elements of a complete theory of HIV transmission.

Causal factors that affect the spread of HIV can be loosely categorized into 'biological' and 'social' categories. Biological factors are generally those that result from the interaction of the infectious agent and the host. Examples of biological factors include the viral titer in the bodily fluids, the duration of the infectious period, host

susceptibility, and time to death. Social factors emerge as a result of the interactions between hosts, or between a host and its social environment. Such factors include sexual practices, number of partners, drug use, and partner choice. The reason for making this distinction is two-fold. First, bio-social distinction provides a nice delineation between the early modeling work that largely focused on biological factors and the later work which has been more focused on social factors. The reason for this shift in focus is that empirical work has been highly successful in defining and limiting the range of biological parameters that are less variable between populations. Secondly, although there are exceptions, such as the geographic distribution of the $\Delta 32$ CCR5 mutation that confers a degree of HIV resistance [63], more of the variance HIV risk between populations is attributable to variance between social factors. Goodreau and Golden have argued that versatility among gay men–the ability to perform both insertive and receptive sexual roles–and role separation in heterosexuals, is one of the primary determinants of the differential burden of HIV in gay and heterosexual populations [25].

The first modeling papers published after HIV was identified as the etiological agent of AIDS [22] [9] were primarily concerned with short-term prediction of the number of AIDS cases [57] [62]. Much of that early work was focused on fitting simple functions to the incidence data to extrapolate the number of case that should be expected in the coming years. These simple predictive models were refined to include variable delays in the time from reporting to the time of diagnosis, which, before the time of highly accurate tests, was based on the presentation of symptoms that might not occur for years after infection [16]. These types of analysis are aimed at making more refined predictions about the burden of AIDS, but not about the causal factors that generate those cases. The prediction of HIV cases is approached as a case of

phenomenological curve fitting and refinement; this type of 'black box' conceptualization of HIV dynamics does little to add to the broader theoretical understanding of HIV transmission.

The first study of the dynamics of HIV was published about 3 years after the co-discovery of HIV [5]. Anderson et al. published a preliminary study of a causal model of HIV transmission in the context of not knowing many of the key epidemiological parameters that we considered well-established today. The authors note that they do not know the latent period (time from exposure to infectiousness), the mortality rate of the infected, the duration of the infectious period, the persistence of virus in the infected, and the transmissibility of the virus. Anderson et al. assume that the time from seroconversion to AIDS is a linearly increasing function of time and that infectiousness is lifelong. Drawing on a very limited data set of sexual behavior in gay men in the UK and USA, they find a distribution of partnership rate (number of partners, not the number of contacts) to have both a high average and high variance. These data were fit to a Gamma distribution that they argue is still poor, but is an improvement over previous fits.

Much of the initial work in Anderson et al. (1986) involved modeling variability in parameters whose distributions in various populations we can now characterize. However, they highlight both the role of heterogeneity in partnership rates and how that heterogeneity affects the epidemic. They find that increasing heterogeneity in the distribution of contact rates reduces the magnitude of the epidemic and the time to peak incidence. When the variance of the contact rate distribution is much greater than the mean (which is the case for the contact rates studied in chapter 2), an increasingly small number of individuals account for an increasingly large proportion of contacts. Likewise, even at very high average contact rates, a large proportion of

the population are making very few contacts (i.e. nearly abstinent). Similar work from around the same time shows that for a given value of R_0, increasing variance in the distribution of contact rates leads to lower fraction of infected individuals [55]. The conclusion is clear: differences in mean partnership rates are insufficient to explain differences in infection rates in different populations; heterogeneity in contact rates also matters to HIV transmission.

Before a 'final' synthesis of the theory of HIV transmission can occur, we first need to determine all of the relevant factors that are causing the patterns of HIV transmission. In the literature, a lineage of work exists from the late 1980's that focuses on relaxing the assumptions of previous models to identify where and how data needs to be collected to better understand the spread of HIV. Anderson et al. (1986) conclude that 'as the intensity of epidemiological research increases, simple models will be of less value once the various unknowns and complications that have been hinted at in this paper are defined and quantified.' Certainly this is true; however, many of the 'various unknowns' mentioned by Anderson et al. are now known, yet our picture of HIV transmission dynamics are not yet complete. As I use very simple models in this thesis and find a unique heterogeneity that has large population-level effects, I have to disagree about the value of simple models.

Anderson et al. (1986) identified that heterogeneity in risk behavior is a key determinant of the spread of HIV, Hyman and Stanley further defined the pattern of mixing between those groups as another important factor. Hyman and Stanley (1988) proposed a model that relaxed several of the simplifying assumptions of previous models, including variable infectivity and structured partner choice [36]. They found that variation in the natural history of infectivity, when controlling for total infectivity, could have major influence on the early trajectory of the epidemic. When

relatively more of the infectivity density occurs nearer to the time of infection, time to peak incidence occurs sooner. Increased density towards the end of the epidemic, on the other hand, had little influence on the time to peak incidence. Hyman and Stanley also found that when individuals preferentially mix with individuals of similar risk status rather than drawing partners at random slows significantly, but the total number of infected cases actually increases. Later work on the natural history of infection has lead to a consensus about the typical infectivity profile; however, the nature of sexual mixing is still largely unknown.

Using theoretical analysis, Jacquez et al. argue that the patterns of sexual mixing are a key determinant of the spread of HIV [39]. They identified three kinds of basic mixing patterns: restricted, proportional, and preferred. Restricted mixing is when individuals draw their sex partners from only within their own risk groups. The assumption is that each risk group has their own mixing site exclusive to them. Within each site individuals are homogeneous and independent of other groups. Proportional mixing occurs when individuals with different contact rates mix at one common site. The rate at which individuals from different groups contact one another is proportional to the product of their contact rates [34]. In a proportional mixing setting, most of the contacts are assortative for high-risk individuals as high-risk individuals are more likely to contact other high-risk individuals. However, contacts made by low-risk individuals are likely to be disassorative as contacts made by low-risk individuals are also likely to be with a high-risk individual. Proportional mixing is a very common assumption in models of HIV transmission not necessarily because it is accurate, but rather because this assumption provides an amount of mathematical convenience. The third mixing pattern, preferred mixing, reduces restricted and proportional mixing to a single common framework by allowing individuals to re-

serve some proportion of all contacts for a restricted mixing site with the remainder going to a common site. The preference parameter is 'universal' in that all classes of individuals have the same preference for mixing at the restricted site. This assumption was relaxed in later work [38]. Jacquez et al. found that transitioning from restricted to proportionate mixing could produce risk ratios of 9-fold in some cases. These effects occur because, as more contacts are made at the common site, more of the high infection potential in the high risk group is 'bleeding' over into the lower risk group [48]. Even if the low risk group is below the epidemic threshold, the continued exposure to high risk infectives at the common site produces significant infection levels in low risk individuals.

In general, the work up to this point has assumed that sexual contacts and sexual partnerships are indistinguishable in that all contacts are instantaneous and otherwise identical. Per-act probabilities of transmission are easier to estimate than per-partnership probabilities of transmission simply because a sex act is much easier to define than a sexual partnership. However, in reality, many sexual acts will occur in the context of a partnership rather than as 'one-offs' between strangers. The simple thought experiment of imagining how HIV would be incapable of spreading in a population with life-long, strict monogamy reveals how partnership dynamics matter to the spread of HIV. Using straightforward simulation methods, Morris and Kretzschmar found that the degree of concurrency (overlapping relationships) is a major determinant of the incidence rate. Holding the total number of contacts constant, they found that concurrency could increase the incidence rate by up to 10-fold and greatly increases the early rise of the epidemic also [60]. However, concurrency is both difficult to define and estimate [33], and it is not immediately clear how to interpret estimated levels of partnership concurrency in men (10-26%) [2] [27] with

respect to HIV dynamics.

There are many of other axises along which models can be made increasingly realistic, but I have introduced what I believe to be the most important classes of assumptions made by the homogeneous SI model that need to be addressed to make a useful and parsimonious model of HIV transmission: heterogeneity in behavior, heterogeneity in mixing patterns, and heterogeneity in partnership dynamics. In this thesis I explore an previously unappreciated dynamic, heterogeneity in individual-level contact rates over time, which I refer to as contact rate volatility. This work can be thought of as the initial work exploring one aspect of the temporal dimension of behavioral heterogeneity.

1.3 Contact rate volatility is a kind of non-differential behavior change

Change in sexual behavior is generally understood to be a differential change in one's state with respect to sexual risk behaviors. For example as movement between insertive and receptive sexual roles [11], movement between core and non-core groups [49], or movement between age groups [50]. This type of behavior change implies a specified change in one's risk, moving from insertive-only to receptive-only states increases fundamentally increases one's risk as receptive sex is generally regarded to be much riskier than insertive sex. I am going to formalize contact rate volatility as a more general kind of behavior change that does not imply changes in state nor a specified differential change in risk behavior.

The first element of a more formal representation of contact rate volatility that needs to be specified is the probability of having a specific contact rate (i.e. the shape of the container in the above metaphor). I will refer to this as the contact rate heterogeneity (CRH) model. Most probability density or mass functions will

probably suffice as a CRH model, although it makes conceptual sense to restrict oneself to functions defined only for positive values – negative contact rates may, in principle, not be a problem but they do make the analyses a bit harder to explain. The second element of contact rate volatility is the model of volatility itself. In the above metaphor, we imagined the temperature of the gas as the single 'volatility' parameter. The simplest formalization is a single rate parameter, the rate at which individuals re-draw their contact rates with probabilities specified by the CRH model. As the volatility parameter increases (i.e. the gas becomes hotter), individuals more rapidly re-draw their contact rates as specified by the CRH model; likewise, the length of time over which an individual's behavior is constant becomes shorter and shorter.

This kind of contact rate volatility is non-differential in that the CRH model is constant over time and that there are no changes to an individual's state that precipitate changes to the contact rate. For example, Blower et al. model changes to an individual's risk in terms of changes of state between receptive-only and insertive-only sexual preference states [11]. In this case, risk is volatile in that it is variable over time, but it is also differential in that the risk of infection is fundamentally different in each state. If, in our stateless volatility model, we make the assumption that the hazard of infection is proportional to the contact rate then see a similar variability in the risk of HIV infection with out having to specify the distribution of individuals over a state space. This kind of stateless, non-differential behavior change allows us to study the general effects of behavior change in a highly abstract way that may help to understand the general mechanisms of how behavior change affects HIV transmission dynamics.

1.4 Basic epidemiology of HIV

HIV is a retrovirus transmitted primarily though sexual contact, sharing of injection works among intravenous drug users, perinatally, and iatrogenically though sub-optimal medical practices. Sexual contact is the dominant route of transmission. In the United states of the estimated 56,300 new cases that occurred in 2006, only 6600 (12%) were attributable to non-sexual routes of transmission [31]. The degree of heterosexual versus homosexual transmission varies quite a bit by country. In the United States, heterosexual transmission only accounted for 33% of incident cases in 2006 [31] despite heterosexual identified persons representing the vast majority of the population.

The physiology of HIV infection is very complex but can be divided into four epidemiologically distinct phases: latent, acute, chronic, and final. Immediately following infection, the virus remains latent for a few days [77], where it shows little signs of amplification. The virus then begins to amplify rapidly, reaching peak levels in the blood about 17 days after infection and in semen about 30 days after infection. Patients in this second, acute, stage of infection can present with a clinical viremia and found to be more likely to report non-specific symptoms of viremia such as fever and malaise than chronically infected patients [17]. Levels then decline 170-fold to a nadir occurring about 10 weeks after infection [65]. After viral titers drop, the patient enters an asymptomatic, chronic stage of infection that is highly variable in duration, as short as 8 years in untreated individuals. During this time viral titers are low, and may even be undetectable in patients with good clinical management [40]. The final stage is characterized by an increasing viral titer and a weakening of the immune system leading eventually to AIDS and then death. Life expectancy in

the final stage is about one year.

The changes in viral titer corresponds to changes in the per act probability of transmission. The highest quality estimates of the per act probability of transmission come from a study of 235 monogamous, HIV-discordant heterosexual couples in Uganda. Serodiscordant couples were identified from a large prospective cohort of 15,127 individuals. The number of sex acts and the infection status of the negative partner over time were ascertained. Because of the study design, where individuals were enrolled as they became infected, the authors were able to classify each contact that the uninfected partner was making as being either with an acute, chronic, or late stage HIV infected partner [74]. Further analysis of the data from this study showed the per-act probability of transmission to be 35 to 61 times higher for contact made with an acutely infected person compared to a chronically infected person [66] and about 7 times higher for final compared to chronic stage.

1.5 Stage contribution is a central question in the dynamics of HIV

The high degree of variance in transmission probability over the course of infection leads to the natural question of what proportion of new infections can be attributed to each infection stage. The timing of transmissions can be a key determinant of the types of interventions will be effective in reducing transmission of HIV. For example, the drop in HIV incidence in the late 1990s [31] was in part due to the the development of highly effective anti-retroviral therapy, a treatment that greatly reduces viral titers, effectively eliminating late-stage HIV. Likewise, interventions that reduce the time from infection to diagnosis by increasing HIV testing rates work on the assumption that diagnosed chronically infected individuals will transmit less that infected individuals unaware of their sero-status. Such an intervention will be

largely ineffective if most new infections come from acutely infected persons. In spite of the centrality of this issue, stage contribution is difficult to determine empirically.

The contribution of each stage is a product of both the per act probability of transmission for each stage but as well as the probability that a contact will be made with an infected person at a given stage. Though the former quantity can be measured in a somewhat direct way (e.g. the Uganda study [74]), however, the latter is virtually impossible to measure directly. Estimates of the relative contribution of each infection stage are based on models of sexual behavior and transmission and can be quite contentious [68]. At the beginning of an epidemic, the incidence rate is much larger than the progression rate from acute to chronic, creating a large pool of individuals in the acute stage. This disproportionately large pool of acutely infected individuals at the beginning of an epidemic drives the initial exponential increase in the number of infected individuals [37]. Beyond that point, opinions on the relative contributions of each infection stage are highly divergent. Koopman et al. showed how the effectiveness of interventions focused on the acute stage are highly influenced by age structured contact rates and movement between high and low-risk behavior groups. Using simulation they found conditions under which acute stage intervention got the system below threshold yet only accounted for 20% of the total contagiousness [49]. Another study by Rapatski et al. of HIV in San Fransisco found conflicting results: they concluded that greater than 90% of transmission came from contact with late stage infected individuals.

This discrepancy may be explained by a key difference between those two studies: in Koopman et al. individuals move between high and low risk groups, creating a kind of state-based temporal volatility in contact rates, while the Repatski et al. study did not. If contact rate volatility can explain the highly divergent conclusions of these

two papers, it may help the design of future prevention programs by highlighting the importance of sexual behavior volatility.

CHAPTER II

Short-term volatile sexual behavior in a prospective dataset

2.1 Introduction

The stability/volatility of sexual behavior is an important determent of key HIV transmission dynamics and is under-studied. Most models of HIV transmission assume that an individual's risk behavior is stable over time. The validity of this assumption is the topic of this chapter. There are many ways that an individual's risk behavior can be unstable: long-term changes in secular trends, contextual behavior modification [58], or self-motivated risk reduction. Also, there are many types of risk behavior. Blower et al. modeled the probabilities of transitioning between insertive-only, receptive-only, versatile, and abstinent sexual roles in a cohort of gay men [11]. Alam et al. further modeled the effects of these transitions on HIV transmission dynamics, finding volatility in sexual role to increase both the prevalence of HIV and the proportion of infections coming from individuals with primary stage HIV infection [4]. The proportion of infection coming from the acute stage has been argued as a possible determinant of the efficacy of pre-exposure prophylaxis [64].

In this paper we look for evidence of volatility in contact rates over a two-year period in a cohort of at-risk gay men. We calculate the likelihood of the following contact types: unprotected oral receptive, unprotected anal receptive, and unpro-

tected anal insertive given two stability models that assume constant risk behavior, and two volatility models that assume variable risk behavior. Strong evidence for volatility in a subset of the population is found for each contact type.

2.2 Materials and methods

2.3 Formal expression of the problem

The top panel of figure 2.1 illustrates the concept of contact rate volatility in terms of a prospective-cohort-type study of sexual behavior. A set of individuals, i, are enrolled in a prospective cohort. At times m the individuals are measured defining a sequence of observational periods n of length δ. Measurement consists of a survey asking about the frequency of sexual contacts, ϕ, that occurred over the previous interval. The data are the set of counts of sexual contacts $\phi_{i,n}$ for each individual. The value of $\phi_{i,n}$ is a Poisson random variable with a rate determined by the integral of the contact rate over that interval, $f(\chi_n) = \int_{t_n}^{t_n+\delta} \chi_{i,t}$. The number of contacts in a given interval converges in distribution to the integral of the contact rate over the interval, $\phi_{i,n} \stackrel{d}{=} Poiss(f(\chi_n))$. The fundamental problem is that the contact rate is unobservable and must be inferred from the observed data.

2.3.1 Behavioral models

The difference between volatility and stability models is the assumptions concerning the form of $f(\chi_n)$; stability models assume $f(\chi)$ is stable over time, volatility model assume $f(\chi)$ is a function of time. We consider two possible stability models, homogeneous and gamma heterogeneous, and two volatility models, interval independence and ontological volatility (o-volatility). In the homogeneous model, we assume that all individuals make sexual contact at one common rate $f(\chi) = \chi\delta$. In this model, contact rates are stable and homogeneous. The gamma heterogeneous

model relaxes the assumption of homogeneous contact rates by assigning individuals unique contact rates, χ_i, such that $\Pr(X = \chi_i) \sim \Gamma(k, \theta)$; in this model $f(\chi) = \chi_i \delta$. The interval independence model assumes that contact rates are re-drawn at the beginning of each observational period with probability $\Pr(X = \chi_{i,n}) \sim \Gamma(k, \theta)$, but are assumed to be constant over the duration of each observational period. This is equivalent to assuming that observational periods are statistically independent of one another. In this model $f(\chi) = \chi_{i,n} \delta$. The bottom panel of figure 2.1 illustrates the o-volatility model. In this model, individuals re-draw their contact rates randomly at a constant rate, ρ. This re-drawing generates a sequence of behavioral phases over which their sexual contact rate is constant. Given the large number of individuals reporting no volatility over the course of the study (discussed below), we also include an additional heterogeneity parameter α giving the probability that individual behavior is constant over the course of the study, regardless of the volatility parameter. We refer to this model as ontological because it specifies that actual mechanism of volatility independently of the structure of the data. This is unlike the interval independence model which assumes that contact rates are volatile, but are dependent on non-causal factors (the duration of observational periods in the study). Unlike the interval independence model, the o-volatility model can account for correlation between the number contacts in each observational period as behavioral phases can extend into multiple observational periods. In this model $f(\chi)$ is given by the average contact rate over the interval which is the contact rate of each behavioral phase multiplied by its duration.

2.3.2 Behavioral data

The data we use in this chapter comes from the Centers for Disease Control and Prevention Collaborative Seroincidence Study. In 1992 this multi-site study

enrolled three large cohorts of HIV-susceptible men in Chicago, Denver, and San Fransisco. Eligibility criteria included report of any anal sex, report of any non-HIV sexually transmitted infection, and, in San Fransisco, report of receptive oral sex with ejaculation. The study enrolled and followed susceptible individuals for a period of up to 24 months. Observation stopped due to follow-up loss, a positive HIV test, or completion of the study. At enrollment, individuals were asked the number of times they had sex in the previous six months. At six month intervals, individuals where asked again about their sexual behavior in the previous six months for up to four total interviews.

The dataset the we received (via Dr. Eric Vittinghoff) excluded any individuals who had reported any injection drug use. We further restricted the data to individuals who completed all 24 months of observation. For those individuals, we extracted three variables: $\phi^x_{i,n}$, the number of unprotected receptive oral contacts with individuals of unknown HIV status of the i^{th} individual in the n^{th} observational period, $\phi^y_{i,n}$, the number of unprotected receptive anal contacts with individuals of unknown HIV status of the i^{th} individual in the n^{th} observational period, and $\phi^z_{i,n}$, the number of unprotected insertive anal contacts with individuals of unknown HIV status of the i^{th} individual in the n^{th} observational period.

2.3.3 Parameter estimation and likelihood of the data

Homogeneous, gamma heterogeneous, and interval independence models

The homogeneous model has one common contact rate, χ, for all individuals, for which the MLE is simply the sample mean $\chi = \bar{\phi} = \frac{1}{n+i} \sum^{i,n} \phi_{i,n}$. The likelihood of the data given this model is $L(D) = \prod^{i,n} Pois(\phi_{i,n}|\chi)$.

The gamma heterogeneous model assumes each individual has a unique contact rate given by the average contact rate over the four observational periods, $\chi_i = \bar{\phi}_i =$

$\frac{1}{n}\sum^{n}\phi_{i,n}$. In the data, $\bar{\chi_i}$ frequently equals zero, which is undefined by $\Gamma(k,\theta)$. An additional parameter, τ, accounts for this by giving the probability that an individual has no contacts over the course of the study. The likelihood of the data given the gamma heterogeneous model is

$$L(D) = \prod^{i} \begin{cases} \tau & \text{if } \bar{\phi_i} = 0 \\ (1-\tau)\Gamma(\bar{\phi_i}|k,\theta)\prod^{n} Pois(\phi_{i,n}|\chi_i) & \text{if } \bar{\phi_i} > 0 \end{cases} \tag{II.1}$$

The interval independence model assumes that each observational period is statistically independent of one another. A simple Poisson process with a gamma distributed prior gives a negative binomial posterior [76]. The likelihood of the data given the interval independence model is $\prod^{i,n} NegBin(\phi_{i,n}|k,\frac{1}{\theta+1})$. The maximum likelihood parameter estimates for this and the gamma heterogeneity model were obtained numerically using the downhill simplex method.

o-volatility model

Calculating the likelihood of the data under the assumptions of the o-volatility model is more complex. In the 3 previous models we made the simplifying assumption that behavioral was stable over the course of the whole study (homogeneous and gamma heterogeneous models) or behavior change occurred a pre-specified times (interval independence model). The additional complexity of the o-volatility model means that there is no closed-form expression for the likelihood of the data. We use Monte Carlo simulation to estimate the maximum likelihood parameters under the volatility model.

To simulate the likelihood of the data given the behavioral parameters, $L(D|\hat{k},\hat{\theta},\hat{\rho},\hat{\rho})$, we numerically simulate the maximum likelihood of the data $L(D|\Theta,\Lambda)$ where Λ is

the population of simulations on which the estimate is based, and Θ represents the behavioral parameters. The following steps define the simulation process:

1. given Θ^*
2. for i in D:
   (a) simulate a sequence of behavioral intervals, r_i^*, with probability $\Pr(r|\Theta^*) \sim Exp(\rho*)$
   (b) simulate a sequence of contact rates,$\chi_{i,n,r}^*$ over r_i^*, with probability $\Pr(\chi|\Theta^*) \sim Gamma(k^*, \theta^*)$
   (c) calculate the average contact rate in each observational period, $\chi_{i,n}^*$, given $r_{i,n}^*$ and $\chi_{i,n,r}^*$
   (d) for each datum, calculate the probability of $\phi_{i,n}$ as $L^*(d_{i,n}|\Theta^*, \lambda) = Pois(\phi_{i,n}|\chi_{i,n}^*)$
   (e) take the product over n to get the likelihood $L^*(d_i|\Theta^*, \lambda) = \prod^n L^*(d_{i,n}|\Theta^*, \lambda)$
   (f) repeat until a population of q simulations are generated
   (g) calculate the likelihood of the individual data as the average over the set of simulations, $L^*(d_i) = \frac{1}{q} \sum^q L^*(d_i|\Theta^*, \lambda_q)$
3. the likelihood of the data is $L^(D) = \prod^i L^*(d_i)$

To do this we randomly selected 34,000 random parameter vectors from the parameter space (reasonable limits were set by previous experience with the model), simulating the likelihood of each vector with $q = 100$. Local non-parametric regression curves were fit to the joint likelihood distribution using the loess function with standard arguments in R. The likelihood surface was estimated by the value of the regression surface evaluated over a gird defined by 50 evenly spaced points over the range of each parameter (a total of 50^4 evaluations). The MLE of each parameter

was defined as the parameter vector with the maximum likelihood value. Confidence intervals for the parameters were calculated using the profile method [71].

2.4 Results

2.4.1 Qualitative

The full dataset included 1883 participants. 111 (6%) individuals were observed for 1 period, 329 (18%) individuals were observed for 2 periods, 561 (30%) were observed for 3 periods, and 882 (47%) individuals were observed for all 4 periods. Loss to seroconversion accounted for 52/1001 (5%) of those observed for less than 4 periods. 52 individuals seroconverted during 35880 total months of observation for an incidence rate of 0.0174 per person-year. Hall et al. estimated about 19,000 MSM cases of HIV in 1992 [31], which, given the incidence rate of 0.0174 per person-year, corresponds to an estimated at-risk population of about 1.09 million. That is, a population of 1.09 million at-risk men would generated 19,000 infections in year given the incidence rate estimated from the data. Comparison of incidences and incidence rates requires estimation of the size of the at-risk population. Such estimations are very difficult to make for abstract, dynamic, and highly personal categories such as sexual identity. However, the estimated at risk population size of 1.09 million represents about 0.4% of the United States population in 1992. That proportion seems low, but is not an unreasonable estimate of the number of sexually active, at-risk gay men in the US in 1992.

The estimated incidence rate is also consistent with incidence rates reported in the state of Florida in 2010. Lieb et al. reported an incidence rate of 0.01312 per person-year for Hispanics and reported incidence rate of 0.03608 per person-year for blacks in Florida [51]. The estimated rate of 0.0174 is between the estimates rates for Hispanics and blacks in Florida. The consistency of the incidence rate estimate with

other reported incidence rates and the reasonable estimate of the at-risk population size given the calculated incidence in 1992 suggest that the behavioral data do not have an inordinate number of higher or lower risk individuals that would be expected.

The individuals who were lost to follow-up reported more abstinence and, in general, fewer and less variable contacts than those who were observed for all four periods. Individuals who were observed for 1, 2, 3, or 4 intervals reported an average of 5.7, 7.1, 8.7, 9.2 contacts per 6 months with average standard deviations of 0, 3.7, 6.1, and 7.4 respectively. Individuals observed for fewer periods were more likely to be abstinent over the course of their time observed. This is expected as an individual observed for 4 waves has to maintain abstinence for longer to have zero average contacts. However, the proportion of individuals observed for four periods that have no contacts (13%) is much higher than is expected assuming the proportion of zeros in individuals observed for 1 wave is an estimate of the probability of making no contacts during an observational period ($\left(\frac{37}{111}\right)^4 \approx 0.01$). A similar pattern is observed in each of the individual contact types.

One possible explanation for the difference in those who are lost to follow-up and those who are not is differential investment. Individuals who are followed for all four periods may identify more strongly with the gay community (i.e. be more 'out') making them more sexually open and more motivated to continue with the study for the broader well-being of the community. Another possible explanation is that individuals who were observed for all four periods are in long-term relationships and are therefore the most stable and motivated to participate. This, however, does not hold water if we consider the pattern of excess variance in those observed for 2, 3, or 4 periods. Excess variance is the multiplicative-scale variance beyond what is expected from a homogeneous Poisson process; excess variance greater than

1 suggests higher than expected volatility. The excess variance is 1.92, 4.27, 5.95 for individuals observed for 2, 3, or 4 periods. Ostensibly, sexual behavior is more consistent and less volatile in the context of a long-term monogamous partnership. If individuals who are observed for all 4 periods are in longer-term, monogamous relationships, the excess variances should be lowest in that category, which is the opposite of what is observed.

The data suggest evidence for both stable and volatile contact rates. Table 2.1 shows the average and standard deviation of the number of contacts stratified by the number of periods in which no contacts were reported (abstinent intervals) for individuals observed for all four periods. The behavioral patterns with the higher degrees of abstinence were the most common: total abstinence is the most common pattern, abstinence in all but 1 interval is the second most common, and so on. If we dichotomize behavior into abstinent and active states, then the most common behavioral pattern (total abstinence) supports the idea of stable short-term sexual behavior; however, the second most common pattern (abstinence in all but one interval) suggests a highly volatile short-term sexual behavior. The latter pattern is, by definition, the pattern with the highest possible variance given a mean contact rate.

The data imply that URO contacts are more 'episodic' than URA contacts. In each activity level, URO has the highest excess variance, the variance above the exception for a homogeneous Poisson process. A high excess variance suggests a high degree of variability in the contact rate, producing episodes of higher and lower risk. Of all contact types, the highest average contact rate is URO for individuals active in all four periods; the lowest average contact rate is URO for individuals abstinent in all but one period. For URO, the average contact rate in a given interval is

4.49 times higher for individuals who maintain activity in all four periods compared to those who report activity in only one period. This high discrepancy is consistent with the observation that oral sex is commonly practiced with both casual and steady partners while both insertive and receptive anal sex without a condom is reserved for steady partners [61]. The high and highly variable rates of URO in individuals with activity in all periods could be accounted for by a combination of stable contacts with a steady partner and volatile contacts with shorter-term concurrent partners while activity separated by periods of abstinence could be accounted for by limited 'one-off' encounters. The latter scenario is also supported by the fact that, in individuals that only reported abstinence in all but one period, made, on average, 3.25 URO contacts in that six month active period. The high frequency of the abstinent-in-all-but-one-period pattern and the relatively low number of contacts made in that period is more consistent with brief episodes of higher risk behavior where partners are randomly selected for 'one offs'. Also, the increasing excess variance with increasing average contact rates also implies some combination of stability and volatility generating the observed contact rate patterns.

Figure 2.2 shows the distribution of the number of contacts reported in any observational period. As expected, we observe highly right-skewed contact rates. Zeros are not shown in the histogram, but no contacts were reported in a large majority of observational periods (84%, 88%, and 81% for oral receptive, anal receptive, and anal insertive respectively). The unprotected insertive anal sex contact rate has the lowest heterogeneity between individuals. High contact rate individuals (above 95%) accounted for for 70% of all unprotected insertive anal contacts. For both oral and anal unprotected receptive sex, high contact rate individuals accounted for 90% of all contacts.

Figure 2.3 shows the trajectory of each individual's contact rate over the course of the study. The number of contacts made appears to be highly volatile with variations occurring, for some individuals, over multiple orders of magnitude. However, most individuals report no contacts over the course of the entire study (67%, 72%, and 58% for unprotected oral receptive, unprotected anal receptive, and unprotected anal insertive, respectively), suggesting some amount of movement between abstinent and sexually active states over short intervals. Figure 2.4 show the joint trajectories of multiple contact types. Unprotected oral receptive contacts seem to be the most independent contact type. In other words, variance in the number of oral contacts does not necessarily correspond to an increase in other contact types. Anal contacts seem to be more correlated with an increase in a reported number or receptive anal contacts corresponding to a similar increase in the reported number of insertive anal contacts.

Simple stochasticity accounts for at least some of the between- and within-individual variance observed in the data. Figure 2.5 shows the variance in the reported number of contacts as a function of the average number of contacts for each contact type. The black line in each plot represents the expected relationship if everyone in the population had stable contact rates. All three contact types clearly show a kind of similar structural deviation from the expectation for a simple Poisson process with individuals with the highest contact rates having much higher variation in the reported number of contacts. The upper limit on the relationship between mean and variance in each plot is the theoretical maximum variance for a given mean. This occurs when an individual has zero contacts in all but one interval.

Figure 2.6 shows mean-variance relationship for a sample of individuals simulated under the homogeneous, gamma heterogeneous, and interval independent models.

Increasing heterogeneity of the contact rate distribution increases the spread of the points but does not capture the deviation from the Poisson expectation in the observed data (figure 2.5). Adding volatility increases the slope of the variance relative to the mean making the pattern more similar in appearance to the pattern in the observed data. As the probability that contact rates are re-drawn between each interval increases, fewer and fewer points lie on the line defining the theoretical maximum variance. This is due to the fact the probability of observing multiple very small contact rates is much smaller than the probability of drawing one very small contact rate.

2.4.2 Quantitative

The initial qualitative analysis suggests that contacts rates to be both heterogeneous and volatile. However, while qualitative analysis can provide insight, it is fundamentally subjective. For a more objective analysis we calculate the likelihood of the data under a sequence of formal models. The intent is two-fold: examine the change in likelihood of the data under a sequence of models that step-wise introduce contact rate heterogeneity between individuals and among individuals over time, and estimate the marginal likelihoods of each model to get an image of the relative contribution of each parameter to the likelihood of the data.

Conceptually, there are three classes of models that we will consider: the homogeneous model that assumes homogeneity of contact rates both between and among individuals, the Gamma heterogeneity model that assumes Gamma distributed heterogeneity in contact rates between individuals but homogeneity in contact rates among individuals, and, finally, the volatility models that assumes Gamma distributed heterogeneity in contact rates between individuals and heterogeneity in contact rates among individuals. Each model relaxes an assumption of the previous model there-

fore giving us an idea of the relative contribution of each modeling assumption to the likelihood of the data.

While the homogeneous model is *sui generis*, there are innumerable ways of expressing heterogeneity in contact rates both between and among individuals. Contact rate heterogeneity between individuals is specified first by a distribution of individuals into subpopulations and then by distributions of contact rates within each of those subpopulations. We found that a single Gamma distribution gave a reasonable fit to the data; adding additional subpopulations would almost certainly increase the fit to the observed data, but the subsequent analysis makes clear that the class of heterogeneity models is almost certainly inferior to the class of volatility models.

The class of volatility models extends the class of heterogeneity models by requiring an explicit specification of how contact rates change over time. To keep things simple, we choose the two most simple volatility models to represent the general class of volatility models. The interval dependence model assumes that contact rates are independently drawn at the beginning of each behavioral period from a single Gamma distribution. This requires an additional set of parameters for the times at which individuals are observed. These parameters are fixed a priori by the study design and are not estimated from the data. This model can be though of as a 'null' volatility model in that any volatility model that gives likelihoods lower than the interval independence model give no extra information about the process that generated the data than is inherent in the study design. The o-volatility model is the simplest model that specifies an actual mechanism by which volatility is produced. Comparisons between the homogeneous, Gamma heterogeneous model, and the interval dependence model give a general sense of which behavioral assumptions best explain the structure of the data. Further, comparison of the interval indepen-

dence model to the o-volatility model gives an idea of how much better, if at all, the o-volatility model is than the null, interval independence model.

Volatility is a superior explanation of the data. The ML parameters and likelihoods of the data for the homogeneous, Gamma heterogeneous, and interval independence, in tables 2.2, 2.3, and 2.4 respectively. The trend in likelihoods is clear, the Gamma heterogeneity model is a better explanation of the data than the homogeneity model and the interval independence model is better than either of the other two models. This conclusion is consistent with the analysis in the previous section.

The comparison of the interval independence models and the o-volatility model is clear for URO and URA contacts. The simple o-volatility model is a significantly better explanation of the data than the interval independence model. The picture for UIA is not quite as clear. The Akaike information criterion (AIC) is a commonly used measure of the relative goodness of fit of multiple models of the data. AIC is not a formal statistical test in the sense of the likelihood ratio test, but rather a guide for balancing model complexity and accuracy [3]. The difference in AIC for the two models is 34 (favoring the o-volatility model). AIC differences greater than 10 from the minimum are assumed to imply no support for the baseline model [13].

The general pattern in all three contact types is that a large proportion of the population is invariant over the two year period, with the remainder experiencing short behavioral phases (high volatility). The data cannot speak to the causes of the very high volatility in this population, but these results are consistent with the qualitative assessment of volatility that shows a small sub-population of individuals with highly variable contact rates.

2.5 Discussion

In this chapter I have shown that rates of unprotected sexual contact are heterogeneous. At the level of observational periods, 81-88% of all 6 month observational periods show no sexual contact; the maximum contact rates are 140, 159, 224 for oral receptive, anal receptive, and anal insertive, respectively. At the individual level, abstinence was less frequent. The proportion reporting abstinence was more divergent between contact types: individuals were the least likely to abstain from insertive anal sex (58%) and most likely to abstain from receptive anal sex (72%). The abstinence rates in our study are much higher than those reported in one of the first large studies of homosexual behavior that reported only 3% of the study population had no partners in the previous year [75]. However, that study was from the pre-AIDS era, and it asked about the number of total sexual partners rather than the number of unprotected sexual contacts, although condom use in urban gay populations was almost non-existent pre-AIDS [54]. The rate of abstinence from unprotected receptive anal sex (72%) is consistent with other estimates (63%) from the same time period [21].

Unprotected sexual contacts are also volatile. Visual inspection of the data show a pattern of reported contacts for unprotected oral receptive and unprotected anal insertive that appeared to be highly variable. The fluctuations for unprotected anal receptive contacts seems to be of a smaller magnitude than the other contact types. It is possible that receptive anal sex was occurring at a steady rate with longer-term partners and that oral receptive and anal insertive sex, which are perceived to be less risky, were occurring more sporadically with casual partners. However, a larger proportion of individuals reported the pattern of having no sexual contacts in all

but one period (156/882, 151/882 and 198/882 for oral receptive, anal receptive, and oral insertive) than expected. The explanation for this pattern is very important. The data have no measurements of the partnership status of each contact, but one explanation for the high frequency of such an extreme pattern would be the new formation of a relatively short-term sexual partnership. Contacts made within long-term partnerships can have very different implications for HIV transmission dynamics than those made without [10] [45].

The volatility of unprotected sexual contact rates is further supported by the likelihood analysis. The log likelihoods of both of the stability models (homogeneous and gamma heterogeneous) are so much larger (farther from 0) than the two volatility models that no special consideration is needed to rule them out as reasonable explanations of the data. Likewise, the o-volatility model is clearly the best explanation (of the considered models) for unprotected oral receptive and unprotected anal receptive contacts. For unprotected insertive anal sex the picture is less clear; the data are more likely under the o-volatility model, but that model requires two more parameters to specify than does the interval independence model. The difference in the Akaike information criterion (AIC) for the two models is 34 (favoring the o-volatility model). AIC differences greater than 10 from the minimum are assumed to imply no support for the model [13]. Technically, the o-volatility is the best model for all three contact types.

The o-volatility model is a better fit to the data than the interval independence model for two reasons: 1) the o-volatility model explicitly models the probability that an individual's behavior will be invariant over the course of the study, making the likelihood of the abstainers much higher. In the interval independence model each interval with no reported contacts is treated as an independent datum, which

given the highly heterogeneous contact rates could lead to very low likelihoods for abstinent intervals under this model; 2) The o-volatility model accounts for some of the correlation between observational periods in a single individual. The interval independence model accounts for the odd pattern where an individual has no contacts in all but one interval; however, for the vast majority of individuals contact rates in one interval predicted contact rates in the subsequent intervals to some degree.

The MLE parameters of the o-volatility model suggest that in each case the majority of individuals are invariant over the course of the study (two years). In the remaining population, contact rates are highly volatile. The average duration of a behavioral interval is less than the duration of an observational period for all three contact types, and, for unprotected anal receptive sex the average behavioral interval is only 1.4 months. It may seem absurd that, on average, every 1.4 months, individual's re-draw their contact rate from a highly heterogeneous distribution (the standard deviation is 4-fold greater than the mean). This interpretation is, however, overreaching. Contact rates are latent variables than can not be directly observed, and a high contact rate in no way implies that an individual is making a large number of contacts during that period. In fact, the opposite is true. As behavioral phases become very short with respect to the duration of an observational period, the number of contacts reported in a observational period is actually less variable. This phenomena is explainable by the properties of the sampling distribution of the mean. The expected number of contacts made by an individual over a observational period is equal to the average contact rate during that period. According to the MLE of the volatility of unprotected insertive anal contact, the average number of behavioral phases that will occur during an observational period is $6/1.4 \approx 4.3$; the mean contact rate over that period is given by the sampling distribution of the mean

$\bar{x} \stackrel{d}{=} N(\mu, \frac{\sigma^2}{n})$ where n is the number of samples (i.e. behavioral intervals experienced in each observational period); the variance of the observed number of contacts is inversely proportional to the volatility rate.

A review of the literature did not find any studies that used comparable methods to estimate contact rate volatility parameters. Blower et al. estimated a sequence of transition matrices giving the probability of men in the Amsterdam cohort (1984-1988) switching between insertive-only, receptive-only, versatile, and abstinent sexual roles with respect to anal sex [11]. They found the highest rates of transition from all other states into the abstinent state, however, the system exhibits a large degree of role switching overall. An abstinent individual (no risk) has a 7% chance over a six month period of moving into the receptive-only state (highest risk). They also found that the transmission matrix was stable over time concluding '...the process of risk behavior change appeared to be time independent' which supports the similar assumption we made in this chapter that the distribution of contact rates is stable over time. We cannot directly compare the degree of volatility observed in this chapter and that measured by Blower et al. as they are measuring two fundamentally different aspects of sexual risk (contact rates, and contact types). An analysis that incorporated both changes in contact rates and in sexual role should be possible with the data at hand.

What we presented here is only a very small aspect of the larger issue of the dynamics of sexual behavior. We are limited by the scope of the data in that we can only address dynamic changes that occur at the scale of the study. The dynamics of sexual behavior are probably affected by a wide range of forces acting at multiple time scales. An ideal data set to study the dynamics of sexual behavior would include more detailed questions about the partnership aspects of each contact and would

sample individuals at different rates (i.e. have variable observational periods) to get at fluctuations that were occurring at shorter or longer time scales.

2.5.1 Data to further clarify the extent and nature of contact rate volatility

The data from the CDC Collaborative HIV Seroincidence Study is limited in multiple ways. The data from the study does not include any information about the relationship status of each contact. Without this key information we only have a limited amount of logic and speculation available to dissect how volatility occurs in the broader context of casual and longer-term partnerships. Ideally, a dataset would include not only the additional stratification of contacts into those made with casual or longer-term partners but also the identity of whom the contacts were made. With that kind of data a very clear image of how sexual behavior changes over time could be constructed. However, confidentially issues and recall biases probably make such a dataset impossible. The second key issue is that a single prospective study with periodic sampling is only able to address volatility that occurs on a scale that is comparable to the frequency of measuring in the study design. For example, for a study with a fixed number of observational periods that measures its participants on a monthly period captures volatility that occurs on the scale of months but would miss variation that occurs on the scale of years. Determining the 'natural' scales of contact rate volatility can only be determined by analysis of multiple studies that sample individuals at variable frequencies. Any prospective study of sexual behavior in gay men could, in theory, contribute to the assessment of contact rate volatility. This could include any study such as the CDC Collaborative HIV Seroincidence Study that were incidence rate surveys that also measured behavioral patterns.

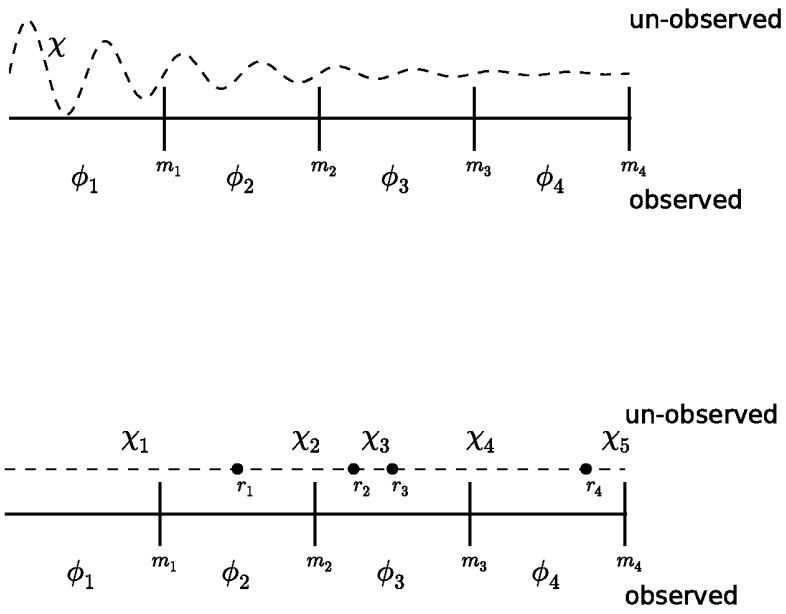

Figure 2.1: Illustration of contact rate volatility and the interval independence model The top figure depicts the concept of contact rate volatility in the context of a prospective cohort. The solid black line represents the duration of the study wherein measurement of the number of contacts, ϕ, in the previous interval are measured at times m. The dashed line indicates the unobservable contact rate that determines ϕ. The bottom panel illustrates the interval independence model which assumes that individuals re-draw their contact rates from the population distribution at a constant rate. The average contact rate in an observational period is a function of the duration of behavioral phases and the contact rates over those phases.

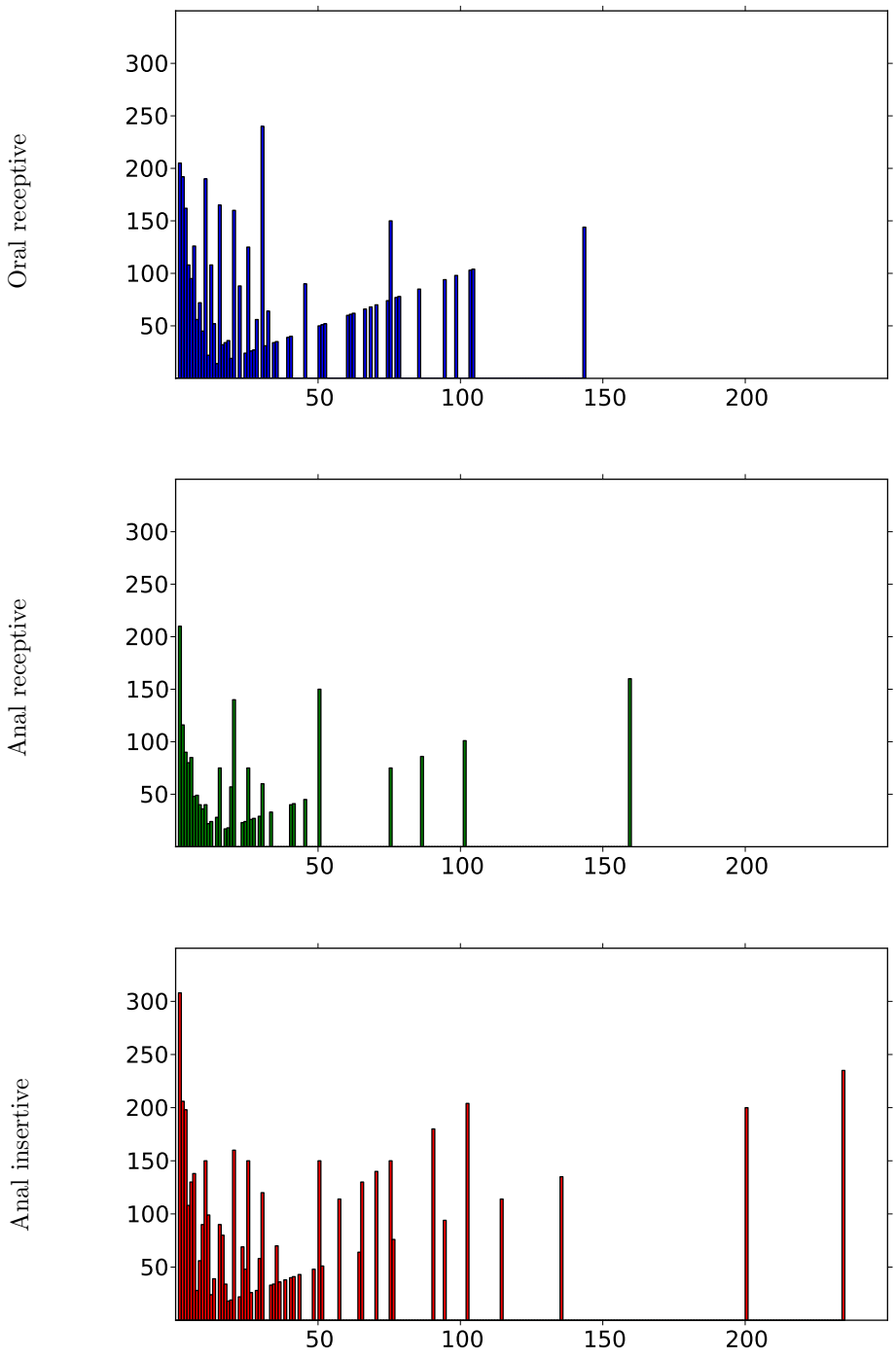

Figure 2.2: Empirical oral and anal contacts rates
This figure shows the number of oral and anal contacts reported over 6 month intervals for 882 gay men observed for a period of two years. The data are weighted by the value of the contact rates such that the denominator is the total number of contacts made by individuals in each category. Even though individuals with the highest contact rates are a small fraction of the population, they represent a large proportion of the total number of contacts.

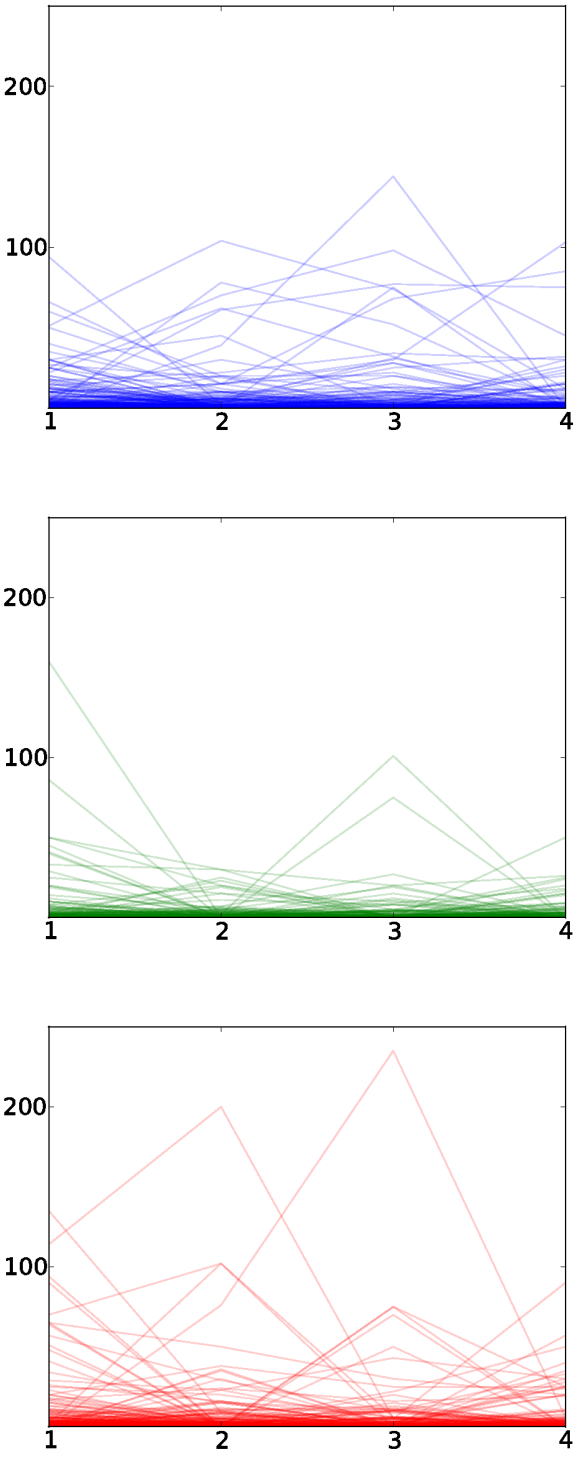

Figure 2.3: Number of contact contacts over a two year period
This figure shows the number of URO (top), URA (middle), UIA (bottom) on the y-axis reported over the previous six months for a total of two years of observation at a sequence of 4 observational periods (x-axis). Each line represents the trajectory of one individual.

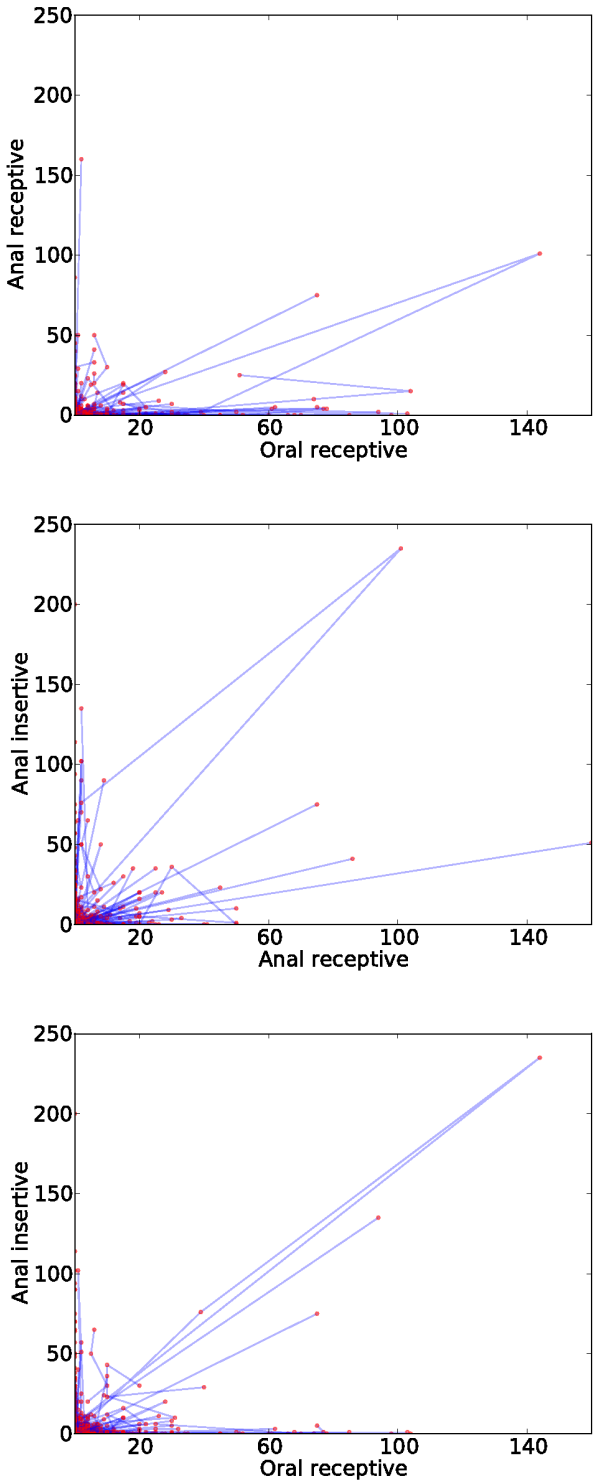

Figure 2.4: Number of contact contacts over a two year period
This figure shows the joint trajectory of URO and URA (top panel), URA and UIA (middle panel), URO and UIA (bottom panel). Each line represents the trajectory of a single individual with a red dot representing an observation.

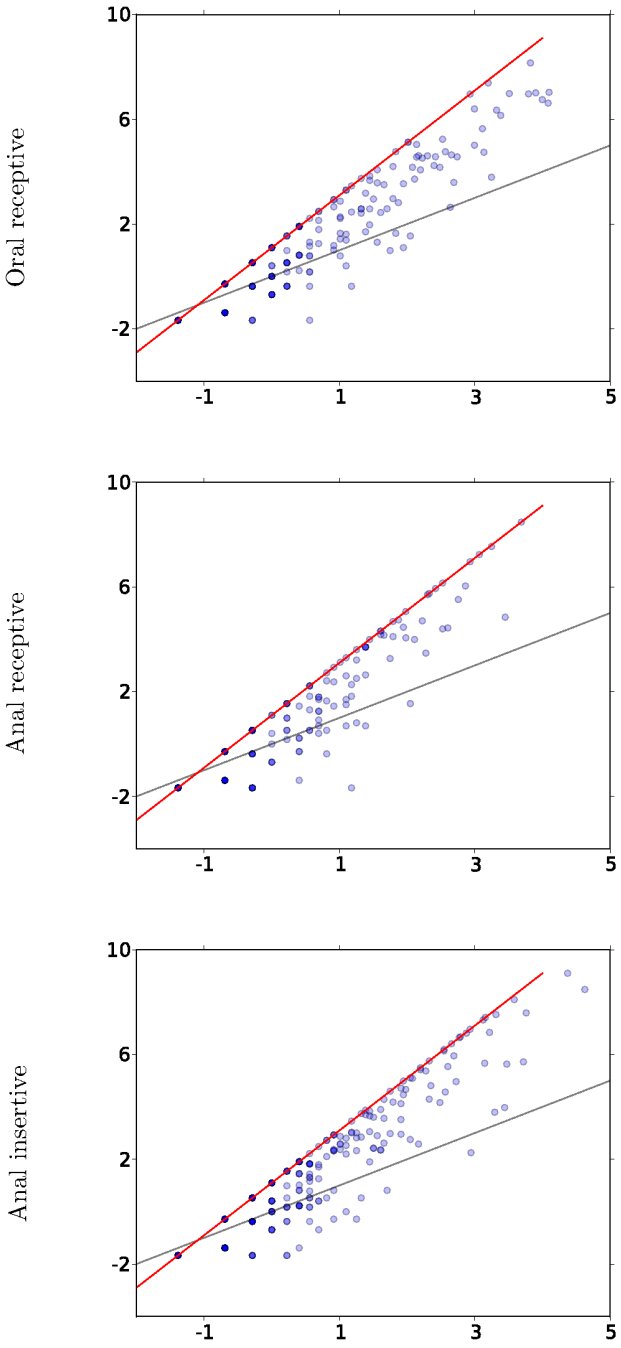

Figure 2.5: Empirical log-scale excess variance
This plot show the log-log scale variance and mean of oral and anal contacts of 882 men. Each point indicates the mean (x-axis) and variance (y-axis) on the log-scale of a single individual individual observed for four six-month periods. The black line shows the expected variance given the mean assuming a constant rate Poisson process. If such a model was a good fit, we should see points evenly distributed above and below the red line. However for contact types there seems to be a distinct trend for variances above the expectation for a Poisson model especially for higher contact rates. The red line show the maximum variance (where all of the contacts occur in one observational period) for each value of the mean.

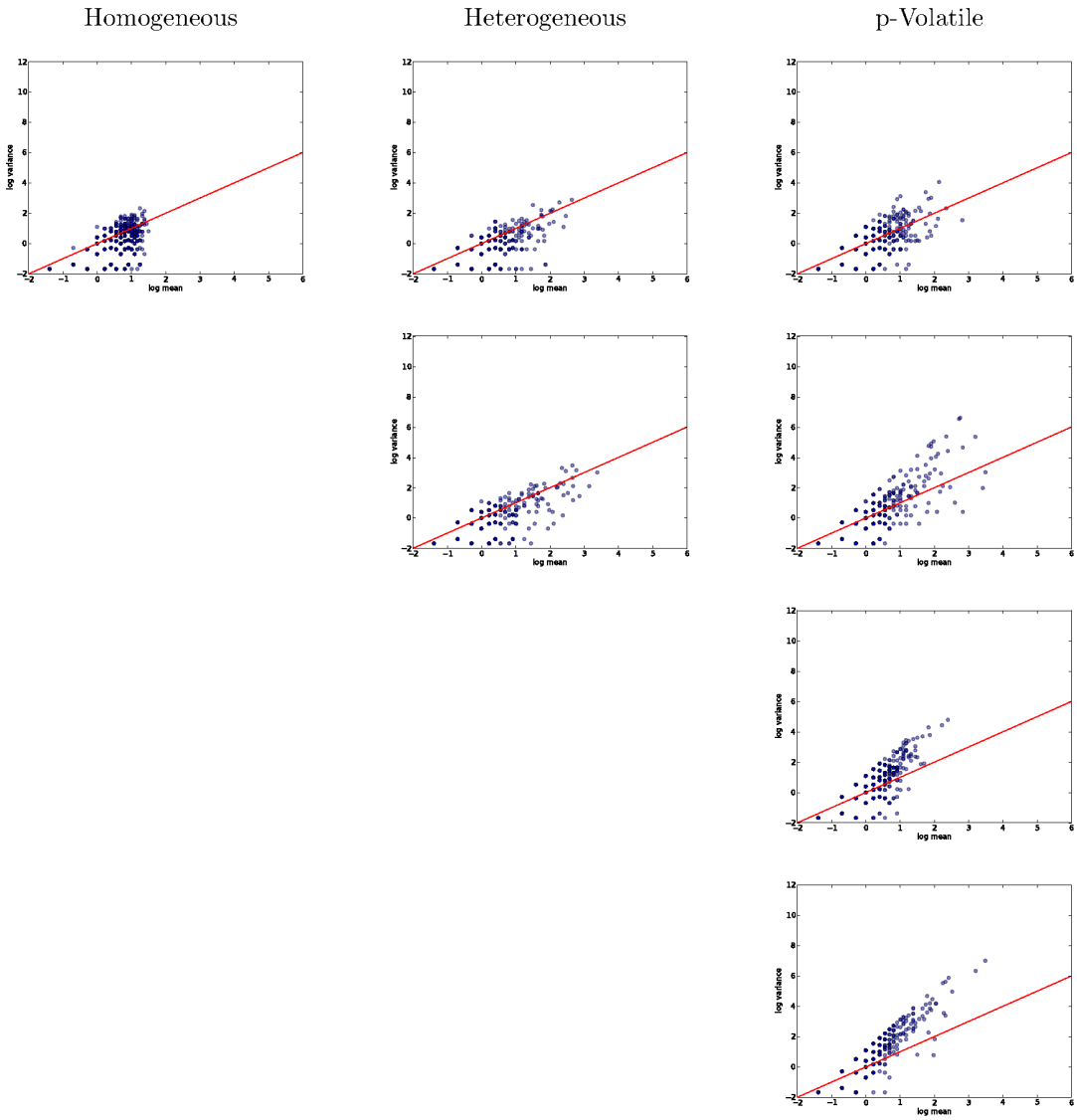

Figure 2.6: Simulated log-scale excess variance
This plot shows the log-log scale variance (y-axis) and mean (x-axis) of observed contacts for 800 simulated individuals. According to either a homogeneous, heterogeneous, or volatile contact rate model, for each individual four contact frequencies were determined corresponding to the four observational intervals in the original dataset. The average number of contacts was fixed at two for each simulation. In the homogeneous and volatile models contact rates are assumed to be log normally distributed with either high (10) or low (2) variances. In the volatility model each individual has a probability of re-drawing their contact rate from the population distribution p before each observation. Both high ($p = 0.75$) and low ($p = 0.25$) volatility was considered. Each point represents one individual.

type	zeros	count	mean contacts	mean std.
URA	0	20	5.64	3.12
	1	28	2.72	2.44
	2	46	2.94	4.02
	3	151	1.45	2.51
	4	637	0	0
UIA	0	39	10.90	6.07
	1	51	5.07	5.93
	2	86	3.55	4.86
	3	198	1.21	2.10
	4	508	0	0
URO	0	40	14.71	8.98
	1	44	5.02	4.83
	2	54	2.84	3.43
	3	156	0.82	1.41
	4	588	0	0

Table 2.1: Average and standard deviations of the number of contacts stratified by number of abstinent intervals
Contact types are unprotected receptive anal (URA), unprotected insertive anal (UIA), unprotected receptive oral (URO). Only individual observed for all 4 observational periods (882) are included. Individuals were stratified based on the number of intervals with no reported contacts. The reported values are average number of contacts made by an individual in a six month period and the average individual standard deviation.

data	χ	log likelihood
URO	1.23	-13240
URA	0.62	-7575
UIA	1.39	-15029

Table 2.2: Homogeneous model parameter estimates and likelihood
The MLE average contact rate per six-month period, χ, in contacts per month and likelihood of the data in the homogeneous model. Contact types are assumed to be independent of one another.

data	mean	std. dev.	τ	log likelihood
URO	3.71	5.23	0.67	-7777
URA	2.21	2.84	0.72	-6268
UIA	3.29	4.50	0.58	-10016

Table 2.3: Gamma heterogeneous parameter estimates and likelihood
The MLE average contact rate per six-month period and the standard deviation of the contact rate, the probability that an individual has a zero contact rate, and the likelihood of the data in the gamma heterogeneous model. Contact types are assumed to be independent of one another.

data	mean	std. dev.	log likelihood
URO	1.24	5.37	-2997
URA	0.62	2.91	-2150
UIA	1.39	5.37	-3387

Table 2.4: interval independence parameter estimates and likelihood
The MLE average contact rate per six-month period and the standard deviation of the contact rate, and likelihood of the data in the interval independence model. Contacts types are assumed to be independent of one another.

data	mean	std. dev.	$1/\rho$	α	log likelihood
URO	1.81 (1.11, 1.20)	4.6 (4.3, 4.9)	4.47 (3.7, 5.2)	0.69 (0.67, 0.71)	-2730
URA	1.17 (1.11, 1.20)	5.2 (4.9, 5.6)	2.16 (1.77, 2.92)	0.61 (0.61, 0.63)	-1943
UIA	1.72 (1.68, 1.78)	6.9 (6.7, 6.9)	1.38 (1.38, 1.38)	0.53 (0.51, 0.53)	-3378

Table 2.5: o-volatility parameter estimates and likelihood
The MLE estimates of the average contact rate per six-month period, the standard deviation of the contact rate, the volatility parameter, the proportion of non-volatile individuals, and likelihood of the data in the interval independence model. Values in parentheses are the bounds of the 0.95 confidence intervals.

CHAPTER III

Contact rate volatility and the epidemic threshold

3.1 Summary

The basic reproduction number, R_0, is a basic statistic that allow us to describe the conditions under which an epidemic can occur given a theoretical model of transmission. In this chapter we use the theory of stochastic branching processes to derive R_0 in three basic transmission models of HIV with contact rate volatility: a one-stage SIR model, a two-stage SIR model, and a two-stage SIR model with two contact types.

3.2 Introduction

The basic reproductive number, R_0, is the expected number of infected persons that an average infected person will generate in a fully susceptible population. In a homogeneous SI model, where the hazard of infection is proportional to the absolute or relative frequency of infected individuals, the knowledge of R_0 can also give other quantities, such as the endemic prevalence. Though those relationships do not always exist in more complex models, R_0 will always correspond to the initial exponential growth rate of an epidemic via the serial interval and given strict eradication criteria– if control measures lower R_0 below one, then the infection cannot persist in the population. In this way, R_0 can be thought of as a theoretical measure of the level

of difficulty of infection control as a function of its transmission parameters [7]. Many HIV transmission parameters such as the relative acute and chronic stage transmission rates are thought of as fixed. However, other parameters are variable, including those describing social contexts of sexuality, heterogeneity in the contact rate distribution or the patterns of sexual mixing. HIV transmission is complex; no one simple model is likely to capture all of the relevant aspects of HIV transmission, so having access to R_0 for a wide rage of theoretical models provides an essential link between theory and intervention design. In this paper we use branching processes to calculate R_0 in a sequence of transmission models that relax the common assumption that an individual's contact rates are constant over time.

Contact rate heterogeneity has been long recognized as a key deviation from the type of homogeneous SI models that are generally used to model sexually transmitted infections [6]. Empirical measurements of sexual contact rates show much more extreme values than are predicted by assumptions of homogeneous behavior; many people abstain from sex while others are more active [30]. These more active people comprise a much larger proportion of the total sexual contacts, leading to a highly skewed distribution of sexual contact rates. We will refer to this phenomena of higher than expected variance in the contact rate distribution as contact rate heterogeneity (CRH). Anderson and May worked out a closed form expression for the basic reproductive number in an SIR model with simple contact rate heterogeneity [7], showing that R_0 increases with increasing variance in the contact rate distribution. This implies that, all other things being equal, contact rate heterogeneity lowers the epidemic threshold and makes elimination of the disease more challenging.

We extend the assumption of heterogeneous contact rates between individuals to include heterogeneity within a single individual over time, which we call contact

rate volatility (CRV). In a stochastic system, individuals will show some degree of apparent contact rate volatility due to simple stochastic variance, which will be Poisson distributed and relatively small. However, an individual's contact rate may also be changing over time, adding a structural component to the variance in the number of contacts individuals make over time. This type of contact rate volatility can take many forms: slow-changing secular trends, faster periodic fluctuations that may correspond to relationship dynamics, or brief aperiodic episodes of high-risk behavior. We focus on a single type of unstructured contact rate volatility where the probability of drawing a given contact rate is constant over time and individuals re-draw their contact rates at a constant rate. In this paper we use the theory of branching processes to derive R_0 in a series of models including a model with multiple contact types, multiple stages of infection, and heterogeneous, volatile contact rates.

3.3 Materials and methods

The work in this chapter is based on the theory of stochastic branching processes [8]. A branching process is a Markov process that models the probability of observing a population of size $\Pr(s)$, at time $n+1$, given the number of individual, s, at time n and some known probability distribution of offspring generation for each individual $\{1, 2, 3, \cdots\}$. Galton and Watson applied these methods to study the probability of extinction of aristocratic surnames [42]. Branching processes can likewise be used to describe the number of infected individuals in the next generation given the current number of infected individuals. Branching processes are analytically treated through the use of generating functions.

3.3.1 Generating functions

The derivations in this section are based on the use of probability generating functions, PGFs, which can be used to describe the properties of functions of independent random variables. For a discreet random variable, X, with probability mass function $\Pr(x)$ the PGF of X is defined as

$$G_X(z) = E(z^X) = \sum_x^{\infty} \Pr(x) z^x \tag{III.1}$$

and as $\int \Pr(x) z^x dx$ for a continuous random variable. In this paper we use distributions with relatively simple PGFs.

PGFs have some nice properties that make the calculation of arithmetic operations on sequences of independent random variables straightforward. Some of those properties are listed below.

- $G_X(1) = \Pr(x_1)1^{x_1} + \Pr(x_2)1^{x_2} + \cdots + \Pr(x_f)1^{x_f} = 1$
- If a random variable has a known moment generating function, $M_X(z)$, then its PGF is given by $G_X(z) = M_X(\log(z))$.
- $G'_X(1) = 0\Pr(0)1^0 + 1\Pr(1)1^1 + \cdots + n\Pr(n)1^n = \mathbb{E}(X)$
- The higher central moments can be calculated by further differentiation and algebraic manipulation. For example $\mathbf{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2$
- The sum of a sequence of independent random variables, $S = X_1 + X_2 + \cdots + X_n$ can be represented as the product of n generating functions: $G_S(z) = E(z^{\sum_i^n X_i}) = \prod_i^n G_{X_i}(z)$
- If n is a random variable, then the sum of random variables, $S = X_1 + X_2 + \cdots + X_n$, is generated by $G_{S_n}(z) = G_n(G_s(z))$

For example, if the number of sexual partners that I have over a period of time is a Poison random variable, X, generated by $G_X(z) = e^{-\lambda+\lambda z}$ and the probability that any one of my partners is infected is a Bernoulli random variable, Y, generated by $G_Y(z) = ((1-p)+pz)$, then the average number of infected contacts that I have over an interval is generated by:

$$G_X(G_Y(z)) = G_N(z) = e^{-\lambda+\lambda((1-p)+pz)} \tag{III.2}$$

$$G'_N((z)) = \lambda p e^{-\lambda+\lambda(1-p+pz)} \tag{III.3}$$

$$G'_N((1)) = \mathbb{E}(N) = \lambda p. \tag{III.4}$$

This result illustrates how PGFs can be used to describe functions of random variables and how central moments of functions of random variables can be calculated using PGFs. For many HIV models, generating functions can be used to derive the central moments of the number of transmissions that are generated by a newly infected person. In models were this is possible, R_0 can be found using the properties of generating functions by either analytical or numeric methods.

3.3.2 Contact rate heterogeneity model

We model CRH as $\chi_i \stackrel{d}{=} \Gamma(k,\theta)$ that the probability that a new individual has contact rate χ_i is $\Pr(\chi = \chi_i) = \chi^{k-1}\frac{exp(-\chi/\theta)}{\Gamma(k)\theta^k}$. We chose to model CRH as a Γ random variable for two primary reasons. First, the Gamma distribution has a convergent closed form for its probability generating function, making it easier to find closed form expressions for R_0, where possible. We also considered using a log-normal model, but this was rejected as there is no simple form for the log-normal PGF that is valid over the domain of its parameters, and this would have required extensive numerical

simulation to obtain results. Second, the Gamma distribution phenomenologically captures several of the features of observed contact rate distributions such as long tails and a high degree of positive skew. Also, at certain parameterizations (at non-zero integer values of the shape parameter) this type of CRH can be simulated using the traditional compartmental models that are common in epidemiology.

3.3.3 Contact rate volatility model

We model contact rate volatility in the simplest way possible. Individuals experience periods of stable behavior, i.e. their contact rates are constant over the duration of a behavioral interval. At the beginning of a new interval, an individual re-draws new contact rates from the distribution defined by the CRH model. Termination of a behavioral interval occurs at the same rate for all individuals such that the length of behavioral intervals are Poisson distributed. The entire contact rate volatility model is specified by a single rate parameter. We conceptualize contact rate volatility as being unstructured making it fundamentally different from other types of temporal variance in contact rates such as age-structured contact rates. A volatility model could be made more complex by considering multiple types of behavioral intervals with different lengths, potentially correlated contact rates, or more complex. However, more complex models not only get away from the fundamental effects of volatility on R_0, but they also make closed form solutions less probable.

3.3.4 The transmission model

To illustrate how contact rate volatility produces its effects we calculate R_0 for three successively more realistic models of HIV transmission. For clarity, we will refer to these models by their number. Each model is based on a basic homogeneous SI model.

1. An SI model with contact rate heterogeneity and volatility, one infectious stage, and one contact type
2. An SI model with contact rate heterogeneity and volatility, two infectious stages, and one contact type
3. An SI model with contact rate heterogeneity and volatility, two infectious stages, and two contact types

Model 1

Individuals exist in three possible states: susceptible, indicated by S; infected, indicated by I; and removed indicated by R. N includes individuals in S, I, or R. Contacts are instantaneous and symmetric. The model parameters are listed in table 3.1. The time evolution of the system is represented by the following independent rate equations where $S^* = \sum^{i \in S} \chi_{i,t}$ and $I^* = \sum^{i \in I} \chi_{i,t}$

$$\xrightarrow{\epsilon} S_\chi \tag{III.5}$$

$$S_\chi \xrightarrow{\zeta} I_\chi \tag{III.6}$$

$$S \xrightarrow{\omega S} R \tag{III.7}$$

$$I \xrightarrow{\omega I} R \tag{III.8}$$

$$I \xrightarrow{\delta I} R \tag{III.9}$$

where $\zeta = S^* \beta \frac{I^*}{N^*}$.

Model 2

In model 2, N individuals exist in four possible states: susceptible, indicated by S; acutely infected, indicated by A; chronically infected, indicated by C; and removed indicated, by R. The model parameters are listed in table 3.2. The time

evolution of the system is represented by the following independent rate equations where $S^* = \sum^{i \in S} \chi_{i,t}$, $A^* = \sum^{i \in A} \chi_{i,t}$, and $C^* = \sum^{i \in C} \chi_{i,t}$.

(III.10) $$\xrightarrow{\epsilon} S_\chi$$

(III.11) $$S_\chi \xrightarrow{\zeta} A_\chi$$

(III.12) $$S \xrightarrow{\omega S} R$$

(III.13) $$A \xrightarrow{\delta_a S} C$$

(III.14) $$A \xrightarrow{\omega A} R$$

(III.15) $$C \xrightarrow{\delta_c C} R$$

(III.16) $$C \xrightarrow{\omega C} R$$

where $\zeta = S^*(\beta_a \frac{A^*}{N^*} + \beta_c \frac{C^*}{N^*})$.

Model 3

Model 3 builds on model 2 by adding oral and anal contacts. As before, all contacts are instant and symmetric. However, individuals have separate oral and anal contact rates that are associated with different transmission probabilities. N individuals exist in four possible states: susceptible, indicated by S; acutely infected, indicated by A; chronically infected, indicated by C; and removed indicated by R. The model parameters are listed in table 3.3. The time evolution of the system is represented by the following independent rate equations where the superscripts o indicates oral, a indicates anal and $S^{\{o,a\}} = \sum^{i \in S} \chi_{i,t}^{\{o,a\}}$, $A^{\{o,a\}} = \sum^{i \in A} \chi_{i,t}^{\{o,a\}}$, and

$C^{\{o,a\}} = \sum^{i \in C} \chi_{i,t}^{\{o,a\}}$

(III.17) $$\xrightarrow{\epsilon} S_{\chi}$$

(III.18) $$S_{\chi_o} \xrightarrow{\zeta_o} A_{\chi_o}$$

(III.19) $$S_{\chi_a} \xrightarrow{\zeta_a} A_{\chi_a}$$

(III.20) $$S \xrightarrow{\omega S} R$$

(III.21) $$A \xrightarrow{\delta_a S} C$$

(III.22) $$A \xrightarrow{\omega A} R$$

(III.23) $$C \xrightarrow{\delta_c C} R$$

(III.24) $$C \xrightarrow{\omega C} R$$

where $\zeta^o = S^o(\beta_a \frac{A^o}{N^o} + \beta_c \frac{C^o}{N^o})$ and $\zeta^a = S^a(\beta_a \frac{A^a}{N^a} + \beta_c \frac{C^a}{N^a})$.

3.3.5 Verification of the methods

We verify our results using two methods. First, we recover an established R_0 for a transmission model without contact rate volatility. Second, we use individual-based simulations to directly measure the number of secondary cases generated by an infected index case.

Anderson and May [7] calculated R_0 for an SI model with heterogeneous contact rates as

(III.25) $$R_0 = \frac{\langle \chi^2 \rangle}{\langle \chi \rangle} \frac{\beta}{\lambda}$$

where β is the probability of transmission, λ is the removal rate, and $\langle \chi^2 \rangle$ and $\langle \chi \rangle$ are the second and first raw moments of the contact rate distribution, respectively. Assuming $\chi \sim Gamma(k, \theta)$ and noting that $\langle \chi^2 \rangle$ and $\langle \chi \rangle$ can be written as $m + \sigma^2/m$ in terms of the first, m, and second, σ^2, central moments, then R_0 can be defined as

(III.26) $$R_0 = \theta(k+1)\frac{\beta}{\lambda}.$$

The hazard of infection is proportional to the contact rate such that the PGF for the contact rate of newly infected persons is $G_{\widetilde{\chi}}(z) = \frac{zG'_{\chi}(z)}{G'_{\chi}(1)}$ where $G_{\chi}(z) = (1-\theta\log(z))^{-k}$ is the PGF of the gamma distribution. The average infectious period is the reciprocal of the removal rate, $1/\lambda$, over which contact rates are generated by $G'_{\widetilde{\chi}}(z)$, β of which are sufficient to transmit. Then

$$R_0 = \frac{G'_{\widetilde{\chi}}(1)\beta}{\omega+\delta} \tag{III.27}$$

$$= \frac{\theta + k\theta}{\omega+\delta}\beta. \tag{III.28}$$

Both methods produce the same R_0 for an SI model with gamma-distributed contact rates.

Results were also confirmed by simulation. For each model an individual based simulation was written in Java (details for each simulation are found in the appendix) and simulated using Gillespie's exact algorithm. R_0 was measured by seeding the population and selecting the first incident infected as the index case. All further transmissions were disabled and the number of sufficient contacts made by the index case were counted. For each parameter set the results were averaged over 300 runs.

3.4 Results

3.4.1 The basic reproduction number in model 1

From the time of infection, t_0, to the time of removal, t_f, an infected individual re-draws their contact rate at times, $t_1, \cdots, t_f$. We consider separately the first behavioral interval and the remaining $f-1$ intervals. After the first interval the number of subsequent intervals is a geometric random variable that is generated by

$$G_I(z) = \frac{\phi}{1-(1-\phi)z}. \tag{III.29}$$

The length of an interval, $T = t_i - t_{t-1} \sim Exponential(\lambda)$. A uniformly selected random individual's contact rate, χ_i, over interval t_j is a Γ random variable generated

by

$$G_\chi(z) = (1 - \theta \log(z))^{-k}. \tag{III.30}$$

However, the hazard of infection is proportional to the contact rate [28] such that the contact rate of newly infected persons, $\widetilde{\chi}$ is generated by

$$G_{\widetilde{\chi}}(z) = \frac{zG'_\chi(z)}{G'_\chi(1)}. \tag{III.31}$$

In the first behavioral interval, (t_0, t_1), the number of contacts made by a newly infected person is Poisson random variable generated by

$$G_{\widetilde{C}}(z; T, \widetilde{\chi}) = e^{-T\widetilde{\chi}(1-z)}. \tag{III.32}$$

Integrating over T and $\widetilde{\chi}$ yields $G_{\widetilde{C}}(z)$ for which there is no closed form expression. However, $G_{\widetilde{C}}(z)$ can be written as

$$G_{\widetilde{C}}(z) = \int \lambda e^{-\lambda T} \int \Pr(\widetilde{\chi}) e^{(z-1)^{\widetilde{\chi}T}} \, dT \, d\widetilde{\chi} \tag{III.33}$$

$$= \int \Pr(T) G_{\widetilde{\chi}}(e^{T(z-1)}) \, dT. \tag{III.34}$$

Likewise, the number of contacts for every over interval after the first is generated by

$$G_C(z) = \int \Pr(T) G_\chi(e^{T(z-1)}) \, dT. \tag{III.35}$$

Using the properties of generating functions we can now define a generating function of a random variable H for the number of contacts made over the course of an infection.

$$G_H(z) = G_{\widetilde{C}}(z) G_I(G_C(z)). \tag{III.36}$$

For clarity, we will differentiate between H, the number of contacts made over the course of an infection, and $\hat{H}$, the number of transmissions generated over the course

of an infection. The product of the expectation of H and the probability of transmission β is the basic reproduction number

$$R_0 = \mathbb{E}(H)\beta = G'_H(1)\beta \tag{III.37}$$

$$= \left[G'_{\widetilde{C}}(z)G_{\widetilde{C}}(z)G_I(G_C(z)) + G_{\widetilde{C}}(z) + G'_I(G_C(z))G'_C(z)\right]_{z=1}\beta \tag{III.38}$$

$$= \left(G'_{\widetilde{C}}(1) + G'_I(1)G'_C(1)\right)\beta \tag{III.39}$$

$$= \left(\frac{\theta(k+1)}{\lambda} + \frac{1-\phi}{\phi}\frac{\theta k}{\lambda}\right)\beta \tag{III.40}$$

$$= \left(\frac{\theta(k+1)}{\lambda} + \frac{\rho}{\delta+\omega}\frac{k\theta}{\lambda}\right)\beta. \tag{III.41}$$

A step-by-step derivation of $G'_{\widetilde{C}}(1)$ and $G'_C(1)$ is given in the appendix VI.

The variance of R_0 can also be calculated, which involves taking the second derivative of $G_H(1)$. We need take a bit more care with the probability of transmission in the second derivative. First, we will formally define a generating function for $\hat{H}$ as the number of actual transmissions generated by a newly infected person in a fully susceptible population. The probability of transmission given contact is a Bernoulli random variable generated by $G_p(z) = (1-\beta) + \beta z$. Then $G_{\hat{H}}(z) = G_H(G_p(z))$ and $G'_{\hat{H}}(z) = G'_H(G_p(z))G'_p(z)$.

$\mathbf{Var}(R_0)$ can be calculated by further differentiation of $G'_{\hat{H}}(z)$ as $G''_{\hat{H}}(1) + G'_{\hat{H}}(1) - [G'_H(1)]^2$. Beginning with the second derivative of $G_{\hat{H}}(1)$

$$G''_{\hat{H}}(1) = [G''_H(G_p(z))G'_p(z)G'_p(z) + G'_H(G_p(z))G''_p(z)]_{z=1} \tag{III.42}$$

$$= G''_H(1)G'_p(1)G'_p(1) \tag{III.43}$$

$$= G''_H(1)\beta^2 \tag{III.44}$$

$$= \left(G''_{\widetilde{C}}(1) + G''_I(1)G'_C(1) + G'_I(1)G''_C(1)\right)\beta^2 \tag{III.45}$$

$$= \left(\frac{(k+2)(k+1)\theta^2}{\lambda} + \frac{2(\phi-1)^2}{\phi^2}\frac{k\theta}{\lambda} + \frac{1-\phi}{\phi}\frac{(k+1)k\theta^2}{\lambda}\right)\beta^2 \tag{III.46}$$

A step-by-step derivation of $G''_{\widetilde{C}}(1)$ and $G''_{C}(1)$ is given in appendix VI. We can then write down a complete expression for the variance of R_0 as

$$\mathbf{Var}(R_0) = G''_{\hat{H}}(1) + G'_{\hat{H}}(1) - \left[G'_{\hat{H}}(1)\right]^2 \tag{III.47}$$

3.4.2 The basic reproduction number in model 2

Model 2 expands on model 1 by including acute and chronic stages of transmission to account for the large spike in viral titer that occurs during acute infection. Inclusion of an additional infection stage is straightforward. We just need to consider additional interval duration rates for acute and chronic stage infectives and the possibility that an infected individual enters the chronic stage before the first behavioral interval terminates. The model parameters are listed in table 3.2.

From the time of infection, t_0, to transition to chronic infection, t_a, an acutely infected individual re-draws their contact rate at times $t_1, \cdots, t_{a-1}$. The total number of behavioral intervals experienced in the acute stage, $a-1$, is a geometric random variable that is generated by

$$G_{I_a}(z) = \frac{\phi_a}{1-(1-\phi_a)z}. \tag{III.48}$$

The remaining time infected t_a, to removal, t_f, a chronically infected individual re-draws their contact rate at times $t_{a+1}, \cdots, t_f$. The total number of behavioral intervals experienced in the chronic stage, $f-1$, is a geometric random variable that is generated by

$$G_{I_c}(z) = \frac{\phi_c}{1-(1-\phi_c)z}. \tag{III.49}$$

The length of behavioral intervals for acute infectives, $T_{t<a}$, is distributed as $Exponential(\lambda_a)$ while for chronic infectives, $T_{t>a}$, is distributed as $Exponential(\lambda_c)$.

The addition of acute and chronic stages requires that we consider two possible ways that an infected individual can progress from acute to chronic infection. If the first behavioral interval extends into the chronic stage, $t_1 > t_a$, then the contact rate in the first interval in the chronic stage is generated by $G_{\widetilde{c}}(z)$. However, if an individual re-draws their contact rate before progressing to the chronic stage, their contact rate will be generated by $G_c(z)$. In the latter case an individual will generate

$$\mathbf{E}(\hat{H}_1) = \left(G'_{\widetilde{c}}(1) + G'_{I_a}(1)G'_c(1)\right)\beta_a + G'_{I_c}(1)G'_c(1)\beta_c \tag{III.50}$$

$$= \left(\frac{\theta(k+1)}{\lambda_a} + \frac{\rho}{\delta_a+\omega}\frac{k\theta}{\lambda_a}\right)\beta_a + \frac{\rho}{\delta_c+\omega}\frac{k\theta}{\lambda_c}\beta_c \tag{III.51}$$

new infections. Otherwise they will generate

$$\mathbf{E}(\hat{H}_2) = G'_{\widetilde{c}}(1)\beta_a + \left(G'_{\widetilde{c}}(1) + G'_{I_c}(1)G'_c(1)\right)\beta_c \tag{III.52}$$

$$= \frac{\theta(k+1)}{\lambda_a}\beta_a + \left(\frac{\theta(k+1)}{\lambda_c} + \frac{\rho}{\delta_c+\omega}\frac{k\theta}{\lambda_c}\right)\beta_c \tag{III.53}$$

new infections.

An acute interval can terminate in one of three ways: re-drawing contact rates, progression to chronic stage, or removal. The probability that an individual re-draws their contact rate at least once before stage progression or removal is the proportion that the contact rate volatility parameter comprises of the total acute interval termination rate, $\frac{\rho}{\rho+\delta_a+\omega} = \frac{\rho}{\lambda_a}$. The basic reproduction number can then be written as as weighted sum of the expectations of $\hat{H}_1$ and $\hat{H}_2$:

$$R_0 = \frac{\rho}{\lambda_a}\mathbf{E}(\hat{H}_1) + \left(1 - \frac{\rho}{\lambda_a}\right)\mathbf{E}(\hat{H}_2) \tag{III.54}$$

The variance of R_0 can also be written as a weighted sum of partial reproduction functions defined as

$$\mathbf{Var}(R_0) = \frac{\rho}{\lambda_a}\mathbf{Var}(\hat{H}_1) + \left(1 - \frac{\rho}{\lambda_a}\right)\mathbf{Var}(\hat{H}_2)$$

where

$$\mathbf{Var}(\hat{H}_1) = G''_{\hat{H}_1}(1) + G'_{\hat{H}_1}(1) - \left[G'_{\hat{H}_1}(1)\right]^2,$$
$$\mathbf{Var}(\hat{H}_2) = G''_{\hat{H}_2}(1) + G'_{\hat{H}_2}(1) - \left[G'_{\hat{H}_2}(1)\right]^2.$$

To finish this calculation we need $G''_{\hat{H}_1}(1)$ and $G''_{\hat{H}_2}(1)$

$$\text{(III.55)} \quad G''_{\hat{H}_1}(1) = \left(G''_{\widetilde{C}}(1) + G''_{I_a}(1)G'_C(1) + G'_{I_a}(1)G''_C(1)\right)\beta_a^2 + \left(G''_{I_c}(1)G'_C(1) + G'_{I_c}(1)G''_C(1)\right)\beta_c^2$$

$$\text{(III.56)} \quad = \left(\frac{(k+2)(k+1)\theta^2}{\lambda_a} + \frac{2(\phi_a-1)^2}{\phi_a^2}\frac{k\theta}{\lambda_a} + \frac{1-\phi_a}{\phi_a}\frac{(k+1)k\theta^2}{\lambda_a}\right)\beta_a^2 + \left(\frac{2(\phi_c-1)^2}{\phi_c^2}\frac{k\theta}{\lambda_c} + \frac{1-\phi_c}{\phi_c}\frac{(k+1)k\theta^2}{\lambda_c}\right)\beta_c^2$$

$$\text{(III.57)} \quad G''_{\hat{H}_2}(1) = \left(G''_{\widetilde{C}}(1)\right)\beta_a^2 + \left(G''_{\widetilde{C}}(1) + G''_{I_c}(1)G'_C(1) + G'_{I_c}(1)G''_C(1)\right)\beta_c^2$$

$$\text{(III.58)} \quad = \left(\frac{(k+2)(k+1)\theta^2}{\lambda_a}\right)\beta_a^2 + \left(\frac{(k+2)(k+1)\theta^2}{\lambda_c} + \frac{2(\phi_c-1)^2}{\phi_c^2}\frac{k\theta}{\lambda_c} + \frac{1-\phi_c}{\phi_c}\frac{(k+1)k\theta^2}{\lambda_c}\right)\beta_c^2$$

A step-by-step derivation of $G''_{\widetilde{C}}(1)$ and $G''_C(1)$ is given in appendix VI.

3.4.3 The basic reproduction number in model 3

Model 3 expands on model 2 by adding oral and anal contacts that are made independently of one another. Multiple routes of transmission add the complication of having to consider the possibility that an average infected individual became infected by either oral or anal sex. R_0 can be calculated though algebraic manipulation of the partial reproduction functions for individuals infected by either oral and anal sex. The approach that I take to derive R_0 is to write down expressions for the partial reproduction functions as a two-by-two matrix and then solve for the dominant eigenvalue, which is the average number of infections an average newly infected person will generate [19].

There are four total partial reproduction functions indicating the number of infections of each type generated by a person who was infected by either oral or anal sex. I will use the notation R_{xy} to indicate how the individual became infected ($x \in \{a, o\}$) and the type of transmissions that they generate ($y \in \{a, o\}$), where a indicates anal and o indicates oral. Continuing the logic used in the derivation of R_0 in model 2 where R^1 represents the number of infections generated by a new infected who does re-draw their contact rate before progressing to the chronic stage and R^2 represents the number of new infections generated by a new infected who does re-draw their contact rate before progressing to the chronic stage. Then, modifying III.51 gives

$$R^1_{oo} = \left(\frac{\theta^o(k^o+1)}{\lambda_a} + \frac{\rho}{\delta_a+\omega}\frac{k^o\theta^o}{\lambda_a} \right)\beta^o_a + \frac{\rho}{\delta_c+\omega}\frac{k^o\theta^o}{\lambda_c}\beta^o_c \tag{III.59}$$

and modifying III.53 gives

$$R^2_{oo} = \frac{\theta^o(k^o+1)}{\lambda_a}\beta^o_a + \left(\frac{\theta^o(k^o+1)}{\lambda_c} + \frac{\rho}{\delta_c+\omega}\frac{k^o\theta^o}{\lambda_c} \right)\beta^o_c \tag{III.60}$$

which after weighting as in III.54

$$R_{oo} = \frac{\rho}{\lambda_a}R^1_{oo} + \left(1 - \frac{\rho}{\lambda_a}\right)R^2_{oo}. \tag{III.61}$$

Because contact rates are assumed to be independent, the timing of the first re-sampling does not effect the total number of anal contacts, greatly simplifying the generating function to

$$\begin{aligned} R_{oa} &= G'_I(1 + G_{C_a}(1)) \\ &= \left(1 + \frac{\rho}{\delta_a+\omega}\right)\frac{k^a\theta^a}{\lambda_a}\beta^a_a + \left(1 + \frac{\rho}{\delta_c+\omega}\right)\frac{k^a\theta^a}{\lambda_c}\beta^a_c. \end{aligned} \tag{III.62}$$

The partial reproduction equation for the number of anal transmissions by an individual infected by anal sex is given by

$$R^1_{aa} = \left(\frac{\theta^a(k^a+1)}{\lambda_a} + \frac{\rho}{\delta_a+\omega}\frac{k^a\theta^a}{\lambda_a}\right)\beta^a_a + \frac{\rho}{\delta_c+\omega}\frac{k^a\theta^a}{\lambda_c}\beta^a_c \tag{III.63}$$

$$R^2_{aa} = \frac{\theta^a(k^a+1)}{\lambda_a}\beta^a_a + \left(\frac{\theta^a(k^a+1)}{\lambda_c} + \frac{\rho}{\delta_c+\omega}\frac{k^a\theta^a}{\lambda_c}\right)\beta^a_c \tag{III.64}$$

$$R_{aa} = \frac{\rho}{\lambda_a}R^1_{aa} + \left(1-\frac{\rho}{\lambda_a}\right)R^2_{aa}. \tag{III.65}$$

The partial reproduction equation for the number of oral transmissions by an individual infected by anal sex is given by

$$\begin{aligned} R_{ao} &= G'_I(1+G_{C_o}(1)) \\ &= \left(1+\frac{\rho}{\delta_a+\omega}\right)\frac{k^o\theta^o}{\lambda_a}\beta^o_a + \left(1+\frac{\rho}{\delta_c+\omega}\right)\frac{k^o\theta^o}{\lambda_c}\beta^o_c. \end{aligned} \tag{III.66}$$

R_0 is the dominant eigenvalue of the two-by-two matrix of the partial reproduction functions. The eigenvalues are defined as

$$\frac{R_{aa}+R_{oo}}{2} \pm \frac{\sqrt{4R_{ao}R_{oa}+(R_{aa}-R_{oo})^2}}{2} \tag{III.67}$$

3.5 Results

3.5.1 The effect of volatility on R_0

In model 1, increasing volatility reduces R_0 (figures 3.1, 3.2 and 3.3). The first behavioral interval lasts, on average $\frac{1}{\rho+\delta+\lambda}$, which shortens as ρ increases. When $\rho = 0$ the index case experiences $\frac{\rho}{\delta+\omega} = 1$ behavioral interval over which they have an average contact rate of $(k+1)\theta$. The average contact rate in the first behavioral interval is θ-times larger than the average contact rate in the general population.

When the average contact rate in the general population is low and the variance is high ($k << \theta$), the average contact rate in the newly infected is much higher than in the general population greatly increasing R_0. If $\rho > 0$, the probability that the first behavioral interval (mean contact rate is $(k+1)\theta$) terminates in re-drawing a new contact rate from the general population distribution (mean rate is $k\theta$) is greater than zero. The length of behavioral phases is an exponential random variable with mean $\frac{1}{\rho}$. As volatility increases, the duration of the first interval decreases, reducing the number of contacts made over the course of the infectious period.

Increasing volatility also reduces R_0 in model 2, however, this reduction is mitigated when compared to a model 1 (figure **??**). We have parameterized model 2 to correspond to the natural history of HIV with a short, highly infectious acute stage and a much longer, less infectious chronic stage with about 40% of the infection potential coming from the acute stage. In this model increases in volatility differentially reduce the contact rate in the chronic stage compared to the acute stage. The odds that the first re-drawing from the population distribution of contact rates occurs in the acute versus chronic stage is equal to $\frac{\rho}{\lambda_a}$. For HIV, $\lambda_a = 2$, therefore the the probability of maintaining the average contact rate $(k+1)\theta$ over the entire acute stage is greater than 50% for $\rho > 2$; unless average behavioral interval are very short, the average contact rate of an index case will be elevated above the population average over the course of the acute stage.

In the kind of simple models that we used in this chapter volatility, in its extreme values, provides a link between contact rate heterogeneity with no volatility ($\rho = 0$) and contact rate homogeneity ($\rho = \infty$). When $\rho = 0$, R_0 reduces to $\frac{(k+1)\theta}{\lambda}\beta$ which is equivalent to the general formulation for R_0 in a transmission system with heterogeneous contact rates presented by Anderson and May [7]. As $\lim_{\rho\to\infty} R_0 =$

$\frac{k\theta}{\lambda}\beta$, which is equivalent to R_0 in a homogeneous system where the average contact rate is $k\theta$; that is, if ρ is very large, a model where $k = 1$ and $\theta = 10$ and another where $k = 0.1$ and $\theta = 100$ are equivalent with respect to R_0 to a homogeneous model where $X = 10$ even though the variance of the contact rate distribution is 10-fold greater in the latter model.

The link between homogeneous and heterogeneous contact rates at the extreme ends of volatility helps to explain the relationship between of R_0 to contact rate heterogeneity and contact rate volatility (figure 3.4. The variance of R_0 decreases with increasing contact rate volatility and increases with increasing contact rate heterogeneity for a given mean contact rate. The reduction in variance with increasing volatility can be understood as a type of regression to the mean. The variance of the mean contact rate is inversely proportional to the number of behavioral intervals experienced over the course of an infection. Contact rate volatility makes individuals, on average, more similar over time as multiple behavioral phases smooth out differences between behavioral extremes. Without volatility, in a population with highly heterogeneous contact rates (low mean, high variance) such as that observed in the previous chapter, the variance of R_0 is inflated by the highly unlikely but very large contact rates in the long tail of the contact rate distribution. Put in other words, in a system with high contact rate heterogeneity but no volatility, almost all individuals have contact rates that are insufficient to start an epidemic. However, in the subpopulation that does have sufficiently high contact rates, the average contact rate is very high–much higher than would be required to start an epidemic. In a system with volatility, the variance is reduced because the probability of maintaining a very high contact rate for a given period of time decreases with increasing volatility.

The theoretical interpretation of the variance of R_0 is not as well-established at the

interpretation of R_0 itself. The variance of R_0 is related to the probability of die-out in the very early stages of the epidemic. As the variance in the number of transmissions generated by an index case in a fully susceptible population increases, the probability that, in a specific instance, that case will generate less than the average number of new cases increases. At the beginning of an epidemic, a single infected individual failing to generate more than one new infection can cause stochastic die-out of the epidemic even when $R_0 > 1$. The reduction in R_0 with increasing volatility is faster than the corresponding reduction in the variance of R_0 leading to a net reduction in the signal-to-noise ratio with increasing volatility (figure 3.5).

3.6 Discussion

To my knowledge, this is the first study of contact rate volatility and R_0 that explicitly focuses on isolating the effects of volatility on R_0. Diekmann et al. formalized a method for the calculation of R_0 in a generalized heterogeneous transmission system as the dominant eigenvalue of the next generation operator [20]. The approach can be though of as formalizing an expression for the infectivity, A, as a function of time from infection, τ, and then integrating over τ to get an expression for R_0, $\int A(\tau)d\tau = R_0$. The next generation operator simply gives the value of the $\int A_h(\tau)d\tau$ for each possible heterogeneity state; if we assume that transmission is dependent only on the heterogeneity state of the infector, then deriving the next generation matrix can be (in some simple cases) straightforward. This method requires either the assumption of a deterministic system or invocation of the law of large numbers (LLN) [32].

The use of branching processes to describe the early infectious dynamics is advantageous for several reasons. Explicit inclusion of the stochasticity inherent in real

transmission systems allows us to treat the quantity of interest as a random variable rather than simply a scalar, such as an expectation under the assumption of LLN. First, generating functions for the random number of infections generated by a typical infected can be used to not only get the expectation (R_0), but also the higher order moments. When the population of infectives is very small, such as during the early epidemic period, the system is dominated by stochastic effects, and R_0 may not give a very complete picture of what to expect in terms of the rate of epidemic spread and the probability of the disease actually taking hold in the population [59].

R_0 alone does not necessarily get at the actual information that we are interested in, such as the answer to the question of how hard an infection will be to eliminate from a population. Consider the result from Anderson and May that $R_0 \propto \mu_c + \sigma_c^2/\mu_c$ in a simple SI model with heterogeneous contact rates, where μ_c is the average contact rate and σ_c^2 is the variance of the contact rate distribution [7]. This relation implies that while holding the average contact rate constant, increasing the variance of the contact rate distribution, increases R_0 and, by the standard interpretation of R_0, makes controlling the epidemic more difficult. However, as demonstrated in this chapter, increasing the variance of the contact rate distribution in the absence of volatility also greatly increases the $\mathbf{Var}(R_0)$. The basic reproduction number is actually highest when the mean is very low and the variance is very high, such as was observed in chapter 2. The implication is that the infection is most difficult to control when a very small proportion of individuals account for almost all of the contacts.

A simple thought experiment shows that R_0 itself only give part of the picture. For Gamma distributed contact rates the variance can be made arbitrarily high for a given average contact rate. As the variance increases the median value decreases.

There is no closed form for the median of a Gamma distribution, but this can be shown numerically. As the variance increases, more of the population has lower contact rates with an increasingly small minority with very large rates. In a finite population, as the variance of the contact rate distribution increases, R_0 becomes large, but the population of individuals responsible for the large R_0 becomes smaller and smaller. The small population is much more susceptible to stochastic die out simply due to its small size. The variance of R_0 tells part of this story. For a given value of R_0, increasing variance of R_0 means that more of the density of the number of infections generated by a typical infected is sub-threshold. That is, most of infected individuals will generate less than one new infection. Volatility reduces R_0 and also the variance of R_0. The immediate conclusion that volatility either reduces the probability of an epidemic or makes an epidemic easier to control is not necessarily supported by this analysis. Further work along these likes should use simulation to correlate R_0 and its variance to both the probability that an epidemic takes hold in a population given a singular introduction and the amenability of an epidemic to be controlled by simulated interventions.

The functional form that the contact rate distribution takes is also important. Liljeros et al. found that a cross-sectional distribution of the number of sexual partners from a large Swedish cohort followed a power law distribution with scaling exponent of 2.3 for males with more than 5 reported lifetime partners [53]. Sexual networks with power-law distributed node distributions are referred to as scale-free networks.

May and Loyd show that for infinite population sizes, scale-free networks do not show threshold behavior; for any non-zero transmission probability an epidemic can occur [56]. The lack of threshold behavior emerges as a result of the underlining

distribution of node degrees having divergent variance (i.e. the variance goes to infinity as the number of nodes becomes large). However, May and Loyd show that in finite populations, R_0 is naturally bound by the maximum number of contacts made by any one individual. Given our assumption of proportional mixing and continuously distributed contact rates, this parameter corresponds to the population size, which we did not specify but must take some finite value. The population size is an upper bound on R_0 regardless of the form of the contact rate distribution. If we assumed power law distributed contact rates, $R_0 = N$ where N is the population size; however, the variance of R_0 would also be very large and subject to the issues brought up in the previous paragraphs.

The concept of volatility does not fit in very well with the work on scale-free networks. Networks are often though of as static, or at least static over some period of time. The power law distribution found by Liljeros et al. was for the total number of lifetime partners. Any volatility in either the number of partnerships or the number of contacts is lost in such an aggregate measure. Even in a dynamic network, it is hard to imagine how volatility as we have implemented it here could be integrated into a network model. The problem is that an individual's actual contact rate is limited by the number of available contacts in the network at any point in time. Contact rates could be conceptualized as a preferred number of partners that would govern whether or not an individual would accept a new partner or terminate an existing partnership. A potentially more fruitful direction for this line of research would be on to conceptualize volatility in the broader context of stable and casual partners.

Parameter	Description
θ	gamma shape parameter for contact rate distribution
k	gamma scale parameter for contact rate distribution
ρ	contact rate volatility parameter
$\chi_{i,t}$	is the contact rate of the i^{th} individual at time t
δ	death rate of infected individuals
ω	natural removal rates
ϵ	entry rate
β	transmissibility per contact
$\lambda = \omega + \rho + \delta$	rate that an interval terminates
$\phi = 1 - \rho/\lambda$	probability that an interval terminates in removal

Table 3.1: Transmission parameters for model 1

Parameter	Description
θ	gamma shape parameter for contact rate distribution
k	gamma scale parameter for contact rate distribution
ρ	contact rate volatility parameter
$\chi_{i,t}$	is the contact rate of the i^{th} individual at time t
δ_a	rate of progression from acute to chronic stage
δ_c	rate of progression from chronic stage to death
ω	natural removal rates
ϵ	entry rate
β_a	transmissibility per acute contact
β_c	transmissibility per chronic contact
$\lambda_a = \omega + \rho + \delta_a$	rate that an acute interval terminates
$\lambda_c = \omega + \rho + \delta_c$	rate that an chronic interval terminates
$\phi_a = 1 - \rho/\lambda_a$	probability that an acute interval ends in removal or progression
$\phi_c = 1 - \rho/\lambda_c$	probability that a chronic interval ends in removal

Table 3.2: Transmission parameters for model 2

Parameter	Description
θ^o	gamma shape parameter for the oral contact rate distribution
k^o	gamma scale parameter for the oral contact rate distribution
θ^a	gamma shape parameter for the anal contact rate distribution
k^a	gamma scale parameter for the anal contact rate distribution
ρ	contact rate volatility parameter
$\chi^o_{i,t}$	is the oral contact rate of the i^{th} individual at time t
$\chi^a_{i,t}$	is the anal contact rate of the i^{th} individual at time t
δ_a	rate of progression from acute to chronic stage
δ_c	rate of progression from chronic stage to death
ω	natural removal rates
ϵ	entry rate
β^o_a	transmissibility per acute oral contact
β^a_a	transmissibility per acute anal contact
β^o_c	transmissibility per chronic oral contact
β^a_c	transmissibility per chronic anal contact
$\lambda_a = \omega + \rho + \delta_a$	rate that an acute interval terminates
$\lambda_c = \omega + \rho + \delta_c$	rate that an chronic interval terminates
$\phi_a = 1 - \rho/\lambda_a$	probability that an acute interval ends in removal or progression
$\phi_c = 1 - \rho/\lambda_c$	probability that a chronic interval ends in removal

Table 3.3: Transmission parameters for model 3

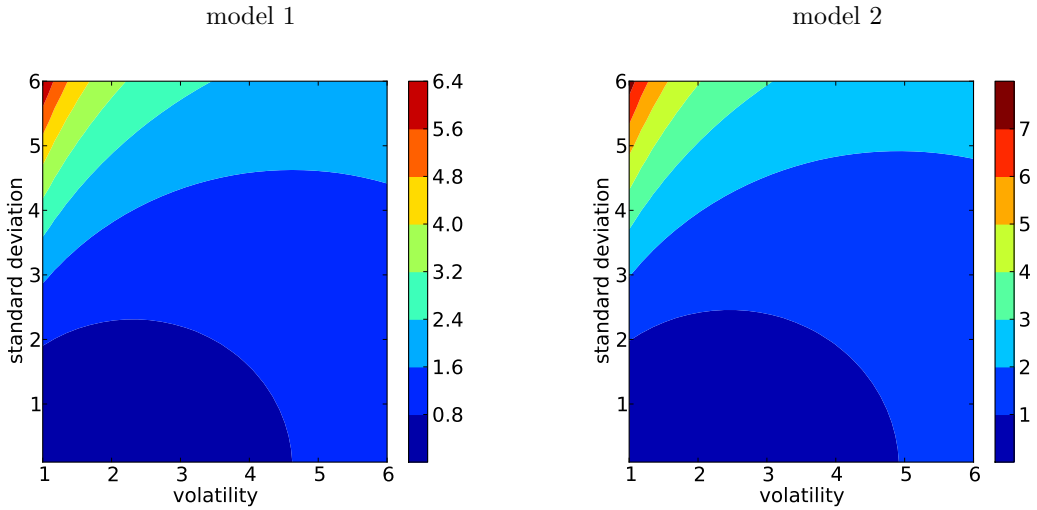

Figure 3.1: R_0 in models one and two without volatility
This plot shows the effect of changing the mean and standard deviation of the Gamma distributed contact rates under the assumption of no volatility ($\rho = 0$). Model 1 is on the left and model 2 is on the right. The acute stage of infection was, on average, 2 months long with 0.05 probability of transmission per contact; the chronic stage of infection was, on average, 142 months long with 0.001 probability of transmission per contact. The single infection stage in model 1 lasts, on average, 144 months with 0.0016 probability of transmission per contact. R_0 is highest when the mean is low and the variance is high due to a very small population of highly active individuals.

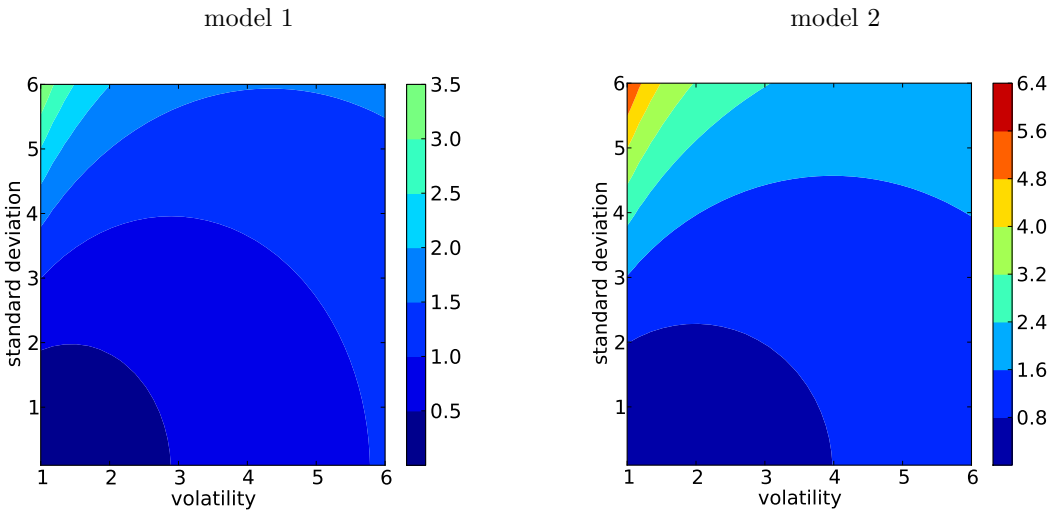

Figure 3.2: R_0 in models one and two with low volatility
This plot shows the effect of changing the mean and standard deviation of the Gamma distributed contact rates under the assumption of low volatility ($\rho = \frac{1}{120}$). Model 1 is on the left and model 2 is on the right. The acute stage of infection was, on average, 2 months long with 0.05 probability of transmission per contact; the chronic stage of infection was, on average, 142 months long with 0.001 probability of transmission per contact. The single infection stage in model 1 lasts, on average, 144 months with 0.0016 probability of transmission per contact. The reduction in R_0 observed in model 1 is mitigated in model 2 due to the high infectivity of the acute stage and the low probability of re-drawing contact rates before progression from the acute to chronic stage.

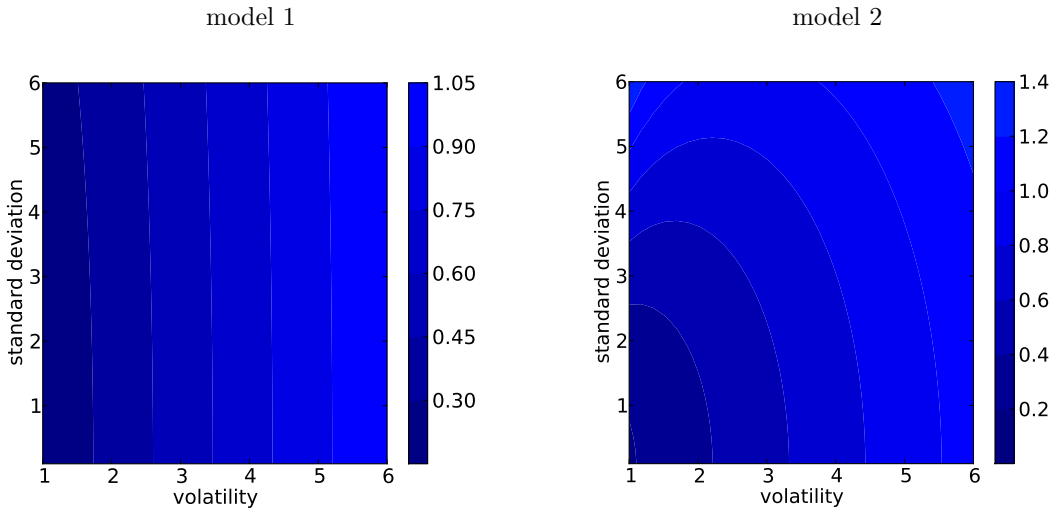

Figure 3.3: R_0 in models one and two with very high volatility
This plot shows the effect of changing the mean and standard deviation of the Gamma distributed contact rates under the assumption of low volatility ($\rho = \frac{1}{1}$). Model 1 is on the left and model 2 is on the right. The acute stage of infection was, on average, 2 months long with 0.05 probability of transmission per contact; the chronic stage of infection was, on average, 142 months long with 0.001 probability of transmission per contact. The single infection stage in model 1 lasts, on average, 144 months with 0.0016 probability of transmission per contact. At this level of volatility, the average behavioral period is only 1 month long which is shorter than the average duration of the acute stage (2 months) making R_0 similar in both models. The plot for model 1 shows how, at very high levels, volatility completely eliminates the effects of contact rate heterogeneity.

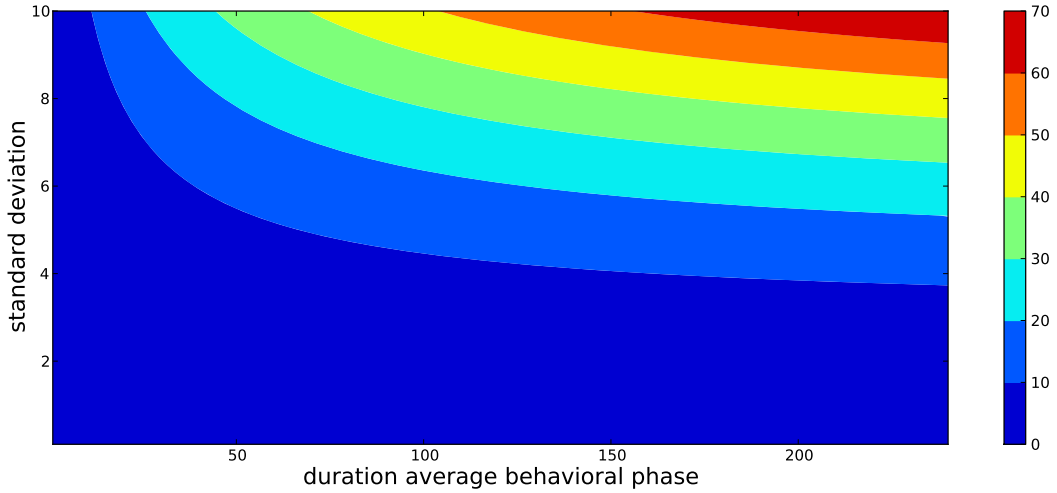

Figure 3.4: Effect of contact rate heterogeneity and volatility on the standard deviation of R_0

This plot shows the effect of the standard deviation of the contact rate distribution (y-axis) and the average duration of a behavioral phase ($\frac{1}{\rho}$) (x-axis) for a fixed average contact rate (4) on variance of R_0 in model 1. The single infection stage in model 1 lasts, on average, 144 months with 0.0016 probability of transmission per contact. Contact rate volatility reduces the variance of R_0 while contact rate heterogeneity increases the variance of R_0. At very high levels of volatility increasing variance in the distribution of contact rates does not increase the variance in the number transmissions an average case makes.

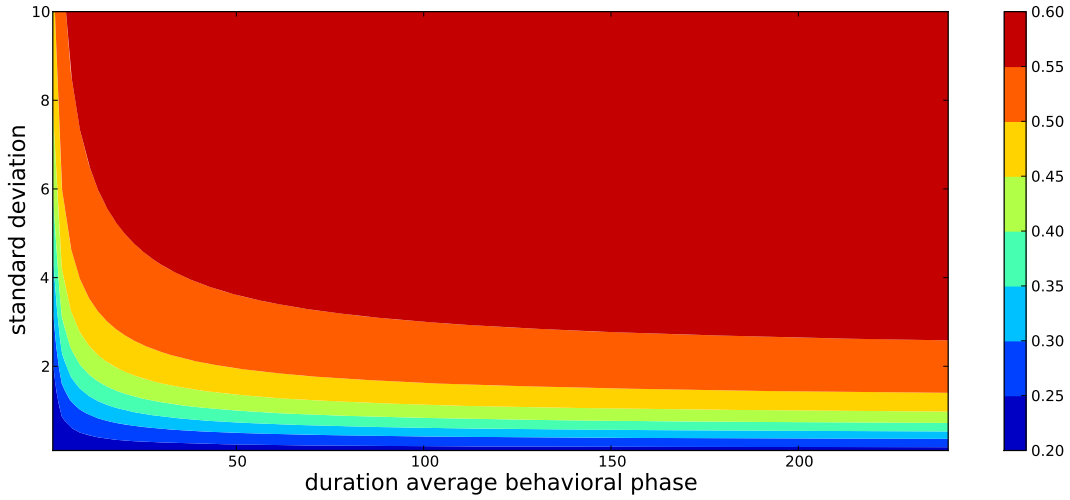

Figure 3.5: Effect of contact rate heterogeneity and volatility on the signal-to-noise ratio of R_0

This plot shows the effect of the standard deviation of the contact rate distribution (y-axis) and the average duration of a behavioral phase ($\frac{1}{\rho}$) (x-axis) for a fixed average contact rate (4) on signal-to-noise ratio, $\frac{\mu}{\sigma}$, of R_0 in model 1. The single infection stage in model 1 lasts, on average, 144 months with 0.0016 probability of transmission per contact. Volatility reduces both R_0 and the variance of R_0; this plot shows that as volatility increases the reduction R_0 is greater than the reduction in the variance of R_0

CHAPTER IV

Short-term volatility in sexual behavior promotes acute stage transmission of HIV

4.1 Summary

The stability of individual level contact rates over time is a common assumption in models of HIV transmission. Using individual-level simulation we relax that assumption by introducing time-variable (volatile) contact rates. Volatile contact rates reduce the probability that an epidemic occurs yet greatly increase the population risk of HIV if an epidemic occurs, all other things being equal. The increased risk is mediated almost entirely through increased acute stage transmission.

4.2 Introduction

The emergence of new HIV prevention methods such as pre-exposure prophylaxis places new emphasis on the role of theory development and modeling in the design of public health interventions [70]. Contrasted to other proposed methods such as the test-and-treat approach [15], pre-exposure prophylaxis acts by directly protecting susceptibles rather than limiting the contagiousness of the chronically infected. The relative efficacy of such programs is determined largely by the role that the short-lived but highly contagious acute stage of infection plays in sustaining the HIV epidemic. If acute stage transmission is key to the maintenance of infection levels, then pre-

exposure prophylaxis can be highly effective. However, the current state of the theory of HIV transmission does not yet understand the range of causal mechanisms that can promote acute stage transmission in real populations, and thus is insufficient to predict the expected efficacy of programs such as pre-exposure prophylaxis.

Previous work on the topic of acute stage transmission dynamics includes theoretical investigations of contact structure [48], heterogeneity in contact rates [6], partnership behavior [45], and sexual role dynamics [4], among other factors, and their ability to promote or retard population-level acute stage HIV transmission. Some of these effects can be quite strong, increasing acute-state incidence by over 40-fold (unpublished data). Studies that found a significant role for acute stage transmission often included some kind of sexual behavior change over time, whereas studies that suggested parity between infection stages [1] or epidemics driven by late-stage infection [67] did not. Failure to include dynamic sexual behavior in transmission models may lead to an underestimation of the role of primary stage HIV transmission.

In this paper we use individual-based simulation to investigate the effects and mechanisms of behavior change on the transmission of HIV. We formalize a highly abstract model of sexual behavior change as a stateless, non-differential change in sexual contact rates. We refer to this as contact rate volatility. Approaching this problem in a highly abstract way allows us to address the issue of behavior change in general without having to specify the nature and risk differences associated with changing sexual behavior.

4.3 Materials and methods

We implement an individual-based model of CRV as a Markov chain where the state and history of each individual is tracked and measured over the course of

a simulation run. This allows us to consider both the effects and mechanisms of CRV as functions of the simulation model parameters. Our general approach to the exploration of this model is to restrict the model parameters to reasonable values and simulate the state of the system and its individual-level history over a grid defined on the parameter space. As the effects and mechanism became clearer, additional simulations were performed for increased resolution and clarity. The results of this study are a synthesis of the primary points discovered by this process.

4.3.1 Behavioral model

The behavioral model specifies the sexual behavior of individuals in the model. It has two components: the contact rate heterogeneity (CRH) model and the contact rate volatility (CRV) model.

Contact rate heterogeneity model

The contact rate heterogeneity (CRH) model is a probability density or mass function that assigns probabilities to possible contact rates. The simplest model of CRH would assign one of two possible contact rates to individuals with probability p and $1 - p$ respectively. The value of those rates and probability of being assigned one rate or the other constitute the CRH model. We would like a more flexible CRH model so we assume that contact rates, X, are log-normally distributed in the population such that

$$\Pr(X = \chi) = \frac{1}{\chi\sqrt{2\pi\sigma^2}}\, e^{\frac{(\log\chi - \mu)^2}{2\sigma^2}}.$$

The log-normal distribution was chosen because it is defined for positive real numbers and displays the highly skewed nature that is commonly observed in sexual frequency data. However, more importantly, the log-normal distribution is parameterized in such a way that the mean can be held constant while the variance is free to vary

over a wide range from nearly homogeneous (low variance) to highly heterogeneous (high variance). In this way we can isolate the effects of the interaction of volatility and heterogeneity independently of the mean contact rate, which will also affect the transmission dynamics.

Contact rate volatility model

This work focuses on a specific type of sexual behavior change we refer to as contact rate volatility (CRV). CRV is stateless in that there is no corresponding change of state that produces the behavior change. CRV is also non-differential in that the probability of having a high or low contact rate is constant over time such that all individuals are equally likely to have any given contact rate. Taken together, CRV produces variable individual-level sexual behavior histories yet homogeneous population-level behavior. An individual's behavioral history can be thought of as a sequence of f intervals of length D, $\{d_1, d_2, \cdots, d_f\}$, within which an individual's contact rate is constant. By definition, the final interval, d_f, terminates in either infection-mediated death or natural removal. All proceeding intervals terminate in re-drawing the contact rate from X, thus defining a new behavioral interval. With increasing volatility an individual experiences more frequent and shorter behavioral intervals preceding the final interval. This simple, time homogeneous CRV model requires only one parameter to specify, ρ, the rate at which individuals re-draw their contact rate from X. Assuming the rate of natural removal and the infection mediated death rate are small with respect to ρ, then D is approximately distributed as $Exponential(\rho)$. As ρ approaches 0 the average length of a behavior interval approaches infinity, at which point there is no volatility.

There are more complex conceptualizations of CRV; for example, we might assume $\Pr(X = \chi)$ is some complex function of time such that individuals draw their contact

rates from different distributions at different points in their existence or the evolution of the whole system. However, for clarity, we restrict our analysis to a simple time homogeneous model of CRV.

4.3.2 Transmission model

The transmission model is based on a simple SIR model [7] [43] with the additions of 2 infectious stages, heterogeneous contact rates, and contact rate volatility. Contacts are assumed to be symmetric and instantaneous. The model parameters are given in table 4.2. Individuals exist in one 1 of 3 states: susceptible (S), acutely infected (A), or chronically infected (C); $N = S + A + C$. The model was implemented in the Java programing language as a continuous-time individual-based model. The trajectory of the system was determined by Gillespie's exact algorithm [24]. The model events and their rates are listed in table 4.3.

4.3.3 Second generation analysis

Some of our conclusions are based on the analysis of a single second generation infected in an otherwise fully susceptible population. This type of analysis allows us to measure the basic reproduction number and the relative contribution of single infected to the force of infection as a function of the volatility and heterogeneity parameters. In this analysis a population is seeded with 20/10000 acutely infected individuals. The first susceptible infected by one of the seeds is marked as the index case, after which all further transmissions are disabled ensuring a nearly fully susceptible population. The total number of transmissions the index would have made and their contact rates at the time of infection and removal are recorded. For each parameter set this process is repeated 2000 times to get stable averages.

4.4 Results

CRV has 3 primary effects: reduction in R_0, increase in population risk, and increase in acute-stage transmissions.

Individual level

We simulated the number of infections generated by a infected index case under conditions of no, low, and high volatility with low, medium, high standard deviations of the contact rate distribution. Figure 4.1) shows that increasing volatility reduces R_0 and that the reduction in R_0 is larger for the most heterogeneous contact rate distributions. As illustrated in figure 4.2, the reduction in R_0 is caused by a reduction in the contract rates of infected index cases over the course of their infection with increasing volatility. Anderson and May proved that increasing variance in the contact rate distribution for a given mean increases R_0 [7]. That is, an epidemic is more probable for a given average contact rate if the variance of the distribution of those contacts is greater. This is due to the fact that in a population with heterogeneous contact rates, some sub-population may have a sufficiently high contact rate to maintain an epidemic even if the average contact rate is too low to maintain such an epidemic in a comparable homogeneous population. Their proof assumed that contact rates are stable over time. Our model of CRV violates the assumption of stable contact rates, demonstrating that this well-known result is not robust to realistic violation of its assumptions.

Individuals become infected with probability proportional to their contact rate such that infected individuals will, on average, have higher contact rates than the general population (illustrated in figure 4.3). Without volatility, infected persons maintain their above average contact rates over the course of their infection. How-

ever, as contact rates become more volatile (increasing ρ) the probability that an infected individual re-draws their contact rate before they are removed increases. Because the mean contact rate is lower in the general population than in the population of the newly infected, increasing volatility reduces the force of infection.

Population level

The dynamics become more complex when we allow secondary infections. Whereas the last section shows that with increasing CRV causes a reduction in R_0 and force of infection at the individual-level, here, at the population-level we see increasing transmission with increasing CRV. Figure 4.4 shows the endemic prevalence as a function of volatility and the standard deviation of the population distribution of contact rates (holding the mean contact rate constant). If there is little variability in the population distribution of contact rates, volatility seems to have little effect on the endemic prevalence. However, if contact rates are highly heterogeneous, volatility greatly increases the average population risk. The large increase in the population risk of HIV is mediated through increased acute stage transmission. Figure 4.5 and table 4.1 show the times between acute and chronic transmissions in the whole population as a function of the contact rate heterogeneity and CRV parameters.

Volatility reduces the average contact rate of a single infected in an otherwise fully susceptible population. However, when we allow secondary infections, the average contact rates of the infector and infected is much higher with increasing volatility (illustrated in figure 4.6). This apparent contradiction, that volatility can reduce the force of infection from a single infected individual yet increase the average contact rate of a newly infected individual, is due to contact rate volatility increasing the availability of high-risk susceptibles. In the absence of volatility, as an epidemic progresses, susceptibles with the highest contact rates become infected and the av-

erage contact rate of susceptibles plummets. High contact rate individuals become available through both entry of new susceptibles and through extant susceptibles re-drawing their contact rates. Taken together, individuals with high contact rates become available at a rate proportional to $\epsilon + \rho S$. In the absence of volatility ($\rho = 0$), the incidence rate is limited by the rate at which new high risk susceptibles enter the population. At the beginning of an epidemic, the rate at which individuals with high contact rates enter is much slower than the rate at which they become infected causing a reduction in average contact rate of susceptibles. Volatility mitigates this effect by increasing ρS as illustrated in figure **??**. As the length of behavioral intervals becomes very short with respect to the incidence rate, the long-run average contact rate of susceptibles approaches $\mathbf{E}(X)$ as all susceptibles will have re-drawn their rates from X between infection events.

Volatility produces two apparently countervailing effects: increase in the availability of high risk susceptibles and decrease in the force of by an individual infected all other things being equal. These effects do not balance out. We parameterized the natural history of HIV such that 40% of the infection potential comes from the acute stage that lasts about 2 months immediately following infection. The force of infection in the acute stage will be unaffected by volatility unless the average length of behavioral phase is comparable to the duration of the acute stage (i.e. contact rates are highly volatile). However, the even modest rates of volatility help to sustain high contact rates in the susceptibles. High risk susceptibles rapidly become high risk acute infecteds, and, unless the volatility is very high, those elevated contact rates will be sustained throughout the highly infectious acute stage. Therefore volatility increases both the availability of high risk susceptibles and high risk acute stage infecteds. This effect interacts with the elevated infectivity of the acute stage

to generate a greatly elevated acute-stage transmission rate.

4.5 Discussion

In this chapter we have demonstrated that volatility has three major effects: 1) reducing R_0, 2) increasing the prevalence of HIV infection, and 3) increasing the proportion of transmissions coming from acute stage infecteds.

In the previous chapter we showed that R_0 decreases with increasing volatility. The contact rate distribution in that chapter was assumed to be a Γ random variable, whereas in these simulations we assumed log normally distributed contact rates. We simulated R_0 from relatively few number of parameter vectors, however those results are consistent with the results from the previous chapter. The log-normal is similar enough to the Γ distribution that is reasonable to assume the general results from the previous chapter hold in this model also.

The increase in the overall prevalence and the acute-stage incidence associated with increasing CRV is caused by the dual mechanisms of maintenance of high average contact rates in the susceptibles and reduction of contact rates of chronically infected individuals. Contact rate volatility both reduces R_0 and increases the prevalence of infection; to our knowledge, volatility is the only transmission parameter that can produce both of these effects.

In this chapter we conceptualized contact rate volatility as a kind of time homogeneous process where all individuals re-draw their contact rates from an unchanging probability distribution at a constant rate. We are unaware of any other studies that have explored this exact type of model. The closest model found was by Koopman et al. who modeled the effects of age structure and movement between high and low-contact rate risk groups on the role of acute stage HIV [49].

In that paper they present a highly structured model with age specific contact rates, age preferred mixing (individuals in each age group have a preference for contacting individuals only within their own age group), and volatility in contact rates, conceptualized as movement between more and less active states. Contact rates were as much as 8.5-fold higher in the more active state compared to the less active state, although this ratio varied for each age group. They concluded that the increased role for primary stage transmission observed in their model was due to the fact that 'transmissions during early infection get linked to individuals who are more likely to transmit during early infection, and the effect of increased transmission probabilities during infection is exponentiated across generations of transmission.' They also noted that increasing the rate of fluctuation increases the role of primary stage infection even further.

The model that I present in this chapter can be thought of as a simplification of the model of Koopman et al. I have essentially removed the age and mixing structure, expanded the concept of high and low contact rate groups to a full distribution of contact rates, and simplified volatility to a single process (re-drawing contact rates as opposed to movement to and from high and low contact rate groups). Simplifying the model allows us to show that contact rate volatility, in the absence of age and mixing structure, can increase the population risk and the relative contribution of acute stage infecteds.

Contact rate volatility may also explain the discrepancy between Koopman et al. and another paper that studied the relative contributions of each infection stage. Rapatski et al. also modeled the relative contributions of each infection stage using a deterministic compartmental model with six risk groups each with its own contact rate [67]. The risk-group specific contact rates were specified by both data and

model analysis. Mixing among risk groups was assumed to be proportional (as opposed to age structured), which they referred to as their 'bathhouse' assumption. They also assumed that behaviors are constant over time, which they justified by noting that a 1969 study of gay male sexual behavior found that nearly 30% of gay men reported having over 1000 lifetime sexual partners. Their argument was that such high numbers of sexual partners could only be reached by decades of high sexual activity. This point, although true, is irrelevant. Contact rate volatility as it is conceptualized here in our work could allow for both the accumulation of a large number of lifetime partners while still promoting primary stage transmission. Rapatski et al. made the error of assuming that volatility only affects infected individuals and failed to consider that volatility also increases the availability of high risk susceptibles regardless of the mean contact rate. As high risk susceptibles become infected they become high risk acutely infected individuals, which by their 'bathhouse' assumption, would account for a disproportional proportion of contacts.

Contact rate volatility as we have conceptualized it here could also be thought of as the limiting case of a dynamic network model with a static degree distribution, where the duration of partnerships tends to zero. In the real world, contact rate volatility is embedded in the system of enduring and short-term partnerships. We are not aware of any sexual network studies of HIV transmission that includes both enduring partnerships and contact rate volatility. Volatility in the context of enduring partnerships could be realized in terms of a fluctuating preference of the number of concurrent partners. The degree of concurrency of partnerships in dynamic network models without volatility has been shown to either amplify or retard the proportion of transmissions coming from the acute stage under different modeling assumptions [45]. Embedding volatility in the context of enduring partnerships

is a key step in incorporating volatility into the broader modeling literature.

Contact rate volatility could also be further simplified by separately considering contagiousness and susceptibility effects. In this simple model, contact rate heterogeneity can be though of as a combination of contagiousness and susceptibility effects, as increased contact rates both put susceptibles at higher risk of infection (increased susceptibility) and infected individuals at higher risk of transmission (increased contagiousness). Most of the work on contagiousness and susceptibility effects comes from the theoretical vaccine literature, where it has been argued that even vaccines that that do not protect individuals may actually reduce contagiousness enough to block transmissibility at the population level [46]. In the context of HIV dynamics, variability in contagiousness and susceptibility could result from temporary co-morbid STI infection. STI infections are known to increase both the probability of becoming infected with HIV [14] and the amount of virus shed by co-infected infected individuals, especially men with symptomatic urethritis [69].

Volatility in susceptibility would generate the 'supply-side' effects observed in this chapter by providing a constant supply of high-risk susceptibles in spite of high infection rates. Both increases in contact rates by the mechanism of volatility and increases in susceptibility due to co-infection increase the risk that a given susceptible will become infected over the period of increased susceptibility. However, assuming homogeneous, stable contact rates, volatile susceptibility will not affect the inference of which infectious stage is generating most of the new infections. The increase in acute-stage transmission caused by contact rate volatility is the result of the assumption of proportionate mixing; contacts made by susceptibles are more likely to be with an acute-stage infectives who are very likely to have very high contact rates. Volatility in susceptibility in an otherwise homogeneous system will not produce a

similar effect. The most susceptible individuals rapidly become infected, but they do not contribute to an increased force of infection. In total, volatility in susceptibility causes an increase in the overall incidence, but leaves the relative stage contribution unchanged.

In the absence of other heterogeneities, volatility as it is implemented in this chapter would have no significant effect on either the incidence rate or the relative stage contributions. Absence in the fluctuation of susceptibility either as a direct effect or as volatility in contact rates precludes the increased incidence rates that are cause by the increased availability of susceptibles. Unlike contact rate volatility, contagiousness volatility (isolated from susceptibility effects) is not correlated to infection risk in susceptibles. If individuals were more likely to be in a highly contagiousness state at the moment of infection, the high contagiousness and high viral titer during the short acute stage would interact to increase the acute stage incidence. However, in the absence of such a correlation, the system behavior will be dominated by the average contagiousness, which is unchanged by volatility. A practical conclusion of this work is to question the notion that cross-sectional data are sufficiently informative to understand the relative contribution of each infection stage and the role that they are playing in sustaining the HIV epidemic. In chapter 2.5.1 we showed that the most risky kinds of contacts are highly volatile in a subset of the population. Future work on contact rate volatility should incorporate estimates of volatility from multiple prospective cohorts into increasingly realistic models of volatility to assess the degree of increased acute-stage transmission.

A practical conclusion of this work is to question the notion that cross-sectional data are sufficiently informative to understand the relative contribution of each infection stage and the role that they are playing in sustaining the HIV epidemic. In

chapter 2.5.1 we showed that the most risky kinds of contacts are highly volatile in a subset of the population. Future work on contact rate volatility should incorporate estimates of volatility from multiple prospective cohorts into increasingly realistic models of volatility to assess the degree of increased acute-stage transmission.

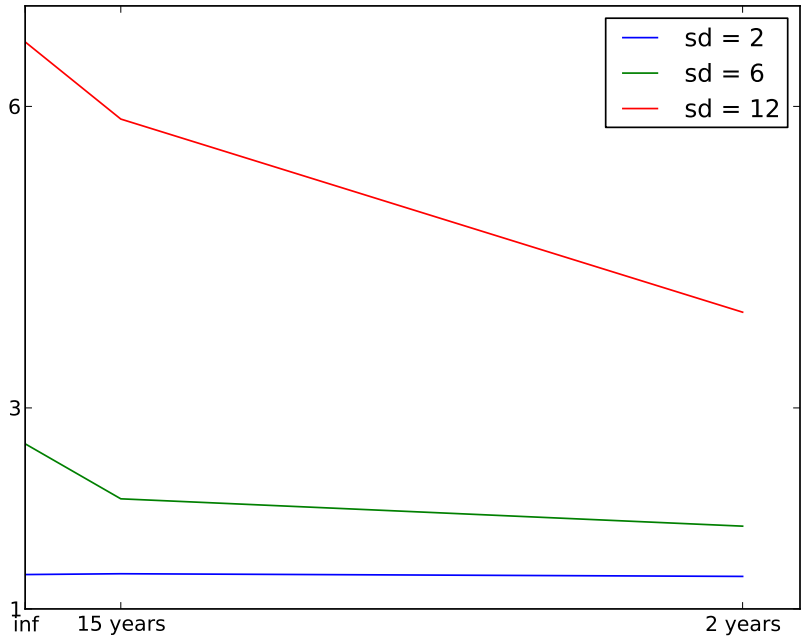

Figure 4.1: Volatility decreases R_0
This plot shows the effect of the average duration of a behavioral period (x-axis) on the average number of infections generated by an average infected in a fully susceptible population (y-axis) all other parameters being equal. The standard deviation (sd) is varied for each case while holding the mean contact rate equal to 6 for all simulations. For each point, the simulation was seeded with a small number of infected individuals. The first infected individual was marked as the index case and all further secondary transmissions where disabled (i.e. to keep the population fully susceptible no new transmission where allowed). The number of transmission that would have occurred was measured along with the contact rate of the index case at the moments of infection and removal either by death or natural removal. Each point is the average of 2000 simulations.

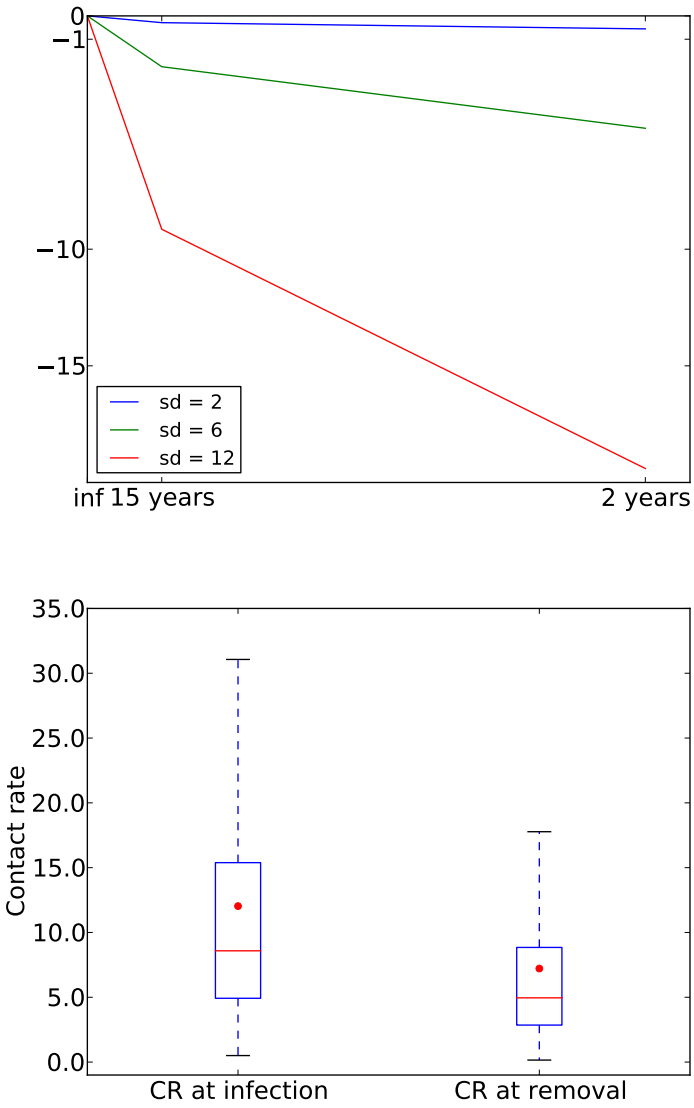

Figure 4.2: Volatility decreases the force of infection
The left-hand plot shows the difference in the contact rate of an infected individual at the moments of infection and removal (y-axis) as a function of the standard deviation (sd) of the contact rate distribution and the average duration of a behavioral interval (x-axis). The mean contact rate is fixed for all simulations at 6. The right-hand plot shows the distribution of the contact rate at the moments of infection and removal where the standard deviation of the contact rate distribution is 6 and the the average duration of a behavioral period is 2 years (values beyond the interquartile range are omitted for clarity). The red dot indicates the mean. Each point is the mean of 2000 observations.

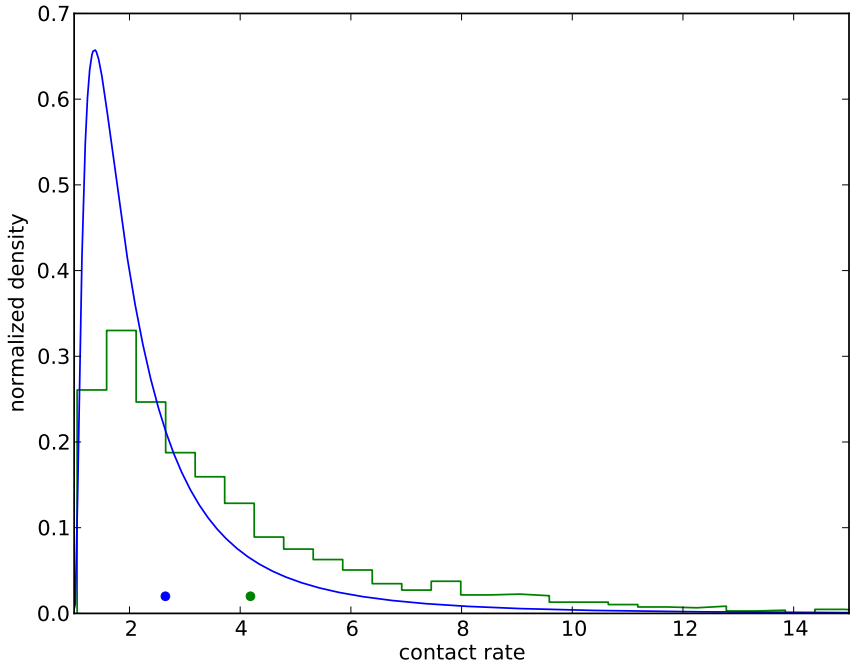

Figure 4.3: Contact rates in the general and infected populations
This figure illustrates the distribution of contact rates of newly infected individuals assuming a constant population size and constant transmission rate. The blue line is the population distribution of contact rates, $X \sim \log N(1, 1)$, while the green line is a histogram of 2000 random variates selected from X with probability proportional to χ. The blue and green dots indicate the means of the population distribution of contact rates in the general and newly infected populations respectively.

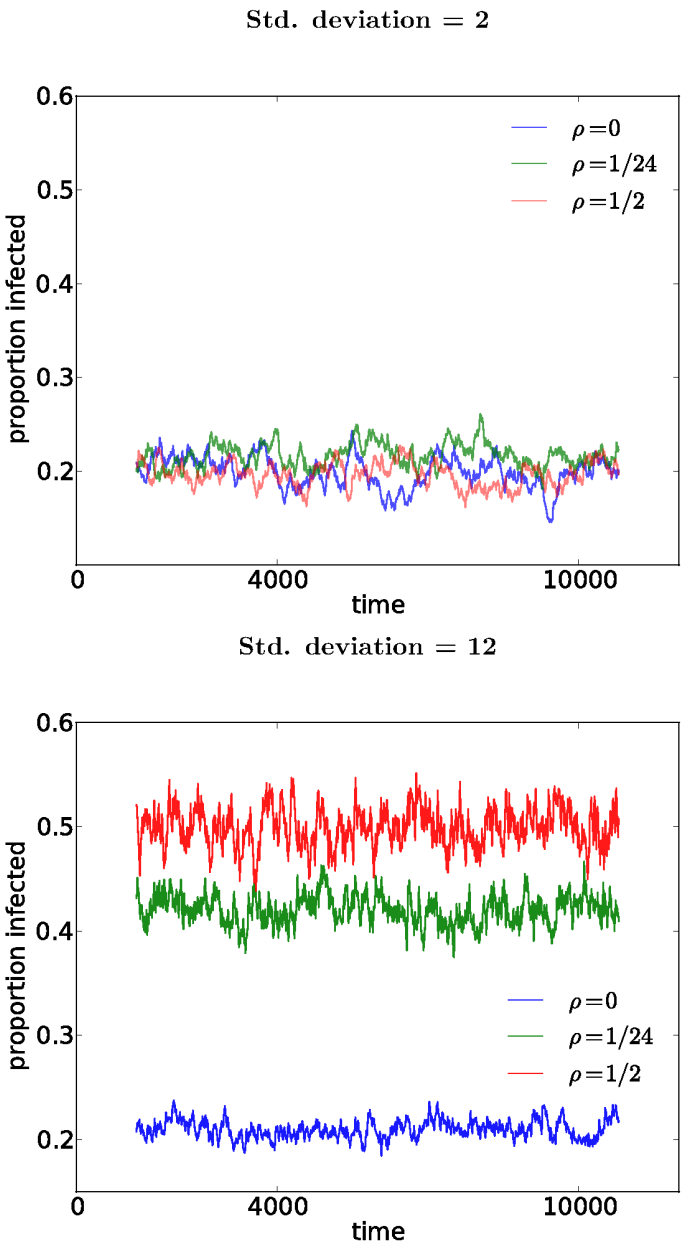

Figure 4.4: Volatility increases the risk of HIV
This plot shows the effect of volatility (the inverse of ρ is the average duration of a behavioral interval) on the endemic prevalence all other parameters being equal. Each line represents one run of the simulation for 11000 months with the first 1000 months discarded as burn-in. The y-axis shows the proportion of individuals infected at a given time (x-axis). In the plot on the left the standard deviation of the contact rate distribution ($\log N(1.74, 0.32)$) is low (2) showing the very limited effect of volatility in the presence of low contact rate heterogeneity. In the right-hand plot the standard deviation of the contact rate distribution ($\log N(0.99, 1.27)$) is high (12). The mean contact rate is held constant at 6 in all simulations.

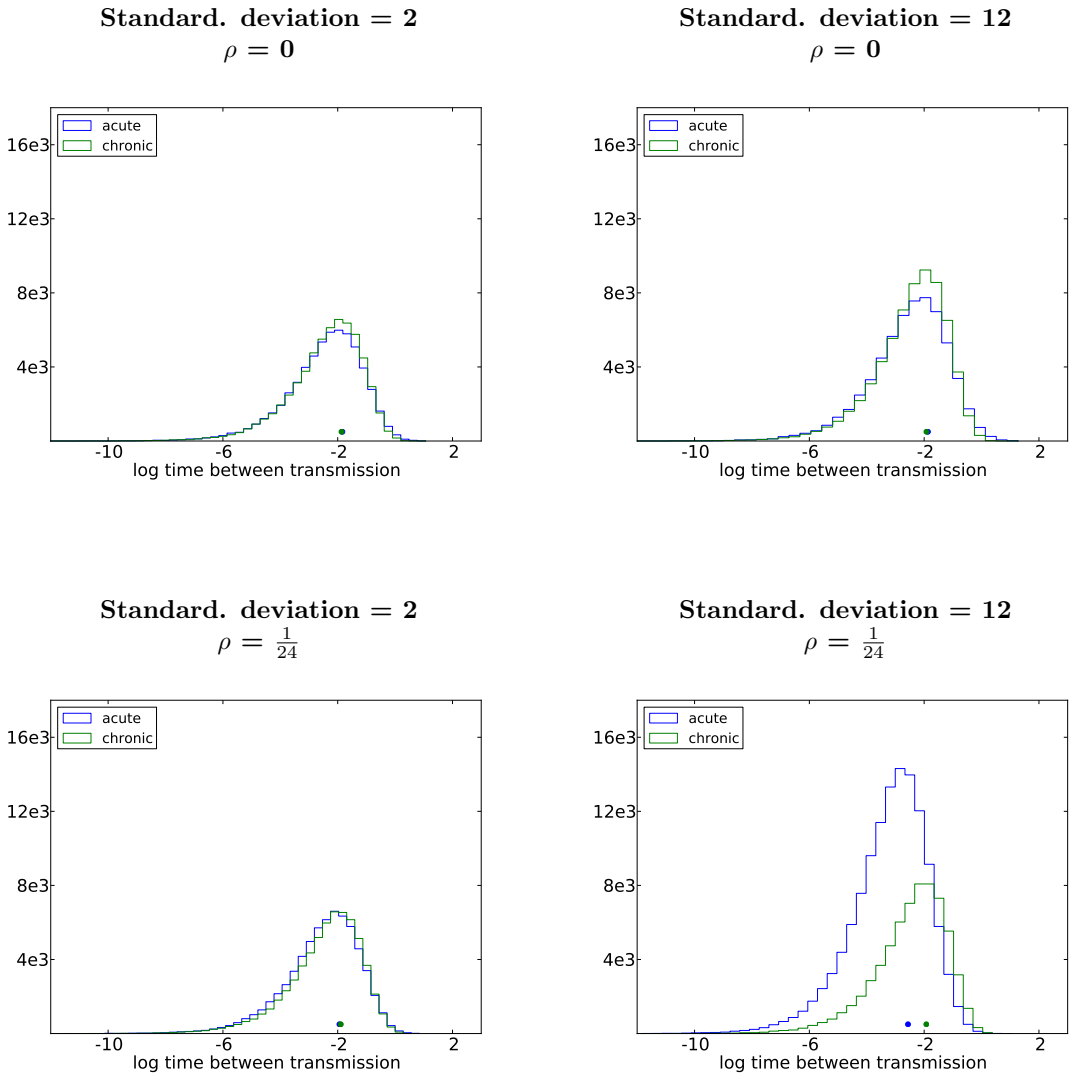

Figure 4.5: Histograms of time between transmissions in whole population
This plot shows effects of volatility (the inverse of ρ is the average duration of a behavioral interval) on the log-scale time between acute and chronic transmissions, all other parameters being equal. The x-axis is the log-scale time between transmissions in the whole population and the y-axis is the frequency over 600 months. The inverse of the average time between transmissions is the incidence rate. In the plot on the left the standard deviation of the contact rate distribution ($\log N(1.74, 0.32)$) is low (2) while in the right-hand plot the standard deviation ($\log N(0.99, 1.27)$) is high (12). The simulations are parameterized such that the force of infection from the acute and chronic stages are comparably equal given homogeneous contact rates (acute stage is 50 times more infectious but 60 times as short). The mean contact rate is held constant at 6 in all simulations. The dots at the bottom of the plot indicate the average time between transmissions.

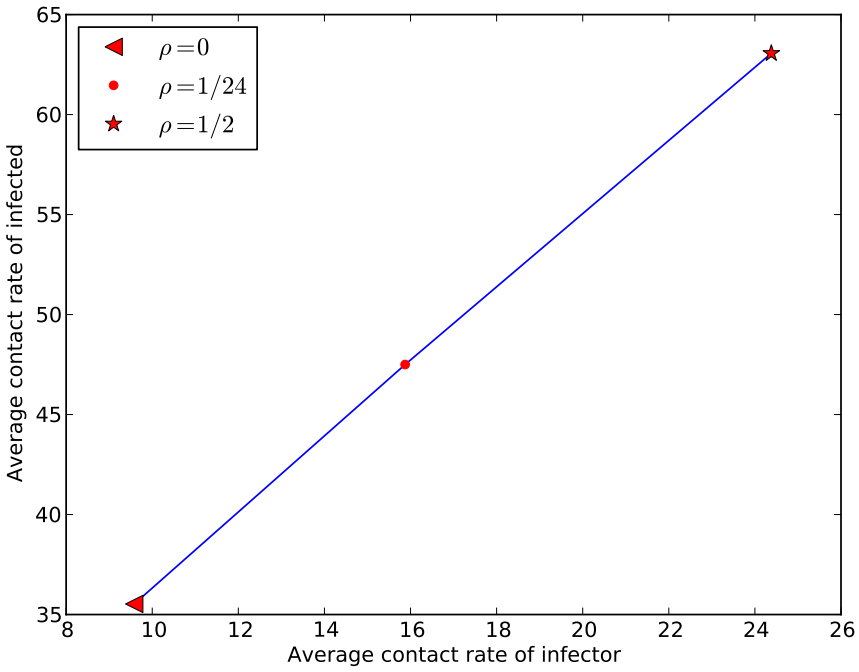

Figure 4.6: Increasing volatility increases the average contact rate of infector and infected at the moment of transmission

This figure illustrates the effect of volatility ($\frac{1}{\rho}$ is the average duration of a behavioral phase) on the average contact rates of infectors and infected at the moment of transmission over 11000 months of simulation time at pseudo-equilibrium. For each transmission in the simulation the average contact rate of the infector (the one transmitting) and the infected (the one transmitted to) is recorded. The mean and standard deviation of X are fixed at 6 and 12 respectively. Despite the fact that volatility reduces the average contact rate of infected individuals, the measured contact rates of the infector are actually higher at higher levels of volatility.

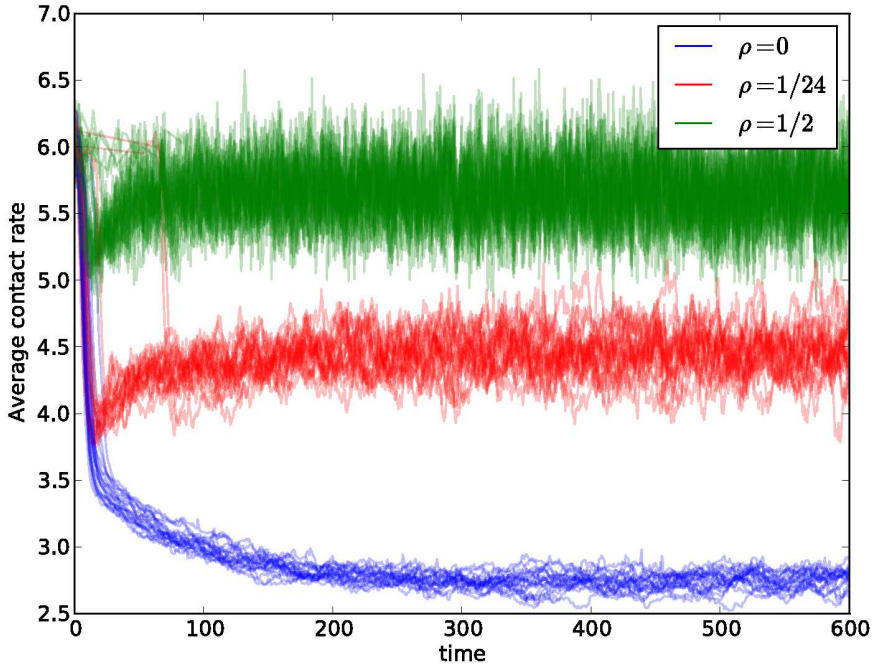

Figure 4.7: Volatility mitigates infection-mediated reduction in susceptible contact rates

This plot shows the effect of volatility ($\frac{1}{\rho}$ is the average duration of a behavioral phase) on the average contact rate of susceptibles all other parameters being equal. Each line represents one run of the simulation for 600 months. Simulations were repeated 20 times for each parameter set fixing the mean (6) and standard deviation (12) of the contact rate distribution. Without volatility (blue lines), the average contact rate in the susceptibles rapidly drops off as high risk individuals rapidly become infected. The rate of decrease slows as the average contact rate approaches a pseudo-equilibrium. With volatility (green and red lines) average contact also rates rapidly drop-off as high-risk individuals become infected. However, the drop-off is brief as high-risk susceptibles are rapidly replaced by former low-risk susceptibles through the process of volatility.

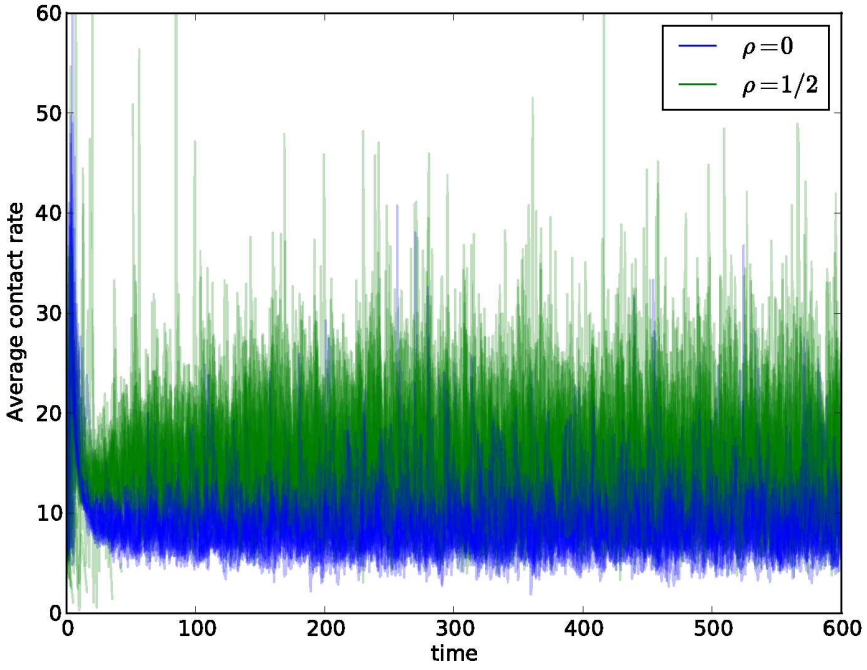

Figure 4.8: Volatility amplifies average contact rates in the acutely infected
This plot shows the effect of volatility ($\frac{1}{\rho}$ is the average duration of a behavioral phase) on the average contact rate of acutely infected all other parameters being equal. Each line represents one run of the simulation for 600 months. Simulations were repeated 20 times for each parameter set fixing the mean (6) and standard deviation (12) of the contact rate distribution. Without volatility (blue lines), the average contact rate is slightly above the population mean. With volatility (green lines), the average contact rates are elevated and show a highly 'spiky' character. The case of medium volatility was omitted for clarity.

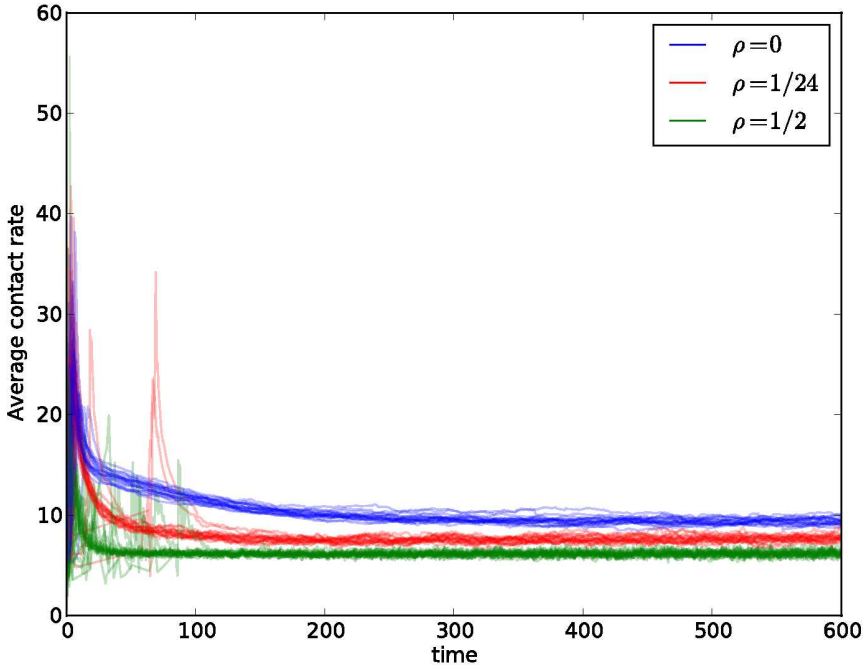

Figure 4.9: Volatility reduces average contact rates in the chronically infected
This plot shows the effect of volatility ($\frac{1}{\rho}$ is the average duration of a behavioral phase) on the average contact rate of acutely infected all other parameters being equal. Each line represents one run of the simulation for 600 months. Simulations were repeated 20 times for each parameter set fixing the mean (6) and standard deviation (12) of the contact rate distribution. Without volatility (blue lines), the average contact rate slowly equilibrate to a value slightly above the population average. With volatility (green and red lines), the average contact rates rapidly equilibrate; with high levels of volatility, the average contact rates are approximately equal to the population average.

Standard deviation	ρ	Δ acute	Δ chronic	Ratio acute
2	0	0.16(0.0007)	0.15(0.0006)	0.97
2	1/24	0.14(0.0006)	0.15(0.0006)	1.04
2	1/2	0.16(0.0007)	0.16(0.0006)	1.02
12	0	0.15(0.0008)	0.15(0.0006)	0.95
12	1/24	0.07(0.0002)	0.15(0.0006)	1.89
12	1/2	0.07(0.0002)	0.16(0.0006)	2.36

Table 4.1: Average time between acute and chronic transmissions
The average time between acute, Δ acute, and chronic transmissions, Δ chronic, and the ratio of the number of acute stage to chronic stage transmissions. In each simulation the average contact rate is held constant at 6. Each parameter set was simulated for 1000 months; the first 200 months of each simulation were discarded as burn-in.

Parameter	Description	Value
μ	log-normal location parameter contact rate	variable
σ^2	log-normal scale parameter contact rate	variable
$X \sim \log N(\mu, \sigma^2)$	probability of contact rates	variable
δ_a	acute to chronic progression rate	0.5 month^{-1}
δ_c	chronic to death progression rate	1/12 years^{-1}
ω	natural removal rate	1/30 years^{-1}
ϵ	entry rate	27.8 per month
β_c	probability of transmission chronic stage	0.001
β_a	probability of transmission acute stage	0.05
ρ	volatility parameter	variable

Table 4.2: Transmission Model parameters

Event	**Rate**	**Action**
New susceptible enters	ϵ	$S \to S + \phi_i()$
Susceptible exits population	$S\omega$	$S \to S - \phi_i(S)$
Acute infected exits population	$A\omega$	$A \to A - \phi_i(A)$
Chronic infected exits population	$C\omega$	$C \to C - \phi_i(C)$
Primary to chronic transition	$A\delta_a$	$A \to A - \phi_i(A), C \to C + \phi_i(A)$
Chronic to death transition	$C\delta_c$	$C \to C - \phi_i(C)$
Behavioral phase terminates	$N\rho$	
Transmission by acute stage	$\widehat{S}_t \frac{\widehat{A_t}}{\widehat{N_t}} \beta_a$	$S \to S - \phi_i(S),\ A \to A + \phi_i(S)$
Transmission by chronic stage	$\widehat{S}_t \frac{\widehat{C_t}}{\widehat{N_t}} \beta_c$	$S \to S - \phi_i(S),\ A \to A + \phi_i(S)$

Table 4.3: Events rates

The events comprising the individual-based transmission model. The notation $\widehat{Z_t}$ indicates the sum of individual contact rates of individuals in state Z at time t, $\widehat{Z_t} = \sum^{i \in Z} \chi_{i,t}$. $\phi_i(Z)$ indicates a randomly selected individual in state Z and $\phi_i(Z)$ indicates an individual in state Z selected with probability proportional to either their contact rate.

CHAPTER V

Conclusion

5.1 Summary

In this thesis I have demonstrated that 1) contact rate volatility is a better explanation of a prospective dataset of risky sexual behavior in gay men compared to a model of contact rate heterogeneity without volatility. 2) Contact rate volatility reduces both the mean and the variance in the number of infections generated by a newly infected individual in a fully susceptible population. And, 3) Contact rate volatility increases both the endemic prevalence and the acute stage transmission rate. I know of no other dynamic that has the properties of both reducing R_0 and increasing the endemic prevalence of HIV.

Volatility proved to be a superior explanation of a prospective dataset of sexual contact count data compared to a heterogeneous model without volatility (stability). An additional heterogeneity parameter, giving the probability that an individual's behavior is stable over the two-year course of the study, was required to get superior likelihoods compared to the stability model. A large proportion (greater than 50%) of individuals for each contact type were estimated to have stable contact rates while the remaining population experienced rather short behavioral intervals. Individuals with the highest average contact rates also had the highest variability in the number

of contacts reported in each observational period of the study suggesting that a volatility model where the degree of volatility is proportional to the average contact rate might be more appropriate. Nevertheless, the data support the idea that sexual behavior is not necessarily stable even over short periods.

All of the work in chapters 3.6 and 4.5 are predicated on the assumption that the hazard of infection is proportional to an individuals per-act contact rate. I defined contacts as instantaneous, symmetric, 'one-off' type events. In this context, the assumption of proportional infection hazard seems reasonable as each contact represents a unique, randomly determined partner. This idea is also supported by the data. Vittinghoff et al. used the same dataset that I used in chapter 2.5.1 to estimate the per-act probability of transmission for certain contact types [72]. Each type of contact was associated with a non-zero probability of transmission. However, the data do not specify whether or not contacts were made with a long-term partner or not. In a situation where nearly every contact was made with long-term (almost) monogamous partners, the risk of infection would not be proportional to the per-act contact rate but rather closer to the partnership formation rate [44].

If we accept the assumption of proportional hazards, then volatility increases the epidemic threshold, all other things being equal. This effect is simple to understand in a fully susceptible population as the average contact rates of newly infected individuals are higher than in the general population (by assumption their risk of infection is higher); when infected individuals re-draw their contact rates, they are drawing them from the lower population distribution of contact rates. On the other hand, volatility also reduces the variance in the basic reproduction number, R_0. This suggests that a singular introduction of HIV into a volatile system where $R_0 > 1$ is more probable to generate an epidemic than an introduction in a non-volatile system

with a comparable R_0. The reduction in the variance of of R_0 caused by increasing volatility is explainable by regression to the mean [18]: as total number of behavioral intervals increases the average contact rate approaches the mean. Over the long run, volatility actually reduces the effect of contact rate heterogeneity by bringing the average contact rates closer to the mean (technically the variance of the sample mean is given by the sampling distribution and is inversely proportional to the volatility parameter).

While volatility reduces the force of infection in a model with one infection stage, this effect is largely mediated with the additional of another infection stage. Unless volatility is very high, individuals maintain their elevated contact rates throughout the acute stage, re-drawing their contact rates some time during the chronic stage. Volatility, unless very high, reduces the force of infection in the chronic but not acute stage of infection.

The effect of volatility is exactly opposite in susceptibles. Going back to the metaphor from the introduction, imagine a column filled with ping-pong balls; the height of the column represents the probability of being removed from the column (infection). As we remove ping-pong balls, the average level of the balls slowly drops as the balls at the top are picked off (average contact rate of susceptibles drops over time). However, if we were to blow a stream of air into the bottom of the column such that the ping-pong balls were bounding about (volatility) and repeat the process, the average level would not drop nearly as much as before. As we picked off the balls that happened to be at the top of the column, the average level would drop a small amount; however, the spots at the top of the column would quickly be occupied by other balls as they were blown about by the stream of air. Volatility mitigates the reduction in the contact rates of susceptibles caused by infection by providing a

constant stream of new high-risk susceptibles.

These effects taken together show that volatility can significantly promote acute stage infection independent of explicit contact or population structure.

5.2 Simplifying assumptions

I chose to use a highly simplified model of contact rate volatility to focus in on its basic effects. I have shown that failure to consider contact rate volatility can, in theory, produce misleading inferences, especially with respect to questions of stage dominance. The downside of using very simple models is that isolating one particular dynamic requires making potentially very strong assumptions about the ignorability of other potential dynamics. The models that I have used throughout this thesis are all very similar and make many strong assumptions. I saw three key assumptions that were not strictly required but allowed me to focus in on the effects of volatility: independence of behavioral phases, uniformity of partner selection (proportional mixing), and population homogeneity.

5.2.1 Independence of behavioral phases

Assuming the independence of behavior phases simplified the calculation of R_0 in chapter 3.6 and made the individual-based simulations in chapter 4.5 more computationally tractable. However behavior change is almost certainly not independent of the current behavioral state: a high-risk individual is probably more likely to maintain a high-risk state than a low-risk individual entering a high-risk state. Behavioral dependence could take multiple forms. A simple first step would be to include a correlation between the current contact rate and future contact rates. Positive correlation implies increased stability in sexual behaviors as individuals are more likely to draw their contact rates similar to their current rate. This would mitigate both the reduc-

tion in the force of infection caused by volatility and the increase in the availability of high risk susceptibles. It is difficult to predict how those effects would play out without formal analysis. Strong negative correlation implies sequential periods of high and then low risk. These type of dynamics would almost certainly amplify the effects observed in chapter 4.5. Adding correlation is a simple first step in relaxing the assumption of independent behavioral phases.

5.2.2 Population homogeneity

In chapters 3.6 and 4.5, I assumed a homogeneous population. Again, this was largely for mathematical and simulation convenience. However, the volatility model from chapter 2.5.1 actually suggests a heterogeneous population with the majority having stable contact rates and the minority having highly volatile contact rates. Certainly, this kind of heterogeneity would need to be included in models that seek to make formal inference; however, for the type of exploratory analysis that I presented in this thesis, population-level heterogeneity in contact rate distributions is not a significant factor. Any analysis that attempted to estimate the fraction of infections from each infection stage would have to include other potentially heterogeneous factors such as co-infection with other STIs that are known to increase both susceptibility and transmissibility which could potentially increase chronic stage transmission.

5.2.3 Proportional mixing

I intentionally assumed proportional mixing to isolate the effects of volatility from the effects of contact structure that were present in Koopman et al [49]. This assumption allowed me to show that volatility, independent of contact structure, can increase the proportion of transmissions coming from acutely infected individuals.

Proportional and unstructured mixing also implies that individuals who were infected by an acutely or chronically infected individuals are indistinguishable with respect to their risk behaviors, and, therefore, the number subsequent transmissions they will generate. This means that, in the simple volatility models that I used in this thesis, stage-specific intervention effects are fully determined by stage-specific incidence rates. However, the interaction of contact structure obliviates the intuitive relationship between stage-specific intervention effects and stage-specific incidence rates by generating correlation between the transmission stage and contact rate. To see this, imagine a model where high-risk individuals only mix with other high risk individuals and individuals move between high and low-risk states at some constant rate. The flow of susceptibles from the low-risk group into the high-risk group increases the contact rate of acutely infected individuals in the same manner as described in chapter 4.5 meaning individuals in the high-risk state are more likely to be infected by an acutely infected individual. In this system, being infected by a acutely infected individual is correlated with having a higher contact rate at the time of infection. The positive correlation between stage of transmission and contact rate of the infector means that the total number of infections attributable to a single acute stage transmission is not captured by the simple proportion of transmissions coming from acutely infected individuals. Put in other words, the effect of intervening on acute stage infection is more efficient because acutely infected individuals are infecting people with high contact rates that amplify the total number of infections that are prevented by blocking a single acute stage transmission.

Relaxing the assumption of proportional mixing is the highest priority action for expanding the volatility model. There are a large number of ways of relaxing this assumption. Keeping with the KISS (keep it simple, stupid) principle, the best way

to proceed is to assume that some proportion of contacts as a function of the contact rate itself are made at a high-risk site. Individuals with the highest contact rates make most of their contacts at the high-risk site. This model would capture the basic features of volatility in the context of contact structure at the cost of only one additional parameter.

5.3 Implications

5.3.1 Prevention

The discovery of a high degree of volatility has significant implications for the control of the HIV epidemic among gay men. Prevention modalities can grouped into two basic categories: those that protect susceptibility and those that reduce infectivity. The former targets susceptible populations while the latter targets infected individuals. Condoms are a common type of prevention that actually fall into both categories; they both protect susceptibility and reduce the infectivity depending on who is using them.

Recently, two prevention modalities moved to the fore as potential novel approaches to the increasing problem of HIV in gay men. Pre-exposure prophylaxis (PrEP) uses low doses of anti-retrovirals to prevent infection in at-risk men. Currently there are multiple trials of PrEP underway, and the initial results are positive [12]. Alternatively, test-and-treat (TaT) strategies target reducing the interval between infection and diagnosis by increasing testing rates with a range of strategies [15]. Diagnosis implies treatment that means lower viral loads and less transmission.

The key difference between PrEP and TaT–and between modalities focused on susceptibility or infectivity in general–is the stage from which transmissions are prevented. PrEP keeps a person who would have otherwise become infected, susceptible. That individual will not transmit in either the acute or chronic stage because they

never become infected. On the other hand, if we assume that diagnosed individuals will no longer transmit after their diagnosis, TaT works by reducing the time from infection to diagnosis and therefore blocks any transmissions that would have otherwise occurred in that interval. The very short duration of the acute stage makes it almost impossible for TaT to prevent acute stage transmissions. The difference between PrEP and TaT is that PrEP can prevent acute stage transmissions while TaT can not. PrEP will work best in populations with high levels of acute stage transmissions.

This work has at least one practical implication for the future work on the dynamics and prevention of HIV: cross sectional data on sexual risk behavior are not necessarily sufficient to model the dynamics of HIV, specifically concerning the issue of acute versus chronic stage dominance. The models that I used in this thesis are very simple and make a large number of very strong simplify assumptions. I can not make any claims about the strength of these effects in the real world as transmission dynamics are much more complex, involving contact structure [47], relationship dynamics [26], and various other factors. However, I can say that, on a fundamental level, contact rate volatility can have strong effects with respect to the evaluation of population risk and stage dominance and that failure to first answer the question 'are contact rates stable over time in my population of interest?' can produce misleading results.

Going back to the discrepancies between Koopman et al. that argued for a central role for acute stage transmission [49] and Rapatski et al. [67] that concluded that almost all transmission were attributable to final stage transmission. Although Rapatski et al. were modeling a specific population and Koopman et al. made no such claim, the conclusions of these two papers are incompatible. The work that I

have presented here suggests that the difference between the theoretical conclusions of Koopman et al. and the empirical conclusions of Rapatski et al. can be attributed to the formers inclusion of a kind of contact rate volatility. Rapatski et al. assumed that because some men had high numbers of lifetime sex partners that partnership rates must be stable over time. However, this is not an answer to the question of whether or not contact rates are volatile; individuals can accrue a large small number of lifetime sexual partners with or without volatility in the partnership rate as the long-term averages are driven by the mean partnership rate that masks short-term variability, which has been shown to be important here.

5.4 Future directions

In the introduction I spoke of contact rate volatility in terms of an abstraction of the notion that that 'context matters': our tolerance to risk is informed by what kind of people we are but also by the places, both physical and mental, where we act. Who we are is constant, but where we act is variable. My approach to this question was to remove as much complexity as possible in an attempt to isolate the fundamental issue of volatile sexual behavior. This allowed me to keep the focus on the core concept of volatile sexual behavior. That focus on simplicity allowed me to show that even simple short-term, volatile sexual behavior can have major impacts on the dynamics of HIV.

A natural extension to this work would be to specify the nature of volatility in the context of partnership dynamics. Contact rate volatility as I have conceptualized it in this thesis can be thought of as the limiting case of a broader partnership process where the duration of partnerships approaches zero. I imagine that some of high degree of volatility observed in chapter 2.5.1 may be explained by the formation of

new partnerships. The epidemiological implication of a sudden change in contact rates would most likely be dampened by a highly monogamous partnerships.

In estimating the volatility parameters in chapter 2.5.1 I ignored a significant amount of structural information in the dataset. Specifically of interest is the correlation structure between contact types. The data have twelve types of measured contacts. Modeling the joint distribution of contact types could be quite informative and would be a unique contribution to the literature on sexual behavior. In the way that I have estimated the parameters currently, they can only be operationalized into a model as fully independent processes.

The analytical work in chapter 3.6 can also be extended. Recently, methods have been developed to define low-dimensional systems of differential equations that can be used to describe epidemic dynamics on configuration model networks with arbitrary degree distributions [73]. Integrating the concepts of volatility and behavior change into such models would allow a much more computationally efficient way to explore how network structure, behavior change, and partnership dynamics all play a role in generating new HIV infections.

CHAPTER VI

Derivations of $G'_C(1)$ and $G'_{\widetilde{C}}(1)$

$G_C(z)$ has no closed form, yet we never work with that function directly. In general, we only ever work with the derivative of $G_C(z)$ evaluated at $z = 1$, which does have a closed form for the gamma distribution. The following is a step-by-step derivation of that form. First we integrate over the probability of the interval length, T, and the contact rate, X, to get an expression for $G_C(z)$:

$$G_C(z) = \int \Pr(T) \int \Pr(\chi) e^{(z-1)^{(\chi T)}} \, dT \, d\chi \tag{VI.1}$$

which by the law of logarithms can be re-written as

$$G_C(z) = \int \Pr(T) \int \Pr(\chi) e^{T(z-1)^{\chi}} \, dT \, d\chi. \tag{VI.2}$$

By the definition of PGFs, the expression $\int \Pr(\chi) e^{T(z-1)^{\chi}} d\chi$ is formally equivalent to $G_\chi(e^{T(z-1)})$ which upon substitution gives

$$G_C(z) = \int \Pr(T)(1 - \theta T(z-1))^{-k} \, dT. \tag{VI.3}$$

Expanding $\Pr(T)$

$$G_C(z) = \int \lambda e^{-\lambda T}(1 - \theta T(z-1))^{-k} \, dT. \tag{VI.4}$$

The value of that integral would have to be approximated if we wanted to work with that function directly. However we are only interested in the derivative of the

function w.r.t z evaluated at $z = 1$

$$G'_C(1) = \left[\frac{d}{dz}\int \lambda e^{-\lambda T}(1-\theta T(z-1))^{-k}\, dT\right]_{z=1}. \tag{VI.5}$$

The order of integration and differential can be reversed by Leibniz's rule

$$G'_C(1) = \int \left[\frac{d}{dz}\lambda e^{-\lambda T}(1-\theta T(z-1))^{-k}\right]_{z=1} dT. \tag{VI.6}$$

Rearranging gives

$$G'_C(1) = \int \lambda e^{-\lambda T}\left[\frac{d}{dz}(1-\theta T(z-1))^{-k}\right]_{z=1} dT. \tag{VI.7}$$

Evaluating the inner derivative

$$\frac{d}{dz}\left[(1-\theta T(z-1))^{-k}\right]_{z=1} = \left[-k(1-\theta T(z-1))^{-(k+1)} \quad \frac{d}{dz}(1-\theta T(z-1))\right]_{z=1} \tag{VI.8}$$

$$= \left[-k(1-\theta T(z-1))^{-(k+1)}(-\theta T)\right]_{z=1} \tag{VI.9}$$

$$= k\theta T. \tag{VI.10}$$

Substituting $k\theta T$ in gives

$$G'_C(1) = \int \lambda e^{-\lambda T} k\theta T dT \tag{VI.11}$$

which, after rearranging is

$$G'_C(1) = \int T\lambda e^{-\lambda T} k\theta dT. \tag{VI.12}$$

The quantity $\int T\lambda e^{-\lambda T} dT$ is the average interval duration which can be written in terms of the model parameters as $\frac{1}{\lambda}$. The final result is then

$$G'_C(1) = \frac{\theta k}{\lambda} \tag{VI.13}$$

which in plain English is the mean contact rate, $k\theta$, multiplied by the mean interval duration, $\frac{1}{\lambda}$

The derivation of $G'_{\widetilde{C}}(1)$ follows the same steps as before. However, the substitution in VI.3 is with $G_{\widetilde{\chi}}(e^{T(z-1)})$ which changes inner derivative in VI.9 to

$$\frac{d}{dz} \quad \frac{(1+T\theta(1-z))^{-k}}{1+T\theta(1-z)} \tag{VI.14}$$

Differentiation by the quotient rule gives $\theta(k+1)$. Continuing the logic in the previous derivation gives a final result of

$$G'_{\widetilde{C}}(1) = \frac{\theta(k+1)}{\lambda} \tag{VI.15}$$

The second derivative of $G_C(1)$ can be calculated by replacing VI.8 with the second deravative. Begining from step VI.8 we have

$$G''_C(1) = \int \lambda e^{-\lambda T} \left[\frac{d^2}{dz^2}(1-\theta T(z-1))^{-k} \right]_{z=1} dT \tag{VI.16}$$

$$= \int \lambda e^{-\lambda T} \left[(k+1)kT^2\theta^2(T(\theta - z\theta)+1)^{-(k+2)} \right]_{z=1} dT \tag{VI.17}$$

$$= \int \lambda e^{-\lambda T}(k+1)kT^2\theta^2 dT \tag{VI.18}$$

which after rearranging becomes

$$G''_C(1) = (k+1)k\theta^2 \int \lambda T^2 e^{-\lambda T} dT. \tag{VI.19}$$

The quantity $\int \lambda T^2 e^{-\lambda T} dT$ is integrable by substituting $\psi = \lambda T$:

$$\int \lambda T^2 e^{-\lambda T} dT = \int \lambda \left(\frac{\psi}{\lambda}\right)^2 e^{-\psi} dT \tag{VI.20}$$

$$= \int \lambda \left(\frac{\psi}{\lambda}\right)^2 e^{-\psi} \frac{d\psi}{\lambda} \tag{VI.21}$$

$$= \frac{1}{\lambda^2} \int \psi^2 e^{-\psi} d\psi \tag{VI.22}$$

$\int \psi^2 e^{-\psi} d\psi$ is the Gamma function $\Gamma(z+1) = \int \psi^z e^{-\psi} d\psi = \Gamma(3) = 2$, therefore $\int \lambda T^2 e^{-\lambda T} dT = \frac{2}{\lambda^2}$. Substituting in we get a final expression

$$G''_C(1) = \frac{2(k+1)k\theta^2}{\lambda^2}. \tag{VI.23}$$

The second derivative of $G_{\widetilde{C}}(1)$ proceeds in the same fashion as above.

$$G''_{\widetilde{C}}(1) = \int \lambda e^{-\lambda T} \left[\frac{d^2}{dz^2} \quad \frac{(1+T\theta(1-z))^{-k}}{1+T\theta(1-z)} \right]_{z=1} \tag{VI.24}$$

$$= \int \lambda e^{-\lambda T} \left[(k+2)(k+1)T^2\theta^2(T(\theta - z\theta)+1)^{-(k+3)} \right]_{z=1} \tag{VI.25}$$

$$= \int \lambda e^{-\lambda T}(k+2)(k+1)T^2\theta^2 \tag{VI.26}$$

$$= \int \frac{2(k+2)(k+1)T^2\theta^2}{\lambda^2} \tag{VI.27}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] L. J Abu-Raddad and I. M Longini Jr. No HIV stage is dominant in driving the HIV epidemic in sub-Saharan africa. *Aids*, 22(9):1055, 2008.

[2] Adaora A. Adimora, Victor J. Schoenbach, and Irene A. Doherty. Concurrent sexual partnerships among men in the united states. *Am J Public Health*, 97(12):2230–2237, December 2007.

[3] H. Akaike. Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1):3–14, 1981.

[4] S.J. Alam, E. Romero-Severson, J.-H. Kim, G. Emond, and J.S. Koopman. Dynamic sex roles among men who have sex with men and transmissions from primary HIV infection. *Epidemiology*, 21(5):669–675, 2010.

[5] R.M. Anderson, G.F. Medley, R.M. May, and A.M. Johnson. A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *Mathematical Medicine and Biology*, 3(4):229–263, 1986.

[6] Roy M. Anderson. The epidemiology of HIV infection: Variable incubation plus infectious periods and heterogeneity in sexual activity. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 151(1):66–98, January 1988. ArticleType: research-article / Full publication date: 1988 / Copyright © 1988 Royal Statistical Society.

[7] Roy M. Anderson, Robert M. May, and B. Anderson. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, USA, September 1992.

[8] K. B. Athreya. T. e. harris and branching processes. *The Annals of Probability*, 39(2):429–434, March 2011.

[9] F Barré-Sinoussi, J C Chermann, F Rey, M T Nugeyre, S Chamaret, J Gruest, C Dauguet, C Axler-Blin, F Vézinet-Brun, C Rouzioux, W Rozenbaum, and L Montagnier. Isolation of a t-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.)*, 220(4599):868–871, May 1983. PMID: 6189183.

[10] C. T. Bauch. A versatile ODE approximation to a network model for the spread of sexually transmitted diseases. *Journal of mathematical biology*, 45(5):375–395, 2002.

[11] S. M Blower, G. J.P van Griensven, and E. H Kaplan. An analysis of the process of human immunodeficiency virus sexual risk behavior change. *Epidemiology*, 6(3):238, 1995.

[12] S.P. Buchbinder and A. Liu. Pre-Exposure prophylaxis and the promise of combination prevention approaches. 2011.

[13] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer, 2nd ed. 2002. corr. 3rd printing edition, July 2002.

[14] David M Butler, Davey M Smith, Edward R Cachay, George K Hightower, Charles Thomas Nugent, Douglas D Richman, and Susan J Little. Herpes simplex virus 2 serostatus and viral loads of HIV-1 in blood and semen as risk factors for HIV transmission among men who have sex with men. *AIDS (London, England)*, 22(13):1667–1671, August 2008. PMID: 18670228.

[15] V. Cambiano, A.J. Rodger, and A.N. Phillips. Test-and-treat: The end of the HIV epidemic? *Current Opinion in Infectious Diseases*, 24(1):19–26, 2011.

[16] D.R. Cox and G.F. Medley. A process of events with notification delay and the forecasting of AIDS. *Philosophical transactions of the Royal Society of London. Series B: Biological sciences*, 325(1226):135–145, 1989.

[17] E S Daar, S Little, J Pitt, J Santangelo, P Ho, N Harawa, P Kerndt, J V Glorgi, J Bai, P Gaut, D D Richman, S Mandel, and S Nichols. Diagnosis of primary HIV-1 infection. los angeles county primary HIV infection recruitment network. *Annals of Internal Medicine*, 134(1):25–29, January 2001. PMID: 11187417.

[18] C E Davis. The effect of regression to the mean in epidemiologic and clinical studies. *American Journal of Epidemiology*, 104(5):493–498, November 1976. PMID: 984023.

[19] K. Diekmann et al. The basic reproduction ratio for sexually transmitted diseases: I. theoretical considerations. *Mathematical biosciences*, 107(2):325–339, 1991.

[20] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio r 0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

[21] M. L Ekstrand, R. D Stall, J. P Paul, D. H Osmond, and T. J Coates. Gay men report high rates of unprotected anal sex with partners of unknown or discordant HIV status. *Aids*, 13(12):1525, 1999.

[22] R C Gallo, S Z Salahuddin, M Popovic, G M Shearer, M Kaplan, B F Haynes, T J Palker, R Redfield, J Oleske, and B Safai. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science (New York, N.Y.)*, 224(4648):500–503, May 1984. PMID: 6200936.

[23] Jose Ortega Gasset, Evelyn Rugg, and Diego Marin. *Meditations on Quixote: Translated from the Spanish by Evelyn Rugg and Diego Marin Introduction and Notes by Julian Marias.* University of Illinois Press, February 2000.

[24] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, December 1977.

[25] Steven M Goodreau and Matthew R Golden. Biological and demographic causes of high HIV and sexually transmitted disease prevalence in men who have sex with men. *Sexually Transmitted Infections*, 83(6):458 –462, October 2007.

[26] P. M. Gorbach. Transmission of STIs/HIV at the partnership level: Beyond Individual-Level analyses. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 80(90003):15iii–25, 2003.

[27] Pamina M. Gorbach, Lydia N. Drumright, and King K. Holmes. Discord, discordance, and concurrency: Comparing individual and Partnership-Level analyses of new partnerships of young adults at risk of sexually transmitted infections. *Sexually Transmitted Diseases*, 32(1):7–12, 2005.

[28] R H Gray, M J Wawer, R Brookmeyer, N K Sewankambo, D Serwadda, F Wabwire-Mangen, T Lutalo, X Li, T vanCott, and T C Quinn. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in rakai, uganda. *Lancet*, 357(9263):1149–1153, April 2001. PMID: 11323041.

[29] Edward C. Green, Daniel T. Halperin, Vinand Nantulya, and Janice A. Hogle. Uganda's HIV prevention success: The role of sexual behavior change and the national response. *AIDS and Behavior*, 10(4):335–346, 2006.

[30] P.N. Halkitis, S. Brockwell, D.E. Siconolfi, R.W. Moeller, R.D. Sussman, P.J. Mourgues, B. Cutler, and M.M. Sweeney. Sexual behaviors of adolescent emerging and young adult men who have sex with men ages 13-29 in new york city. *Journal of Acquired Immune Deficiency Syndromes*, 56(3):285–291, 2011.

[31] H. I Hall, R. Song, P. Rhodes, J. Prejean, Q. An, L. M Lee, J. Karon, R. Brookmeyer, E. H Kaplan, M. T McKenna, et al. Estimation of HIV incidence in the united states. *JAMA: the journal of the American Medical Association*, 300(5):520, 2008.

[32] J. A. P. Heesterbeek and K. Dietz. The concept of ro in epidemic theory. *Statistica Neerlandica*, 50(1):89–110, 1996.

[33] Stéphane Helleringer, Hans-Peter Kohler, Linda Kalilani-Phiri, James Mkandawire, and Benjamin Armbruster. The reliability of sexual partnership histories: implications for the measurement of partnership concurrency during surveys. *AIDS*, 25(4):503–511, 2011.

[34] Herbert W. Hethcote and James W. Van Ark. Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs. *Mathematical Biosciences*, 84(1):85–118, May 1987.

[35] King K. Holmes, Ruth Levine, and Marcia Weaver. Effectiveness of condoms in preventing sexually transmitted infections. *Bulletin of the World Health Organization*, 82(6):454–461, June 2004. PMID: 15356939 PMCID: 2622864.

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

[37] J.A. Jacquez, J.S. Koopman, C.P. Simon, and I.M. Longini Jr. Role of the primary infection in epidemics of HIV infection in gay cohorts. *Journal of Acquired Immune Deficiency Syndromes*, 7(11):1169–1184, 1994.

[38] John A Jacquez, Carl P Simon, and James Koopman. Structured mixing: heterogeneous mixing by the definition of activity groups. page 301–315. Springer-Verlag New York, Inc., New York, NY, USA, 1989. ACM ID: 90448.

[39] John A. Jacquez, Carl P. Simon, James Koopman, Lisa Sattenspiel, and Timothy Perry. Modeling and analyzing HIV transmission: the effect of contact patterns. *Mathematical Biosciences*, 92(2):119–199, December 1988.

[40] Dj Jevtović, D Salemović, J Ranin, O Dulović, D Ilić, and B Brmbolić. The prognosis of highly active antiretroviral therapy (HAART) treated HIV infected patients in serbia, related to the time of treatment initiation. *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, 47(2):131–135, February 2010. PMID: 20006540.

[41] J.H. Jones and M.S. Handcock. Social networks: Sexual contacts and epidemic thresholds. *Nature*, 423(6940):605–606; discussion 606, 2003.

[42] D. G Kendall. Branching processes since 1873. *Journal of the London Mathematical Society*, 1(1):385, 1966.

[43] W.O. Kermack and A.G. McKendrick. Contributions to the mathematical theory of epidemics - i. *Bulletin of Mathematical Biology*, 53(1-2):33–55, 1991.

[44] A.A. Kim, T. Hallett, J. Stover, E. Gouws, J. Musinguzi, P.K. Mureithi, R. Bunnell, J. Hargrove, J. Mermin, R.K. Kaiser, A. Barsigo, and P.D. Ghys. Estimating HIV incidence among adults in kenya and uganda: A systematic comparison of multiple methods. *PLoS ONE*, 6(3), 2011.

[45] J.-H. Kim, R.L. Riolo, and J.S. Koopman. HIV transmission by stage of infection and pattern of sexual partnerships. *Epidemiology*, 21(5):676–684, 2010.

[46] James S. Koopman and Roderick J. Little. Assessing HIV vaccine effects. *American Journal of Epidemiology*, 142(10):1113 –1120, November 1995.

[47] James S Koopman, Carl P Simon, John A Jacquez, and Tae Sung Park. Selected contact within structured mixing with an application to HIV transmission risk from oral and anal sex. page 316–348. Springer-Verlag New York, Inc., New York, NY, USA, 1989. ACM ID: 90449.

[48] James S. Koopman, Carl P. Simon, and Chris P. Riolo. When to control endemic infections by focusing on High-Risk groups. *Epidemiology*, 16(5):621–627, 2005. ArticleType: research-article / Full publication date: Sep., 2005 / Copyright © 2005 Lippincott Williams & Wilkins.

[49] J.S. Koopman, J.A. Jacquez, G.W. Welch, C.P. Simon, B. Foxman, S.M. Pollock, D. Barth-Jones, A.L. Adams, and K. Lange. The role of early HIV infection in the spread of HIV through populations. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 14(3):249–258, 1997.

[50] J. Li and F. Brauer. *Continuous-time age-structured models in population dynamics and epidemiology*, volume 1945. 2008.

[51] Spencer Lieb, Stefanie White, Becky L Grigg, Daniel R Thompson, Thomas M Liberti, and Stephen J Fallon. Estimated HIV incidence, prevalence, and mortality rates among racial/ethnic populations of men who have sex with men, florida. *Journal of Acquired Immune Deficiency Syndromes (1999)*, 54(4):398–405, August 2010. PMID: 20182358.

[52] F. Liljeros, C.R. Edling, L.A. Nunes Amaral, H.E. Stanley, and Y. Åberg. Social networks: The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[53] Fredrik Liljeros, Christofer R. Edling, Luis A. Nunes Amaral, H. Eugene Stanley, and Yvonne Aberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, June 2001.

[54] J L Martin. The impact of AIDS on gay male sexual behavior patterns in new york city. *Am J Public Health*, 77(5):578–581, May 1987.

[55] R.M. May and R.M. Anderson. Transmission dynamics of HIV infection. *Nature*, 326(6109):137–142, 1987.

[56] Robert M. May and Alun L. Lloyd. Infection dynamics on scale-free networks. *Physical Review E*, 64(6):066112, November 2001.

[57] M McEvoy and H E Tillett. Some problems in the prediction of future numbers of cases of the acquired immunodeficiency syndrome in the UK. *Lancet*, 2(8454):541–542, September 1985. PMID: 2863552.

[58] D. J McKirnan, D. G Ostrow, and B. Hope. Sex, drugs and escape: a psychological model of HIV-risk sexual behaviours. *AIDS care*, 8(6):655–669, 1996.

[59] Joel C. Miller, Bahman Davoudi, Rafael Meza, Anja C. Slim, and Babak Pourbohloul. Epidemics with general generation interval distributions. *Journal of Theoretical Biology*, 262(1):107–115, January 2010.

[60] M. Morris and M. Kretzschmar. Concurrent partnerships and the spread of HIV. *Aids*, 11(5):641, 1997.

[61] Ted Myers, Dan Allman, Liviana Calzavara, Ken Morrison, and et al. Gay and bisexual men's sexual partnerships and variations in risk behaviour. *The Canadian Journal of Human Sexuality*, 8(2):115, 1999.

[62] C. Norman. AIDS trends: Projections from limited data. some aspects of AIDS epidemiology are now well understood; others remain puzzling because of the long latency period after injection. *Science*, 230(4729):1018+1020–1021, 1985.

[63] John Novembre, Alison P Galvani, and Montgomery Slatkin. The geographic spread of the CCR5 ?32 HIV-Resistance allele. *PLoS Biol*, 3(11):e339, October 2005.

[64] A. David Paltiel, Kenneth A. Freedberg, Callie A. Scott, Bruce R. Schackman, Elena Losina, Bingxia Wang, George R. Seage, Caroline E. Sloan, Paul E. Sax, and Rochelle P. Walensky. HIV preexposure prophylaxis in the united states: Impact on lifetime infection risk, clinical outcomes, and Cost-Effectiveness. *Clinical Infectious Diseases*, 48(6):806 –815, March 2009.

[65] C.D. Pilcher, G. Joaki, I.F. Hoffman, F.E.A. Martinson, C. Mapanje, P.W. Stewart, K.A. Powers, S. Galvin, D. Chilongozi, S. Gama, M.A. Price, S.A. Fiscus, and M.S. Cohen. Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection. *AIDS*, 21(13):1723–1730, 2007.

[66] Steven D. Pinkerton. Probability of HIV transmission during acute infection in rakai, uganda. *AIDS and Behavior*, 12(5):677–684, 2007.

[67] B.L. Rapatski, F. Suppe, and J.A. Yorke. HIV epidemics driven by late disease stage transmission. *Journal of Acquired Immune Deficiency Syndromes*, 38(3):241–253, 2005.

[68] B.L. Rapatski, F. Suppe, and J.A. Yorke. Reconciling different infectivity estimates for HIV-1. *Journal of Acquired Immune Deficiency Syndromes*, 43(3):253–256, 2006.

[69] K. Rotchford, A.W. Strum, and D. Wilkinson. Effect of coinfection with STDs and of STD treatment on HIV shedding in genital-tract secretions: Systematic review and data synthesis. *Sexually Transmitted Diseases*, 27(5):243–248, 2000.

[70] J. Stover. HIV models to inform health policy. *Current Opinion in HIV and AIDS*, 6(2):108–113, 2011.

[71] D. J. Venzon and S. H. Moolgavkar. A method for computing Profile-Likelihood-Based confidence intervals. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 37(1):87–94, January 1988. ArticleType: research-article / Full publication date: 1988 / Copyright © 1988 Royal Statistical Society.

[72] E. Vittinghoff, J. Douglas, F. Judson, D. McKirnan, K. Macqueen, and S.R. Buchbinder. Per-contact risk of human immunodeficiency virus transmission between male sexual partners. *American Journal of Epidemiology*, 150(3):306–311, 1999.

[73] E. Volz. Susceptible-infected-recovered epidemics in populations with heterogeneous contact rates. *The European Physical Journal B*, 63(3):381–386, 2008.

[74] M. J Wawer, R. H Gray, N. K Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in rakai, uganda. *Journal of Infectious Diseases*, 191(9):1403, 2005.

[75] B. A.R Weinberg. Homosexualities: A study of diversity among men and women. *New York, Simon and Shuster*, 1978.

[76] C.S. Withers and S. Nadarajah. On the compound poisson-gamma distribution. *Kybernetika*, 47(1):15–37, 2011.

[77] Z.-Q. Zhang, T. Schuler, M. Zupancic, S. Wietgrefe, K. A. Staskus, K. A. Reimann, T. A. Reinhart, M. Rogan, W. Cavert, C. J. Miller, R. S. Veazey, D. Notermans, S. Little, S. A. Danner, D. D. Richman, D. Havlir, J. Wong, H. L. Jordan, T. W. Schacker, P. Racz, K. Tenner-Racz, N. L. Letvin, S. Wolinsky, and A. T. Haase. Sexual transmission and propagation of SIV and HIV in resting and activated CD4+ t cells. *Science*, 286(5443):1353 –1357, November 1999.