Exploring Effect of Individual Risk Behavior Volatility on Model Inference of HIV Transmission Dynamics and HIV Phylogenetic Tree

by

Xinyu Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2017

Doctoral Committee:

Professor Emeritus James S. Koopman, Co-Chair
Assistant Professor Marisa C. Eisenberg, Co-Chair
Professor Edward Ionides
Assistant Professor Rafael Meza
Professor Carl P. Simon

Xinyu Zhang

xinyuz@umich.edu

ORCID iD: 0000-0002-0166-0604

ACKNOWLEDGMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, colleagues, and support from my family and husband.

I would like to express my deepest gratitude to my advisor, Dr. James Koopman, for his excellent guidance, patience, providing me with an excellent atmosphere for doing research and financially supporting my research. I would like to thank Dr. Marisa Eisenberg, who generously offered help whenever I needed support with my research. I would also like to thank Dr. Erik Volz for guiding my research and helping me to develop my background in the field of phylodynamics. I would also like to thank Dr. Edward Ionides, Dr. Carl Simon and Dr. Rafael Meza for giving their precious advices about my research and my thesis.

I would like to thank Dr. Ethan Romero-Severson, Lin Zhong for their wonderful collaboration on the work presented in Chapter II. I would also like to thank Christopher Henry for giving his best suggestions and collaboration on the work presented in Chapter III and Chapter IV.

I would also like to thank my parents. They were always supporting me and encouraging me with their best wishes. Finally, I would like to thank my husband, Siyan Cao. He has been always supportive and cheering me up throughout the years of my dissertation research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

APPENDIX

ABSTRACT

Men who have sex with men (MSM) are heavily affected by HIV infections. However, controlling HIV transmission among MSM population remains a challenging task due to the complexity of the transmission dynamics of HIV. Mathematical models can facilitate understanding such dynamics. They also provide a basis for estimating important epidemiological parameters that can guide public health decision. This thesis advances methods to achieve both of these objectives and makes a substantive advance in the first area. In this dissertation, we relaxed the constancy of individual risk behavior assumption, by allowing individual risk behavior to fluctuate among different levels over time, namely individual risk behavior volatility (risk volatility).

We found that increasing risk volatility considerably increases fraction of transmission from acute HIV infection and prevalence at endemic equilibrium. In addition, we found that increasing risk volatility considerably reduces the minimum required individual effectiveness to eliminate HIV infections of Universal Test and Treat or universally applied pre-exposure prophylaxis (PrEP). Furthermore, our results suggest that increasing risk volatility reduces the extent that a case's risk level at HIV acquisition determines this case's capacity to cause onward transmission later during infection. Consequently, assuming no risk volatility may cause one to overestimate the benefit of prioritizing PrEP efforts to susceptible individuals experiencing high risk as a strategy to eliminate HIV infections. Finally, we explored the possibility of using HIV phylogeny to indicate when risk volatility affects HIV transmission dynamics. Our results suggest that risk volatility has unique and strong impact on phylogeny imbalance and clustering pattern. This implies that risk volatility is potentially identifiable from HIV sequence data.

Research in this thesis contributions to the field of study that uses mathematical models to estimate epidemiological parameters of HIV transmission from two perspectives. Firstly, research in this thesis reveals the importance of evaluating individual risk behavior volatility to enhance robustness of model inference of epidemiological parameters and quantities. Secondly,

research in this thesis suggests HIV sequence data is potentially valuable to improve the identifiability of risk volatility parameter. Therefore, research in this thesis takes significant steps forward to improve mathematical model inference of HIV transmission parameters.

CHAPTER I

Introduction

HIV Epidemic among Men Who have Sex with Men. In the decades from its emergence, great strides have been made in controlling HIV. Antiretroviral treatment has greatly improved and has saved millions of lives. Neonatal transmission and transmission via needles among injecting drug users have been markedly reduced.[1] On the other hand, HIV incidence among men who have sex with men (MSM) continues to rise with increasing rates of treatment having little effect on controlling transmission in this group.[1,2] There are diverse speculations as to why this might be the case including rapid evolution of the virus itself, safe-sex exhaustion, or emergence of versatility in terms of insertive or receptive sex roles among MSM. [3–6] Nonetheless, convincing evidence and arguments as to how different factors contribute to our failure to control HIV transmission among MSM are lacking. The formulations we present here constitute a theory expressing one mechanism through which biological and social factors interact to maintain high infection levels in MSM populations.

Two challenges are recognized in controlling HIV transmission among MSM. First, transmission from acute stage of HIV infection, a brief but highly infectious period, is potentially efficient but hard to estimate.[7] Our ability to understand HIV transmission from early stage of infection is hindered by the fact that we cannot observe HIV transmission patterns directly. In part, this is due to most infections not being detected until long after transmission. That makes it impossible to trace all sexual partners and count directly how many transmissions are occurring during the highly infectious but brief acute HIV infection stage (AHI). It also makes it rare that AHI outbreaks are detected and investigated at the time they are occurring.

Another challenge is to estimate the relationship of individual effectiveness of an intervention strategy and its population effect. Accurate estimation of such relationship requires fully understanding of the HIV transmission dynamics among the target population. This is

particularly important when elimination of HIV infection becomes the ultimate prevention goal. Two biomedical strategies that use antiretroviral medications have been recognized as promising intervention tools to eliminate HIV infections. One is treatment as prevention, where treatment of HIV with antiretroviral medications prevents onward transmission. Among treatment as prevention strategies the most promising and debated strategy is universal test and treat (UT&T), which is to get every individual tested and quickly treated if he or she is tested HIV positive. Previous study has suggested that testing every individual annually and initiating treatment immediately once that individual is tested HIV positive can possibly eliminate HIV infections.[8] Later studies further explore and discuss the promise of UT&T in eliminating HIV infections in various epidemiological context. [10–19] These studies indicate that effect of this optimal strategy might be hindered by many factors such as insufficient testing, failed linkage to care and decreased retention in care (frequent dropout). [10–19]

Another promising biomedical intervention strategy to prevent HIV transmission is Pre-exposure prophylaxis (PrEP). It is a prevention strategy that HIV negative people use antiretroviral drugs before exposure to HIV. [19] Animal studies suggest that PrEP can considerably reduce risk of HIV acquisition.[20,21] This is because PrEP drugs can maintain a high rectal and genital level and act at early life cycle of HIV virus.[19–21] Phase 3 clinical trial show that PrEP can reduce the risk of HIV acquisition among MSM population by 44% if it is orally taken on a daily base.[22] However, in reality, PrEP shows great variation in its effectiveness: low adherence and reduction in other ways of protection can reduce the effectiveness of PrEP.[23,24]

The great variation in the individual effectiveness of these two biomedical intervention strategies implies a great uncertainty in the population effect that either strategy could achieve. This issue has been recognized in earlier studies, in which sensitive analysis is done to explore the range of population effects that UT&T, PrEP or their combination could reach with various levels of individual effectiveness.[13,25–28] However, when eliminating HIV infections is the desired population effect, one would want to know the minimum required individual effectiveness of an intervention strategy to eliminate HIV infections. If most plausible levels of individual effectiveness of a strategy fall below this threshold, it would be hard to reach elimination with that strategy. One of the relevant research question is the how effective the UT&T strategy needs to be in testing and treating individuals to eliminate HIV infections.

Granich et al suggest that testing every individual once a year and immediately initiating treatment once the individual is diagnosed HIV positive can eliminate HIV infections.[8] UNAIDS proposed that HIV will be eliminated if 90% of infected individuals are tested, 90% of people who are tested HIV positive are linked to care and 90% of infected people who are linked to care are successfully treated, a so called 90-90-90 strategy.[29] However, there are speculations that the potential of both strategies to eliminate HIV infections may vary given the diverse epidemiological contexts of HIV transmission.[11,13,16,27,30–34] Therefore, how effective UT&T needs to be at individual level to eliminate HIV infections should be estimated based on the specific epidemiological context.

Given the complexity of HIV epidemics, eliminating HIV infections may require combined efforts of multiple HIV interventions rather than relying on a single strategy. Mathematical studies suggest that UT&T can be combined with PrEP to reach the maximum prevention effects because two strategies can reach two distinct population: infected people and uninfected people.[35–37] However, the population effect that could be achieved with the combination of the two strategies vary greatly among different studies.[25,36,38–40] Therefore, when one desire to eliminate HIV infections in the foreseeable future, it becomes imperative to understand the minimum required individual effectiveness of both strategies to eliminate HIV infections for the target population and plan the allocation of control efforts correspondingly.

Importance of Studying Individual Risk Behavior Volatility. To successfully eliminate HIV infections, one would want to make inference of these two important quantities: transmission from acute HIV infection and minimum required individual effectiveness to eliminate HIV infections of HIV intervention strategies. However, it is difficult to make such inference because HIV transmission dynamics among MSM population is complex and cannot be directly observed in the real world. Therefore, we usually construct mathematical models to test our hypothesis of what is shaping HIV transmission dynamics. A well built model is simple but does not lack important details. These imply two important features of a well built model: robustness of inference and identifiability of model parameters.[41] A model makes robust inference if model's prediction/estimation of a quantity is robust to realistic relaxation of the model's assumptions. This requires that the model formulation does not miss important determinant of the quantity of interest. The first step to examine this issue is to realistically relax each model assumption and

observe how model prediction/estimation of quantity of interest changes. On the other hand, a well built model should have identifiable model parameters. That is, there is a unique correspondence between a set of values of a model parameter and a set of values of model's output.[42] If model output as result of changing one model parameter can be completely reproduced by varying another parameter, this indicates poor model parameter identifiability. Both are necessary steps to determine whether a model is qualified for making inference of epidemiological parameters and variables.

Studies have examined several important behavioral characteristics of MSM population incorporating which can improve the model's ability to predict important epidemiological quantities.[43,44] However, one behavioral characteristic of MSM population is understudied, which is that individual risk behavior fluctuates over time, namely individual risk behavior volatility. Empirical data suggest that this phenomenon exists among MSM population.[45,46] However, for mathematical models of HIV transmission, it is a common assumption that individuals have constant risk behavior over time. Although some studies assume that individual risk behavior can monotonically change as response to public health intervention or prevalence of HIV infection, this does not reflect the same pattern of variation of risk behavior as individual risk behavior volatility.

It is not clear that how relaxing the assumption of constant individual risk behavior over time would affect a deterministic model's inference of the two epidemiological quantities: fraction of transmission from acute HIV infection and required individual effectiveness of an intervention strategy to eliminate HIV infections. Romero-Severson et al (2013) demonstrates that individual risk behavior volatility can increase the fraction of transmission from acute infection and prevalence of HIV infections.[47] Another study by Romero-Severson et al (2014) suggest that individual risk behavior volatility can reduce R_0.[48] However, both studies assume continuous risk behavior distribution and use individual-based models. Therefore, how incorporating individual risk behavior volatility would change the inference of deterministic compartmental model of these important quantities remain unexplored.

Potential Value of HIV Phylogenies in Estimating Individual Risk Behavior Volatility. Nonetheless, if a model's inference of epidemiological quantities is not robust to relaxation of a model assumption, one would want to improve the model structure by accommodating the

relaxed assumption. However, relaxing model assumptions often requires adding model parameters. This leads to decreased identifiability given available data.[42] Therefore, if adding individual risk behavior volatility can considerably alter model inference of quantities of interest, one would need to find data to improve the identifiability of the parameter that quantifies individual risk behavior volatility. One type of data that has such potential value is HIV sequence data. As HIV sequence data become increasingly abundant, studies start to use HIV sequence data to estimate important epidemiological parameters that is otherwise hard to infer with traditional surveillance data.[49–53] This is because that evolution of HIV virus happens at the same time scale of HIV transmission, making it possible to infer the HIV transmission dynamics from the pattern of branching of HIV phylogenetic tree.[54,49,55] Study by Alam et al (2012) suggest that episodic risk, a simplified version of individual risk behavior volatility for people with dichotomous risk behavior, can increase the extent that acute HIV infections cluster in the transmission tree.[56] This implies the potential value of HIV phylogenies in informing the individual risk behavior volatility. However, such value has not been explored.

Motivated by these unanswered research questions, this thesis is designed with the following research aims. First, we aim to explore whether relaxing the assumption that individual risk behavior remains constant over time changes model's inference of transmission from acute infection and prevalence of HIV infections. Second, we aim to explore whether relaxing the assumption that individual risk behavior remains constant over time changes model's inference of the minimum individual effectiveness of UT&T and PrEP strategy to eliminate HIV infections. We divide research under this study aim into two parts. Firstly, in order to explore how risk volatility affects PrEP strategy that is targeted at a specific risk group, we derive the measure of required efforts of targeted controls to eliminate transmissions, type reproduction number, for model with risk volatility. In the second part, we use this measure and R_0 and calculate minimum individual effectiveness of PrEP strategy to eliminate HIV infections. Third, we aim to explore whether signal of individual risk behavior volatility can be detected from HIV phylogenetic tree and whether such signal can be distinguished from other biological or behavior characteristics of MSM population.

As the first step to model individual risk behavior volatility deterministically, we focus on population with dichotomous risk behavior. Specifically, we model individual risk behavior

volatility as that individual risk alternates between high and low level over time. In the following chapters, we call such behavior characteristics as episodic individual risk behavior, abbreviated as episodic risk.

CHAPTER II

Effect of Episodic Risk on HIV Prevalence and the Fraction of Transmission from Acute HIV Infection

Introduction

In the decades from its emergence, great strides have been made in controlling HIV. Antiretroviral treatment has greatly improved and has saved millions of lives. Neonatal transmission and transmission via needles among injecting drug users have been markedly reduced.[1] On the other hand, HIV incidence among men who have sex with men (MSM) continues to rise with increasing rates of treatment having little effect on controlling transmission in this group.[1,2] There are diverse speculations as to why this might be the case including rapid evolution of the virus itself, safe-sex exhaustion, or emergence of versatility in terms of insertive or receptive sex roles among MSM. [3–6] Nonetheless, convincing evidence and arguments as to how different factors contribute to our failure to control HIV transmission among MSM are lacking. The formulations we present here constitute a theory expressing one mechanism through which biological and social factors interact to maintain high infection levels in MSM populations.

Our ability to understand HIV transmission dynamics is hindered by the fact that we cannot observe HIV transmission patterns directly. In part, this is due to most infections not being detected until long after transmission. That makes it impossible to trace all sexual partners and count directly how many transmissions are occurring during the highly infectious but brief acute HIV infection stage (AHI). It also makes it rare that AHI outbreaks are detected and investigated at the time they are occurring. Currently the existence of such outbreaks is detected through genetic clustering [52,57] and phylogenetic relationships.[53,58] As many as 64% of infections detected in the first 6-24 months after infection can be identified as part of genetic clusters.[52,57,59] Quite possibly most clustering involves transmission during AHI but inferring AHI transmission fractions from genetic data require better theory.

This chapter takes a step in that direction by developing a model that elucidates an important mechanism by which acute infections are linked into outbreaks. The ability to use genetic data to test transmission theory has recently been advanced by new methods for incorporating transmission models into coalescent models.[51,60,61] Being able to fit different models to genetic data creates the potential to assess the robustness of inferences about the fraction of transmissions from AHI by realistically relaxing diverse aspects of transmission models.[62] These include partnership formation and duration patterns[63], specific behaviors such as anal or oral sex with or without protection, insertive and receptive behavior patterns[64], different sexual mixing sites[65], and heterogeneous sexual behavior in the population.[66,67] In this chapter, we focus on elaboration of a particular model aspect that has especially strong effects: episodic risk behavior. Another chapter in this symposium illustrated short-term temporal heterogeneity in sexual contact rates in a prospective risk behavior study designed to estimate transmission risks for different types of sexual contacts between MSM (Romero-Severson et al. from this edition). Episodic risk is defined by brief periods of higher risk behavior possibly precipitated by random events such as a change in relationship status, drug use, or as an endogenous aspect of sexual behavior in general.

We modeled episodic risk with flows back and forth between high and low contact rate states. The first paper to do this was Koopman et al. (1997).[68] However, the predominant feeling has been that previous demonstration of high AHI transmission[68,69] was only applicable during the early stages of the epidemic.[70] That is, once infection levels reach an endemic equilibrium, the fraction of transmissions is a simple function of the times one spends in different stages and the transmission probabilities during those stages. [71,72]

In this chapter, we show how episodic risk can amplify AHI transmissions even after endemic equilibrium has been reached. Specifically, we show how behavioral changes in individuals over time can amplify transmission at the population level by amplifying transmission from acute HIV infection. The mechanisms include both increasing the contact rates of recently infected individuals and increasing the number of high risk susceptible individuals with whom high risk infected individuals can make contact.

Model Formulation

In this thesis, we developed a deterministic compartmental model of HIV transmission in men who have sex with men that integrates episodic risk behavior, non-random mixing, and multiple stages of infection. The model has only two risk phases with different contact rates (high and low), two infection stages (acute and chronic), and two sexual mixing sites. Contacts are modeled as instantaneous and symmetrical partnerships with an undefined number of sex acts having no specified direction or pattern (such as insertive or receptive, anal or oral).

The deterministic compartmental model is specified as a system of six differential equations for a MSM population, one for each compartment in the model. A schematic of the model is shown in Figure 1. The characters *S*, *A* and *C* for each compartment refer to the susceptible, acute HIV and chronic HIV stages respectively. The subscripts *H* and *L* refer to the high and low contact rate for each subpopulation respectively. The grey area in Figure 1 depicts the high-risk mixing site where individuals with high contact rate make a proportion of their contacts (υ) exclusively with each other. The dotted area designated as the common mixing site represents the site where individuals with high contact rate make the remaining fraction of their contacts, 1-m, with individuals with low contact rate proportionally. The parameters that govern the rate of each flow are shown in Figure II-1 near the arrow depicting that flow. The values of f_H and f_L are given in Table II-1.

Arrows between subpopulations with identical infection status represent the flow of individuals between these subpopulations due to the movement between high and low-risk phases. ϕ_H represents the rate of turnover from high to low contact rate while ϕ_L refers to the flow in the opposite direction. We assume a constant birth rate indicated by the arrows in Figure II-1 pointing into S_L and S_H. Individuals leave the system either by a process of 'natural removal' that accounts for cessation for sexual activity for any reason (horizontal arrows pointing out of each compartment in Figure II-1) or by death of chronically infected individuals due to AIDS (vertical arrows pointing out of the C_L and C_H compartments). Table II-1 presents a complete list of the model's parameters.

We model episodic risk as movement between phases of high and low sexual activity. The magnitude of difference between the high and low risk phases is parameterized by the ratio of contact rates in the high and low risk phases. Average contact rates in the whole system are

held constant such that an increase in this ratio or the proportion of individuals in the high-risk phase implies a decrease in the low risk contact rate to maintain a constant average contact rate. The default value for the average contact rate (χ) was selected so the system is just above the epidemic threshold under the homogenous risk setting without episodic dynamics.

The natural history of HIV infection is modeled with acute and chronic stages of infection. Contagiousness is abstracted as a higher transmissibility per act during a relatively short acute stage and a lower transmissibility during a much longer chronic stage with first order flows between stages and out of the last stage. As a simplifying assumption, we ignore the rise of contagiousness that has been demonstrated in late infection. This assumption is reasonable in a population with widespread access to antiretroviral treatments that have greatly reduced the prevalence of AIDS and the concomitant rise in viral titer associated with end stage HIV infection. Simulations also suggested that assuming three rather than two infections stages does not fundamentally alter the phenomenon that we illustrate. In our model, we assume the average duration of the acute stage to be 2 months and that of the chronic stage to be 10 years (Pilcher et al. 2004).[73] On average, uninfected individuals will remain sexually active for 40 years.

The contagiousness parameters were adapted from an analysis of the Rakai data by Pinkerton (2008).[74] In our model, we set the baseline transmission probability (β) across the entire course of infection to be 0.003, which is consistent with overall average transmission probabilities across all homosexual sex acts.[6] The possible underestimate of the baseline transmission probability per partnership is compensated for by the fact that we adjusted the average contact rate to be near threshold for the given baseline transmission probability.

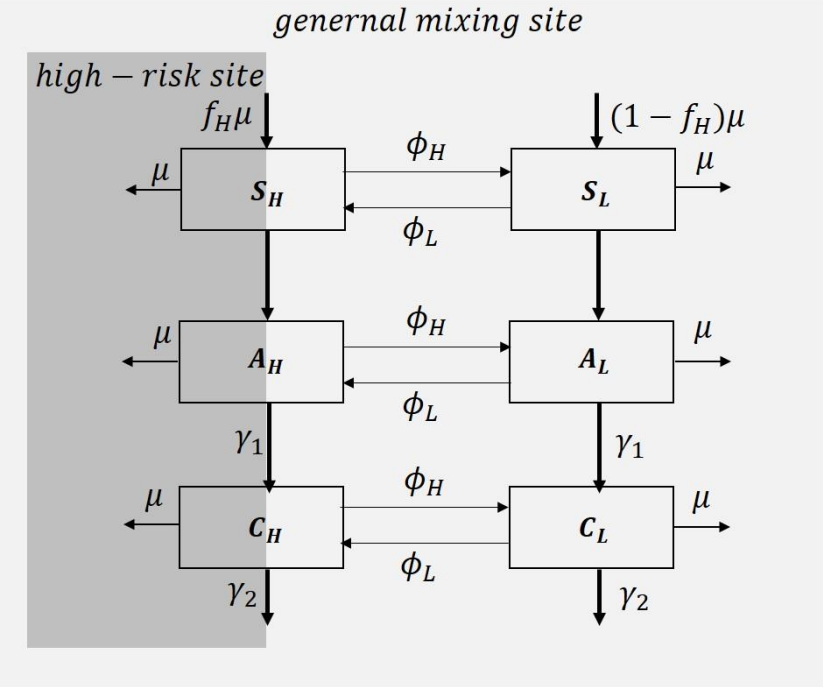

Figure II- 1 Conceptual episodic risk model with contact structure and population composition for model in Chapter II.

The increased contagiousness in the acute stage is governed in the model by the fraction of transmission potential from the acute stage (*X*), which is the expected fraction of transmissions from acute HIV occurring in a homogeneous population at equilibrium. The following presents how we derived the formulation of transmission probability per contact during acute HIV infection, β_1, and during chronic HIV infection, β_2, on the basis of baseline transmission probability, β and *X*.

The average time an infected individual spends in acute infection, denoted as T_1， is calculated as $T_1 = \frac{1}{\gamma_1 + \mu}$ (1). Given that the probability for an acutely infected individual to progress to the chronic stage rather than being removed due to other reasons is $\frac{\gamma_1}{\gamma_1 + \mu}$, the average time an infected individual spends in the chronic stage, denoted as T_2, can be expressed as $T_2 = (\frac{\gamma_1}{\gamma_1 + \mu})(\frac{1}{\gamma_2 + \mu})$ (2).

If the rate of sexual contact in the homogeneous population is denoted as □, the average number of secondary cases generated by an infected individual during acute infection is $\beta_1 \chi T_1$,

and the average number of secondary cases generated by an infector during chronic infection is $\beta_2 \chi T_2$, while the average number of secondary cases generated by an infector during the entire period of infection is $\beta\chi(T_1+T_2)$. As a result, the fraction of transmission potential from acute infection is $X = \frac{\beta_1 \chi T_1}{\beta\chi(T_1+T_2)} = \frac{\beta_1 T_1}{\beta(T_1+T_2)}$ (3), and the fraction of transmission potential from chronic infection is $1-X = \frac{\beta_2 \chi T_2}{\beta\chi(T_1+T_2)} = \frac{\beta_2 T_2}{\beta(T_1+T_2)}$ (4). After substituting equations (1) and (2) into (3), $\beta_1 = X(1+\frac{\gamma_1}{\gamma_2+\mu})\beta$. Using the same logic, we find that $\beta_2 = (1-X)(1+\frac{\gamma_2+\mu}{\gamma_1})\beta$.

Parameterizing the system in this way, we can easily determine the effect of episodic risk on acute stage transmission rates compared to a homogeneous null model. Transmissibility per contact is assumed to be constant during each infection stage. This is determined by the above-mentioned model parameters (see Table II-1).

Table II- 1 List of the episodic risk model parameters and their default values used in Chapter II

Parameter	Value	Unit	Definition
μ	1/(40*12)	/month	Rate of flow of new individuals into sexually active population and also Rate of leaving the sexually active population
γ_1	6	/month	Rate of transitioning from acute to chronic infection
γ_2	1/120	/month	Rate of death from AIDS during chronic infection
β	0.003	/contact	Average transmission probability across stages
X	variable	-	expected fraction of transmissions from acute HIV occurring in a homogeneous population at equilibrium
β_1	$X\beta\left(1+\frac{\gamma_1}{\gamma_2+\mu}\right)$	/contact	Transmission probability during acute stage
β_2	$(1-X)\beta\left(1+\frac{\gamma_2+\mu}{\gamma_1}\right)$	/contact	Transmission probability during chronic stage
f_H	variable	-	Average fraction of population with high contact rate in the absence of HIV
χ	3.4114	/month	Average contact rate in the entire population
r_{HL}	variable	-	Ratio of high contact rate over

			low contact rate
χ_H	$r_{HL}\chi_L$	/month	Contact rate for the high-risk population
χ_L	$\frac{\chi}{1-f_H+f_H r_{HL}}$	/month	Contact rate for the low risk population
v	0.3	-	Fraction of contacts of individuals with high contact rate at the high-risk site
ϕ_H	variable	/month	Rate of transitioning from high contact rate state to low contact rate state
ϕ_L	$\phi_H(\frac{f_H}{1-f_H})$	/month	Rate of transitioning from low contact rate state to high contact rate state
λ_{Hsite}	$\frac{A_H\beta_1+C_H\beta_2}{S_H+A_H+C_H}$	-	Force of infection per contact at the high-risk mixing site
λ_{Gsite}	$\frac{(A_L+(1-v)A_H r_{HL})\beta_1+(C_L+(1-v)C_H r_{HL})\beta_2}{S_L+A_L+C_L+(S_H+A_H+C_H)r_{HL}(1-v)}$	-	Force of infection per contact at the general mixing site

Table II- 2 Differential equations for the deterministic episodic risk compartmental model for Chapter II

$$\frac{dS_H}{dt} = \mu f_H - S_H\chi_H\left(v\lambda_{H_{site}} + (1-v)\lambda_{G_{site}}\right) - \phi_H S_H + \phi_L S_L - \mu S_H$$

$$\frac{dS_L}{dt} = \mu(1-f_H) - S_L\chi_L\lambda_{G_{site}} + \phi_H S_H - \phi_L S_L - \mu S_L$$

$$\frac{dA_H}{dt} = S_H\chi_H\left(v\lambda_{H_{site}} + (1-v)\lambda_{G_{site}}\right) - A_H\gamma_1 - \phi_H A_H + \phi_L A_L - \mu A_H$$

$$\frac{dA_L}{dt} = S_L\chi_L\lambda_{G_{site}} - A_L\gamma_1 + \phi_H A_H - \phi_L A_L - \mu A_L$$

$$\frac{dC_H}{dt} = A_H\gamma_1 - \phi_H C_H + \phi_L C_L - \gamma_2 C_H - \mu C_H$$

$$\frac{dC_L}{dt} = A_L\gamma_1 + \phi_H C_H - \phi_L C_L - \gamma_2 C_L - \mu C_L$$

To better clarify what flows from what stages of infection and what contact rates were sustaining transmission, we divided each of the four compartments of the infected population (A_H, A_L, C_H, C_L) from the episodic risk model (Table II-2) into six compartments. By doing this,

we track the information of location of transmission, contact rate phase and stage of HIV infection of the source infectors. Thus, in this elaborated model, infected individuals now have a 3-letter superscript that refer to the state of the individual who infected them and where they went on to transmit. The first character in the superscripts refers to the site where they became infected, with *R* being the high-risk mixing site and *G* being the common mixing site. The middle letter refers to the contact rate phase of the source infector, with *H* being the high contact rate phase and *L* being the low contact rate phase. The last character represents the stage of HIV infection of the source infector, with *A* being the acute stage and *C* being the chronic stage. The set of differential equations for the dynamic system analysis of the elaborated model is provided in Table II-3.

Model elaboration for dynamic system analysis

Table II- 3 Equations for the elaborated deterministic compartmental model focusing on the source and site of infection in Chapter II

$$\frac{dA_H^{RHA}}{dt} = S_H \chi_H \nu \frac{A_H \beta_1}{S_H + A_H + C_H} - A_H^{RHA}\gamma_1 - \phi_H A_H^{RHA} + \phi_L A_L^{RHA} - \mu A_H^{RHA}$$

$$\frac{dA_H^{RHC}}{dt} = S_H \chi_H \nu \frac{C_H \beta_2}{S_H + A_H + C_H} - A_H^{RHC}\gamma_1 - \phi_H A_H^{RHC} + \phi_L A_L^{RHC} - \mu A_H^{RHC}$$

$$\frac{dA_H^{GHA}}{dt} = S_H \chi_H (1-\upsilon) \frac{A_H (1-\upsilon) r_{HL} \beta_1}{(S_H + A_H + C_H)(1-\upsilon) r_{HL} + S_L + A_L + C_L} - A_H^{GHA}\gamma_1 - \phi_H A_H^{GHA} + \phi_L A_L^{GHA} - \mu A_H^{GHA}$$

$$\frac{dA_H^{GHC}}{dt} = S_H \chi_H (1-\upsilon) \frac{C_H (1-\upsilon) r_{HL} \beta_2}{(S_H + A_H + C_H)(1-\upsilon) r_{HL} + S_L + A_L + C_L} - A_H^{GHC}\gamma_1 - \phi_H A_H^{GHC} + \phi_L A_L^{GHC} - \mu A_H^{GHC}$$

$$\frac{dA_H^{GLA}}{dt} = S_H \chi_H (1-\upsilon) \frac{A_L \beta_1}{(S_H + A_H + C_H)(1-\upsilon) r_{HL} + S_L + A_L + C_L} - A_H^{GLA}\gamma_1 - \phi_H A_H^{GLA} + \phi_L A_L^{GLA} - \mu A_H^{GLA}$$

$$\frac{dA_H^{GLC}}{dt} = S_H \chi_H (1-\upsilon) \frac{C_L \beta_2}{(S_H + A_H + C_H)(1-\upsilon) r_{HL} + S_L + A_L + C_L} - A_H^{GLC}\gamma_1 - \phi_H A_H^{GLC} + \phi_L A_L^{GLC} - \mu A_H^{GLC}$$

$$\frac{dC_H^{RHA}}{dt} = A_H^{RHA}\gamma_1 - C_H^{RHA}\gamma_2 - \phi_H C_H^{RHA} + \phi_L C_L^{RHA} - \mu C_H^{RHA}$$

$$\frac{dC_H^{RHC}}{dt} = A_H^{RHC}\gamma_1 - C_H^{RHC}\gamma_2 - \phi_H C_H^{RHC} + \phi_L C_L^{RHC} - \mu C_H^{RHC}$$

$$\frac{dC_H^{GHA}}{dt} = A_H^{GHA}\gamma_1 - C_H^{GHA}\gamma_2 - \phi_H C_H^{GHA} + \phi_L C_L^{GHA} - \mu C_H^{GHA}$$

$$\frac{dC_H^{GHC}}{dt} = A_H^{GHC}\gamma_1 - C_H^{GHC}\gamma_2 - \phi_H C_H^{GHC} + \phi_L C_L^{GHC} - \mu C_H^{GHC}$$

$$\frac{dC_H^{GLA}}{dt} = A_H^{GLA}\gamma_1 - C_H^{GLA}\gamma_2 - \phi_H C_H^{GLA} + \phi_L C_L^{GLA} - \mu C_H^{GLA}$$

$$\frac{dC_H^{GLC}}{dt} = A_H^{GLC}\gamma_1 - C_H^{GLC}\gamma_2 - \phi_H C_H^{GLC} + \phi_L C_L^{GLC} - \mu C_H^{GLC}$$

$$\frac{dA_L^{RHA}}{dt} = -A_L^{RHA}\gamma_1 + \phi_H A_H^{RHA} - \phi_L A_L^{RHA} - \mu A_L^{RHA}$$

$$\frac{dA_L^{RHC}}{dt} = -A_L^{RHC}\gamma_1 + \phi_H A_H^{RHC} - \phi_L A_L^{RHC} - \mu A_L^{RHC}$$

$$\frac{dA_L^{GHA}}{dt} = S_L\chi_L \frac{A_H\beta_1(1-\upsilon)r_{HL}}{(S_H + A_H + C_H)(1-\upsilon)r_{HL} + S_L + A_L + C_L} - A_L^{GHA}\gamma_1 + \phi_H A_H^{GHA} - \phi_L A_L^{GHA} - \mu A_L^{GHA}$$

$$\frac{dA_L^{GHC}}{dt} = S_L\chi_L \frac{C_H\beta_2(1-\upsilon)r_{HL}}{(S_H + A_H + C_H)(1-\upsilon)r_{HL} + S_L + A_L + C_L} - A_L^{GHC}\gamma_1 + \phi_H A_H^{GHC} - \phi_L A_L^{GHC} - \mu A_L^{GHC}$$

$$\frac{dA_L^{GLA}}{dt} = S_L\chi_L \frac{A_L\beta_1}{(S_H + A_H + C_H)(1-\upsilon)r_{HL} + S_L + A_L + C_L} - A_L^{GLA}\gamma_1 + \phi_H A_H^{GLA} - \phi_L A_L^{GLA} - \mu A_L^{GLA}$$

$$\frac{dA_L^{GLA}}{dt} = S_L\chi_L \frac{C_L\beta_2}{(S_H + A_H + C_H)(1-\upsilon)r_{HL} + S_L + A_L + C_L} - A_L^{GLA}\gamma_1 + \phi_H A_H^{GLC} - \phi_L A_L^{GLC} - \mu A_L^{GLC}$$

$$\frac{dC_L^{RHA}}{dt} = A_H^{RHA}\gamma_1 - C_L^{RHA}\gamma_2 - \phi_L C_L^{RHA} + \phi_H C_H^{RHA} - \mu C_L^{RHA}$$

$$\frac{dC_L^{RHC}}{dt} = A_H^{RHC}\gamma_1 - C_L^{RHC}\gamma_2 - \phi_L C_L^{RHC} + \phi_H C_H^{RHC} - \mu C_L^{RHC}$$

$$\frac{dC_L^{GHA}}{dt} = A_H^{GHA}\gamma_1 - C_L^{GHA}\gamma_2 - \phi_L C_L^{GHA} + \phi_H C_H^{GHA} - \mu C_L^{GHA}$$

$$\frac{dC_L^{GHC}}{dt} = A_H^{GHC}\gamma_1 - C_L^{GHC}\gamma_2 - \phi_L C_L^{GHC} + \phi_H C_H^{GHC} - \mu C_L^{GHC}$$

$$\frac{dC_L^{GLA}}{dt} = A_H^{GLA}\gamma_1 - C_L^{GLA}\gamma_2 - \phi_L C_L^{GLA} + \phi_H C_H^{GLA} - \mu C_L^{GLA}$$

$$\frac{dC_L^{GLC}}{dt} = A_H^{GLC}\gamma_1 - C_L^{GLC}\gamma_2 - \phi_L C_L^{GLC} + \phi_H C_H^{GLC} - \mu C_L^{GLC}$$

Results

Exploring joint effects of high-risk turnover rate and between-risk-group contact rate ratio at the population-level. We numerically solved equations out to equilibrium. The rate of turnover in the high-risk group had strong effects on both the endemic prevalence and fraction of transmission during acute stage.

As the turnover rate of the high-risk group slows from 10 times per month (once every 0.008 years) to once per 1000 months (once every 83.3 years), both the endemic prevalence and the acute transmission rate show an inverted U shape, peaking between turnover rates of 1 per month and 1 per year (Figure II-2). At the extremes of fast or slow turnover, the fraction of acute transmissions is the same as in a homogeneous population. As the ratios of contact rates between high and low risk groups increase, the peak effects of turnover rates increase as well and occur at faster turnover rates.

Increasing the fraction of transmission potential during acute stage while keeping the total transmission potential constant increases both the endemic fraction of transmissions during acute stage and the endemic prevalence (compare panels A and B). Furthermore, the endemic prevalence peaks at a faster turnover rate because given an increased transmission rate during the acute stage, individuals need to spend shorter time at high-risk phase to lift the endemic prevalence to peak. These observed effects hold at all values of the contact rate ratio as shown in Figure II-2.

As the fraction of time that each individual spends at high risk (f_H) increases, both the endemic fraction of transmission during the acute stage and the endemic prevalence decrease at each value of the contact rate ratio (compare panels C and D in Figure II-2). Given our assumption of constant average contact rate, for any given contact rate ratio greater than 1, increasing f_H causes a decrease in contact rates for both risk levels and hence decreases the overall force of infection. When the turnover rate is low enough and the contact rate ratio is high enough, the endemic prevalence starts to decrease as the contact rate ratio increases. Given a slow turnover, high-risk susceptible individuals are infected much faster as compared to the low risk susceptible individuals due to their elevated contact rates causing infection saturation in the high-risk phase. Higher turnover rates supplement the pool of high-risk susceptible individuals with new high-risk susceptible individuals. As f_H increases, each individual spends more time in the high-risk phase. This increases the proportion of contacts between infected individuals causing saturation of infection in the high-risk phase.

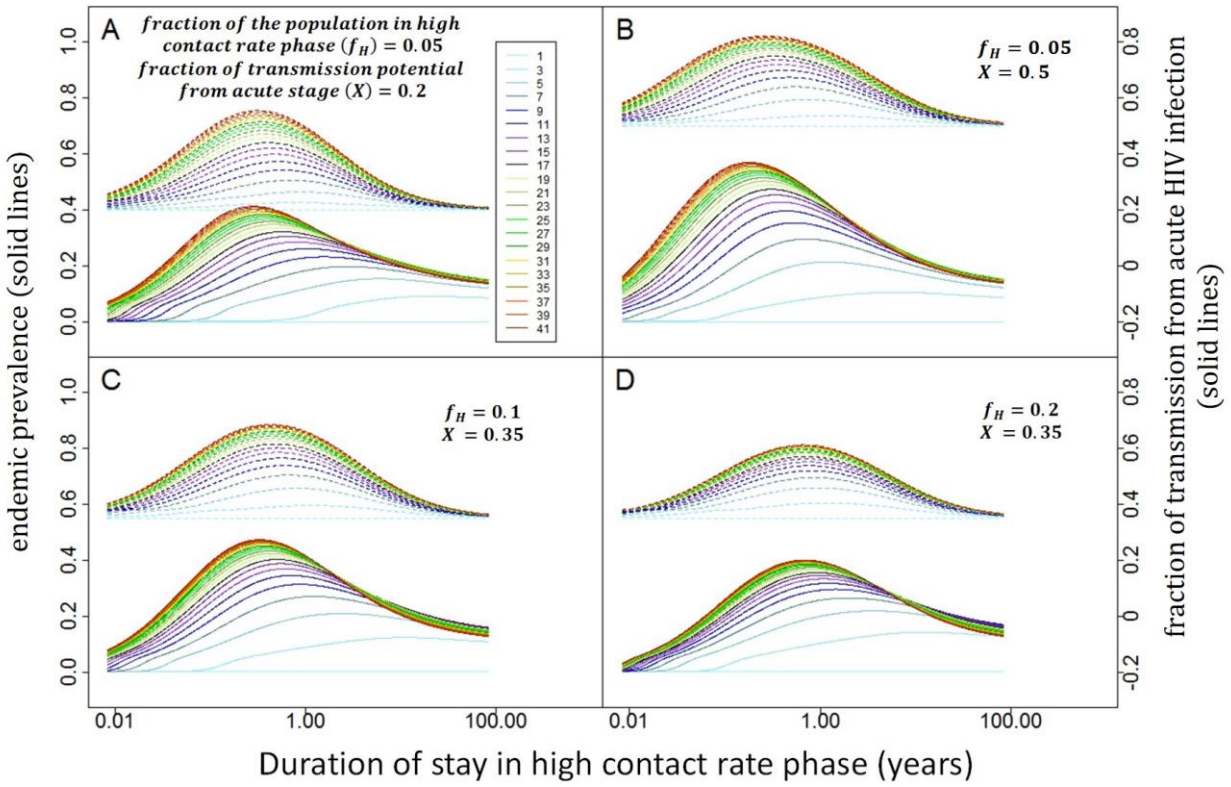

Figure II- 2 Endemic prevalence and fraction of transmission during acute stage of infection as function of log scale of duration of stay at high risk level as the contact rate ratio increases from 1 to 41 by increments of 2 and showed in different colors. The left y axis is for the endemic prevalence and the right y axis is for the fraction of transmission during the acute stage of infection. Parameter values not specified in this figure are consistent with those in Table II-1.

Effect of separate mixing on the endemic prevalence and fraction of transmission by acute stage of infection

Overall, the fraction of contacts made by high-risk infectors at the high-risk site (f_H) showed little impact either on the endemic prevalence or on fraction of transmission during acute stage or on the dynamics of these two outcomes (data not shown). This is unlike the observations in Koopman et al. (2005).[65] Only at lower overall contact rates and very slow turnover between contact rate groups do we see the effects of f_H as noted in Koopman et al (2005).[65] At any other contact rates, a higher fraction of high contact rate group's contacts made at a high-risk mixing site increases the rate of exponential growth in the early epidemic.

Exploration of dynamics underlying high contact rate group turnover effects

We now look into the dynamics of prevalence and the fraction of transmissions from acute HIV (AHI) over a period of 150 years. In Figure II-3 (A and B), we explore these dynamics for the duration of time spent in the high-risk state for ranges from 0.01 to 80 years. In Figure II-

3 (C and D), we explore the dynamics by varying the ratio of contact rates from 1 to 41.

Figure II-3 (A and B) shows that episodic risk increases the contribution of acute HIV early in the epidemic and later at the endemic equilibrium. Faster risk turnover results in the faster replenishment of the susceptible population in the high-risk phase. At the same time, it increases the risk of a low-risk susceptible contacting an infected individual who is still in the highly contagious acute stage. Note that while varying the average duration of high risk phase, we keep the fraction of time that population spend in high risk phase, f_H, constant. Therefore, the increase in replenishment of high risk susceptibles is simply result of increase in turnover rate. For slower turnover rates, the fraction of the transmissions from AHI decreases. This is because of the decrease in the inflow of susceptibles in the high-risk phase. It also increases the duration of the time spent in the high-risk phase by high-risk individuals. Thus, high-risk individuals are likely to progress to the chronic stage by the time they switch to the low-risk phase.

Overall, the prevalence and the fraction of transmissions during the acute stage increase with the ratio of contact rates (see Figure II-3; panels C and D). Increasing contact rate ratios result in increasing high-risk contact rates that in turn increase transmission from high-risk infectors. For high contact rate ratios, the fraction of transmission from acute stage reaches a higher peak level, indicating a greater effect when the high-risk contact rate increases.

Higher peak levels of acute stage transmissions reflect the potential of the acute HIV population to build the epidemic together with the replenishment of the susceptible population. This explains why the fraction of transmissions during acute stage with a higher peak level also has a larger rebound as shown in Figure II-3 (B and D).

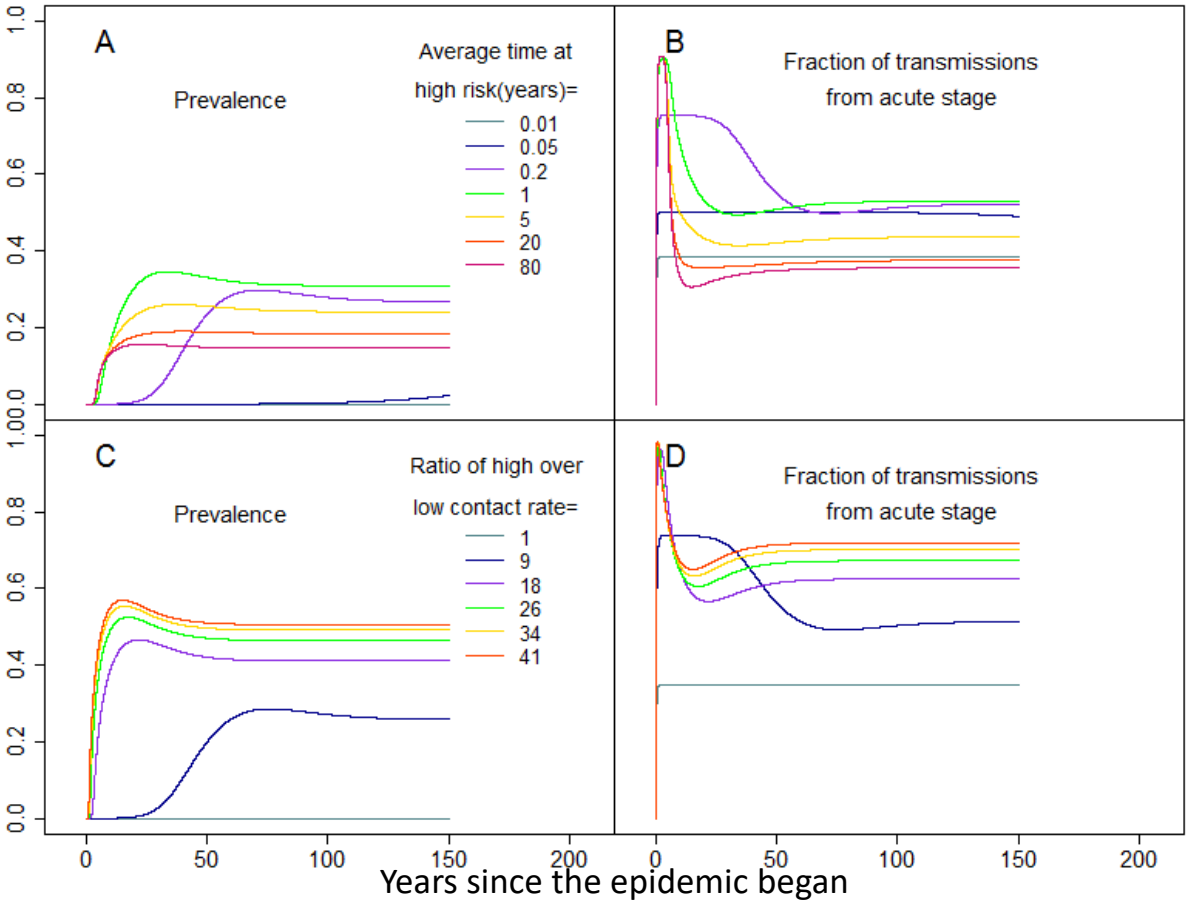

Figure II- 3 Model runs showing the endemic prevalence (Panels A and C) and the fraction of transmissions from the acute stage (Panels B and D) for 150 years. Above: average time spent at high contact rate phase is raised from 0.1 to 80 years. Below: the ratio of high over low contact rate is raised from 1 to 41. In these simulations, the average fraction of time population spends at high risk phase, f_H, is kept at 0.05. Other parameter values not specified in this figure are the same as given in Table II-1.

Effect of risk turnover and contact rate ratios on the contact rate of population with acute stage of HIV

In Figure II-4, we examine the effects of high-risk group turnover rate on the endemic prevalence, the average contact rates of the acute stage population and that of the susceptibles at equilibrium. We explore for a range of the duration of time spent in the high-risk phase from 0.008 years to 83 years and the ratio of contact rates from 1 to 41. As Figure II-4B shows, the turnover rate of the high-risk population has a clear effect on the average contact rates of acute stage infected individuals at equilibrium. Higher turnover rate brings more susceptibles into the high contact rate phase. At higher contact rate ratios, the new susceptible at high contact rate are more likely to be infected and thus faster turnover in the high contact rate phase has even greater effects.

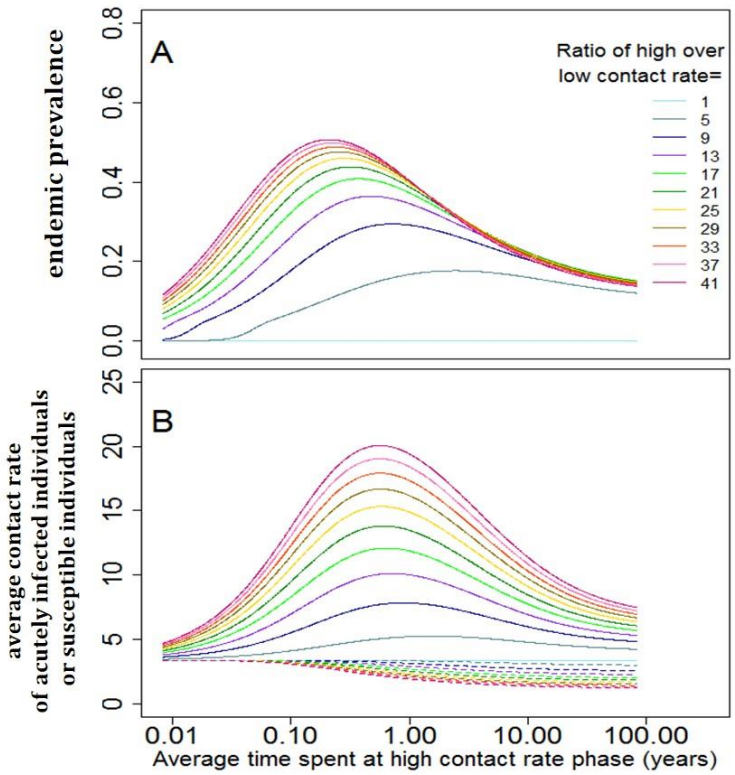

Figure II- 4 Endemic prevalence (A) and average contact rate of acutely infected individuals (solid lines in B) or average contact rate of susceptibles (dashed lines in B) at equilibrium as average time spent at high contact rate phase is varied from 0.01 to 100 years. Parameter values not specified in this figure are the same as given in Table II-1.

The endemic prevalence reaches its peak at a faster turnover rate than the average contact rate of acutely infected individuals (Figure II-4A). This is due to the combined effect of the overall force of infection and average contact rate among the susceptibles (Figure II-4B). Given the high contagiousness of the acute stage infectors, the interaction of high transmissibility and increased contact rates during acute stage drives the overall force of infection.

Tracking sources of infections

By breaking the compartments of infections in our original model into different subgroups, we were able to track the source of infection and mixing site where the transmissions occurred. Figure II-5A and II-5C illustrate the fraction of transmissions during acute stage and chronic stage HIV infection. Figures II-5B and II-5D focus on the contributions of high-risk infectors and low risk infectors. From B and D, although the high contact rate group only comprises of 5% of the total population, they account for 50% or more of all transmissions at contact rate ratios as low as 7. Although there is an effect of the turnover rate on the fraction of infections by individuals in the high-risk phase, the effect is small. The role of the high versus low-risk

infectors is controlled primarily by the contact rate ratio.

From panels A and C in Figure II-5, we see a different role for the contact rate ratio and the turnover rate in determining the fraction of transmission from acute stage infectors. The contact rate ratio has little effect on the fraction of infections from acute infectors at extreme values of the turnover rate. However, at intermediate values of the turnover rate, the contact rates ratio can nearly double the contribution of acute stage infectors. This interaction is caused by the increased prevalence of high-risk, acutely infected individuals. The flow of new susceptibles into the high risk phase generates an increased incidence and therefore an increased prevalence of acute stage, high risk infectors. Because the dynamics are assumed to be mass action, the increased prevalence of acute high-risk infectors means increased contact between susceptibles and acutely infected individuals. The effect diminishes at extremely short average durations of the high risk phase because the system becomes essentially homogeneous. In that case, the system is controlled by the average contact rate rather than the contact rate ratio. At extremely long average durations, on the other hand, the system becomes essentially two non-interacting sub-systems, each of which is homogeneous. In this limit, the contact rate ratio still matters for determining the total incidence, but does not affect the fraction of that incidence that comes from transmissions by acutely or chronically infected individuals.

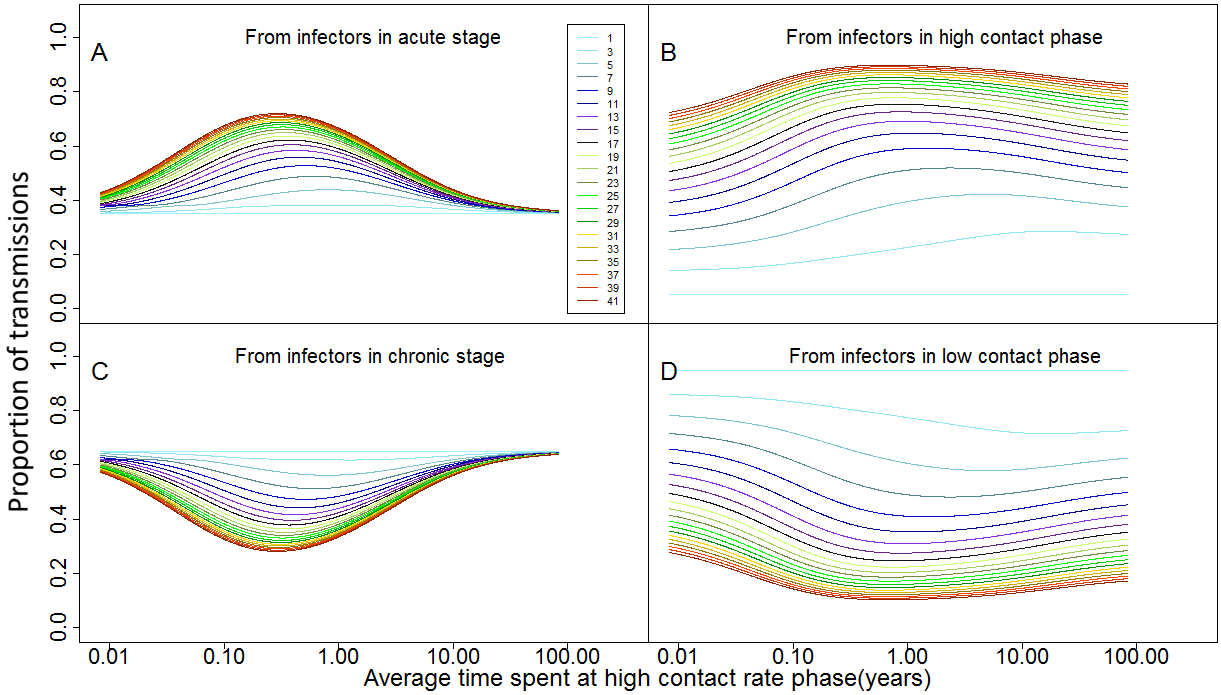

Figure II- 5 Proportion of transmissions at endemic equilibrium generated by acutely infected individuals (A), infected individuals with high contact rate (B), chronically infected individuals (C) or infected individuals with low contact rate (D). In each panel, the ratio of high over low contact rate is raised from 1 to 41 by increments of 2. Parameter values not specified in this figure are the same as given in Table II-1.

In Figure II-6, we further divide the source of infection based on their contact rate phase, stage of infection and location. When the turnover rate is low, neither acute nor chronic individuals are generating many new infections at the high-risk site (panels A and D). This is due to the fact that without turnover to replenish susceptibles in the high risk phase, the high-risk individuals rapidly become infected. Therefore, at the high-risk site almost all contacts are among already infected individuals. However, at the low-risk site, susceptibles with low contact rates are much more likely to contact either an acute (panel B) or a chronic (panel E) high risk infected individual resulting in elevated incidence rates from those classes in the absence of turnover. Also, by assumption, most individuals entering the model are susceptibles with low contact rates, which provides a greater supply of susceptibles to the low risk mixing site.

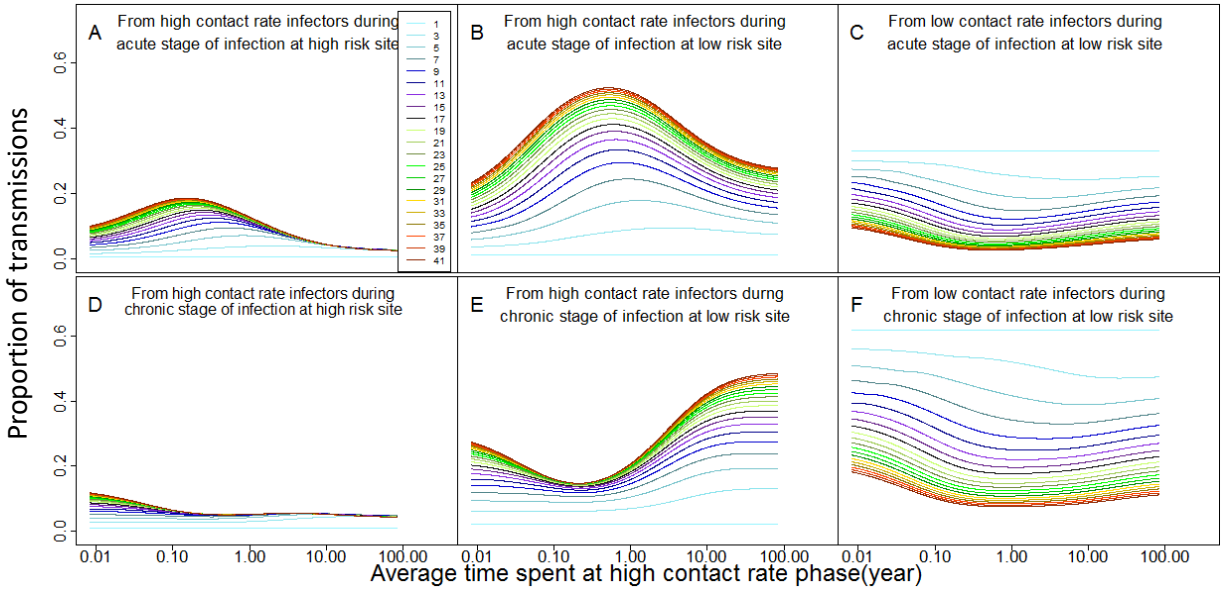

Figure II- 6 Proportion of transmissions at endemic equilibrium generated by different types of infected individuals. In each panel, the high-to-low contact rate ratio is raised from 1 to 41 by increments of 2. Parameter values not specified in this figure are the same as given in Table II-1.

Discussion of Results in Chapter II

Explaining the major effects observed in this Chapter

We have shown that episodic periods of high-risk behavior interspersed with longer periods of lower risk have large effects on the endemic level of HIV infection and the fraction of transmissions from acute HIV infection. We see that prevalence can vary from near zero to very high levels and the fraction of transmissions from acute infection can nearly double as the turnover rate from low-to-high contact rate phases increase and as the time spent with high rates varies. This is despite the fact that the total contacts in the population and the transmission potential for each individual are kept constant during each stage of infection.

Two important elements explain the strong effects mentioned above. The first has to do with the average total contagiousness of individuals during acute and chronic infection. That contagiousness is determined jointly by the number of contacts made and transmission probabilities per contact. Episodic risk increases the association between high contact rates and the high transmissibility of acute infection because an individual is more likely to get infected during a high contact rate period and is more likely to still be in that high contact rate period during their acute HIV infection (AHI). Then, as they progress to the less contagious chronic infection stage, they also pass into a lower contact rate phase. The more strongly that the high transmission stage of infection is correlated with the high contact rate state, the higher will be the average contagiousness across the course of infection. This is reflected in the part of Fig.II-4 that presents the average contact rates during the acute infection stage.

The second part of the explanation for these big effects has to do with the average contact rates of susceptible individuals. When individuals spend a long time in the high contact rate state, they become infected and a sexual contact by them with other infected individuals no longer presents an opportunity to transmit infection. If individuals stay in a low contact rate state, they are more likely to avoid infection. Then when they transit to a high contact rate state, they provide a new source of high contact rate susceptible individuals that the individuals with high contact rates in the acute infection stage can infect. This is seen in the contact rates of susceptible individuals in Fig.II-4. The average contact rate of susceptible individuals is seen to increase considerably as the turnover increases. When the turnover rate is slow, high contact rate individuals are infected and are not replaced by new susceptible individuals. This lowers the

average contact rate of susceptible individuals.

Our previous work on when to focus control on high risk groups led us to believe that a third element would also act to generate episodic risk effects on prevalence and AHI transmissions.[65] That previous work modeled SIR infections. For SIR infections, we saw that when high-risk individuals mixed separately from low risk individuals, it allowed them to build up higher levels of infection so that they then disseminated more infection when they made contact with lower risk individuals. However, that was not the case for the HIV infections modeled here. Whether high contact individuals mixed at a high-risk or a common mixing site, the prevalence and the fraction of transmissions from acute infection were barely changed at all. The difference was not that in the previous work we allowed transmission probabilities rather than contact rates to define high risk. In fact, we saw the same effects of separate high risk mixing in that previous work when we defined high risks by contact. In further support of the cause of high risk (contact rates vs. transmission probabilities), we have observed that mixing makes little difference when high risk is caused by higher transmission probabilities (data not presented). The reasons for this are complex: under some conditions, separate mixing at a high-risk site can lead to a greater dissemination of infection. However, discussion of this effect is beyond the scope of this chapter.

The real-world importance of the described theoretical effects

If the episodic risk effects that we have described occur in the real world to the extent that we suspect they do, then previous model analyses using fixed behaviors for subgroups have underestimated the fraction of transmissions from acute HIV. [7,70] Nevertheless, a key question is how we can assess the extent to which these episodic risk effects are occurring in any particular population. Fluctuation of contact rates was observed in the CDC cohort studied to assess HIV transmission risks from different acts.[75] That study included diverse geographic areas. The existence of episodic risk behavior is less formally supported by experiences of HIV counselors where most people with high-risk behaviors report that they have had longer periods of low risk behaviors.

Our results may be compared to recent modeling studies that have focused on the role of acute infection.[76,77] Eaton et al. (2011) analyze an important phenomenon that, like episodic risk, augments the importance of acute infection.[76] In some sense, entering a high contact rate phase may be akin to entering a concurrent partnership phase, although in our model, we assume instantaneous contacts. On the other hand, the study by Powers et al. (2011) assume fixed risk,

which corresponds to our results when the duration of stay in high risk is very long such that there is virtually no risk turnover.[77] The Bayesian melding approach to model fitting used by Powers et al. (2011) means that the effects of the phenomenon of episodic risk could be absorbed partially into the transmission by stage parameter estimates. It seems that their results would have estimated a larger fraction of transmissions from acute infection if they had included episodic risk in their model. However, this is hard to assess.

Broader evidence should be sought regarding whether there are real world effects of episodic risk corresponding to the theoretical effects we have demonstrated. One approach could be to do more behavioral studies that ask the needed questions. So far, few studies have documented episodic risk behavior. Another approach is to find signatures of episodic risk in both HIV surveillance data and in HIV genetic sequences.[52] When the model in this chapter is simulated as a stochastic discrete individual-based model, we observe that infection trees of acute HIV transmissions are interspersed among more isolated chronic infection transmissions and have size distributions that are roughly comparable to those observed in Montreal.[57] Our initial results show that episodic risk dynamics influence both the size and duration of acute HIV outbreaks providing a possible link between genetic cluster size distributions and episodic risk dynamics. More sophisticated methods that are currently being developed can be used to calculate the likelihood of the genetic data given a transmission model allowing for direct evaluation of the evidence for episodic risk in the genetic data.[51,60,61] Parameter ranges should be further constricted by fitting models to observed patterns of HIV infection over time using surveillance data.

More thorough model explorations may be needed to assess the real world importance of episodic risk effects. More surprises regarding dynamic effects may arise as we analyze the interactions of episodic risk effects with contact pattern effects, partnership duration, insertive-receptive behavior effects, and oral, anal, protected, and unprotected risk pattern effects. To strengthen our inferences that episodic risk behavior greatly increases HIV prevalence and acute HIV transmissions, a thorough exploration of the effects of relaxing the simplifying assumptions in the model presented is called for. For example, realistic distributions for the duration of acute stage and the distribution of virus levels across that time should be explored. In reality, there is a high narrow peak of virus levels a few weeks into infection and then a long period of falling levels before the set point of chronic infection is reached.[73] We hypothesize that the effects we

have observed will be just as strong or stronger under more realistically detailed acute infection timing and contagiousness. Likewise, our initial explorations indicate that a more realistic description of episodic risk behavior in a population will not significantly diminish the effects we have illustrated in this chapter.

The infection control implications of our findings suggest that some HIV transmission control activities may be misdirected. For example, the test and treat strategy by default detects mostly cases after acute infection. Our results indicate that greater focusing on acute infection transmission clusters will be needed to detect more acute infections.[52] Second, behavior modification messages may be similarly misdirected. In our model, we did not incorporate change in sexual behavior over an individual's lifetime because of, e.g. HIV awareness programs. This is because our aim in this chapter is to understand the dynamics of episodic risk behavior under theoretical settings. Nevertheless, if suggestions to decrease risk only work for a period but are overwhelmed by situational pressures, then our model indicates that such suggestions could not only fail to decrease transmission, but might in fact increase transmission.

The broader context of this work and future directions

The analyses presented in this chapter are a needed step toward improving our understanding of HIV transmission dynamics. All perception comes from a combination of data and theory. We do not see patterns when our minds are not prepared to see them. To an extent our ignorance of HIV transmission patterns in MSM is attributable to a dearth of well-structured theory about how different hypothetical transmission system conformations lead to specific transmission patterns and what mechanisms underlie these effects. A body of HIV transmission system theory is growing and thousands of modeling studies of HIV transmission have been published.[78] Still, theory and data are barely beginning to come together in a way that allows us to see clearly how and why acute HIV transmissions cluster, and how this could explain the failure to control the spread of HIV infection in the MSM population.

By relaxing some of the unrealistic assumptions in our model and incorporating realistic details such as oral and anal acts, insertive and receptive behavior, partnership duration, and mixing patterns, we will begin to identify characteristics of both surveillance data and genetic sequence data that indicate what is affecting transmission dynamics. This, we believe, will clarify the role acute infection transmissions plays in the MSM HIV transmission system.

CHAPTER III

Formulating and Interpreting Type Reproduction Numbers for Population with Individual Risk Behavior Volatility

Introduction

One of the most important concerned questions in controlling transmissions of infectious diseases is how much control efforts is required to eliminate infections. The most commonly used measure for the elimination condition is basic reproduction number, R_0. R_0 calculates the expected number of secondary cases a case would cause during the entire infection time if this case was introduced into a fully susceptible population.[79] It is found that for homogeneous population the minimum coverage of vaccination to reach elimination is $1-1/R_0$.[79] However, when the population has heterogeneous risk states and intervention is specifically targeted at a risk group, R_0 is less relevant. In this case, the appropriate threshold measure is the type reproduction number, T. The concept of T is introduced by Roberts and Heesterbeek (2003). It is defined as the expected number of new infections of a specific type that a case of this type would cause during the entire infection time through all transmission paths without intermediate new infections of this type.[80] It is found that T also has the threshold property as R_0: $R_0>1$ and $T>1$, $R_0=1$ and $T=1$, and $R_0<1$ and $T<1$.[80] In addition, Heesterbeek and Roberts (2007) found that the minimum required control efforts of a targeted intervention to eliminate infections can be expressed as a simple function of T.[81] Specifically, as suggested by Heesterbeek and Roberts (2007), the minimum coverage of vaccination of type i susceptible group to reach elimination is a simple function of corresponding type reproduction number Ti, i.e. $1-1/Ti$.[81]

In study by Henry and Koopman (2015), R_0 for model with episodic risk has been clearly formulated and thoroughly discussed.[82] In study by Romero-Severson et al (2012), R_0 as function of risk volatility that incorporates continuously distributed individual risk behavior is also

derived and examined.[48] However, there has not been any study that examines the type reproduction numbers for model with individual risk behavior volatility. As targeting people experiencing higher risk of HIV acquisition has been one of the most common intervention strategy to control HIV transmissions, understanding how risk volatility affects type reproduction number will help clarify how critical control efforts to eliminate HIV transmissions would be different given risk volatility.

Therefore, in this chapter, we formulate type reproduction number, T, for a transmission system of HIV infection among MSM population where individuals have risk volatility. As the first attempt to do so, we focus on a system where individuals alternate risk behavior between a higher level and a lower level, i.e. episodic risk. The aim of this chapter is to understand how individual risk behavior volatility (episodic risk) affects type reproduction numbers and critical control efforts of relevant targeted intervention strategies. In Section 1, we describe the structure of the deterministic model and model differential equations. In Section 2, we use the next generation matrix approach as outlined in Roberts and Heesterbeek (2003) to derive T. In Section 3, we formulate the type reproduction number for model with risk volatility based on its epidemiological meaning. In Section 4, we show that the expression of type reproduction numbers derived using next generation matrix approach is equivalent to that derived using the approach presented in Section 3. In Section 5, we explore that what type of intervention can be informed by type reproduction numbers formulated in Section 3, and discuss how risk volatility affects type reproduction numbers and relevant critical control efforts.

Methods and Results

1.Model Population and Model Structure

In Chapter II, we build an episodic risk model with stages of HIV infection and dichotomous risk states. In this chapter, we focus on this model. However, we make several changes on the basis of model in Chapter II.

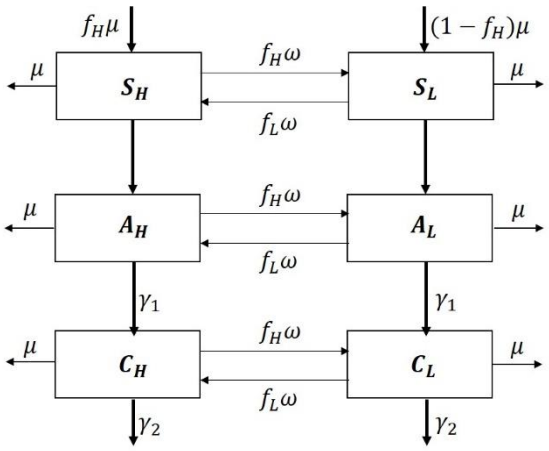

Figure III- 1 Conceptual model with contact structure and population composition for analysis in Chapter III.

Firstly, in Chapter II we used the risk transition rate, Φ_H and Φ_L, to quantify the rate of flows from high risk phase to low risk phase and vice versa. The problem with this is that two parameters are used to specify rate of changing risk levels. Although each can be expressed as a function of the other, using two parameters makes model analysis less tractable. Therefore, in this chapter we use a single risk change parameter instead of these two risk transition parameters to simplify derivation of model variables. Specifically, we adopt the way that Henry and Koopman (2015) modeled episodic risk.[82] They used a single parameter that determines the rate of fluctuation of individual risk between higher level and lower level. This parameter, as introduced by Henry and Koopman (2015), is the rate at which individuals re-select contact rate from the population distribution, namely risk re-selection rate (denoted ω). The probability distribution of population contact rates is determined by the fraction of average time people spend in each risk phase at disease free equilibrium (f_H for high risk phase and f_L for low risk phase).[82] Therefore, the rate of flow from the high-risk to the low-risk subpopulation is $f_L\omega$ and $f_H\omega$ as the rate of flow in the opposite direction. It is also important to note that re-selecting contact rate periodically is not a behavior pattern among MSM population. However, modeling risk volatility this way enables us to use a single parameter to control the rate of risk fluctuation in both directions. In following sections, we name the condition that an individual has high contact rate (or low contact rate) as that this individual is in high risk phase (or low risk phase).

In this chapter, we model assortative mixing by contact rate by allowing individuals to make certain fraction (m) of contact with people at the same risk level, for both people who are experiencing high risk or and people with low risk.[83] Same as described in Chapter II,

individuals proportionately mix with each other and make (1-m) fraction of their contacts at a general mixing site. Compartments' symbols, names and definitions are listed in Table III-1. Model parameters' symbols, default values and meanings are listed in Table III-2.

Table III- 1 Variable symbols, values, units and definitions for Episodic Risk model with Test and Treat in Chapter III

Compartment	Meaning
S_H	Susceptible subpopulation with high contact rate
S_L	Susceptible subpopulation with low contact rate
A_H	Subpopulation at acute stage of HIV infection with high contact rate
A_L	Subpopulation at acute stage of HIV infection with low contact rate
C_H	Subpopulation at chronic stage of HIV infection with high contact rate
C_L	Subpopulation at chronic stage of HIV infection with low contact rate

Table III- 2 Symbols, default values, and definition of model parameters

Parameter	Default values	Unit	Definition
μ	1/40	/year	Rate of removal from the sexually-active population unrelated to HIV due to death or other competing causes. Because we set the equilibrium population in the absence of disease to 1, this is also the (absolute) rate of entry of new individuals into the sexually active population
γ_1	4	/year	Rate of transitioning from acute to chronic infection
γ_2	1/10	/year	Rate of death from AIDS during chronic infection
β_1	variable	Per contact	Per-contact transmissibility during acute stage
β_2	variable	Per contact	Per-contact transmissibility during chronic stage
f_H	0.05	--	Fraction of population that is at high-risk phase at disease-free equilibrium
f_L	$1 - f_H$	--	Fraction of population that is at low-risk phase at disease-free equilibrium
χ_H	variable	/year	Average contact rate of high risk phase
r_{HL}	variable	--	Ratio of contact rate in high risk phase over contact rate in low risk phase
χ_L	$\frac{\chi_H}{r_{HL}}$	/year	Average contact rate of low risk phase
χ	$f_H\chi_H + f_L\chi_H$	/year	Expected population average contact rate at disease free equilibrium
m	variable	--	Fraction of contacts reserved for people experiencing the same level of risk
ω	variable	--	Rate of reselecting contact rate

The differential equations of this episodic risk model are,

$$\dot{S_H} = \mu f_H - \lambda_H S_H - f_H \omega S_H + f_L \omega S_L - \mu S_H$$

$$\dot{S_L} = \mu f_L - \lambda_L S_L - f_L \omega S_L + f_H \omega S_H - \mu S_L$$
$$\dot{A_H} = \lambda_H S_H - f_H \omega A_H + f_L \omega A_L - \gamma_1 A_H - \mu A_H$$
$$\dot{A_L} = \lambda_L S_L - f_L \omega A_L + f_H \omega A_H - \gamma_1 A_L - \mu A_L$$
$$\dot{C_H} = \gamma_1 A_H - f_H \omega C_H + f_L \omega C_L - \gamma_2 C_H - \mu C_H$$
$$\dot{C_L} = \gamma_1 A_L - f_L \omega C_L + f_H \omega C_H - \gamma_2 C_L - \mu C_L$$

Where λ_H and λ_L are force of infections that cause new infections among subpopulation in high risk phase and subpopulation in low risk phase, respectively. Specifically,

$$\lambda_H = \left(\frac{A_H\beta_1 + C_H\beta_2}{S_H + A_H + C_H}\right)\chi_H m + \frac{(A_H\beta_1 + C_H\beta_2)\chi_H + (A_L\beta_1 + C_L\beta_2)\chi_L}{(S_H + A_H + C_H)\chi_H + (S_L + A_L + C_L)\chi_L}\chi_H(1-m) \qquad (1.1)$$

$$\lambda_L = \left(\frac{A_L\beta_1 + C_L\beta_2}{S_L + A_L + C_L}\right)\chi_L m + \frac{(A_H\beta_1 + C_H\beta_2)\chi_H + (A_L\beta_1 + C_L\beta_2)\chi_L}{(S_H + A_H + C_H)\chi_H + (S_L + A_L + C_L)\chi_L}\chi_L(1-m) \qquad (1.2)$$

In equation (1.1), λ_H has two components. The first is $\left(\frac{A_H\beta_1+C_H\beta_2}{S_H+A_H+C_H}\right)\chi_H m$, which is the force of infection that causes new infections among people in high risk phase due to assortative mixing within population in high risk phase. The second is $\frac{(A_H\beta_1+C_H\beta_2)\chi_H+(A_L\beta_1+C_L\beta_2)\chi_L}{(S_H+A_H+C_H)\chi_H+(S_L+A_L+C_L)\chi_L}\chi_H(1-m)$, which is the force of infection that causes new infections among people in high risk phase due to proportionate mixing at the general mixing site. Similarly, λ_L also has two components: force of infection attributed to assortative mixing, $\left(\frac{A_L\beta_1+C_L\beta_2}{S_L+A_L+C_L}\right)\chi_L m$, and force of infection attributed to proportionate mixing at the general mixing site, $\frac{(A_H\beta_1+C_H\beta_2)\chi_H+(A_L\beta_1+C_L\beta_2)\chi_L}{(S_H+A_H+C_H)\chi_H+(S_L+A_L+C_L)\chi_L}\chi_L(1-m)$.

2. Type Reproduction Numbers as Function of Episodic Risk using Next Generation Matrix

As defined in Roberts and Heesterbeek (2003), type reproduction number for host of type i, T_i, calculates the number of new infections among type i host generated through all possible transmission pathways: direction transmission from type i case or transmission from type i case to other types of hosts which end up in a transmission back to type i subpopulation.[80] In Roberts and Heesterbeek (2003), type reproduction numbers are calculated based on next generation matrix (NGM), which we denote K in following sections. Construction of K requires two elements: matrix of new infections, F, and matrix of migration or other transitions, Σ.

Matrix F is the matrix of rates of new infections caused by transmission and it can be written as,

$$F = \begin{bmatrix} \chi_H\beta_1(g_H + mg_L) & \chi_L\beta_1(1-m)g_H & \chi_H\beta_2(g_H + mg_L) & \chi_L\beta_2(1-m)g_H \\ \chi_H\beta_1(1-m)g_L & \chi_L\beta_1(g_L + mg_H) & \chi_H\beta_2(1-m)g_L & \chi_L\beta_2(g_L + mg_H) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Each entry F_{ij} represents rate of transmission from state specified by jth column to state specified by ith row. States from first to forth rows (columns) of matrix F are, acutely infected with high contact rate (A_H), acutely infected with low contact rate (A_L), chronically infected with high contact rate (C_H) and chronically infected with low contact rate (C_L). Note that entries of F in last two rows are zero. This is because these two rows correspond to chronic infections and that new chronic infection cannot be caused through transmission.

In each entry of F, g_H and g_L denote the expected fraction of contacts made by susceptible individuals with high contact rate or by susceptible individuals with low contact rate at disease-free equilibrium, respectively. Specifically, g_H is a function of contact rate ratio r_{HL}, average time population spends in high contact rate phase at disease-free equilibrium, f_H, $g_H = \frac{r_{HL}f_H}{r_{HL}f_H+1-f_H}$. We can also get that $g_L = 1 - g_H = \frac{1-f_H}{r_{HL}f_H+1-f_H}$. We use entry F_{12} as an example to show how we get matrix F. Entry F_{12} is the rate of new infections among subpopulation with high contact rate (A_H) caused by an acutely infected individual with low contact rate (A_L). Given that an individual with low contact rate makes (1-m) fraction of his contacts with individuals experiencing high risk, we can get $F_{12}=\chi_L\beta_1(1\text{-}m)\, g_H$.

Matrix Σ is the matrix of rates migration among different groups,

$$\Sigma = \begin{bmatrix} -(\nu_1 + f_L\omega) & f_H\omega & 0 & 0 \\ f_L\omega & -(\nu_1 + f_H\omega) & 0 & 0 \\ \gamma_1 & 0 & -(\nu_2 + f_L\omega) & f_H\omega \\ 0 & \gamma_1 & f_L\omega & -(\upsilon_2 + f_H\omega) \end{bmatrix}$$

Where ν_1 denote the total rate of leaving acute HIV infection due to death, disease progression or other competing causes of leaving sexually active life, i.e. $\nu_1=\gamma_1+\mu$, and ν_2 denote the rate of leaving chronic HIV infection due to death or other competing causes of leaving sexually active life, i.e. $\nu_2=\gamma_2+\mu$. Also, note that in matrix Σ, there are two types of migrations: progression from acute infection to chronic infection (rate of which is γ_1), and risk transitions (rate of which is $f_L\omega$ from high risk to low risk and $f_H\omega$ from low risk to high risk). Formulations of variables used for matrix manipulation are listed in Table III-3.

Table III- 3 Symbols, function as model parameters and meanings of variables for calculation of next generation matrix

symbols	Function as model parameters	meaning
$\boldsymbol{\upsilon_1}$	$\boldsymbol{\gamma_1+\mu}$	Total rate of leaving acute HIV infection due to death, disease progression or other competing causes of leaving sexually active life
$\boldsymbol{\upsilon_2}$	$\boldsymbol{\gamma_2+\mu}$	rate of leaving chronic HIV infection due to death from AIDS or other competing causes of leaving sexually active life
$\boldsymbol{g_H}$	$\boldsymbol{\frac{r_{HL}f_H}{r_{HL}f_H+1-f_H}}$	expected fraction of contacts made by susceptible individuals with high contact rate at disease-free equilibrium
$\boldsymbol{g_L}$	$\boldsymbol{\frac{1-f_H}{r_{HL}f_H+1-f_H}}$	expected fraction of contacts made by susceptible individuals with low contact rate at disease-free equilibrium

The next generation matrix (NGM) with large domain, K_L, is calculated as $K_L=-F\Sigma^{-1}=F(-\Sigma^{-1})$. Where matrix $-\Sigma^{-1}$ is,

$$-\Sigma^{-1}=\begin{bmatrix} \frac{(\upsilon_1+f_H\omega)}{\upsilon_1(\upsilon_1+\omega)} & \frac{f_H\omega}{\upsilon_1(\upsilon_1+\omega)} & 0 & 0 \\ \frac{f_L\omega}{\upsilon_1(\upsilon_1+\omega)} & \frac{(\upsilon_1+f_L\omega)}{\upsilon_1(\upsilon_1+\omega)} & 0 & 0 \\ \frac{\gamma_1}{\upsilon_1\upsilon_2}\left(\frac{\upsilon_1\upsilon_2}{(\upsilon_1+\omega)(\upsilon_2+\omega)}+\frac{f_H\omega(\omega+\upsilon_1+\upsilon_2)}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right) & \frac{\gamma_1 f_H\omega}{\upsilon_1\upsilon_2(\upsilon_2+\omega)} & \frac{(\upsilon_2+f_H\omega)}{\upsilon_2(\upsilon_2+\omega)} & \frac{f_H\omega}{\upsilon_2(\upsilon_2+\omega)} \\ \frac{\gamma_1 f_L\omega(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_1+\omega)(\upsilon_2+\omega)} & \frac{\gamma_1(f_L\omega(\upsilon_1+\upsilon_2+\omega)+\upsilon_1\upsilon_2)}{\upsilon_1\upsilon_2(\upsilon_1+\omega)(\upsilon_2+\omega)} & \frac{f_L\omega}{\upsilon_2(\upsilon_2+\omega)} & \frac{(\upsilon_2+f_L\omega)}{\upsilon_2(\upsilon_2+\omega)} \end{bmatrix}$$

The states of rows and columns of matrix $-\Sigma^{-1}$ are arranged in the same way as those for matrix F and Σ. States from first to last row (column) of $-\Sigma^{-1}$ are acutely infected with high contact rate (A_H), acutely infected with low contact rate (A_L), chronically infected with high

contact rate (C_H) and chronically infected with low contact rate (C_L), respectively. Each entry of $-\boldsymbol{\Sigma}^{-1}$ is the average time that a case who entered state of corresponding column spends in state of corresponding row. For example, entry $(-\boldsymbol{\Sigma}^{-1})_{11}$ is $\frac{(\upsilon_1+f_H\omega)}{\upsilon_1(\upsilon_1+\omega)}$. It calculates the expected time that a case who was in high risk phase at infection (state A_H) spends in high risk phase during acute infection (state A_H). In addition, $(-\boldsymbol{\Sigma}^{-1})_{33}$ is $\frac{(\upsilon_2+f_H\omega)}{\upsilon_2(\upsilon_2+\omega)}$. It calculates the expected time that a case who progresses to chronic infection while in high risk phase (state C_H) spends in high risk phase during chronic infection (state C_H). Also note that the four entries in the upper right block of $-\boldsymbol{\Sigma}^{-1}$ are zeros. This is because that when a case has progressed to chronic infection, this case cannot go back to acute infection.

Given the matrix F and matrix $-\boldsymbol{\Sigma}^{-1}$, we can obtain the next generation matrix with large domain, K_L,

$$K_L = F(-\Sigma^{-1}) = \begin{bmatrix} \chi_H\beta_1(g_H+mg_L)\frac{(\upsilon_1+f_H\omega)}{\upsilon_1(\upsilon_1+\omega)} + \chi_L\beta_1(1-m)g_H\frac{f_L\omega}{\upsilon_1(\upsilon_1+\omega)} + \chi_H\beta_2(g_H + mg_L)\frac{\gamma_1}{\upsilon_1\upsilon_2}\left(\frac{\upsilon_1\upsilon_2}{(\upsilon_1+\omega)(\upsilon_2+\omega)} + \frac{f_H\omega(\omega+\upsilon_1+\upsilon_2)}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right) + \chi_L\beta_2(1-m)g_H\frac{\gamma_1 f_L\omega(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_1+\omega)(\upsilon_2+\omega)} & \chi_H\beta_1(g_H+mg_L)\frac{f_H\omega}{\upsilon_1(\upsilon_1+\omega)} + \frac{(\upsilon_1+f_L\omega)}{\upsilon_1(\upsilon_1+\omega)}\chi_L\beta_1(1-m)g_H + \frac{\gamma_1 f_H\omega}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\chi_H\beta_2(g_H+mg_L) + \frac{\gamma_1(f_L\omega(\upsilon_1+\upsilon_2+\omega)+\upsilon_1\upsilon_2)}{\upsilon_1\upsilon_2(\upsilon_1+\omega)(\upsilon_2+\omega)}\chi_L\beta_2(1-m)g_H & \chi_H\beta_2(g_H+mg_L)\frac{(\upsilon_2+f_H\omega)}{\upsilon_2(\upsilon_2+\omega)} + \frac{f_L\omega}{\upsilon_2(\upsilon_2+\omega)}\chi_L\beta_2(1-m)g_H & \chi_H\beta_2(g_H+mg_L)\frac{f_H\omega}{\upsilon_2(\upsilon_2+\omega)} + \frac{(\upsilon_2+f_L\omega)}{\upsilon_2(\upsilon_2+\omega)}\chi_L\beta_2(1-m)g_H \\ \chi_H\beta_1(1-m)g_L\frac{(\upsilon_1+f_H\omega)}{\upsilon_1(\upsilon_1+\omega)} + \chi_L\beta_1(g_L+mg_H)\frac{f_L\omega}{\upsilon_1(\upsilon_1+\omega)} + \chi_H\beta_2(1-m)g_L\frac{\gamma_1}{\upsilon_1\upsilon_2}\left(\frac{\upsilon_1\upsilon_2}{(\upsilon_1+\omega)(\upsilon_2+\omega)} + \frac{f_H\omega(\omega+\upsilon_1+\upsilon_2)}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right) + \chi_L\beta_2(g_L+mg_H)\frac{\gamma_1 f_L\omega(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_1+\omega)(\upsilon_2+\omega)} & \chi_H\beta_1(1-m)g_L\frac{f_H\omega}{\upsilon_1(\upsilon_1+\omega)} + \chi_L\beta_1(g_L+mg_H)\frac{(\upsilon_1+f_L\omega)}{\upsilon_1(\upsilon_1+\omega)} + \chi_H\beta_2(1-m)g_L\frac{\gamma_1 f_H\omega}{\upsilon_1\upsilon_2(\upsilon_2+\omega)} + \chi_L\beta_2(g_L+mg_H)\frac{\gamma_1(f_L\omega(\upsilon_1+\upsilon_2+\omega)+\upsilon_1\upsilon_2)}{\upsilon_1\upsilon_2(\upsilon_1+\omega)(\upsilon_2+\omega)} & \chi_H\beta_2(1-m)g_L\frac{(\upsilon_2+f_H\omega)}{\upsilon_2(\upsilon_2+\omega)} + \chi_L\beta_2(g_L+mg_H)\frac{f_L\omega}{\upsilon_2(\upsilon_2+\omega)} & \chi_H\beta_2(1-m)g_L\frac{f_H\omega}{\upsilon_2(\upsilon_2+\omega)} + \chi_L\beta_2(g_L+mg_H)\frac{(\upsilon_2+f_L\omega)}{\upsilon_2(\upsilon_2+\omega)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The states of rows and columns of K_L are again arranged in the same way as those of matrix F and $\boldsymbol{\Sigma}$. Entry $(K_L)_{ij}$ is the expected number of new infections at state i caused by transmission from a case since this case enters state j till end of infection time. Entries in two bottom rows of K_L are zeros because new chronic infections cannot be caused through infection. K_L has the same dimensionality as that of matrix F and $\boldsymbol{\Sigma}$. Its entries include the expected number of secondary cases generated since a case just begins infection time, i.e. at state-at-infection, till end of

infection time (entries $(K_L)_{11}$, $(K_L)_{12}$, $(K_L)_{21}$ and $(K_L)_{22}$). Its entries also include the expected number of secondary cases a case generates during later stage of infection (entries $(K_L)_{13}$, $(K_L)_{14}$, $(K_L)_{23}$ and $(K_L)_{24}$), which cannot be entered through infection, i.e. non-state-at-infection. However, the next generation matrix used for reproduction numbers (basic reproduction number and type reproduction number) is a square matrix that only involves the state that cases can enter by infection, i.e. state-at-infection. This is because reproduction numbers are only concerned with generation of new infections through transmission. Therefore, next generation matrix (NGM) K for reproduction numbers is part of K_L where rows and columns are state-at-infection (entries $(K_L)_{11}$, $(K_L)_{12}$, $(K_L)_{21}$ and $(K_L)_{22}$). As shown by Diekmann et al (2010), K_L can be transformed into next generation matrix K through following matrix manipulation,[84]

$$K = E' K_L E$$

Where E is an auxiliary matrix consisting of unit column vectors e_i, for all i where ith row of matrix F (or K_L) does not have all zero entries,

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The purpose of this matrix manipulation is to single out all rows and columns that correspond to the states-at-infection (acutely infected in high risk phase and acutely infected in low risk phase). By solving $K = E' K_L E$, we obtain,

$$K = \begin{bmatrix} \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(g_H + mg_L) & \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_L(1-m)g_H \\ + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_H\chi & + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_H\chi \\ \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(1-m)g_L & \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_L(g_L + mg_H) \\ + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_L\chi & + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_L\chi \end{bmatrix}$$

After the matrix manipulation, the dimensionality of NGM K is reduced to 2 by 2, where only rows and columns that correspond to states at infection are reserved. This is the major

difference between NGM K and NGM with large domain, K_L. Specifically, entry K_{ij} is the expected number of new infections with risk level i caused by a case who was at risk level j at time of infection. For example, if a case had high contact rate at time of infection, then transmissions from this case to susceptibles with high contact rates contribute to entry K_{11} even after this case re-selects a low contact rate later during infection time. According to Roberts and Heesterbeek (2003), type reproduction number of subpopulation at ith risk level, T_i, can be calculated as,[80]

$$T_i = e'K(I - (I - P)K)^{-1}e \qquad (2.1)$$

Where e denotes the unit vector for which $e_i = 1$ and $e_j = 0$ for $j \neq i$. Matrix I is identity matrix and matrix P is projection matrix where $P_{ii} = 1$ and $P_{kj} = 0$ when $k \neq i$ or $j \neq i$. As defined by Roberts and Heesterbeek (2003), the "type" in the definition of type reproduction number, T, is "type" at infection, or state-at-infection.[80] This is also reasonable since T is calculated on the basis of NGM K, which describes the generation of cases specified by cases' state-at-infection.

The rationale of calculation of T_i as shown in equation (2.1) has been introduced in Roberts and Heesterbeek (2003).[80] Briefly speaking, vector Ke calculates the expected number of new infections of type i as result of direct transmission from a case of type i. Vector $(I - P)K$ calculates the extent to which other types of hosts contribute to generation of new infection of type i during the process that infection is passed to other hosts and back to type i subpopulation. Vector $(I - (I - P)K)^{-1}$ takes into account all the possibilities: infection is not passed to other types of hosts, infection is passed to other types of hosts and back to type i subpopulation, and infection is passed to other types of hosts and passed among them for one to endless rounds before ending up in transmission back to type i subpopulations. Using equation (2.1), we get the type reproduction numbers for subpopulation defined by the state-at-infection,

$$T_{H_{Inf}} = K_{11} + \frac{K_{21}K_{12}}{1-K_{22}} \qquad (2.2)$$

$$T_{L_{Inf}} = K_{22} + \frac{K_{12}K_{21}}{1 - K_{11}} \qquad (2.3)$$

Subscripts 'H_{Inf}' and 'L_{Inf}' indicate that it is the type reproduction number for cases who are in higher risk phase at time of infection or lower risk phase at time of infection, respectively.

It is important to note that type reproduction number for cases of a particular type is only epidemiologically meaningful when the other types of groups alone cannot sustain ongoing transmission. That says, $T_{H_{Inf}}$ is only epidemiologically meaningful when low-risk-at-infection subpopulation (susceptible population who currently have lower contact rate and infected people who had lower contact rate at infection) alone cannot sustain ongoing transmission. The threshold value for the ability of low-risk-at-infection subpopulation to sustain ongoing transmission alone is the largest dominant eigenvalue of NGM K (R_0) where all entries of K that involve high-risk-at-infection subpopulation are zero (K_{11}, K_{12}, K_{21}). A little algebra shows that this threshold is K_{22}. That says, $T_{H_{Inf}}$ is epidemiologically meaningful only when $K_{22}<1$. Similarly, the threshold value that quantifies the ability of high-risk-at-infection subpopulation to sustain ongoing transmission alone is the largest dominant eigenvalue of K where entries that involve low-risk-at-infection states (K_{12}, K_{21}, K_{22}) are zero, which is K_{11}. That says, $T_{L_{Inf}}$ is epidemiologically meaningful only when $K_{11}<1$.

3. Derive Type Reproduction Numbers as Function of Episodic Risk Based on Their Epidemiological Meanings

As shown in last section, each entry of NGM K is a complicated function of the risk re-selection rate, ω. Since each type reproduction number is a function of entries of NGM K, they are also a complicated function of ω. This prevents us from better interpreting the effect of episodic risk on type reproduction numbers. Therefore, in this section we derive type reproduction numbers without matrix manipulation, but only based on epidemiological meaning of type reproduction numbers given by Roberts and Heesterbeek (2003).[80] In order to keep our argument clear, we divide this section into four subsections, so that each subsection describes a key step towards successful derivation of type reproduction numbers.

The complexity of episodic risk model arises from the fact that a model case can transmit HIV infection in both risk phases. This makes it hard to understand how many new infections at a specific state-at-infection a case with the same state-at-infection would cause if this case

fluctuates between two risk phases during the entire infection time. The measure that would help to clarify this is the average time that a case spends in a risk phase at a stage of infection. As we fix the length of each stage of infection, this measure is determined by the probability that a case acts in a risk phase at a stage of infection. We name such probability as state probability. In the first subsection (3.1.), we derive the probability a model case is in a risk phase during a stage of infection. In the second subsection (3.2.), we further discuss the properties of state probabilities derived in subsection (3.1.), and rewrite them in a way that has clearer epidemiological meaning and also more easily adapted to arbitrary number of stages of infection. The third subsection (3.3.) describes the expected number of new infections a model case cause during entire infection time based on the case's states probability as given in subsection 3.1 and subsection 3.2. The forth subsection (3.4.) describes how elements depicted in first three subsections contribute to construction of type reproduction numbers, and relationship of formulation of type reproduction numbers with the next generation approach in Section 2.

3.1. **Probability that a Model Case is in a Risk Phase during Infection Time Given Episodic Risk.** Our model population has dichotomous risk states (contact rates) and two stages of infection (acute HIV infection and chronic HIV infection). Therefore, there are four possible states at which a case can transmit HIV infection: acutely infected in high risk phase (A_H), acutely infected in low risk phase (A_L), chronically infected in high risk phase (C_H), and chronically infected in low risk phase (C_L). Given episodic risk, a model case can be in any of these four states, no matter which state this case was in at HIV acquisition. Since each state has a unique transmission potential, one needs to know the probability that a case is in each state to calculate total expected new infections a model case would cause.

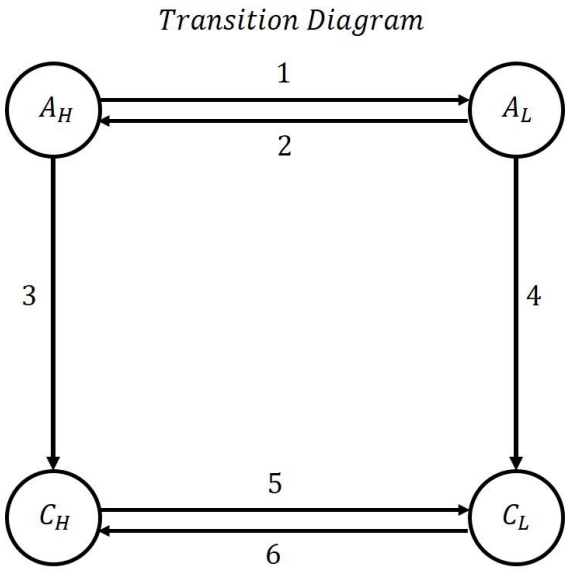

Figure III- 2 Diagram to show all routes through which a model case can transit among states: acutely infected in high risk phase (A_H), acutely infected in low risk phase (A_L), chronically infected in high risk phase (C_H) and chronically infected in high risk phase (C_L).

As shown in Fig.III-2, there are six paths that a case can transit among these states. Among them, four paths are due to risk transitions: from high risk phase to low risk phase while acutely infected (path '1' from A_H to A_L) or while chronically infected (path '5' from C_H to C_L), and risk transition in the opposite direction while acutely infected (path '2') or chronically infected (path '6'). The remaining two paths are due to disease progression: progression from acute infection to chronic infection while in high risk phase, i.e. A_H to C_H (path '3'), and progression from acute infection to chronic infection while in low risk phase, i.e. A_L to C_L (path '4').

By specifying the rate of transitioning through each path, we can get the probability that a case changes state through each path. In earlier sections, we denote the rate that an acutely infected case leaves acute infection as ν_1, then $\nu_1 = \gamma_1 + \mu$, where γ_1 is the rate of progressing from acute infection to chronic infection and μ is the rate of leaving sexually active life due to other competing causes. Similarly, the rate that a chronically infected case leaves chronic infection ν_2 is, $\nu_2 = \gamma_2 + \mu$, where γ_2 is the rate of leaving chronic infection due to death from AIDS. As described earlier, the rate of transition from high risk phase to low risk phase is $f_L\omega$, and risk transition in opposite direction is $f_H\omega$, where ω is the rate at which a case re-selects contact rate from the population distribution at disease free equilibrium (probability f_H to select high contact rate and probability f_L =1- f_H to selection low risk phase). After specifying these

rates, we can calculate the probability that a case transit from one state to another state through one of the six paths as shown in Fig.III-2.

Let $T_{p_i q_j}$ be the probability that a model case transits from risk level p at i stage of infection to risk level q at j stage of infection, then the probability that a case transits from high risk phase to low risk phase during acute infection (path '1' in Fig.III-2), $T_{H_1 L_1}$, is

$$T_{H_1 L_1} = \frac{f_L \omega}{\nu_1 + f_L \omega} \qquad (3.1.1)$$

Similarly, we can also get the probability that a case transits from low risk phase to high risk phase during acute infection (path '2' in Fig.III-2),

$$T_{L_1 H_1} = \frac{f_H \omega}{\nu_1 + f_H \omega} \qquad (3.1.2)$$

In addition, the probability that a case transits from high to low risk phase during chronic infection (path '5' in Fig.III-2) is,

$$T_{H_2 L_2} = \frac{f_L \omega}{\nu_2 + f_L \omega} \qquad (3.1.3)$$

While the probability that a case transits from low to high risk phase during chronic infection (path '6' in Fig.III-2) is,

$$T_{L_2 H_2} = \frac{f_H \omega}{\nu_2 + f_H \omega} \qquad (3.1.4)$$

In addition, the probability that a case progresses from acute infection to chronic infection while in high risk phase (path '3' in Fig.III-2), $T_{H_1 H_2}$, is the probability that this case progresses to chronic infection (at rate γ_1) instead of transitioning to low risk phase (at rate $f_L \omega$) or leaving sexually active life (at rate μ). Therefore, the formulation of $T_{H_1 H_2}$ is,

$$T_{H_1 H_2} = \frac{\gamma_1}{\gamma_1 + \mu + f_L \omega} = \frac{\gamma_1}{\upsilon_1 + f_L \omega} \qquad (3.1.5)$$

Similarly, the probability that a case progresses from acute infection to chronic infection while in low risk phase (path '4' in Fig.III-2), $T_{L_1 L_2}$, is

$$T_{L_1 L_2} = \frac{\gamma_1}{\upsilon_1 + f_H \omega} \qquad (3.1.6)$$

The probability that a case is in each state (A_H, A_L, C_H and C_L) during the entire infection time can be calculated based on these transition probabilities. Firstly**, let $P_{p_i q_j}$ be the**

probability that a case who enters ith stage of infection while in risk phase p is in risk phase q during jth stage of infection. Suppose a model case was in high risk phase at HIV acquisition. This case first enters compartment A_H through HIV acquisition. The probability that this case is in high risk phase during acute infection A_H, $P_{H_1H_1}$, is the combination of probabilities of following situations: first, this case stays at state A_H without any risk transition; second, this case transits to low risk phase during acute infection and transits back to high risk phase later during acute infection. Calculation of probability of first situation is straightforward, which is 1 minus the probability of risk transition to low risk phase, $(1 - T_{H_1L_1})$. Calculation of probability in the second situation is more complicated. This is because that this case can end up in high risk phase after multiple rounds of back and forth risk transitions. As described earlier, we suppose that this case enters state A_H at HIV acquisition. Therefore, in order to transition back to high risk phase during acute infection, this case can either transit to low risk phase and back to high risk phase, probability of which is $T_{H_1L_1}T_{L_1H_1}$, or transition back to high risk phase after two rounds of risk transitions, probability of which is $(T_{H_1L_1}T_{L_1H_1})^2$. Theoretically, this case can go through risk transitions for endless rounds before ending up in high risk phase during acute infection. Therefore, the probability that case who enters A_H state at HIV acquisition and is in A_H state after at least one round of risk transition is $(1 - T_{H_1L_1})\{T_{H_1L_1}T_{L_1H_1} + (T_{H_1L_1}T_{L_1H_1})^2 + \cdots + (T_{H_1L_1}T_{L_1H_1})^n + \cdots\}$. If we combine this probability with the probability that this case has not had any risk transition $(1 - T_{H_1L_1})$, we get,

$$P_{H_1H_1} = (1 - T_{H_1L_1})\left\{1 + T_{H_1L_1}T_{L_1H_1} + (T_{H_1L_1}T_{L_1H_1})^2 + \cdots + (T_{H_1L_1}T_{L_1H_1})^n + \cdots\right\} = (1 - T_{H_1L_1})\sum_{i=0}^{\infty}(T_{H_1L_1}T_{L_1H_1})^i$$

.

Given that $T_{H_1L_1} < 1$ and $T_{L_1H_1} < 1$, $T_{H_1L_1}T_{L_1H_1} < 1$. $P_{H_1H_1}$ can be further written as,

$$P_{H_1H_1} = (1 - T_{H_1L_1})\sum_{i=0}^{\infty}(T_{H_1L_1}T_{L_1H_1})^i = \frac{(1 - T_{H_1L_1})}{1 - T_{H_1L_1}T_{L_1H_1}} \quad (3.1.7)$$

Based on equation (3.1.1) and (3.1.2), we can substitute $T_{H_1L_1}$ with $\frac{f_L\omega}{\nu_1 + f_L\omega}$ and $T_{L_1H_1}$ with $\frac{f_H\omega}{\nu_1 + f_H\omega}$, and obtain,

$$P_{H_1H_1} = \frac{\left(1 - \frac{f_L\omega}{\nu_1 + f_L\omega}\right)}{1 - \left(\frac{f_L\omega}{\nu_1 + f_L\omega}\right)\left(\frac{f_H\omega}{\nu_1 + f_H\omega}\right)} = \frac{\nu_1 + f_H\omega}{\nu_1 + \omega} \qquad (3.1.8)$$

Although expression of $P_{H_1H_1}$as shown in (3.1.8) is derived using the transition probabilities, one may also interpret it as follows. The denominator $(\nu_1 + \omega)$ is the total rate that a case who was in high risk phase at HIV acquisition takes an action during acute infection: either leaving acute infection at rate ν_1 or re-selecting contact rate at rate ω. Numerator $\nu_1 + f_H\omega$ is the total rate of events that result in being in high risk phase during acute infection: not re-selecting contact rate at all during acute infection (at rate ν_1), or re-selecting high contact rate again (re-selecting contact rate at rate ω and selecting high contact rate with probability f_H). Their ratio gives the probability $P_{H_1H_1}$.

The probability that a case who was in high risk phase at HIV acquisition is in low risk phase during acute infection, $P_{H_1L_1}$, can also be calculated with same reasoning, except that this case needs to first transit to low risk phase, probability of which is $T_{H_1L_1}$. After transitioning to low risk phase, this case may just stay in low risk phase probability of which is $\left(1 - T_{L_1H_1}\right)$, or transit back to low risk phase after one to endless rounds of back and forth risk transitions, probability of which is $\left(1 - T_{L_1H_1}\right)\sum_{i=1}^{\infty}\left(T_{H_1L_1}T_{L_1H_1}\right)^i$. By summing up all these possibilities, we obtain,

$$P_{H_1L_1} = T_{H_1L_1}\left\{\left(1 - T_{L_1H_1}\right) + \left(1 - T_{L_1H_1}\right)\sum_{i=1}^{\infty}\left(T_{H_1L_1}T_{L_1H_1}\right)^i\right\} = T_{H_1L_1}\left(1 - T_{L_1H_1}\right)\sum_{i=0}^{\infty}\left(T_{H_1L_1}T_{L_1H_1}\right)^i = \frac{T_{H_1L_1}\left(1 - T_{L_1H_1}\right)}{1 - T_{H_1L_1}T_{L_1H_1}} \qquad (3.1.9)$$

Based on equation (3.1.1) and (3.1.2), we can substitute $T_{H_1L_1}$ with $\frac{f_L\omega}{\nu_1 + f_L\omega}$ and $T_{L_1H_1}$ with $\frac{f_H\omega}{\nu_1 + f_H\omega}$, and obtain,

$$P_{H_1L_1} = \frac{f_L\omega}{\nu_1 + \omega} \qquad (3.1.10)$$

Based on equation (3.1.8) and (3.1.10) we can obtain,

$$P_{H_1H_1} + P_{H_1L_1} = 1 \quad (3.1.11)$$

This relationship is reasonable since a case can only either be in high risk phase or low risk phase during acute infection.

With same reasoning, we also obtain the probability that a model case who entered state A_L at HIV acquisition is in high risk phase during acute infection, $P_{L_1H_1}$,

$$P_{L_1H_1} = \frac{T_{L_1H_1}(1 - T_{H_1L_1})}{1 - T_{H_1L_1}T_{L_1H_1}} = \frac{f_H\omega}{\nu_1 + \omega} \quad (3.1.12)$$

and the probability that a model case who entered state A_L at HIV acquisition is in low risk phase during acute infection, $P_{L_1L_1}$,

$$P_{L_1L_1} = \frac{(1 - T_{L_1H_1})}{1 - T_{H_1L_1}T_{L_1H_1}} = \frac{\nu_1 + f_L\omega}{\nu_1 + \omega} \quad (3.1.13)$$

Based on formulation of $P_{L_1H_1}$ and $P_{L_1L_1}$, we also obtain,

$$P_{L_1H_1} + P_{L_1L_1} = 1 \quad (3.1.14)$$

Situations become more complicated when we calculate the probability that a model case is in a risk phase during chronic HIV infection. This is because as a case progresses through more stages of infection, the number of possible states a case can be in before ending up in the state of interest also increases. For example, a model case who was in high risk phase at HIV acquisition can be in high risk phase during chronic infection as result of progressing to chronic infection while in high risk phase (A_H->C_H), or as result of transitioning to low risk phase, progressing to chronic infection and transitioning back to high risk phase (A_H-> A_L ->C_L-> C_H). Therefore, to calculate the probability that a model case who was in high risk phase at HIV acquisition is in high risk phase during chronic infection, $P_{H_1H_2}$, we need to examine every possible pathway that leads a case from state A_H to state C_H. The major difference between the two paths in above example is the risk phase in which a case progresses to chronic infection. To take this step into account during probability calculation, we define a **probability $P_{p_1(k_2)q_2}$, which is the probability that a case who was in risk phase p at HIV acquisition, progresses**

to chronic infection while in risk phase k and is in risk phase q during chronic infection. Brackets are added to the middle subscript to indicate that this is an intermediate state.

We start from $P_{H_1(H_2)H_2}$, which is the probability that a case who was in high risk phase at HIV acquisition, progresses to chronic infection while in high risk phase and is in high risk phase during chronic infection. This probability has two components. First is the probability that this case progresses to chronic infection while in high risk phase. We denote such probability as $P_{H_1(H_2)}$, where subscript in brackets again represents the intermediate state this case goes through. Calculation of $P_{H_1(H_2)}$ is similar as calculation of $P_{H_1H_1}$, except that this time the case progresses to chronic infection instead of staying in acute infection. $P_{H_1(H_2)}$ is the product of probability that an acutely infected case in high risk phase progress to chronic infection, $T_{H_1H_2}$, and probability that this case is in high risk phase during acute infection after zero to endless rounds of risk transitions, $\sum_{i=0}^{\infty}\left(T_{H_1L_1}T_{L_1H_1}\right)^i$,

$$P_{H_1(H_2)} = T_{H_1H_2}\sum_{i=0}^{\infty}\left(T_{H_1L_1}T_{L_1H_1}\right)^i = \frac{T_{H_1H_2}}{1 - T_{H_1L_1}T_{L_1H_1}} \qquad (3.1.15)$$

After a little algebra, one can also find out the relationship between $P_{H_1(H_2)}$ and $P_{H_1H_1}$,

$$P_{H_1(H_2)} = P_{H_1H_1}T_{12} \qquad (3.1.16)$$

Where

$$T_{12} = \frac{\gamma_1}{\nu_1} \qquad (3.1.17)$$

which is probability that a case progresses from acute infection to chronic infection regardless of risk phase this case was in during acute infection. This is because $P_{H_1(H_2)}$ can also be interpreted as the probability that a case progresses to chronic infection,T_{12}, given that this case is in high risk phase during acute infection, probability of which is $P_{H_1H_1}$.

Since we have found the first component of $P_{H_1(H_2)H_2}$, $P_{H_1(H_2)}$, we move forward to the second component of $P_{H_1(H_2)H_2}$. It is the probability that given a case progressed to chronic infection while in high risk phase, this case is in high risk phase during chronic infection, $P_{H_2H_2}$.

The rationale of calculating $P_{H_2H_2}$ is the same as that of $P_{H_1H_1}$. The only difference is that $P_{H_2H_2}$ is for chronic infection, so risk transition probabilities are those that occur during chronic infection, $T_{H_2L_2}$ and $T_{L_2H_2}$. Using the reasoning of calculating $P_{H_1H_1}$(equation (3.1.7)), we obtain,

$$P_{H_2H_2} = \left(1 - T_{H_2L_2}\right)\left\{1 + T_{H_2L_2}T_{L_2H_2} + \left(T_{H_2L_2}T_{L_2H_2}\right)^2 + \cdots + \left(T_{H_2L_2}T_{L_2H_2}\right)^n + \cdots\right\}$$
$$= \left(1 - T_{H_2L_2}\right)\sum_{i=0}^{\infty}\left(T_{H_2L_2}T_{L_2H_2}\right)^i \quad (3.1.18)$$

Based on equation (3.1.3) and (3.1.4), we substitute $T_{H_2L_2}$ with $\frac{f_L\omega}{\nu_2+f_L\omega}$ and $T_{L_2H_2}$ with $\frac{f_H\omega}{\nu_2+f_H\omega}$, and obtain,

$$P_{H_2H_2} = \frac{\nu_2 + f_H\omega}{\nu_2 + \omega} \qquad (3.1.19)$$

As indicated by earlier argument, $P_{H_1(H_2)H_2}$ is the combination of $P_{H_1(H_2)}$ and $P_{H_2H_2}$, so we obtain,

$$P_{H_1(H_2)H_2} = P_{H_1(H_2)}P_{H_2H_2}$$

According to equation (3.1.16) one can substitute $P_{H_1(H_2)}$ with $P_{H_1H_1}T_{12}$, so,

$$P_{H_1(H_2)H_2} = P_{H_1H_1}T_{12}P_{H_2H_2} \qquad (3.1.20)$$

Another pathway that leads a case from state A_H (at HIV acquisition) to state C_H is that this case progresses to chronic infection while in low risk phase and transition to high risk phase during chronic infection. We denote the probability for this pathway to occur as $P_{H_1(L_2)H_2}$. Similar as $P_{H_1(H_2)H_2}$, $P_{H_1(L_2)H_2}$ contains two components. The first is the probability that this case progresses to chronic infection while in low risk phase, $P_{H_1(L_2)}$. Calculation of $P_{H_1(L_2)}$ follows the same reasoning of $P_{H_1(H_2)}$, which gives rise to,

$$P_{H_1(L_2)} = T_{H_1L_1}T_{L_1L_2}\sum_{i=0}^{\infty}\left(T_{H_1L_1}T_{L_1H_1}\right)^i = \frac{T_{H_1L_1}T_{L_1L_2}}{1 - T_{H_1L_1}T_{L_1H_1}} \qquad (3.1.21)$$

In equation (3.1.21), risk transition probability $T_{H_1L_1}$ is included because this case was in high risk phase at HIV acquisition and can only be in low risk phase through risk transition. Same as the relationship $P_{H_1(H_2)} = P_{H_1H_1}T_{12}$ (equation (3.1.16)), $P_{H_1(L_2)}$ can also be expressed as,

$$P_{H_1(L_2)} = P_{H_1L_1}T_{12} \qquad (3.1.22)$$

The second component of $P_{H_1(L_2)H_2}$ is the probability that a case is in high risk phase during chronic infection given this case entered chronic infection while in low risk phase $P_{L_2H_2}$. $P_{L_2H_2}$ is the product of the probability that this case transits to high risk phase during chronic infection and does not transit back, $T_{L_2H_2}(1 - T_{H_2L_2})$, and probability that this case is in high risk phase after zero to endless rounds of risk transitions, $\sum_{i=0}^{\infty}\left(T_{H_2L_2}T_{L_2H_2}\right)^i$,

$$P_{L_2H_2} = T_{L_2H_2}\sum_{i=0}^{\infty}\left(T_{H_2L_2}T_{L_2H_2}\right)^i = \frac{T_{L_2H_2}(1 - T_{H_2L_2})}{1 - T_{H_2L_2}T_{L_2H_2}} \qquad (3.1.23)$$

Based on equation (3.1.3) and (3.1.4), we substitute $T_{H_2L_2}$ with $\frac{f_L\omega}{\nu_2+f_L\omega}$ and $T_{L_2H_2}$ with $\frac{f_H\omega}{\nu_2+f_H\omega}$, and obtain,

$$P_{L_2H_2} = \frac{f_H\omega}{\nu_2 + \omega} \qquad (3.1.24)$$

As indicated by earlier argument, probability $P_{H_1(L_2)H_2}$ is the production of $P_{H_1(L_2)}$ and $P_{L_2H_2}$,

$$P_{H_1(L_2)H_2} = P_{H_1(L_2)}P_{L_2H_2}$$

Based on equation (3.1.22), we substitute $P_{H_1(L_2)}$ with $P_{H_1L_1}T_{12}$, which gives rise to,

$$P_{H_1(L_2)H_2} = P_{H_1L_1}T_{12}P_{L_2H_2} \qquad (3.1.25)$$

As discussed earlier, the probability that a case was in high risk phase at HIV acquisition is in high risk phase during chronic infection, $P_{H_1H_2} = P_{H_1(H_2)H_2} + P_{H_1(L_2)H_2}$. If we incorporate equations (3.1.20) and (3.1.25), we obtain,

$$P_{H_1H_2} = P_{H_1(H_2)H_2} + P_{H_1(L_2)H_2} = P_{H_1H_1}T_{12}P_{H_2H_2} + P_{H_1L_1}T_{12}P_{L_2H_2}$$

After a little algebra, we obtain,

$$P_{H_1H_2} = T_{12}\left(\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)} + \frac{f_H\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}\right) \qquad (3.1.26)$$

Similar as reasoning presented above, the probability that a case who was in high risk phase at HIV acquisition is in low risk phase during chronic infection, $P_{H_1L_2}$, can also be divided into two components based on the risk phase in which this case progresses to chronic infection. One is this probability given that the case progresses to chronic infection while in high risk phase, $P_{H_1(H_2)L_2}$. Another is this probability given that this case progresses to chronic infection while in low risk phase, $P_{H_1(L_2)L_2}$. $P_{H_1(H_2)L_2}$ and $P_{H_1(L_2)L_2}$ are calculated with the same reasoning as $P_{H_1(H_2)H_2}$ and $P_{H_1(L_2)H_2}$, with which we obtain,

$$P_{H_1(H_2)L_2} = P_{H_1H_1}T_{12}P_{H_2L_2} \qquad (3.1.27)$$

$$P_{H_1(L_2)L_2} = P_{H_1L_1}T_{12}P_{L_2L_2} \qquad (3.1.28)$$

where $P_{H_2L_2}$ is the probability that a case who enters chronic infection while in high risk phase is in low risk phase during chronic infection, which can be calculated as,

$$P_{H_2L_2} = T_{H_2L_2}(1-T_{L_2H_2})\sum_{i=0}^{\infty}\left(T_{H_2L_2}T_{L_2H_2}\right)^i = \frac{T_{H_2L_2}(1-T_{L_2H_2})}{1-T_{H_2L_2}T_{L_2H_2}}$$

and $P_{L_2L_2}$ is the probability that a case who enters chronic infection while in low risk phase is in low risk phase during chronic infection, which can be calculated as,

$$P_{L_2L_2} = \left(1-T_{L_2H_2}\right)\sum_{i=0}^{\infty}\left(T_{H_2L_2}T_{L_2H_2}\right)^i = \frac{\left(1-T_{L_2H_2}\right)}{1-T_{H_2L_2}T_{L_2H_2}}$$

Based on equation (3.1.3) and (3.1.4), we substitute $T_{H_2L_2}$ with $\frac{f_L\omega}{\nu_2+f_L\omega}$ and $T_{L_2H_2}$ with $\frac{f_H\omega}{\nu_2+f_H\omega}$, and obtain,

$$P_{H_2L_2} = \frac{f_L\omega}{\nu_2+\omega} \qquad (3.1.29)$$

$$P_{L_2L_2} = \frac{\nu_2+f_L\omega}{\nu_2+\omega} \qquad (3.1.30)$$

Given that $P_{H_2H_2} = \frac{\nu_2 + f_H\omega}{\nu_2 + \omega}$ (equation (3.1.19)) and $P_{H_2L_2} = \frac{f_L\omega}{\nu_2 + \omega}$ (3.1.29) we can get that,

$$P_{H_2H_2} + P_{H_2L_2} = 1 \quad (3.1.31)$$

This is reasonable because given that a case enters chronic infection while in high risk phase, this case can only be either in high risk phase or low risk phase during chronic infection.

Similarly, we also obtain,

$$P_{L_2H_2} + P_{L_2L_2} = 1 \quad (3.1.32)$$

Based on equation (3.1.27) and (3.1.28), we obtain the probability that a case who was in high risk phase at HIV acquisition is in low risk phase during chronic infection, $P_{H_1L_2}$,

$$P_{H_1L_2} = P_{H_1(H_2)L_2} + P_{H_1(L_2)L_2} = P_{H_1H_1}T_{12}P_{H_2L_2} + P_{H_1L_1}T_{12}P_{L_2L_2} \quad (3.1.33)$$

Substituting $P_{H_1H_1}$, $P_{H_2L_2}$, $P_{H_1L_1}$ and $P_{L_2L_2}$ with their formulations given by equations (3.1.8), (3.1.29), (3.1.10) and (3.1.30), respectively, we obtain,

$$P_{H_1L_2} = T_{12}\frac{f_L\omega(\nu_1 + \nu_2 + \omega)}{(\nu_1 + \omega)(\nu_2 + \omega)} \quad (3.1.34)$$

Based on formulation of $P_{H_1H_2}$ and $P_{H_1L_2}$ given by equation (3.1.26)and (3.1.34), we can get their sum as,

$$\begin{aligned} P_{H_1H_2} + P_{H_1L_2} &= T_{12}\left(\frac{\nu_1\nu_2}{(\nu_1 + \omega)(\nu_2 + \omega)} + \frac{f_H\omega(\nu_1 + \nu_2 + \omega)}{(\nu_1 + \omega)(\nu_2 + \omega)}\right) + T_{12}\frac{f_L\omega(\nu_1 + \nu_2 + \omega)}{(\nu_1 + \omega)(\nu_2 + \omega)} \\ &= T_{12}\left(\frac{\nu_1\nu_2}{(\nu_1 + \omega)(\nu_2 + \omega)} + \frac{\omega(\nu_1 + \nu_2 + \omega)}{(\nu_1 + \omega)(\nu_2 + \omega)}\right) = T_{12} \quad (3.1.35) \end{aligned}$$

This is reasonable since that a case can only be in high risk phase or low risk phase during chronic infection, as long as this case progresses to chronic infection instead of leaving acute infection due to other reasons, probability of which is T_{12}.

Furthermore, with the same rationale, we can also get the probability that given a case is in low risk phase at HIV acquisition, this case is in high risk phase during chronic infection, $P_{L_1H_2}$,

$$P_{L_1H_2} = P_{L_1(H_2)H_2} + P_{L_1(L_2)H_2} = P_{L_1H_1}T_{12}P_{H_2H_2} + P_{L_1L_1}T_{12}P_{L_2H_2}$$

and the probability that given a case was in low risk phase at HIV acquisition, this case is in low risk phase during chronic infection, $P_{L_1L_2}$,

$$P_{L_1L_2} = P_{L_1(H_2)L_2} + P_{L_1(L_2)L_2} = P_{L_1H_1}T_{12}P_{H_2L_2} + P_{L_1L_1}T_{12}P_{L_2L_2}$$

with a little algebra we obtain,

$$P_{L_1H_2} = T_{12}\frac{f_H\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)} \qquad (3.1.36)$$

$$P_{L_1L_2} = T_{12}\left(\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)} + \frac{f_L\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}\right) \qquad (3.1.37)$$

Based on formulation given by equation (3.1.36) and (3.1.37), we obtain

$$P_{L_1H_2} + P_{L_1L_2} = T_{12} \qquad (3.1.38)$$

Which is also reasonable because a case can only be in high risk phase or low risk phase during chronic infection, as long as this case progresses to chronic infection instead of leaving acute infection due to other reasons, probability of which is T_{12}.

The state probabilities that we have derived in this subsection are summarized in Table III-4.

Table III- 4 Variables used in calculation of probability a model case in a risk phase during a stage of infection

Symbols	Function as other variables	Function as model parameters	Definition
ν_1	--	$\gamma_1 + \mu$	Total rate of leaving acute infection
ν_2	--	$\gamma_2 + \mu$	Total rate of leaving chronic infection
$T_{H_1L_1}$	--	$\frac{f_L\omega}{\nu_1 + f_L\omega}$	Probability of transition from high risk phase to low risk phase during acute infection
$T_{L_1H_1}$	--	$\frac{f_H\omega}{\nu_1 + f_H\omega}$	Probability of transition from low risk phase to high risk phase during acute infection
$T_{H_1H_2}$	--	$\frac{\gamma_1}{\nu_1 + f_L\omega}$	Probability of progressing from acute infection to chronic infection while in high risk phase
$T_{L_1L_2}$	--	$\frac{\gamma_1}{\nu_1 + f_H\omega}$	Probability of progressing from acute infection to

			chronic infection while in low risk phase
$T_{H_2L_2}$	--	$\frac{f_L\omega}{\nu_2+f_L\omega}$	Probability of transition from high risk phase to low risk phase during chronic infection
$T_{L_2H_2}$	--	$\frac{f_H\omega}{\nu_2+f_H\omega}$	Probability of transition from low risk phase to high risk phase during chronic infection
T_{12}	--	$\frac{\gamma_1}{\nu_1}$	Probability of progressing from acute infection to chronic infection
$P_{H_1H_1}$	$\frac{1-T_{H_1L_1}}{1-T_{H_1L_1}T_{L_1H_1}}$	$\frac{\nu_1+f_H\omega}{\nu_1+\omega}$	Probability that a case who is infected while in high risk phase is in high risk phase during acute infection
$P_{H_1L_1}$	$\frac{T_{H_1L_1}(1-T_{L_1H_1})}{1-T_{H_1L_1}T_{L_1H_1}}$	$\frac{f_L\omega}{\nu_1+\omega}$	Probability that a case who is infected while in high risk phase is in low risk phase during acute infection
$P_{L_1H_1}$	$\frac{T_{L_1H_1}(1-T_{H_1L_1})}{1-T_{H_1L_1}T_{L_1H_1}}$	$\frac{f_H\omega}{\nu_1+\omega}$	Probability that a case who is infected while in low risk phase is in high risk phase during acute infection
$P_{L_1L_1}$	$\frac{1-T_{L_1H_1}}{1-T_{H_1L_1}T_{L_1H_1}}$	$\frac{\nu_1+f_L\omega}{\nu_1+\omega}$	Probability that a case who is infected while in high risk phase is in low risk phase during acute infection
$P_{H_2H_2}$	$\frac{1-T_{H_2L_2}}{1-T_{H_2L_2}T_{L_2H_2}}$	$\frac{\nu_2+f_H\omega}{\nu_2+\omega}$	Probability that a case who is infected while in high risk phase is in high risk phase during chronic infection
$P_{H_2L_2}$	$\frac{T_{H_2L_2}(1-T_{L_2H_2})}{1-T_{H_2L_2}T_{L_2H_2}}$	$\frac{f_L\omega}{\nu_2+\omega}$	Probability that a case who is infected while in high risk phase is in low risk phase during chronic infection
$P_{L_2H_2}$	$\frac{T_{L_1H_1}(1-T_{H_1L_1})}{1-T_{H_1L_1}T_{L_1H_1}}$	$\frac{f_H\omega}{\nu_2+\omega}$	Probability that a case who is infected while in low risk phase is in high risk phase during chronic infection
$P_{L_2L_2}$	$\frac{1-T_{L_2H_2}}{1-T_{H_2L_2}T_{L_2H_2}}$	$\frac{\nu_2+f_L\omega}{\nu_2+\omega}$	Probability that a case who is infected while in high risk phase is in low risk phase during chronic infection
$P_{H_1(H_2)H_2}$	$P_{H_1H_1}T_{12}P_{H_2H_2}$	$T_{12}\left(\frac{\nu_1+f_L\omega}{\nu_1+\omega}\right)\left(\frac{\nu_2+f_H\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in high risk phase progresses to chronic infection in high risk phase and is in high risk phase during chronic infection
$P_{H_1(L_2)H_2}$	$P_{H_1L_1}T_{12}P_{L_2H_2}$	$T_{12}\left(\frac{f_L\omega}{\nu_1+\omega}\right)\left(\frac{f_H\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in high risk phase progresses to chronic infection in low risk phase

			and is in high risk phase during chronic infection
$P_{H_1(H_2)L_2}$	$P_{H_1H_1}T_{12}P_{H_2L_2}$	$T_{12}\left(\frac{\nu_1+f_H\omega}{\nu_1+\omega}\right)\left(\frac{f_L\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in high risk phase progresses to chronic infection in high risk phase and is in low risk phase during chronic infection
$P_{H_1(L_2)L_2}$	$P_{H_1L_1}T_{12}P_{L_2L_2}$	$T_{12}\left(\frac{f_L\omega}{\nu_1+\omega}\right)\left(\frac{\nu_2+f_L\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in high risk phase progresses to chronic infection in low risk phase and is in low risk phase during chronic infection
$P_{L_1(H_2)H_2}$	$P_{L_1H_1}T_{12}P_{H_2H_2}$	$T_{12}\left(\frac{f_H\omega}{\nu_1+\omega}\right)\left(\frac{\nu_2+f_H\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in low risk phase progresses to chronic infection in high risk phase and is in high risk phase during chronic infection
$P_{L_1(L_2)H_2}$	$P_{L_1L_1}T_{12}P_{L_2H_2}$	$T_{12}\left(\frac{\nu_1+f_L\omega}{\nu_1+\omega}\right)\left(\frac{f_H\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in low risk phase progresses to chronic infection in low risk phase and is in high risk phase during chronic infection
$P_{L_1(H_2)L_2}$	$P_{L_1H_1}T_{12}P_{H_2L_2}$	$T_{12}\left(\frac{f_H\omega}{\nu_1+\omega}\right)\left(\frac{f_L\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in low risk phase progresses to chronic infection in high risk phase and is in low risk phase during chronic infection
$P_{L_1(L_2)L_2}$	$P_{L_1L_1}T_{12}P_{L_2L_2}$	$T_{12}\left(\frac{\nu_1+f_L\omega}{\nu_1+\omega}\right)\left(\frac{\nu_2+f_L\omega}{\nu_2+\omega}\right)$	Probability that a case who is infected in low risk phase progresses to chronic infection in low risk phase and is in low risk phase during chronic infection
$P_{H_1H_2}$	$P_{H_1(H_2)H_2}+P_{H_1(L_2)H_2}$	$T_{12}\left(\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)}+\frac{f_H\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}\right)$	Probability that a case who is infected in high risk phase is in high risk phase during chronic infection
$P_{H_1L_2}$	$P_{H_1(H_2)L_2}+P_{H_1(L_2)L_2}$	$T_{12}\frac{f_L\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}$	Probability that a case who is infected in high risk phase is in low risk phase during chronic infection
$P_{L_1H_2}$	$P_{L_1(H_2)H_2}+P_{L_1(L_2)H_2}$	$T_{12}\frac{f_H\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}$	Probability that a case who is infected in low risk phase is in high risk phase during chronic infection
$P_{L_1L_2}$	$P_{L_1(H_2)L_2}+P_{L_1(L_2)L_2}$	$T_{12}\left(\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)}+\frac{f_L\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}\right)$	Probability that a case who is infected in low risk phase is in low risk phase during chronic infection

3.2. Understanding State Probabilities. One thing worth noting is the rule by which the probability that a case is in a specific risk phase during a stage of infection is formulated. These probabilities include $P_{H_1H_1}$, $P_{H_1L_1}$, $P_{H_1H_2}$, $P_{H_1L_2}$ for a case who was infected in high risk phase and $P_{L_1L_1}$, $P_{L_1H_1}$, $P_{H_1H_2}$, $P_{H_1L_2}$ for a case who was infected in low risk phase. To better illustrate such rule, we start with the first group, $P_{H_1H_1}$, $P_{H_1L_1}$, $P_{H_1H_2}$, $P_{H_1L_2}$. According to section 3.1., their formulations are,

$P_{H_1H_1} = \frac{\nu_1+f_H\omega}{\nu_1+\omega}$, $P_{H_1L_1} = \frac{f_L\omega}{\nu_1+\omega}$, $P_{H_1H_2} = T_{12}\left(\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)} + \frac{f_H\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}\right)$ and $P_{H_1L_2} = T_{12}\frac{f_L\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}$

where $P_{H_1L_1}$ can be further transformed as,

$$P_{H_1L_1} = \frac{f_L\omega}{\nu_1+\omega} = f_L\left(1-\frac{\nu_1}{\nu_1+\omega}\right) \qquad (3.2.1)$$

Equation (3.2.1) can be interpreted as follows. Term $\frac{\nu_1}{\nu_1+\omega}$ calculates the probability that a case does not re-select contact rate by end of acute infection, so $1-\frac{\nu_1}{\nu_1+\omega}$ calculates the probability that a case has re-selected contact rate at least once before leaving acute infection. Each time a case re-selects contact rate, he always selects low contact rate with probability f_L, so $f_L\left(1-\frac{\nu_1}{\nu_1+\omega}\right)$ calculates the probability that this case (who was in high risk phase at HIV acquisition) selected lower contact rate during acute infection, i.e. $P_{H_1L_1}$.

One can also do similar transformation for $P_{H_1L_2}$,

$$P_{H_1L_2} = T_{12}\frac{f_L\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)} = T_{12}f_L\left(\frac{\omega(\nu_1+\nu_2+\omega)}{(\nu_1+\omega)(\nu_2+\omega)}\right)$$
$$= T_{12}f_L\left(1-\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)}\right) \qquad (3.2.2)$$

In equation (3.2.2), term $\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)}$ is the product of $\frac{\nu_1}{\nu_1+\omega}$ and $\frac{\nu_2}{\nu_2+\omega}$. As discussed earlier, $\frac{\nu_1}{\nu_1+\omega}$ calculates the probability that a case does not re-select contact rate by the end of acute infection. Term $\frac{\nu_2}{\nu_2+\omega}$ calculates such probability during chronic infection. Therefore, $\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)}$ is the probability that a case has not re-selected contact rate since HIV acquisition

and till the end of chronic infection, so $1-\frac{\nu_1\nu_2}{(\nu_1+\omega)(\nu_2+\omega)}$ is the probability that this case has at least re-selected contact rate once by the time this case leaves chronic infection. Parameter f_L is added to adjust for the probability of selecting low contact phase, while T_{12} is included to adjust for the probability of progressing from acute infection to chronic infection.

Formulations of $P_{H_1L_1}$ and $P_{H_1L_2}$ share a common feature: a case who was in high risk phase at HIV acquisition has to re-select contact rate to be in low risk phase. Each time a case selects a contact rate, he has a probability f_L to select low contact rate. Given these two common features, one may infer the rule of formulating $P_{H_1L_i}$, the probability that a case who is infected in high risk phase is in low risk phase at ith stage of infection,

$$P_{H_1L_i} = T_{1i}f_L\left(1-\prod_{j=1}^{i}\left(\frac{\upsilon_j}{\upsilon_j+\omega}\right)\right) \qquad (3.2.3)$$

Where T_{1i} is the probability that a case progresses from first stage of infection to ith stage of infection. For the two stages of infection as modeled in this chapter, $T_{11}=1$ and $T_{12}=\frac{\gamma_1}{\upsilon_1}$. Term $\prod_{j=1}^{i}\left(\frac{\upsilon_j}{\upsilon_j+\omega}\right)$ is the probability that a case has not re-selected contact rate by the time this case leaves ith stage of infection. This term has been discussed in Henry and Koopman (2015), which is calculated as the expected fraction of time a case spends before first risk re-selection during ith stage of infection given that this case has progressed to ith stage of infection.[82] Although here we describe it as a probability, they are essentially the same. Same as Henry and Koopman (2015), we denote it as ψ_i,

$$\psi_i = \prod_{j=1}^{i}\left(\frac{\upsilon_j}{\upsilon_j+\omega}\right) \qquad (3.2.4)$$

When there are two stages of infection, the probability that a case has not re-selected contact rate by the end of acute infection, ψ_1, is

$$\psi_1 = \frac{\upsilon_1}{\upsilon_1+\omega} \qquad (3.2.5)$$

Also, the probability that a case has not re-selected contact rate by the end of chronic infection, ψ_2, is

$$\psi_2 = \prod_{j=1}^{2}\left(\frac{\upsilon_j}{\upsilon_j + \omega}\right) = \left(\frac{\upsilon_1}{\upsilon_1 + \omega}\right)\left(\frac{\upsilon_2}{\upsilon_2 + \omega}\right) \qquad (3.2.6)$$

Furthermore, since we defined $\psi_i = \prod_{j=1}^{i}\left(\frac{\upsilon_j}{\upsilon_j+\omega}\right)$, we can rewrite formulation of $P_{H_1L_i}$ as,

$$P_{H_1L_i} = T_{1i} f_L (1 - \psi_i) \qquad (3.2.7)$$

In subsection (3.1.), we identified the relationship that $P_{H_1H_1} + P_{H_1L_1} = 1$, $P_{H_1H_2} + P_{H_1L_2} = T_{12}$. That says, the sum of probabilities that a case is in a risk phase during ith stage of infection is the probability that this case progresses to ith stage of infection, i.e. $P_{H_1H_i} + P_{H_1L_i} = T_{1i}$. Based on this relationship, we obtain the rule by which $P_{H_1H_i}$ is formulated,

$$P_{H_1H_i} = T_{1i} - P_{H_1L_i} = T_{1i} - f_L T_{1i}(1 - \psi_i) = f_H T_{1i} + f_L T_{1i}\psi_i = T_{1i}(f_H + f_L\psi_i) \qquad (3.2.8)$$

With the same rationale, we obtain the probability that a case who was in low risk phase at HIV acquisition is in high risk phase during ith stage of infection, $P_{L_1H_i}$,

$$P_{L_1H_i} = T_{1i} f_H (1 - \psi_i) \qquad (3.2.9)$$

and the probability that a case who was in low risk phase at HIV acquisition is in low risk phase during ith stage of infection, $P_{L_1L_i}$,

$$P_{L_1L_i} = T_{1i} - P_{L_1H_i} = T_{1i} - T_{1i} f_H (1 - \psi_i) = T_{1i}(f_L + f_H\psi_i) \qquad (3.2.10)$$

Based on the rules by which that state probabilities are formulated, effect of episodic risk on each probability can be more clearly observed. Firstly, $P_{H_1L_i}$, $P_{L_1H_i}$ have negative linear relationships with ψ_i (equation (3.2.7) and equation (3.2.9)) while $P_{H_1H_i}$, $P_{L_1L_i}$ have positively linear relationships with ψ_i (equation (3.3.8) and equation (3.3.10)). When ω increases, ψ_i decreases, the probabilities that a case is in the same risk phase as that at HIV acquisition $P_{H_1H_i}$, $P_{L_1L_i}$ decrease, while the probabilities that a case is in a risk phase different from the one he was in at HIV acquisition, $P_{H_1L_i}$, $P_{L_1H_i}$, increase. When ω is extremely high, ψ_i approaches zero. In

this condition, $P_{H_1H_i}$ and $P_{L_1H_i}$ both approach $f_H T_{1i}$ (panel (A) and (C) of Fig.III-3), while $P_{H_1L_i}$ and $P_{L_1L_i}$ both approach $f_L T_{1i}$ (panel (B) and (D) of Fig.III-3). This means that the probability that a case is in a risk phase at ith stage of infection mainly depends on probability of people being in this risk phase from population distribution (f_H for high risk phase and f_L for high risk phase), and probability of disease progression T_{1i}, but barely on the risk phase that this case was in at HIV acquisition.

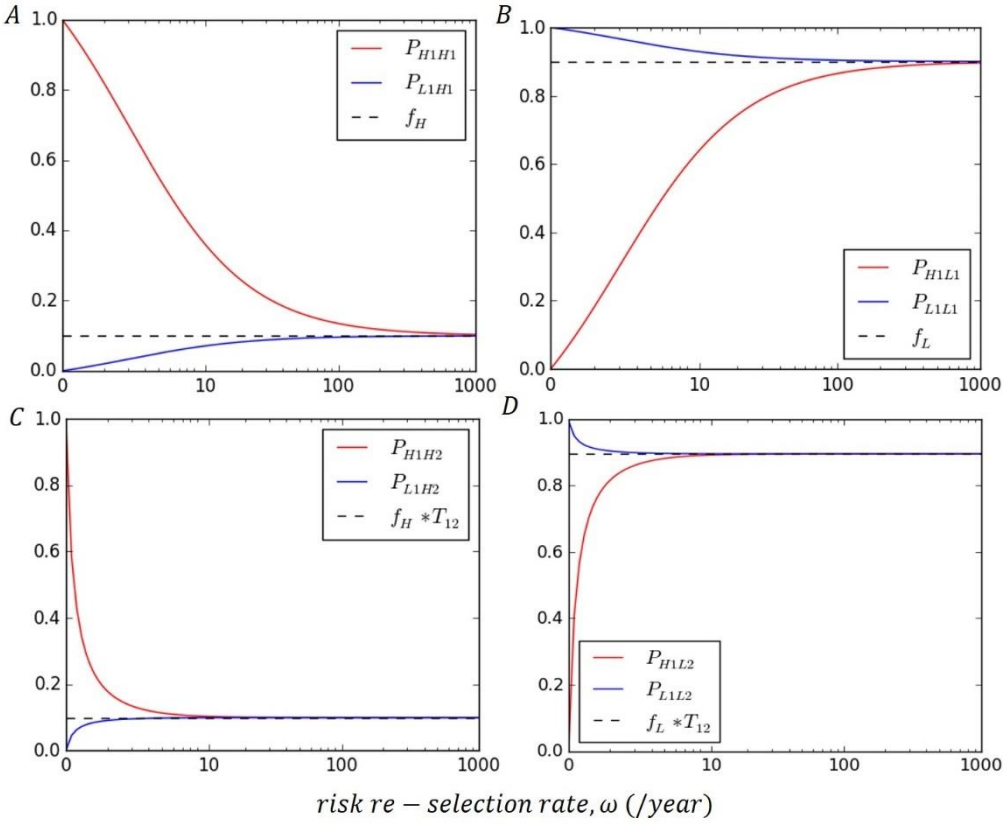

Figure III- 3 Effect of increasing risk re-selection rate, ω, on state probabilities.

Another important feature of state probabilities is that the probability that a case does not re-select contact rate by the time this case leaves ith stage of infection, $\psi_i = \prod_{j=1}^{i} \left(\frac{\upsilon_j}{\upsilon_j+\omega}\right)$, decreases as the number of stages of infection, i, increases. This is because the probability that this case does not re-select contact rate at ith stage of infection is conditional on this case not re-selecting contact rate at earlier stage, ψ_{i-1}, i.e. $\psi_i = \psi_{i-1}\left(\frac{\upsilon_i}{\upsilon_i+\omega}\right)$. This phenomenon agrees with the observation in Henry and Koopman (2015), which suggest that the expected fraction of time a case spend before first risk re-selection is smaller for later stages of infection.[82]

3.3. Expected Number of New Infections along Each Transmission Route that Contributes to Type Reproduction Number. In subsections 3.1. and 3.2., we obtain the probabilities that a model case is in a risk phase during acute infection or during chronic infection given the risk phase this case was in at HIV acquisition. In this subsection, based on the state probabilities we calculate expected number of new infections in a risk phase a case would cause given the risk phase this case was in at HIV acquisition, which will be the elements to formulate type reproduction numbers in later sections.

First, **let R_{pqe} be the total expected number of new infections in p risk phase caused by a case who was infected in q risk phase during entire infection time given episodic risk (indicated by subscript 'e'), and R_{pq_ie} be the expected number of new infections in p risk phase caused by a case who was infected in q risk phase during ith stage of infection given episodic risk.** In order to illustrate how we calculate R_{pqe}, we use R_{HHe} as an example. For a disease with two stages of infection, $R_{HHe} = R_{HH_1e} + R_{HH_2e}$. We start with calculation of R_{HH_1e}. Suppose a case is infected while in high risk phase. During acute infection, this case has $P_{H_1H_1}$ probability to be in high risk phase and $P_{H_1L_1} = 1 - P_{H_1H_1}$ probability to be in low risk phase. Given average duration of acute infection is $\frac{1}{\upsilon_1}$, this case on average spends $P_{H_1H_1}\left(\frac{1}{\upsilon_1}\right)$ time in high risk phase and $\left(1 - P_{H_1H_1}\right)\left(\frac{1}{\upsilon_1}\right)$ time in low risk phase during acute infection. Therefore, during acute infection, this case's expected transmission potential in high risk phase is $P_{H_1H_1}\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H$, and $\left(1 - P_{H_1H_1}\right)\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_L$ in low risk phase.

As we discussed in section 1., while in high risk phase this case makes fraction m fraction of contacts with susceptibles in high risk phase due to assortative mixing and (1-m)g_H fraction of contacts with susceptibles in high risk phase due to proportionate mixing (g_p is the expected fraction of contacts contributed by susceptible individuals in p risk phase at disease-free equilibrium). Therefore, when this case is introduced to a completely susceptible population, he is expected to cause $P_{H_1H_1}\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H(m + (1-\mathrm{m})\mathrm{g}_H) = P_{H_1H_1}\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H(\mathrm{g}_H + \mathrm{mg}_L)$ high-risk new infections while in high risk phase during acute infection. When this case is in low risk phase, he makes (1-m) fraction of his contacts at general mixing site and g_H fraction of these

contacts is made with susceptibles in high risk phase. Hence, this case is expected to cause $P_{H_1L_1}\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_L(1-m)\mathrm{g}_H$ high-risk new infections while in low risk phase during acute infection. Combining the number of new infections, we can obtain the total expected number of new infections in high risk phase a case who is infected in high risk phase cause during acute infection,

$$R_{HH_1e} = P_{H_1H_1}\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H(\mathrm{g}_H + \mathrm{mg}_L) + P_{H_1L_1}\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_L(1-m)\mathrm{g}_H \qquad (3.3.1)$$

In order to make expression of R_{HH_1e} more concise, we introduce two variables: expected number of new infections in high risk phase a case would cause during acute infection if this case spends whole acute infection time in high risk phase (no episodic risk), R_{HH_1}, and expected number of new infections in high risk phase a case would cause during acute infection if this case spends whole acute infection in low risk phase (no episodic risk), R_{HL_1}.

Based on the definitions of R_{HH_1} and R_{HL_1}, we obtain,

$$R_{HH_1} = \left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H(\mathrm{g}_H + \mathrm{mg}_L) \qquad (3.3.2)$$

$$R_{HL_1} = \left(\frac{1}{\upsilon_1}\right)\beta_1\chi_L(1-m)\mathrm{g}_H \qquad (3.3.3)$$

After we substituting $\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H(\mathrm{g}_H + \mathrm{mg}_L)$ with R_{HH_1} and $\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_L(1-m)\mathrm{g}_H$ with R_{HL_1} on right side of equation (3.3.1), we obtain R_{HH_1e} as,

$$R_{HH_1e} = P_{H_1H_1}R_{HH_1} + P_{H_1L_1}R_{HL_1} \qquad (3.3.4)$$

Similarly, we can also get the expected number of new infections in high risk phase this case causes during chronic infection, R_{HH_2e} as.

$$R_{HH_2e} = P_{H_1H_2}\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_H(\mathrm{g}_H + \mathrm{mg}_L) + P_{H_1L_2}\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_L(1-m)\mathrm{g}_H \qquad (3.3.5)$$

In order to simplify expression of R_{HH_2e}, we introduce two additional variables: expected number of new infections in high risk phase a case would cause during chronic infection if this

case spends whole chronic infection time in high risk phase (no episodic risk), R_{HH_2}, and expected number of new infections in high risk phase a case would cause during chronic infection if this case spends whole period of chronic infection in low risk phase (no episodic risk), R_{HL_2}.

Based on their definitions, we obtain,

$$R_{HH_2} = T_{12}\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_H(\mathrm{g}_H + \mathrm{m}\mathrm{g}_L) \qquad (3.3.6)$$

$$R_{HL_2} = T_{12}\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_L(1-m)\mathrm{g}_H \qquad (3.3.7)$$

where T_{12} is added to take into account the probability that a case progresses from acute infection to chronic infection.

Incorporating equation (3.3.6) and (3.3.7) into equation (3.3.5), one can get,

$$R_{HH_2e} = \frac{P_{H_1H_2}}{T_{12}}R_{HH_2} + \frac{P_{H_1L_2}}{T_{12}}R_{HL_2} \qquad (3.3.8)$$

Given R_{HH_1e} and R_{HH_2e}, we can obtain that given episodic risk, the total expected number of new infections in high risk phase a model case who is in high risk phase at HIV acquisition causes, R_{HHe},

$$R_{HHe} = R_{HH_1e} + R_{HH_2e} \qquad (3.3.9)$$

Given equation (3.3.4) and (3.3.8), we obtain,

$$R_{HHe} = R_{HH_1e} + R_{HH_2e} = P_{H_1H_1}R_{HH_1} + P_{H_1L_1}R_{HL_1} + \frac{P_{H_1H_2}}{T_{12}}R_{HH_2} + \frac{P_{H_1L_2}}{T_{12}}R_{HL_2}$$

we can further rewrite formulation of R_{HHe} as,

$$\begin{aligned} R_{HHe} &= P_{H_1H_1}R_{HH_1} + P_{H_1L_1}R_{HL_1} + \frac{P_{H_1H_2}}{T_{12}}R_{HH_2} + \frac{P_{H_1L_2}}{T_{12}}R_{HL_2} \\ &= P_{H_1H_1}R_{HH_1} + \frac{P_{H_1H_2}}{T_{12}}R_{HH_2} + P_{H_1L_1}R_{HL_1} + \frac{P_{H_1L_2}}{T_{12}}R_{HL_2} \end{aligned}$$

After further arranging this formulation, we can more clearly see the rule by which R_{HHe} is formulated,

$$R_{HHe} = P_{H_1H_1}R_{HH_1} + \frac{P_{H_1H_2}}{T_{12}}R_{HH_2} + P_{H_1L_1}R_{HL_1} + \frac{P_{H_1L_2}}{T_{12}}R_{HL_2} = \sum_{i=1}^{2}\frac{P_{H_1H_i}}{T_{1i}}R_{HH_i} + \sum_{i=1}^{2}\frac{P_{H_1L_j}}{T_{1i}}R_{HL_i}$$

Formulation of R_{HH_e} in this way help us to more easily interpret R_{HH_e} in epidemiological context. Specifically, the first term $\sum_{i=1}^{2}\frac{P_{H_1H_i}}{T_{1i}}R_{HH_i}$ calculates the part of R_{HH_e} attributed to transmissions that occur when the case is in high risk phase. The second term $\sum_{i=1}^{2}\frac{P_{H_1L_j}}{T_{1i}}R_{HL_i}$ calculates the part of R_{HH_e} attributed to transmissions that occur when the case is in low risk phase.

However, we are not instantly clear about the effect of episodic risk on R_{HH_e} by checking the above formulation. This is because both R_{HH_i} and R_{HL_i} are complicated functions of model parameters. Therefore, we further transform above formulation of R_{HH_e}. In section 3.2., we have obtained that $P_{H_1H_i} = T_{1i}(f_H + f_L\psi_i)$ and $P_{H_1L_i} = T_{1i}f_L(1-\psi_i)$, where $\psi_i = \prod_{j=1}^{i}\left(\frac{v_j}{v_j+\omega}\right)$ which is the probability that a case has not re-selected contact rate by the time this case leaves ith stage of infection. Hence, replacing $P_{H_1H_i}$ with $T_{1i}(f_H + f_L\psi_i)$ and $P_{H_1L_i}$ with $T_{1i}f_L(1-\psi_i)$, we obtain,

$$\begin{aligned} R_{HHe} &= \sum_{i=1}^{2}\frac{T_{1i}(f_H + f_L\psi_i)}{T_{1i}}R_{HH_i} + \sum_{i=1}^{2}\frac{T_{1i}f_L(1-\psi_i)}{T_{1i}}R_{HL_i} \\ &= \sum_{i=1}^{2}(f_H + f_L\psi_i)R_{HH_i} + \sum_{i=1}^{2}f_L(1-\psi_i)\,R_{HL_i} \end{aligned}$$

After rearranging this equation, we obtain,

$$R_{HHe} = \sum_{i=1}^{2}\psi_i R_{HH_i} + \sum_{i=1}^{2}(1-\psi_i)(f_H R_{HH_i} + f_L R_{HL_i})$$

In above equations, ψ_i calculates the probability that a case has not re-selected contact rate at ith stage of infection. Given this condition of not re-selecting contact rate, if case was in high risk phase at HIV acquisition, he will have 100% possibility of being in high risk phase.

Therefore, at ith stage of infection, this case causes $R_{HH_i}\psi_i * 1 = R_{HH_i}\psi_i$ new infections if he stays in high risk phase. When this case has at least re-selected a contact rate at ith stage of infection, probability of which is $(1-\psi_i)$, this case has f_H probability of selecting high contact rate and f_L probability of selecting low contact rate. Therefore, at ith stage of infection this case causes $(1-\psi_i)\left(f_H R_{HH_i} + f_L R_{HL_i}\right)$ new infections if he re-selects a risk phase.

Based on formulations of R_{HH_1} and R_{HH_2}, we can obtain the rules of formulating R_{HH_i},

$$R_{HH_i} = \frac{T_{1i}\beta_i\chi_H(g_H + mg_L)}{\upsilon_i}$$

Where T_{1i} is the probability that a case progresses from first stage of infection to ith stage of infection: $T_{11} = 1, T_{12} = \frac{\gamma_1}{\upsilon_1}$.

Similarly, we can also obtain the rules of formulating R_{HL_i},

$$R_{HL_i} = \frac{T_{1i}\beta_i\chi_L(1-m)g_H}{\upsilon_i}$$

Based on formulations of R_{HH_i} and R_{HL_i}, and based on relationship $f_H\chi_H + f_L\chi_L = \chi$, and $\frac{f_H\chi_H}{\chi} = g_H$ (where χ is average contact rate at disease-free equilibrium), we can do the following transformation of term $f_H R_{HH_i} + f_L R_{HL_i}$,

$$\begin{aligned} f_H R_{HH_i} + f_L R_{HL_i} &= \frac{f_H T_{1i}\beta_i\chi_H(g_H + mg_L)}{\upsilon_i} + \frac{f_L T_{1i}\beta_i\chi_L(1-m)g_H}{\upsilon_i} = \frac{T_{1i}\beta_i}{\upsilon_i}\left(f_H\chi_H(g_H + mg_L) + f_{L\chi_L}(1-m)g_H\right) \\ &= \frac{T_{1i}\beta_i}{\upsilon_i}\left(f_H\chi_H(g_H + m(1-g_H)) + f_{L\chi_L}(1-m)g_H\right) \\ &= \frac{T_{1i}\beta_i}{\upsilon_i}\left(\left(f_H\chi_H + f_{L\chi_L}\right)(1-m)g_H + f_H\chi_H m\right) = \frac{T_{1i}\beta_i}{\upsilon_i}(\chi(1-m)g_H + g_H m) = \frac{T_{1i}\beta_i\chi}{\upsilon_i}\mathrm{g}_H \end{aligned}$$

Hence, expression of R_{HHe} can be further transformed as,

$$R_{HHe} = \sum_{i=1}^{2}\psi_i R_{HH_i} + \sum_{i=1}^{2}(1-\psi_i)\frac{T_{1i}\beta_i\chi}{\upsilon_i}\,\mathrm{g}_H$$

Note that second term $\sum_{i=1}^{2}(1-\psi_i)\frac{T_{1i}\beta_i\chi}{\upsilon_i}$ only contains average contact rate, χ. In addition, parameter for assortative mixing, m, does not show up either. Earlier in discussion, we

have mentioned that this term calculates the contribution to R_{HHe} after the case re-selected contact rates. We have discussed that if a case re-selects contact rate, he would re-select from population distribution, regardless of the risk phase this case was in at HIV acquisition. This explains why overall average contact rate of this case during the period of risk fluctuation equals population average contact rate (at disease-free equilibrium), χ. The reason that assortative mixing does not affect this term is that a case randomly re-selects contact rate. The increase in contacts due to assortative mixing in one risk phase is counterbalanced by decrease in contacts due to disassortative mixing in the other risk phase. This term indicates that risk re-selection eliminates the impact of risk heterogeneity and assortative mixing on the transmission system.

Given that $R_{HH_i} = \frac{T_{1i}\beta_i}{\upsilon_i}\chi_H(\mathrm{g}_H + \mathrm{mg}_L)$, we can further transform expression of R_{HHe} as follow,

$$
\begin{aligned}
R_{HHe} &= \sum_{i=1}^{2} \psi_i R_{HH_i} + \sum_{i=1}^{2} (1-\psi_i)\frac{T_{1i}\beta_i\chi}{\upsilon_i}\mathrm{g}_H \\
&= \sum_{i=1}^{2} \psi_i \frac{T_{1i}\beta_i}{\upsilon_i}\chi_H(\mathrm{g}_H + \mathrm{mg}_L) + \sum_{i=1}^{2} \frac{T_{1i}\beta_i\chi\mathrm{g}_H}{\upsilon_i} - \sum_{i=1}^{2} \psi_i \frac{T_{1i}\beta_i\chi\mathrm{g}_H}{\upsilon_i} \\
&= \chi_H(\mathrm{g}_H + \mathrm{mg}_L)\sum_{i=1}^{2} \psi_i \frac{T_{1i}\beta_i}{\upsilon_i} + \chi\mathrm{g}_H \sum_{i=1}^{2} \frac{T_{1i}\beta_i}{\upsilon_i} - \chi\mathrm{g}_H \sum_{i=1}^{2} \psi_i \frac{T_{1i}\beta_i}{\upsilon_i} = \\
&= \chi_H(\mathrm{g}_H + \mathrm{mg}_L)\sum_{i=1}^{2} \frac{T_{1i}\beta_i}{\upsilon_i}\left(\frac{\sum_{i=1}^{2} \psi_i \frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2} \frac{T_{1i}\beta_i}{\upsilon_i}}\right) + \chi \sum_{i=1}^{2} \frac{T_{1i}\beta_i}{\upsilon_i}\mathrm{g}_H \\
&- \chi\mathrm{g}_H \sum_{i=1}^{2} \frac{T_{1i}\beta_i}{\upsilon_i}\left(\frac{\sum_{i=1}^{2} \psi_i \frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2} \frac{T_{1i}\beta_i}{\upsilon_i}}\right)
\end{aligned}
$$

Term $\chi\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}$ is the total expected number of new infections a case would cause during the entire infection time if this case has population average contact rate at disease-free equilibrium, χ. It is also R_0 for population with homogeneous contact rate, χ, we denote this term as H,

$$H = \chi\sum_{i=1}^{2}\beta_i T_{1i}\left(\frac{1}{\upsilon_i}\right) \qquad (3.3.10)$$

To understand term $\left(\frac{\sum_{i=1}^{2}\psi_i\frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}}\right)$, it is helpful to consider the meaning of ψ_i and $\frac{T_{1i}\beta_i}{\upsilon_i}$. As discussed earlier, ψ_i is the probability that a case has not re-selected contact rate at ith stage of infection. If all stages of infection have equal contribution to overall transmission potential, then the total probability that a case transmits HIV when he has not re-selected contact rate is $\frac{\sum_{i=1}^{2}\psi_i}{2}$. However, different stages of infection may have different contribution to overall transmission potential. Therefore, probability ψ_i needs to be weighted by the contribution of ith stage of infection to transmission potential. Hence, ratio $\frac{\psi_i\frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}}$ is the probability that a case has not re-selected contact rates and transmits HIV through a contact at ith stage of infection. Sum of $\frac{\psi_i\frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}}$ over all stages of infection, $\left(\frac{\sum_{i=1}^{2}\psi_i\frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}}\right)$, is the probability that a case transmits HIV when this case has not re-selected contact rate during the entire infection time. We denote this term as ψ, so

$$\psi = \frac{\sum_{i=1}^{2}\psi_i\frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}} \qquad (3.3.11)$$

This term has been derived and discussed in Henry and Koopman (2015).[82] However, in study by Henry and Koopman (2015), this term is described as remaining heterogeneity effect, which is the expected fraction of transmission potential attributed before first risk re-selection.[82] Here, we mainly described it as probability. However, given we discussed in a mean field framework, they are essentially the same.

Therefore, given equation (3.3.10) and (3.3.11), we can further simplify expression of R_{HHe} as,

$$R_{HHe} = \chi_H(g_H + mg_L)\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}\left(\frac{\sum_{i=1}^{2}\psi_i\frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}}\right) + \chi\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}g_H$$
$$- \chi\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}g_H\left(\frac{\sum_{i=1}^{2}\psi_i\frac{T_{1i}\beta_i}{\upsilon_i}}{\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}}\right)$$
$$= \psi\chi_H(g_H + mg_L)\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i} + Hg_H - Hg_H\psi$$

The first term $\psi\chi_H(g_H + mg_L)\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}$ can be further simplified by appreciating the fact that $R_{HH_i} = \chi_H(g_H + mg_L)\frac{T_{1i}\beta_i}{\upsilon_i}$,

$$\psi\chi_H(g_H + mg_L)\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i} = \psi\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i}\chi_H(g_H + mg_L) = \psi\sum_{i=1}^{2}R_{HH_i}.$$

Let R_{HH} be the total expected number of secondary high-risk infections a case would cause if this case spends the whole infection time in high risk phase (no episodic risk), so

$$R_{HH} = \sum_{i=1}^{2}R_{HH_i}$$

Substituting $\sum_{i=1}^{2}R_{HH_i}$ with R_{HH}, we obtain,

$$R_{HHe} = \psi\chi_H(g_H + mg_L)\sum_{i=1}^{2}\frac{T_{1i}\beta_i}{\upsilon_i} + Hg_H - Hg_H\psi = \psi R_{HH} + (1-\psi)Hg_H \qquad (3.3.12)$$

With the same rationale, we can obtain expected number of new infections in high risk phase caused by a case who was in low risk phase at HIV acquisition, R_{HLe},

$$R_{HLe} = \sum_{i=1}^{2}\psi_i R_{HL_i} + \sum_{i=1}^{2}(1-\psi_i)(f_H R_{HH_i} + f_L R_{HL_i}) = R_{HL}\psi + (1-\psi)Hg_H \qquad (3.3.13)$$

expected number of new infections in low risk phase caused by a case who was in high risk phase at HIV acquisition, R_{LHe},

$$R_{LHe} = \sum_{i=1}^{2} \psi_i R_{LH_i} + \sum_{i=1}^{2} (1-\psi_i)(f_H R_{LH_i} + f_L R_{LL_i}) = R_{LH}\psi + (1-\psi) H g_L \qquad (3.3.14)$$

and expected number of new infections in low risk phase caused by a case who was in low risk phase at HIV acquisition, R_{LLe},

$$R_{LLe} = \sum_{i=1}^{2} \psi_i R_{LL_i} + \sum_{i=1}^{2} (1-\psi_i)(f_H R_{LH_i} + f_L R_{LL_i}) = R_{LL}\psi + (1-\psi) H g_L \qquad (3.3.15)$$

Where in equation (3.3.14) and (3.3.15), variable R_{LL_i} is the expected number of new infections in low risk phase a case would cause during ith stage of infection if this case spends the whole period of ith stage of infection in low risk phase (no episodic risk), and R_{LH_i} is the expected number of new infections in low risk phase a case would cause during ith stage of infection if this case spends the whole period of ith stage of infection in high risk phase (no episodic risk). Based on their definitions, we obtain,

$$R_{LL_1} = \left(\frac{1}{\upsilon_1}\right) \beta_1 \chi_L (g_L + m g_H) \qquad (3.3.16)$$

$$R_{LL_2} = \left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{1}{\upsilon_2}\right) \beta_1 \chi_L (g_L + m g_H) \qquad (3.3.17)$$

and

$$R_{LH_1} = \left(\frac{1}{\upsilon_1}\right) \beta_1 \chi_H (g_L + m g_H) \qquad (3.3.18)$$

$$R_{LH_2} = \left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{1}{\upsilon_2}\right) \beta_1 \chi_L (g_L + m g_H) \qquad (3.3.19)$$

Formulation of R_{**_1}, R_{**_2}, R_{**} and R_{**_e} are listed in Table III-5.

Table III- 5 Formulations of variables used in calculation of expected number of new infections a model case generates during infection time

Symbol	formulation	meaning
R_{HH_1}	$\left(\frac{1}{\upsilon_1}\right)\beta_1 \chi_H (g_H + m g_L)$	Expected number of new infections in high risk phase caused by a case during acute infection if this case spends the whole acute infection in high risk phase (no episodic risk)
R_{HH_2}	$\left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{1}{\upsilon_2}\right)\beta_2 \chi_H (g_H + m g_L)$	Expected number of new infections in high risk phase caused by a case during chronic infection if this case spends the whole chronic infection

		in high risk phase (no episodic risk)
R_{HH}	$R_{HH_1} + R_{HH_2}$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case spends the whole infection time in high risk phase (no episodic risk)
R_{HL_1}	$\left(\frac{1}{v_1}\right)\beta_1\chi_L(1-m)g_H$	Expected number of new infections in high risk phase caused by a case during acute infection if this case spends the whole acute infection in low risk phase (no episodic risk)
R_{HL_2}	$\left(\frac{\gamma_1}{v_1}\right)\left(\frac{1}{v_2}\right)\beta_2\chi_L(1-m)g_H$	Expected number of new infections in high risk phase caused by a case during chronic infection if this case spends the whole acute infection in low risk phase (no episodic risk)
R_{HL}	$R_{HL_1} + R_{HL_2}$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case spends the whole infection time in low risk phase (no episodic risk)
R_{LH_1}	$\left(\frac{1}{v_1}\right)\beta_1\chi_H(1-m)g_L$	Expected number of new infections in high risk phase caused by a case during acute infection if this case spends the whole acute infection in high risk phase (no episodic risk)
R_{LH_2}	$\left(\frac{\gamma_1}{v_1}\right)\left(\frac{1}{v_2}\right)\beta_2\chi_H(1-m)g_L$	Expected number of new infections in high risk phase caused by a case during chronic infection if this case spends the whole chronic infection in high risk phase (no episodic risk)
R_{LH}	$R_{LH_1} + R_{LH_2}$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case spends the whole infection time in high risk phase (no episodic risk)
R_{LL_1}	$\left(\frac{1}{v_1}\right)\beta_1\chi_L(g_L + mg_H)$	Expected number of new infections in low risk phase caused by a case during acute infection if this case spends the whole acute infection in low risk phase (no episodic risk)
R_{LL_2}	$\left(\frac{\gamma_1}{v_1}\right)\left(\frac{1}{v_2}\right)\beta_2\chi_L(g_L + mg_H)$	Expected number of new infections in low risk phase caused by a case during chronic infection if this case spends the whole chronic infection in low risk phase (no episodic risk)
R_{LL}	$R_{LL_1} + R_{LL_2}$	Expected number of new infections in low risk phase caused by a case during entire infection time if this case spends the whole infection time in low risk phase (no episodic risk)
ψ_1	$\frac{v_1}{v_1 + \omega}$	Probability that a case does not re-select contact rate during acute infection
ψ_2	$\left(\frac{v_1}{v_1 + \omega}\right)\left(\frac{v_2}{v_2 + \omega}\right)$	Probability that a case has not re-selected contact rate at chronic infection
ψ	$\frac{\left(\frac{\beta_1}{v_1}\right)\psi_1 + \left(\frac{\gamma_1}{v_1}\right)\left(\frac{\beta_2}{v_2}\right)\psi_2}{\left(\frac{\beta_1}{v_1} + \left(\frac{\gamma_1}{v_1}\right)\left(\frac{\beta_2}{v_2}\right)\right)}$	Probability that a case transmits HIV infection when this case has not re-selected contact rate
H	$\chi\frac{\beta_1}{v_1} + \chi\left(\frac{\gamma_1}{v_1}\right)\left(\frac{\beta_2}{v_2}\right)$	expected number of new infections a case would cause during entire infection period if the expected contact rate of this case equals population average contact rate, χ
R_{HHe}	$R_{HH}\psi + (1-\psi)Hg_H$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case was in high risk phase at HIV acquisition, given episodic risk
R_{HLe}	$R_{HL}\psi + (1-\psi)Hg_H$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case was in low risk phase at HIV acquisition, given episodic risk
R_{LHe}	$R_{LH}\psi + (1-\psi)Hg_L$	Expected number of new infections in low risk phase caused by a case during entire infection time if this case was in high risk phase at HIV acquisition, given episodic risk
R_{LLe}	$R_{LL}\psi + (1-\psi)Hg_L$	Expected number of new infections in low risk phase caused by a case during entire infection time if this case was in low risk phase at HIV acquisition, given episodic risk

3.4. Formulating Type Reproduction Number Based on Its Epidemiological Meaning.

In the above we have derived the R_{pqe}, which is the expected number of new infections in risk phase p caused by a case who is infected in risk phase q if this case is introduced into a fully susceptible population. With formulations of R_{pqe}, we are able to derive the type reproduction number T based on its epidemiological meaning. In Section 2, we use next generation matrix approach to derive T. We also discussed that in the definition of type reproduction number, "type" refers to the state of cases at time of their infection, or "epidemiological birth". In our model, we define types of cases as the risk phase at "epidemiological birth", i.e. time of HIV acquisition. Based on the definition, type reproduction number, T_i, describe the number of new infections in i risk phase caused by a case who was in i risk phase at HIV acquisition. This implies that there are two subpopulations that intervention targeted at whom may be informed by type reproduction numbers defined this way. First is susceptible population who is in the risk phase of interest. The other is infected population who were in the risk phase of interest at HIV acquisition. This is because reducing transmission to susceptible population currently in risk phase i always reduce new infections in risk phase i. Such population is a reasonable to target since intervention such as pre-exposure prophylaxis (PrEP) usually allocate efforts based on susceptible people's current risk states. The second subpopulation is hard to identify since one needs to trace back to the time of HIV acquisition of infected individuals, a strategy which is less likely to implement so far. Therefore, the goal of defining type reproduction number this way is to explore the potential value of the type reproduction number in informing controls that are targeted at susceptible populations who are experiencing the risk phase of interest, e.g. higher rate of unprotected sexual contact compared to other susceptible individuals.

As defined in Roberts and Heesterbeek (2003), type reproduction number, T_i, calculates the expected number of new infections of type i caused by a case of type i through all transmission routes without intermediate type i new infections.[80] In section 2, we used next generation matrix approach to calculate, $T_{H_{Inf}}$ and $T_{L_{Inf}}$, which are type reproduction numbers for populations who were in high risk phase at HIV acquisition or in low risk phase at HIV acquisition, respectively. However, $T_{H_{Inf}}$ and $T_{L_{Inf}}$ are both functions of entries of NGM K, while each entry of K is complicated a function of risk re-selection rate, ω. Therefore, in this

section, we calculate these two type reproduction numbers based on our interpretation of their epidemiological meanings given by Roberts and Heesterbeek (2003). [80]

We start with $T_{H_{Inf}}$. Based on its definition, it has two components: expected number of new infections in high risk phase caused by a case who was in high risk phase at HIV acquisition during this case's entire infection time, and expected number of new infections in high risk phase caused by a case who was in high risk phase at HIV acquisition through transmission routes that involve cases who was in low risk phase at HIV acquisition during this case's entire infection time. We name the first component as contribution from direct transmission path and the second component as contribution from indirect transmission path. In section 3.3., we have obtained the contribution from direct path, which is R_{HHe}. Therefore, we mainly explain how we calculate contribution from indirect path. We start from the simplest condition, which is that a case who was in high risk phase at HIV acquisition transmits HIV infection to a case who was infected in low risk phase, and result in transmission to new infections in high risk phase.

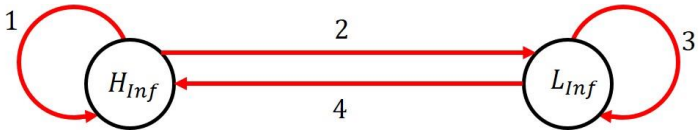

Figure III- 4. Schematic of paths that generate new infections among two compartments: infected people who were in high risk phase at time of infection, H_{Inf}, and infected people who were in low risk phase at time of infection, L_{Inf}.

Fig III-4 illustrates the pathways that generate new infections of each type: high risk phase at HIV acquisition, H_{Inf}, and low risk phase at HIV acquisition, L_{Inf}, where subscript 'Inf' indicates they refer to the risk phase at time of infection. As shown in Fig.III-4, if a case was in high risk phase at HIV acquisition, he entered compartment 'H_{Inf}' through HIV acquisition. We refer to this case as case '0', which indicates that he is the zero generation. Case '0' can cause new infections among H_{Inf} through pathway '1'. In subsection (3.3.), we obtained that expected number of new infections generated through this pathway is R_{HH_e}. In addition, case '0' can also first pass infection to compartment L_{Inf} through pathway '2', by transmitting HIV to susceptible individuals experiencing low risk. In section 3.3., we have obtained that expected number of new infections generated through this pathway '2' is R_{LH_e}. After infection is passed to compartment

L_{Inf}, cases in compartment L_{Inf} can immediately passed infections back to compartment 'H_{nf}' through pathway '4', while each case in L_{Inf} causes R_{HL_e} new infections in compartment 'H_{Inf}'. Therefore, the total expected number of new infections in H_{Inf} resulting from case '0' through pathway '2'->'4' is $R_{LH_e}R_{HL_e}$. In addition, infection can also be passed within compartment L_{Inf} for a round i.e. through pathway '3', before ending up in transmission back to H_{Inf}. The expected number of new infections generated by a case through pathway '3' is R_{LL_e}. In this situation, the transmission route that leads from case '0' to new infections in compartment H_{Inf} is '2'->'3'->'4'. As result, the expected number of new infections in H_{Inf} generated by case '0' is $R_{LH_e}R_{LL_e}R_{HL_e}$. With the same reasoning, we can infer that the expected number of new infections in H_{Inf} caused by case '0' through transmission route that involve n rounds of transmission within compartment L_{Inf} is $R_{LH_e}(R_{LL_e})^n R_{HL_e}$. Theoretically, infection can be passed on within compartment L_{Inf} for endless rounds. After including all the possibilities, the expected number of new infections in compartment H_{Inf} caused by case '0' (who is also infected while in high risk phase) through indirect pathway is

$R_{LH_e}R_{HL_e} + R_{LH_e}R_{LL_e}R_{HL_e} + \cdots + R_{LH_e}(R_{LL_e})^n R_{HL_e} + \cdots = \sum_{i=0}^{\infty} R_{LH_e}(R_{LL_e})^i R_{HL_e}$. When $R_{LL_e} < 1$, we can rewrite this equation as

$\sum_{i=0}^{\infty} R_{LH_e}(R_{LL_e})^i R_{HL_e} = R_{LH_e}R_{HL_e}\sum_{i=0}^{\infty}(R_{LL_e})^i = \frac{R_{LH_e}R_{HL_e}}{1-R_{LL_e}}$.

$T_{H_{Inf}}$ is the sum of contribution from direct path, R_{HH_e}, and contribution from indirect path, $\frac{R_{LH_e}R_{HL_e}}{1-R_{LL_e}}$,

$$T_{H_{Inf}} = R_{HH_e} + \frac{R_{LH_e}R_{HL_e}}{1 - R_{LL_e}} \qquad (3.4.1)$$

With the same rationale, we can also obtain the contribution from direct path to $T_{L_{Inf}}$, R_{LL_e}, and contribution from indirect path to $T_{L_{Inf}}$, $\sum_{i=0}^{\infty} R_{HL_e}(R_{HH_e})^i R_{LH_e} = \frac{R_{HL_e}R_{LH_e}}{1-R_{HH_e}}$, so $T_{L_{Inf}}$ is,

$$T_{L_{Inf}} = R_{LL_e} + \frac{R_{HL_e}R_{LH_e}}{1 - R_{HH_e}} \qquad (3.4.2)$$

One may desire to compare formulations of $T_{H_{Inf}}$ and $T_{L_{Inf}}$ given by equation (3.4.1) and (3.4.2) with their formulations given in Section 2,

$$T_{H_{Inf}} = K_{11} + \frac{K_{21}K_{12}}{1-K_{22}}$$

$$T_{L_{Inf}} = K_{22} + \frac{K_{12}K_{21}}{1-K_{11}}$$

In above expressions, $K_{11}, K_{12}, K_{21}, K_{22}$, are the four entries of NGM K. Therefore, it is helpful to first understand how R_{HHe}, R_{LHe}, R_{HLe} and R_{LLe} are related each entry of NGM K. We present NGM K again to better illustrate its relationship with these four variables,

$$K = \begin{bmatrix} \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(g_H+mg_L) & \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_L(1-m)g_H \\ + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_H\chi & + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_H\chi \\ \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(1-m)g_L & \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_L(g_L+mg_H) \\ + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_L\chi & + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_L\chi \end{bmatrix}$$

As discussed in Section 2, entry K_{ij} is defined as the expected number of cases who is in i risk phase at infection caused by a case who was in j risk phase. We also defined that, the first row(column) of K corresponds to being in high risk phase at HIV acquisition, while the second row(column) corresponds to being in low risk phase at infection. In section 3.3., we also defined R_{pqe} as expected number of cases who were in risk phase p at infection caused by a case who was in risk phase q at infection. Such definitions imply that $K_{11} = R_{HHe}$, $K_{12} = R_{HLe}$, $K_{21} = R_{LHe}$, $K_{22} = R_{LLe}$. In order to confirm whether these relationships are true, we list formulations of K_{11}, K_{12}, K_{21} and K_{22} and transform them one by one.

As shown in K,
$K_{11} = \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(g_H+mg_L) + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_H\chi$. we can transform K_{11} as follows,

$$K_{11} = \left(\frac{\beta_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(g_H + mg_L) + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(\frac{\omega}{\upsilon_1+\omega}\right)g_H\chi$$
$$= \left(\frac{\beta_1}{\upsilon_1}\left(\frac{\upsilon_1}{\upsilon_1+\omega}\right) + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\frac{\upsilon_1\upsilon_2}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(g_H + mg_L) + \left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1(\upsilon_1+\upsilon_2+\omega)}{\upsilon_1\upsilon_2(\upsilon_2+\omega)}\right)\left(1 - \frac{\upsilon_1}{\upsilon_1+\omega}\right)g_H\chi$$
$$= \left(\frac{\beta_1}{\upsilon_1}\left(\frac{\upsilon_1}{\upsilon_1+\omega}\right) + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\frac{\upsilon_1\upsilon_2}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(g_H + mg_L) + \left(\frac{\beta_1}{\upsilon_1}\chi + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\chi\right)g_H - \left(\frac{\beta_1}{\upsilon_1}\chi\frac{\upsilon_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\frac{\upsilon_1}{(\upsilon_2+\omega)}\frac{\upsilon_1}{\upsilon_1+\omega}\chi\right)g_H$$

According to our calculation in section 3.3., the three terms on right side of equation can be transformed as,

$$\left(\frac{\beta_1}{\upsilon_1}\left(\frac{\upsilon_1}{\upsilon_1+\omega}\right) + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\frac{\upsilon_1\upsilon_2}{(\upsilon_1+\omega)(\upsilon_2+\omega)}\right)\chi_H(g_H + mg_L) = \left(\frac{\beta_1}{\upsilon_1}\psi_1 + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\psi_2\right)\chi_H(g_H + mg_L)$$
$$= \frac{\left(\frac{\beta_1}{\upsilon_1}\psi_1 + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\psi_2\right)}{\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}}\chi_H(g_H + mg_L)\left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\right) = \psi R_{HH}$$

$$\left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\right)\chi g_H = Hg_H$$

$$\left(\frac{\beta_1}{\upsilon_1}\frac{\upsilon_1}{\upsilon_1+\omega} + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\frac{\upsilon_1}{(\upsilon_2+\omega)}\frac{\upsilon_1}{\upsilon_1+\omega}\right)\chi g_H = \left(\frac{\beta_1}{\upsilon_1}\psi_1 + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\psi_2\right)\chi g_H$$
$$= \frac{\left(\frac{\beta_1}{\upsilon_1}\psi_1 + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\psi_2\right)}{\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}}\left(\frac{\beta_1}{\upsilon_1} + \frac{\beta_2\gamma_1}{\upsilon_1\upsilon_2}\right)\chi g_H = \psi H g_H$$

We can obtain,

$$K_{11} = \psi R_{HH} + Hg_H - \psi Hg_H = \psi R_{HH} + (1-\psi)Hg_H = R_{HHe}$$

Similarly,

$$K_{12} = \left(\frac{\beta_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{(v_1+\omega)(v_2+\omega)}\right)\chi_L(1-m)g_H + \left(\frac{\beta_1}{v_1} + \frac{\beta_2\gamma_1(v_1+v_2+\omega)}{v_1v_2(v_2+\omega)}\right)\left(\frac{\omega}{v_1+\omega}\right)g_H\chi =$$
$$\left(\frac{\beta_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{(v_1+\omega)(v_2+\omega)}\right)\chi_L(1-m)g_H + \left(\frac{\beta_1}{v_1}\chi + \frac{\beta_2\gamma_1}{v_1v_2}\chi\right)g_H - \left(\frac{\beta_1}{v_1}\chi\frac{v_1}{v_1+\omega} + \right.$$
$$\left.\frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1}{(v_2+\omega)}\frac{v_1}{v_1+\omega}\chi\right)g_H = \left(\frac{\beta_1}{v_1}\frac{v_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1v_2}{(v_1+\omega)(v_2+\omega)}\right)\chi_L(1-m)g_H + \left(\frac{\beta_1}{v_1}\chi + \frac{\beta_2\gamma_1}{v_1v_2}\chi\right)g_H -$$
$$\left(\frac{\beta_1}{v_1}\chi\frac{v_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1}{(v_2+\omega)}\frac{v_1}{v_1+\omega}\chi\right)g_H = \psi R_{HL} + Hg_H - \psi Hg_H = \psi R_{HL} + (1-\psi)Hg_H = R_{HLe}$$

$$K_{21} = \chi_H\beta_1(1-m)g_L\frac{(v_1+f_H\omega)}{v_1(v_1+\omega)} + \chi_L\beta_1(g_L+mg_H)\frac{f_L\omega}{v_1(v_1+\omega)}$$
$$+ \chi_H\beta_2(1-m)g_L\frac{\gamma_1}{v_1v_2}\left(\frac{v_1v_2}{(v_1+\omega)(v_2+\omega)} + \frac{f_H\omega(\omega+v_1+v_2)}{(v_1+\omega)(v_2+\omega)}\right)$$
$$+ \chi_L\beta_2(g_L+mg_H)\frac{\gamma_1f_L\omega(v_1+v_2+\omega)}{v_1v_2(v_1+\omega)(v_2+\omega)} =$$
$$= \left(\frac{\beta_1}{v_1}\frac{v_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1v_2}{(v_1+\omega)(v_2+\omega)}\right)\chi_H(1-m)g_L + \left(\frac{\beta_1}{v_1}\chi + \frac{\beta_2\gamma_1}{v_1v_2}\chi\right)g_L$$
$$- \left(\frac{\beta_1}{v_1}\chi\frac{v_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1}{(v_2+\omega)}\frac{v_1}{v_1+\omega}\chi\right)g_L = \psi R_{LH} + Hg_L - \psi Hg_L$$
$$= \psi R_{LH} + (1-\psi)Hg_L = R_{LHe}$$

$$K_{22} = \left(\frac{\beta_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{(v_1+\omega)(v_2+\omega)}\right)\chi_L(g_L+mg_H)$$
$$+ \left(\frac{\beta_1}{v_1} + \frac{\beta_2\gamma_1(v_1+v_2+\omega)}{v_1v_2(v_2+\omega)}\right)\left(\frac{\omega}{v_1+\omega}\right)g_L\chi$$
$$= \left(\frac{\beta_1}{v_1}\left(\frac{v_1}{v_1+\omega}\right) + \frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1v_2}{(v_1+\omega)(v_2+\omega)}\right)\chi_L(g_L+mg_H)$$
$$+ \left(\frac{\beta_1}{v_1} + \frac{\beta_2\gamma_1(v_1+v_2+\omega)}{v_1v_2(v_2+\omega)}\right)\left(1-\frac{v_1}{v_1+\omega}\right)g_L\chi$$
$$= \left(\frac{\beta_1}{v_1}\left(\frac{v_1}{v_1+\omega}\right) + \frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1v_2}{(v_1+\omega)(v_2+\omega)}\right)\chi_L(g_L+mg_H) + \left(\frac{\beta_1}{v_1}\chi + \frac{\beta_2\gamma_1}{v_1v_2}\chi\right)g_L$$
$$- \left(\frac{\beta_1}{v_1}\chi\frac{v_1}{v_1+\omega} + \frac{\beta_2\gamma_1}{v_1v_2}\frac{v_1}{(v_2+\omega)}\frac{v_1}{v_1+\omega}\chi\right)g_L = \psi R_{LL} + Hg_L - \psi Hg_L$$
$$= \psi R_{LL} + (1-\psi)Hg_L = R_{LLe}$$

In summary, after some algebra, we found that each entry of NGM K is equivalent to the expected number of new infections in risk phase of corresponding row generated by a case who was in risk phase of corresponding column: $K_{11} = R_{HHe}, K_{12} = R_{HLe}, K_{21} = R_{LHe}, K_{22} = R_{LLe}$.

If we substitute K entries with each R_{**e} that they are equivalent to, we can transform the two type reproduction numbers as,

$$T_{H_{Inf}} = K_{11} + \frac{K_{21}K_{12}}{1-K_{22}} = R_{HHe} + \frac{R_{LHe}R_{HLe}}{1-R_{LLe}}$$

$$T_{L_{Inf}} = K_{22} + \frac{K_{12}K_{21}}{1-K_{11}} = R_{LLe} + \frac{R_{HLe}R_{LHe}}{1-R_{LLe}}$$

which match the formulation that we obtain in this section (equation (3.4.1) and equation (3.4.2)). This confirms that our derivation of $T_{H_{Inf}}$ and $T_{L_{Inf}}$ based on their epidemiological meanings are correct.

4. Using Type Reproduction Number for Targeted Controls that Reduces Acquisition Risk

In this section, we explore how type reproduction numbers can be used to inform targeted controls to reduce acquisition risk. In subsection (3.4.), we obtain the formulation of the two type reproduction numbers,

$$\mathrm{T_{H_{Inf}}} = \mathrm{R_{HH_e}} + \frac{\mathrm{R_{LH_e}R_{HL_e}}}{1-\mathrm{R_{LL_e}}} \qquad \text{(equation (3.4.1))}$$

$$\mathrm{T_{L_{Inf}}} = \mathrm{R_{LL_e}} + \frac{\mathrm{R_{HL_e}R_{LH_e}}}{1-\mathrm{R_{HH_e}}} \qquad \text{(equation (3.4.2))}$$

Where subscript 'Inf' is added behind the risk phase 'H' or 'L' because that risk phases at time of infection are used to category people's types. As discussed earlier, this is because reproduction number describes the expected number of new infections a case generates of his/her own type, which is the type at the time of infection, or "epidemiological birth". We also discussed that a population in risk phase p at "epidemiological birth" includes two subpopulations: susceptible people who are currently in the risk phase p and can potentially enter infection while in risk phase p, and infected people who were infected while in risk phase p. The former subpopulation can be potential target population for intervention that reduces susceptibility, given that interventions usually assess people's risk status based on recent or current risk behavior rather than risk behavior in the past. However, the later subpopulation can be hard to identify because it is hard to assess people's risk status in the past, especially exactly at the time of HIV acquisition. Therefore, in this section, we focus on the former subpopulation

and examine how type reproduction numbers may inform control strategies that are targeted at them.

Based on potential risk phase at HIV acquisition, susceptible populations can be categorized as those experiencing high risk phase (who will also enter infection while in high risk phase), or experiencing low risk phase (who will also enter infection in low risk phase). In section 1., we name the first susceptible population as 'S_H' and the second as 'S_L'. Suppose a control strategy is targeted at all 'S_H' people, and this control strategy can reduce each 'S_H' individual acquisition risk by fraction δ_H. Then all transmissions to 'S_H' individuals will be reduced by fraction δ_H. Variables that quantify such transmissions include, R_{HH_e} and R_{HL_e}. After this control strategy is implemented, R_{HH_e} is reduced to $R_{HH_e}(1-\delta_L)$ while R_{HL_e} is reduced to $R_{HL_e}(1-\delta_L)$. Then $T_{H_{Inf}}$ after this control strategy is implemented, $T_{H_{Inf}C}$ is,

$$T_{H_{Inf}C} = R_{HH_e}(1-\delta_H) + \frac{R_{LH_e}R_{HL_e}(1-\delta_H)}{1-R_{LL_e}} = (1-\delta_H)\left(R_{HH_e} + \frac{R_{LH_e}R_{HL_e}}{1-R_{LL_e}}\right)$$
$$= (1-\delta_H)T_{H_{Inf}}$$

The critical δ_H to eliminate HIV transmissions is one at which $T_{H_{Inf}C}$ equals 1. By setting $T_{H_{Inf}C} = (1-\delta_H)T_{H_{Inf}} = 1$, we can solve critical δ_H to eliminate HIV transmissions as,

$$\delta_H = 1 - \frac{1}{T_{H_{Inf}}} \qquad (4.1)$$

If this control strategy is targeted at susceptible population who are in low risk phase. 'S_L', and suppose that this strategy reduces each 'S_L' individual acquisition risk by fraction δ_L. In this case, all transmissions to 'S_L' individuals will be reduced by fraction δ_L. Variables that quantify such transmissions include, R_{LH_e} and R_{LL_e}. After this control strategy is implemented, R_{LH_e} is reduced to $R_{LH_e}(1-\delta_L)$ while R_{LL_e} is reduced to $R_{LL_e}(1-\delta_L)$. Therefore, $T_{L_{Inf}}$ after this control strategy is implemented, $T_{L_{Inf}C}$ is,

$$T_{L_{Inf}C} = R_{LL_e}(1-\delta_L) + \frac{R_{HL_e}R_{LH_e}(1-\delta_H)}{1-R_{HH_e}} = (1-\delta_L)\left(R_{LL_e} + \frac{R_{HL_e}R_{LH_e}}{1-R_{HH_e}}\right) = (1-\delta_L)T_{H_{Inf}}$$

By setting $T_{L_{Inf}C} = (1 - \delta_L)T_{L_{Inf}} = 1$, we can solve critical δ_L to eliminate HIV transmissions as,

$$\delta_L = 1 - \frac{1}{T_{L_{Inf}}} \qquad (4.2)$$

As shown by equation (4.1) and (4.2), δ_H is positively associated with $T_{H_{Inf}}$ while δ_L is positively associated with $T_{L_{Inf}}$. One can interpret such relationship in epidemiological context: type reproduction number T_i describes the number of new infections of type i in each generation that is necessary to sustain ongoing transmissions. If T_i is large, there are many new infections of type i subpopulation per generation to sustain ongoing transmission, which indicates each type i new infection plays a small role in sustaining ongoing transmission. An extreme case would be other types of subpopulation alone can sustain ongoing transmission. In this case T_i becomes infinity, each type i new infection has almost zero importance in sustaining ongoing transmission, i.e. 1/infinity. In this case, blocking transmissions to type i new infection cannot eliminate transmissions. This also implies that in the same transmission system, if T_i of type i is smaller than T_j of type j, each type i new infection plays a more important role in sustaining ongoing transmission than each type j new infection. This sounds counterintuitive, but it is important to address that type reproduction number does not quantify the ability of corresponding type of population alone to cause onward transmission of. Rather, it is the combination of contribution from all types of subpopulation to generating new infections of corresponding type.

5. Effect of Episodic Risk on Type Reproduction Numbers and Critical Control Efforts of Relevant Targeted Controls

In subsection (3.4.), we derived the formulations of type reproduction numbers as function of episodic risk. In this section, we examine how type reproduction numbers change as risk re-selection rate, ω, varies.

In subsection (3.4.), we have obtained the formulations of type reproduction numbers as function of episodic risk, which are,

$$T_{H_{Inf}} = R_{HH_e} + \frac{R_{LH_e}R_{HL_e}}{1 - R_{LL_e}} \qquad \text{(equation (3.4.1))}$$

$$T_{L_{Inf}} = R_{LL_e} + \frac{R_{HL_e}R_{LH_e}}{1 - R_{HH_e}} \qquad \text{(equation (3.4.2))}$$

We also obtained R_{**_e} in subsection (3.3.), which are,

$$R_{HHe} = R_{HH}\psi + (1-\psi)Hg_H$$

$$R_{HLe} = R_{HL}\psi + (1-\psi)Hg_H$$

$$R_{LHe} = R_{LH}\psi + (1-\psi)Hg_L$$

$$R_{LLe} = R_{LL}\psi + (1-\psi)Hg_L$$

In above four equations, R_{pq_e} denote the expected number of new infections in p risk phase caused by a case who was in q risk phase at HIV acquisition, while R_{pq} denotes the expected number of new infections in p risk phase caused by a case who stays in q risk phase during the entire infection time. That says, R_{pq} is calculated assuming no episodic risk. Each R_{pq_e} can be expressed in term, $R_{pq}\psi + (1-\psi)Hg_p$, where ψ is the probability that a case transmits HIV infection when this case has not re-selected contact rate since HIV acquisition. Therefore, term $R_{pq}\psi$ calculates the contribution of transmissions that occur during period before first re-selection while $(1-\psi)Hg_p$ calculates contribution of transmissions that occur during period after first re-selection. If R_{pq_e}s are substituted with their expressions $R_{pq}\psi + (1-\psi)Hg_p$, $T_{H_{Inf}}$ and $T_{L_{Inf}}$ can also be written as,

$$T_{H_{Inf}} = R_{HH}\psi + (1-\psi)Hg_H + \frac{\left(R_{LH}\psi + (1-\psi)Hg_L\right)\left(R_{HL}\psi + (1-\psi)Hg_H\right)}{1 - \left(R_{LL}\psi + (1-\psi)Hg_L\right)} \qquad (5.1)$$

$$T_{L_{Inf}} = R_{LL}\psi + (1-\psi)Hg_L + \frac{\left(R_{HL}\psi + (1-\psi)Hg_H\right)\left(R_{LH}\psi + (1-\psi)Hg_L\right)}{1 - \left(R_{HH}\psi + (1-\psi)Hg_H\right)} \qquad (5.2)$$

As shown in equation (5.1) and (5.2), both numbers are complicated function of ψ, so are also complicated functions of risk re-selection rate, ω, since $\psi = \frac{\left(\frac{\beta_1}{\upsilon_1+\omega}\right)+\beta_2\gamma_1\frac{\upsilon_1}{(\upsilon_1+\omega)}\frac{\upsilon_2}{(\upsilon_2+\omega)}}{\left(\frac{\beta_1}{\upsilon_1}+\left(\frac{\gamma_1}{\upsilon_2}\right)\left(\frac{\beta_2}{\upsilon_2}\right)\right)}$.

However, we can obtain values of $T_{H_{Inf}}$ and $T_{L_{Inf}}$ under two extreme conditions, ω=0 and ω->∞ and gain insight of how increase of ω alters $T_{H_{Inf}}$ and $T_{L_{Inf}}$ from one level to the other level.

When ω=0, $\psi = 1$, so type reproduction numbers become their counterparts for population without episodic risk,

$$T_H = R_{HH} + \frac{R_{HL}R_{LH}}{1 - R_{LL}} \qquad (5.3)$$

$$T_L = R_{LL} + \frac{R_{HL}R_{LH}}{1 - R_{HH}} \qquad (5.4)$$

When ω->∞, ψ approaches zero, so type reproduction numbers under this extreme condition are,

$$T_{H_{Inf}} = Hg_H + \frac{Hg_L Hg_H}{1 - Hg_L} = \frac{Hg_H}{1 - Hg_L}$$

$$T_{L_{Inf}} = Hg_L + \frac{Hg_L Hg_H}{1 - Hg_H} = \frac{Hg_L}{1 - Hg_H}$$

We denote T_h and T_l as the $T_{H_{Inf}}$ and $T_{L_{Inf}}$ under this extreme condition, respectively,

$$T_h = Hg_H + \frac{Hg_L Hg_H}{1 - Hg_L} = \frac{Hg_H}{1 - Hg_L} \qquad (5.5)$$

$$T_l = Hg_L + \frac{Hg_L Hg_H}{1 - Hg_H} = \frac{Hg_L}{1 - Hg_H} \qquad (5.6)$$

As shown in above equations, heterogeneity of model cases' contact rates is not reflected in T_h and T_l at all. Instead, a model case's transmission potential is that expected for corresponding population with homogeneous contact rate, H. The only variables that reflect heterogeneity of contact rates are g_H and g_L, which are expected fraction of contacts contributed by susceptible individuals in high risk phase or susceptible individuals in low risk phase at disease-free equilibrium. Episodic risk does not affect them for two reasons: firstly, calculation of type reproduction numbers assumes a case is introduced into a fully susceptible population at disease-free equilibrium; secondly, no matter what is the risk re-selection rate, susceptible

individuals always re-select contact rate from population distribution at disease-free equilibrium. In addition, assortative mixing (quantified by parameter m) is not reflected in T_h and T_l either. As discussed in section 3., this is because as a model case periodically re-selects contact rate from population distribution, so the increase in contacts with group of susceptible people due to assortative mixing when this case is in one risk phase will be compensated by the decrease in contacts with the same group of susceptibles after this case transits to the other risk phase.

As shown Fig.III-5, as ω increases from zero towards very high level, $T_{H_{Inf}}$ gradually change from T_H to T_h while $T_{L_{Inf}}$ gradually change to T_l (panel (B) of Fig.III-5, T_L is not plotted because when ω=0, high risk subpopulation can sustain ongoing transmission alone, so T_L is not epidemiological meaningful).

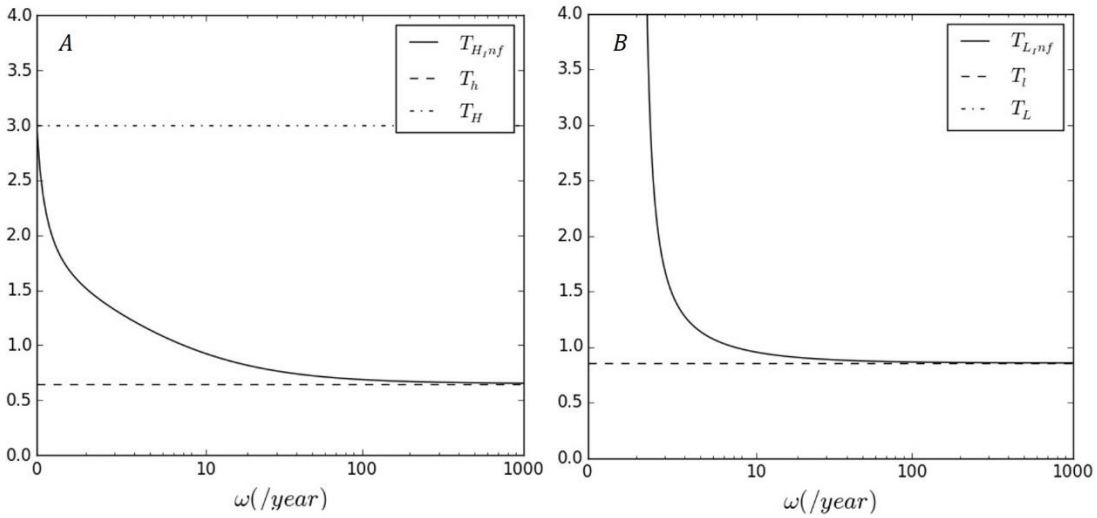

Figure III- 5. Effect of increasing risk re-selection rate, ω, on type reproduction numbers. In this simulation, $\beta_1 = 0.0548/contact$, $\beta_2 = 0.003/contact$, m=0.5, $\chi_H = 193/year$ and $\chi_L = 28/year$. Other parameter are set at default values as shown in Table III-2.

In Section 4, we discussed that the critical fractional reduction in acquisition risk of susceptible individuals experiencing high risk to eliminate HIV infections, δ_H, is a simple function of $T_{H_{Inf}}$,

$$\delta_H = 1 - \frac{1}{T_{H_{Inf}}} \qquad \text{(equation (4.1))}$$

While the critical fractional reduction in acquisition risk of susceptible individuals experiencing low risk to eliminate HIV infections, δ_L, is a simple function of $T_{L_{Inf}}$,

$$\delta_L = 1 - \frac{1}{T_{L_{Inf}}} \qquad \text{(equation (4.2))}$$

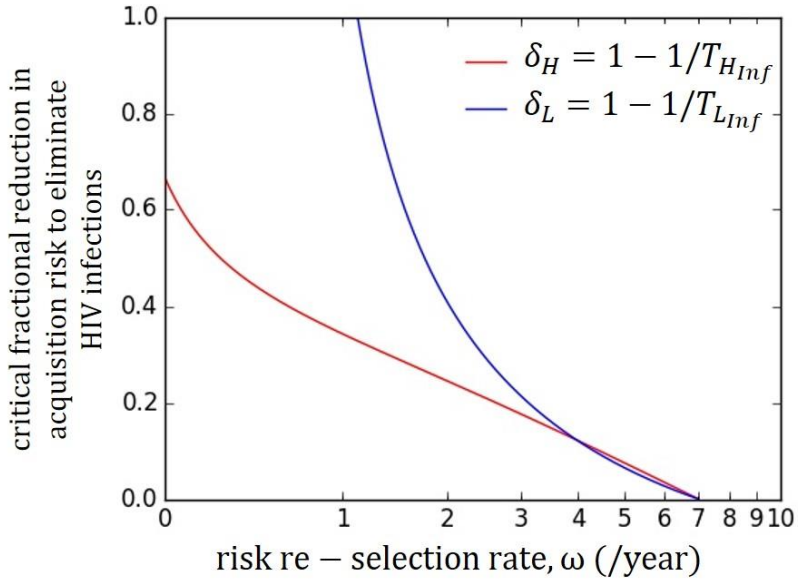

Figure III- 6 **Effect of increasing risk re-selection rate, ω, on critical fractional reduction in acquisition risk of susceptible individuals experiencing high risk, δ_H, to reach elimination and critical fractional reduction in acquisition risk of susceptible individuals experiencing low risk, δ_L, to reach elimination. In this simulation, $\beta_1 = 0.0548/contact$, $\beta_2 = 0.003/contact$, $\chi_H = 193/year$ and $\chi_L = 28/year$, m=0.5. Other parameters are set at their default values as shown in Table III-2.**

Figure III-6 illustrates how increasing risk re-selection rate, ω, changes the critical fractional reduction in acquisition risk of susceptible individuals experiencing high risk to eliminate HIV infections, $\delta_H = 1 - 1/T_{H_{Inf}}$, and its counterpart for susceptible individuals experiencing low risk, $\delta_L = 1 - 1/T_{L_{Inf}}$. As shown in Figure III-6, when ω is zero or relatively low δ_L does not show up. This is because when there is slow risk fluctuation, cases who were infected in high risk phase at HIV acquisition has high probability of staying in high risk phase to transmit HIV infections and have adequate potency to sustain ongoing transmission alone. When ω reaches a moderate level (right above 1/year), δ_L becomes epidemiologically meaningful. This is because cases who were in high risk phase at HIV acquisition have relatively small probability of staying in high risk phase, and cannot sustain ongoing transmission alone. When ω is relatively high, δ_H and δ_L are close to each other (Fig.III-6). This is because when there is fast risk fluctuation, cases change risk levels shortly after HIV acquisition. During period of risk fluctuation model cases periodically select contact rate from population distribution,

regardless of risk phase this was case in at HIV acquisition. Therefore, the risk phase that a case was in at HIV acquisition has limited impact on cases' potency of causing onward transmissions. In this condition, the fraction of transmission to susceptibles experiencing high risk that needs to be blocked to reach elimination, δ_H, is more similar as the fraction of transmission to susceptibles experiencing low risk that needs to be blocked to reach elimination, δ_L.

Discussion of Results in Chapter III

In this chapter, we derived type reproduction numbers for a model with individual risk behavior volatility, i.e. episodic risk. We derived their formulations with both next generation matrix approach introduced by Roberts and Heesterbeek (2003) and based on their epidemiological meanings.[80] We showed that the formulations derived with two approaches match each other.

We demonstrated that increasing individual risk behavior volatility greatly changes type reproduction numbers and the critical control efforts of targeted controls that reduces acquisition risk to eliminate HIV transmissions. When there is zero or low level of risk volatility, elimination may not be achieved by targeting susceptible individuals experiencing low risk. By contrast, when there is high level of risk volatility, it requires similar fractional reduction in acquisition risk of susceptible individuals experiencing high risk and that of susceptible individuals experiencing low risk to reach elimination (Fig.III-6). The mechanism is that as risk volatility increases, a case's potency of causing onward transmissions less depends on the risk phase a model case was in at HIV acquisition. When risk volatility is large, model cases have large probability of changing risk levels shortly after their HIV acquisition, and periodically select contact rate from the population distribution at disease free equilibrium. During period of risk fluctuation, a model case's expected contact rate becomes population average contact rate at disease free equilibrium, which is irrelevant of his contact rate at HIV acquisition.

This is the first time that type reproduction number is formulated for model with individual risk behavior volatility. In Henry and Koopman (2015), R_0 for model with episodic risk is formulated and discussed. A concept "remaining risk heterogeneity effect" is introduced in their study, which is defined as the expected fraction of transmission potential attributed to the

period before a case's first risk transition since HIV acquisition. In this chapter, we derived a quantity which is formulated in same way as the "remaining risk heterogeneity effect". We define it as the probability that a case transmit HIV infections before changing risk level for the first time since HIV acquisition. We found that the two quantifies are essentially the same. In another paper by Romero-Severson et al (2014), R_0 is formulated for a model with risk volatility assuming continuously distributed individual risk states.[48] Both study suggest that risk volatility can reduce the impact of risk heterogenetiy on the transmission system, which agree with what we observe in this chapter.

In order to illustrate calculation of state probabilities and type reproduction numbers, we used two diagrams: transmission diagram and transition diagram. These two diagrams are seemingly similar as the life cycle graph that earlier studies use based on graph theory to derive R_0.[85,86] However, they are fundamentally different. Life cycle graph is an intermediate step towards calculating R_0. Each node of life cycle graph is treated as one trivial node. During the process of calculating R_0, each node is eliminated by joining arcs in and out of it, a process called graph reduction. The process of graph reduction is visualizing the matrix manipulation process that would be less understandable without the graph. By contrast, transmission diagram presented in this chapter is used to visualize the transmission routes that contribute to formulation of type reproduction numbers. No graph reduction is applied and it is not directly related to matrix manipulations either.

We demonstrated that type reproduction numbers derived based on their epidemiological meanings are equivalent to their formulations derived using next generation matrix approach. This is not a surprising finding. This is because each entry of next generation matrix calculates the expected number of new infections in next generation through each transmission route, which

are also the elements that we calculated based on state probabilities. However, formulating type reproduction numbers based on state probabilities is more advantageous than using next generation matrix approach from two perspectives. Firstly, state probabilities are formulated in the way that they can be adapted to model with more stages of infection. By contrast, calculation using next generation matrix will be less efficient as number of stage of infection increases. This is because that dimensionality of transmission matrix F and transition matrix Σ are determined by the number of stages of infection and risk levels. Secondly, the properties of state probabilities and the rules by which they are formulated (section (3.1.) and section (3.2.) in this chapter) help clarify the effect of risk volatility on type reproduction numbers which would not be revealed by next generation matrix.

CHAPTER IV

Effect of Episodic Risk on the Minimum Effectiveness of Universal Test and Treat and Pre-exposure Prophylaxis to Eliminate HIV Infections

Introduction

Despite increasing efforts to control transmission of HIV, the incidence of HIV infection remains stable in the U.S.[2,87] Men who have sex with men (MSM) are the risk group in which most annual new infections have occurred. This is the only risk group where incidence is still increasing.[87–89]

Two biomedical strategies that use antiretrovirals have been recognized as promising intervention tools to prevent HIV transmission. One is Pre-exposure prophylaxis (PrEP), where HIV negative people use chemoprophylaxis before exposure to HIV. [19] PrEP can efficaciously protect individuals from acquisition of HIV infection.[22] However, PrEP shows great variation in effectiveness: low adherence and reduction in other ways of protection can reduce the effectiveness of PrEP.[23,24]

Another is treatment as prevention, where treatment of HIV with antiretrovirals prevents onward transmission. Among treatment as prevention strategies the most promising and debated strategy is universal test and treat (UT&T), which is to get every individual tested and quickly treated if he or she is tested HIV positive. Previous study has suggested that testing every individual annually and initiating treatment immediately once that individual is tested HIV positive can possibly eliminate HIV infections.[8] Later studies further explore and discuss the promise of UT&T in eliminating HIV infections in various epidemiological context. [10–19] These studies indicate that effect of this optimal strategy might be hindered by many factors such as insufficient testing, failed linkage to care and decreased retention in care (frequent dropout). [10–19]

Simulation studies show that episodic risk, a simplified individual risk behavior volatility where individuals alternate between high and low risks, can dramatically change level of endemic prevalence of HIV and fraction of transmission from acute infection.[90] Furthermore, a mathematical model analysis shows that the effective treatment rate needed to reach elimination can considerably differ when individuals change their risk behavior at different rates.[82] However, the effect of episodic risk behavior on individual effectiveness of HIV intervention required to reach elimination remains understudied.

Wise allocation of control efforts is an essential step towards elimination of HIV. It is believed that prioritizing PrEP efforts to people with high behavior risk is an optimal strategy compared to randomly distributing efforts among the whole population. For example, studies have shown that concentrating PrEP efforts more on people with the highest risk can result in greater averted HIV new infections than randomly distributing similar amount of effort to the whole population.[38,91] In addition, studies suggested that focusing PrEP efforts on high-risk susceptible people would be more cost-effective than general PrEP strategy.[38,92] However, these analyses assume that a susceptible individual experiencing high risk would remain at the same risk level. That entails the assumption that the amount of downstream transmission a susceptible individual could cause once infected depends only on current risk behavior and not on future risk changes. Our analysis shows why and how this assumption leads to erroneous inferences.

Motivated by these unanswered research questions, we pursue three study aims in this chapter. First, we seek to understand how and why individual risk behavior volatility changes the required individual effectiveness of PrEP or UT&T to reach elimination. Second, we seek to understand how and why individual risk behavior volatility differentially changes the required individual effectiveness to reach elimination when PrEP is focused on susceptible individuals who are experiencing high risk and when PrEP is allocated to the general population. Third, we analyze how the behavioral or biological characteristics of HIV transmission cause differences in how public health agencies should implement PrEP or UT&T.

Methods

Formulation of the Episodic-Risk Model with Universal Test and Treat

We build the model with process of universal test and treat on the basis of model described in Chapter III. This is a model that explores effect of episodic risk on required effectiveness of

UT&T to eliminate HIV infection, and was introduced by Henry and Koopman (2015).[82] In this model, the operation of test and treat strategy is abstracted as the process that all infected people are tested, recruited and effectively treated at certain rate. We assume that people who are effectively treated have their infectiousness permanently reduced to zero.

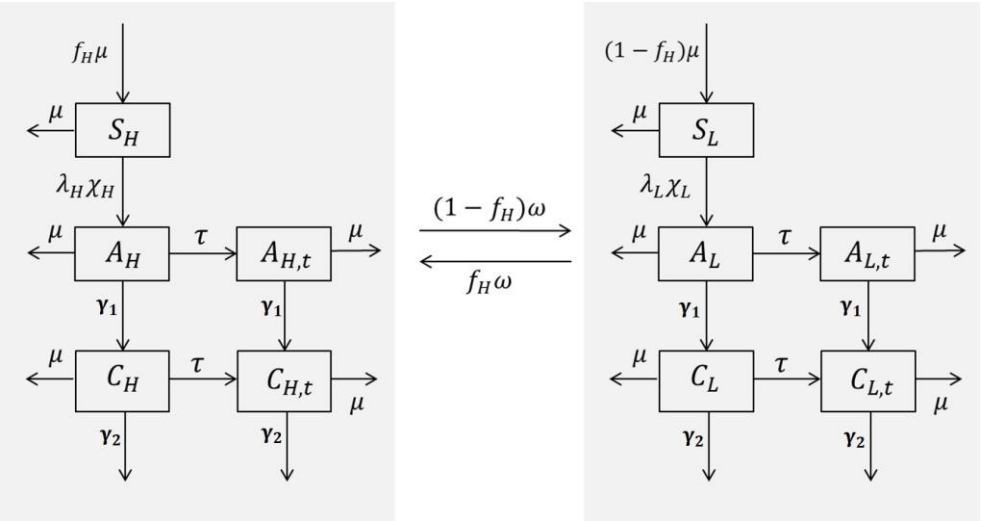

Figure IV- 1. Conceptual model with contact structure and population composition for analysis in Chapter IV

A schematic of the model is shown in Fig.IV-1. The letters S, A, and C in the compartment labels denote compartments containing subpopulations that are susceptible, at the acute stage of infection, and at the chronic stage of infection, respectively. The subscripts H and L denote subpopulations with a high or low contact rate, respectively. We assume that all contacts are instantaneous and have no partnership duration. Compartments with subscript 't' are added to model the process of UT&T treatment. These compartments denote people who are tested and effectively treated. We adopt the way of modeling UT&T in Henry and Koopman (2015), which is to abstract the entire treatment cascade by a single parameter, τ.[82] The parameter τ is defined as the rate that people get effectively treated and have their contagiousness permanently reduced to zero. Therefore, infected people can be categorized as either effectively treated or not effectively treated. The second group includes people who were not tested, tested HIV positive but not linked to care, or linked to care without suppression of virus levels. We assume that people who are not effectively treated have the same contagiousness per contact no matter if they have been tested or not. We also assume that the series of steps from testing to treatment does not induce change in contact rate. The meaning of each compartment is listed in Table IV-1.

Same as in Chapter III, we model assortative mixing by contact rate by allowing individuals to make certain fraction (m) of contact with people in the same risk phase, for both people who are experiencing high risk or and people with low risk.[83] In addition, individuals proportionately mix with each other and make (1-m) fraction of their contacts at a general mixing site. If not specified, m is set to zero and random mixing is assumed in default setting.

In Fig.IV-1, arrows between the two patches with grey shade represent the flow of individuals between the two risk groups when they change their contact rates. Same as Chapter III, we use a single risk change parameter instead of these two risk transition parameters to simplify derivation of model variables. Specifically, we adopt the way that Henry and Koopman (2015) models episodic risk. It is to specify a single parameter, varying which can change the extent to which individual risk is episodic. Such parameter, as introduced by Hendy and Koopman (2015), is the rate at which individuals re-select level of risk from the population distribution, namely risk re-selection rate (denoted ω). People re-select risk group at certain rate (denoted ω), with probability of choosing risk group proportional to the fraction of average time people spend in corresponding risk group (f_H for high-risk subpopulation and f_L for low-risk subpopulation).[82] Therefore, the rate of flow from the high-risk to the low-risk subpopulation is $f_L\omega$ and $f_H\omega$ as the rate of flow in the opposite direction.

In Fig.IV-1, arrows from a compartment without subscript 't' to one with a subscript 't' represent flows due to effective treatment. There is a constant rate of initiating sex, which is also the (per capita) rate of leaving the sexually active population due to competing risk, as indicated by the two arrows on the top and six arrows going out from the sides, respectively. Arrows between compartments with the same risk status indicate infection (from S to A) or progression from acute infection to chronic (from A to C). The arrows at the bottom indicate removal from the sexually active population due to AIDS. The definitions and defaults for parameters are given in Table IV-1.

Equations of this model are listed in Table AI-1 in section I-i of Appendix I.

Table IV- 1 Variable symbols, values, units and definitions for Episodic Risk model with Test and Treat in Chapter IV

Compartment	Meaning
S_H	Susceptible subpopulation with high contact rate
S_L	Susceptible subpopulation with low contact rate

A_H	Subpopulation at acute stage of HIV infection with high contact rate
A_L	Subpopulation at acute stage of HIV infection with low contact rate
$A_{H,t}$ *	Subpopulation at acute stage of HIV infection with high contact rate who are effectively treated (continuousness permanently reduced to zero)
$A_{L,t}$ *	Subpopulation at acute stage of HIV infection with low contact rate who are effectively treated (continuousness permanently reduced to zero)
C_H	Subpopulation at chronic stage of HIV infection with high contact rate
C_L	Subpopulation at chronic stage of HIV infection with low contact rate
$C_{H,t}$ *	Subpopulation at chronic stage of HIV infection with high contact rate who are effectively treated (continuousness permanently reduced to zero)
$C_{L,t}$ *	Subpopulation at chronic stage of HIV infection with low contact rate who are effectively treated (continuousness permanently reduced to zero)

*compartments not shown in previous chapters

Table IV- 2 Parameter symbols, values, units and definitions for model in Chapter IV

Parameter	Value	Unit	Definition
μ	1/40	/year	Rate of removal from the sexually-active population unrelated to HIV. Because we set the equilibrium population in the absence of disease to 1, this is also the (absolute) rate of entry of new individuals into the sexually active population
γ_1	4	/year	Rate of transitioning from acute to chronic infection
γ_2	1/10	/year	Rate of death from AIDS during chronic infection
β	variable	/contact	Average per-contact transmissibility across both stages of infection
ζ^*	variable	-	Relative transmissibility of acute stage of infection (ratio of per contact transmissibility during acute infection over per contact transmissibility during chronic infection)
$\boldsymbol{\beta_1}$	$\frac{\beta\zeta(\gamma_1+\gamma_2+\mu)}{\gamma_1+\zeta(\gamma_2+\mu)}$	/contact	Per-contact transmissibility during acute stage of HIV infection
$\boldsymbol{\beta_2}$	$\frac{\beta(\gamma_1+\gamma_2+\mu)}{\gamma_1+\zeta(\gamma_2+\mu)}$	/contact	Per-contact transmissibility during chronic stage of HIV infection
f_H	variable	-	Fraction of population that is at high-risk phase at disease-free equilibrium
f_L	$1-f_H$	-	Fraction of population that is at low-risk phase at disease-free equilibrium
χ	variable	/year	Population average contact rate at disease free equilibrium
r_{HL}	variable	-	Ratio of average contact rate at high-risk phase over average contact rate at low-risk phase

χ_H	$\frac{\chi r_{HL}}{f_H r_{HL} + f_L}$	/year	Average contact rate in high risk phase
χ_L	$\frac{\chi}{f_H r_{HL} + f_L}$	/year	Average contact rate in low risk phase
m	variable	-	Fraction of contacts reserved for people experiencing the same level of risk
τ^*	variable	/year	Effective treatment rate of universal test and treat
ω	variable	/year	Rate of reselecting risk group

*parameters not shown in previous chapters

Model Analysis. The goal of our study is to examine the effect of episodic risk on the minimum effectiveness of pre-exposure prophylaxis (PrEP) or Universal test and treat (UT&T) required to eliminate HIV infections. We specifically examine the scenarios when UT&T is operated alone, when PrEP is operated alone and when UT&T is combined with PrEP.

To maximize conceptual clarity, we first constrained ourselves to a PrEP strategy that either affects the entire susceptible population equally or exclusively affects susceptible population who have high contact rate. We name the former strategy as universal PrEP and the latter as focused PrEP. At this step of analysis, we assume the whole susceptible population is affected by the intervention uniformly. For universal PrEP, we define the individual effectiveness as the fractional reduction in average risk of HIV acquisition per contact. Focused PrEP reduces the risk of HIV acquisition per contact of every individual experiencing high risk equally. We define the individual effectiveness of focused PrEP as a fractional reduction in the risk of HIV acquisition per contact of people who are experiencing high risk. For both PrEP strategies, we assume equal proportionate effects, so that we can derive model relationships that clarify the dynamics behind intervention effects.

To further relate our analysis with public health control in the real world, we relax the above assumptions and allow that PrEP may not cover 100% of the target population. We also assume that given a specific coverage, PrEP can be either randomly distributed among the whole susceptible population, or prioritized to all susceptible individuals experiencing high risk while the remaining efforts are distributed evenly among the rest of population, i.e. susceptible individuals experiencing low risk in our model. We name the former as general PrEP strategy and latter as high-risk-prioritized PrEP. In addition, we assume that prioritization only happens

in terms of coverage. Therefore, we assume either strategy evenly affect the covered individuals. Again, we define the individual effectiveness of PrEP (general or high-risk-prioritized) as the fractional reduction in risk acquisition per covered susceptible individual.

Measure of Minimum Required Individual Effectiveness of PrEP to Reach Elimination

To calculate the minimum required individual effectiveness to reach elimination for universal PrEP strategy, we first calculate the threshold measures which quantify the control efforts of elimination. For population-targeted PrEP, this measure is R_0. It is the expected number of secondary cases generated by a case among a completely susceptible population during his entire period of infection.[66] When $R_0<1$, ongoing transmission cannot be sustained, and elimination is thus achieved. Therefore, for a population with homogeneous contact rate, the minimum required individual effectiveness of population-targeted PrEP is the fractional reduction in the risk of HIV acquisition per partnership of every susceptible individual, δ, that reduces R_0 to 1. By setting that $\delta R_0=1$, We get that,

$$\delta = 1 - \frac{1}{R_0} \qquad (1)$$

We use the formulation of R_0 for episodic risk model which is introduced by Henry and Koopman (2015). For population with random mixing, R_0 for episodic risk model can be expressed as,

$$R_0 = \mathrm{H}\left(1 + \psi \frac{var(\chi)}{\chi^2}\right)$$

where H is the R_0 expected for corresponding population with homogeneous contact rate, χ. $var(\chi)$ is the variance of population contact rates. Variable ψ is named "remaining risk heterogeneity effect" in Henry and Koopman (2015). It calculates the expected fraction of transmission potential attributed to the period before a case's first risk re-selection. The formulation of R_0 that incorporates assortative mixing has also been derived in Henry and Koopman (2015), and that formulation is more complicated. However, after incorporating assortative mixing, the basic structure that "remaining risk heterogeneity effect" ψ controls the extent to which ratio $\frac{var(\chi)}{\chi^2}$ boosts value of R_0 still holds. In Chapter III, we have also derived variable ψ as the probability that a case transmits HIV when this case has not re-selected contact rate yet. Although we derive and discuss this variable as probability, it is essentially the same as

the "remaining risk heterogeneity effect". This is because we both did the formulation and interpretation in the mean field framework.

For PrEP strategy targeted at susceptible people at a specific risk level, we calculate the measure of control efforts to eliminate diseases for interventions targeted at a subset of population. Such measure is introduced by Roberts and Heesterbeek (2003) and is named the type reproduction number.[80] As suggested by Roberts and Heesterbeek (2003), type reproduction number for host of type i, T_i, calculates the expected number of new infections of type i that a type i case generates through all possible transmission chains with no intermediate type i new infection when that case is introduced into a susceptible population. In Chapter III, we have formulated type reproduction numbers for the episodic risk model. In Chapter III, we also found that the minimum required individual effectiveness of an intervention strategy that is targeted at every suscepetible individuals experiencing high risk, δ_H is,

$$\delta_H = 1 - \frac{1}{T_{H_{Inf}}}$$

Where $T_{H_{Inf}}$ is the type reproduction number of population who are in high risk phase at HIV acquisition. Such population include susceptible individuals experiencing high risk (who will enter infection while in high risk phase). Therefore, $T_{H_{Inf}}$ can be used to calculate the minimum required individual effectiveness of PrEP targeted at susceptible individuals experiencing high risk. We include formulation of $T_{H_{Inf}}$ in the section I-ii of Appendix I.

Measure of Minimum Required Individual Effectiveness of PrEP to Reach Elimination Given a Fixed PrEP Coverage

As before, we assume that general PrEP and high-risk-prioritized PrEP may not cover the whole susceptible population. Therefore, we first need to specify the coverage of both strategies, which we denote κ. When general PrEP can reach fraction κ of the susceptible population, the acquisition risk of the susceptible population decreases by fraction $\kappa\delta_g$, where δ_g is the fractional reduction in acquisition risk per covered individual. Therefore, the minimum required individual effectiveness of general PrEP to reach elimination, is the level of δ_u at which R_0 equals 1. By setting $R_0(1- \kappa\delta_g)=1$, we get that $\delta_g=(1-1/ R_0)/ \kappa$.

Furthermore, as before, we assume that high-risk-prioritized PrEP will first cover all susceptible individuals experiencing high risk, and then allocate the remaining efforts randomly among the susceptible individuals experiencing low risk. Therefore, to calculate its minimum required individual effectiveness to eliminate HIV infections, we need to know how many low-risk susceptible individuals can be covered by the remaining PrEP efforts. This requires knowledge of prevalence of susceptible individuals experiencing each level of risk. To simplify the analysis, we assume that epidemic has reached endemic equilibrium before implementation of PrEP. Therefore, we simulate the model until endemic equilibrium and record the fraction of susceptible individuals experiencing high risk or low risk, denoted as p_{sh} and p_{sl}, respectively. Thus, after high-risk-prioritized PrEP allocates effort to all susceptible individuals experiencing high risk, the remaining PrEP efforts can cover $\kappa - p_{sh,}$ of the susceptible population. This results in $(\kappa - p_{sh})/ p_{sl}$ fraction of susceptible individuals experiencing low risk who are covered. With coverage of each susceptible subpopulation specified, we find the fractional reduction in risk acquisition per covered susceptible individual that reduces R_0 to 1, δ_{hp}. Given that two subpopulations are unevenly affected, δ_{hp} cannot be expressed as a concise function of reproduction numbers. Therefore, we resort to numerical simulation to solve it.

Note that general PrEP and high-risk-prioritized PrEP become less different as the coverage of PrEP increases. This is because with high coverage, high-risk-prioritized PrEP can also cover most susceptible individuals who are experiencing low risk. These two strategies are same when PrEP can cover all susceptible individuals.

The target population and minimum required individual effectiveness of the four PrEP strategies (universal, focused, general and high-risk-prioritized) are summarized in Table IV-3.

Table IV- 3 Type of PrEP Strategies, corresponding target population and minimum individual effectiveness to eliminate HIV infections examined in Chapter IV

Type of PrEP strategies	Target population	Minimum individual effectiveness to eliminate HIV infections
Universal PrEP	The whole susceptible population	$1-1/R_0$
Focused PrEP	Susceptible people who are experiencing high risk	$1-\frac{1}{T_{H_{Inf}}}$
General PrEP	Randomly selected susceptible individuals with *given coverage of*	$(1-1/R_0)/\kappa$*

	PrEP	
High-risk-prioritized PrEP	All susceptible individuals who are experiencing high risk and randomly selected susceptible individuals who are experiencing low risk with *given coverage of PrEP*	Numerical simulation

*coverage of PrEP strategy

Results

Results are divided into five sections. The first section presents the results when UT&T is implemented alone. The second section presents results when universal PrEP is implemented alone. The third section presents that how episodic risk differently affect universal PrEP and focused PrEP when they are implemented alone or combined with UT&T. The fourth section presents the results for the scenario that general PrEP or high-risk-prioritized PrEP are combined with UT&T. The fifth section presents how some model parameters changes effects of episodic risk on the individual effectiveness required to reach elimination of PrEP and UT&T.

Effect of Episodic Risk on Minimum Required Effective Treatment Rate of Universal Test and Treat to Reach Elimination

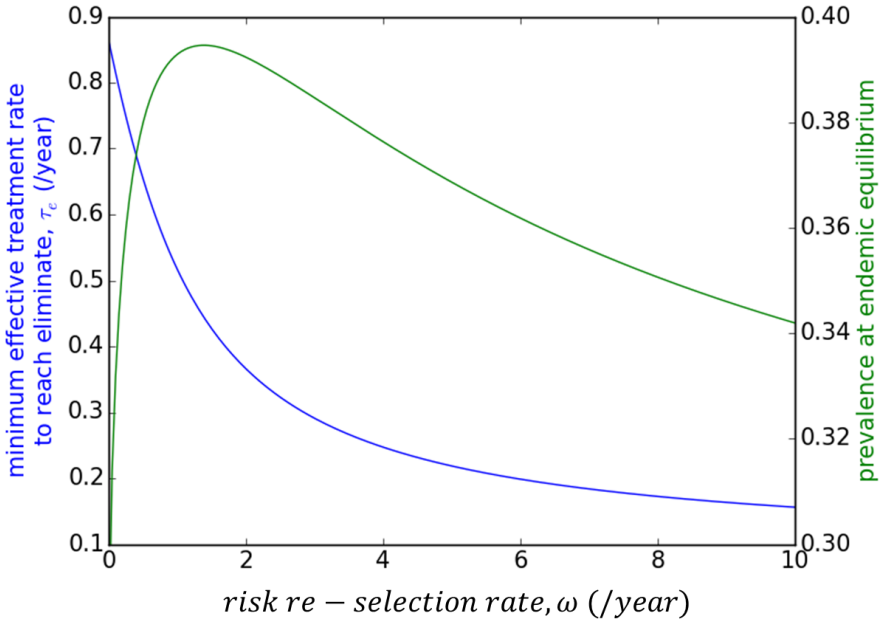

Figure IV- 2 Effect of risk re-selection rate, ω, on minimum required effective treatment rate of universal test and treat to eliminate HIV, τ_e.

As shown in Fig.IV-2, episodic risk has different effects on endemic prevalence (green curve) and the minimum effective treatment rate to reach elimination, τ_e, (blue curve). As the rate at which individuals re-select their risk behavior category, ω, increases, endemic prevalence first increases and then decreases while τ_e decreases greatly. To understand the different effect of episodic risk on τ_e and endemic prevalence, first consider how episodic risk changes R_0. As suggested by Anderson and May, for a homogeneous population, the effective contact rate is the average contact rate.[66] As contact rate becomes heterogeneous, the effective contact rate that contributes to R_0 becomes the sum of average contact rate and ratio of variance in contact rate over average contact rate.[66] Therefore, risk heterogeneity increases R_0 from its expected value for a homogeneous population. However, the Anderson and May analysis assumes that cases do not change their risk levels. When individuals change risk status, there is a decreased correlation between the current risk level of cases and their risk level at time of HIV acquisition. Hence, the effect of risk heterogeneity that increases R_0 decreases.[82] The faster individuals change risk level, the less likely cases are to stay at the risk level where they acquired HIV, and the less risk heterogeneity boosts value of R_0. Consequently, at higher value of ω, R_0 is lower and a lower effective treatment rate is needed to reduce R_0 to 1.

Although the effects of this reduction on the endemic prevalence are somewhat more complicated, they are broadly similar, with the result that a sufficiently high risk re-selection rate can reduce the endemic prevalence as well. The second aspect of episodic risk on the transmission system is that when an individual's risk alternates between high and low levels over time, high-risk susceptible people who have become infected can be replaced by new susceptible people transitioning into the high-risk group from the low-risk group. In contrast to a population with static contact heterogeneity, this acts as an additional source of new high-risk susceptible people and alleviates the saturation of infection among high-risk subpopulation. The more volatile risk behavior is, the more the infection saturation is alleviated, and the less endemic prevalence is suppressed due to contact heterogeneity. Therefore, when ω is relatively low, increase in ω results in reduced infection saturation which boosts level of endemic prevalence. As ω becomes sufficiently high, it causes noticeable reduction in overall transmission potential

and reduces the endemic prevalence. As result, endemic prevalence first increases and then decreases when ω increases from zero to 10/year.

Effect of Episodic Risk on Minimum Required Individual Effectiveness of Universal PrEP to Reach Elimination

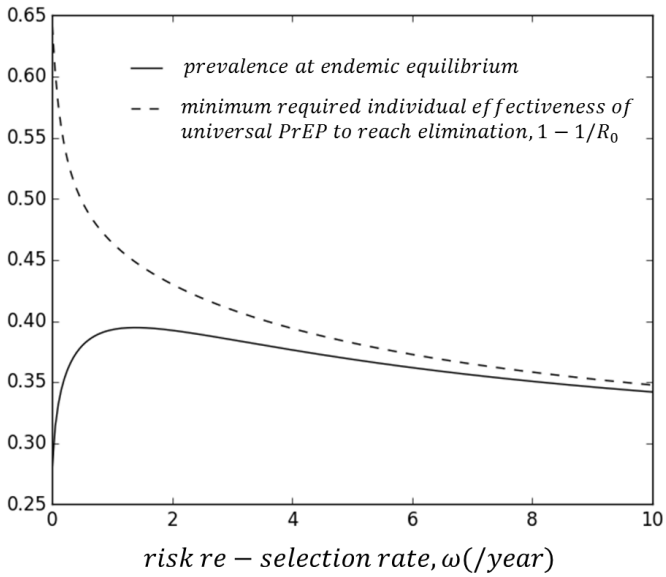

Figure IV- 3 Effect of risk re-selection rate, ω, on minimum required individual effectiveness of universal PrEP strategy to eliminate HIV, $1-\frac{1}{R_0}$, and level of prevalence at endemic equilibrium.

Fig.IV-3 illustrates how episodic risk differently changes the minimum individual effectiveness of universal PrEP to reach elimination, $1\text{-}1/R_0$ (dashed curve) and prevalence at endemic equilibrium (solid curve). As shown in Fig.IV-3, when ω is low, $1\text{-}1/R_0$ is relatively high but endemic prevalence is relatively low. When ω reaches about 10/year, these two quantities become almost equal. To understand such phenomena, consider how episodic risk affects the transmission system. As mentioned, episodic risk reduces the extent to which risk heterogeneity boosts R_0. On the other hand, episodic risk changes endemic prevalence from two perspectives: it alleviates infection saturation within the high-risk subpopulation and reduces the extent to which risk heterogeneity increases overall transmission potential. Therefore, essentially episodic risk reduces the impact of risk heterogeneity on the transmission system. Consequently, as ω increases endemic prevalence and $1\text{-}1/R_0$ both converge to their expected levels for a homogeneous population. In a homogeneous population, the level of endemic prevalence, p,

equals the minimum coverage of vaccination to reach elimination i.e. p=1-1/R_0. [45] Consequently, p and 1-1/R_0 gradually approach each other when ω increases.

Effect of Episodic Risk on Minimum Required Individual Effectiveness of Focused PrEP Strategy to Reach Elimination.

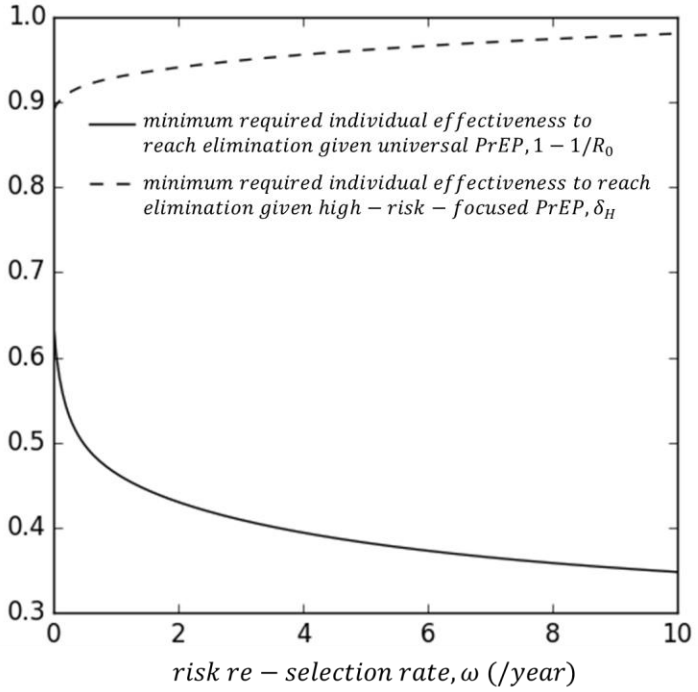

Figure IV- 4 Effect of risk re-selection rate, ω, on minimum required individual effectiveness to eliminate HIV of high-risk-focused PrEP strategy (dashed curve) and universal PrEP strategy (solid curve).

Fig.IV-4 illustrates how change in risk re-selection rate, ω, differently alters the minimum individual effectiveness to reach elimination of focused PrEP and universal PrEP. As shown in Fig.IV-4, as ω increases the minimum individual effectiveness of high-risk focused PrEP needed to eliminate HIV infection, δ_H, increases while the minimum individual effectiveness of low-risk universal PrEP needed to eliminate HIV infection, 1-1/R_0, decreases. When there is no episodic risk, i.e. ω=0, 1-1/R_0 is about 0.64 while δ_H is about 0.9. This indicates that when ω=0, HIV infections can be eliminated by 64% reduction in acquisition risk of every susceptible individual or by 90% reduction in acquisition risk of susceptible individuals who are experiencing high risk. In other words, 90% of susceptible individual experiencing high risk contributes 64% of overall transmission potential. As ω increases, δ_H becomes increasingly higher than 1-1/R_0. This indicates that a higher percentage of susceptible people experiencing high risk contributes a

lower percentage of overall transmission potential. That says, as ω increases each susceptible individual who is experiencing high risk has decreased contribution to overall transmission potential. The mechanisms behind these effects can be appreciated by realizing that when there is episodic risk, i.e. nonzero ω, cases only stay at the risk level at their HIV acquisition until they re-select a new risk level and cycle between risk states (as illustrated in Fig. IV-1). During this period, all cases re-select contact rates from the population distribution (high contact rate at probability f_H and low contact rate at probability f_L) at the rate ω, regardless of their contact rates at HIV acquisition. Therefore, cases who were at high risk at HIV acquisition only have elevated transmission potential during the period before first risk re-selection. As ω increases, cases spend less time before first risk re-selection. As result, the potential contribution of susceptible individuals experiencing high risk to ongoing transmission less surpass that of susceptible individuals experiencing low risk. As result, increase in ω reduces the potential contribution of high risk susceptible individuals to ongoing transmission.

Effect of Episodic Risk on Minimum Required Individual Effectiveness of PrEP to Reach Elimination-Combined With Universal Test and Treat

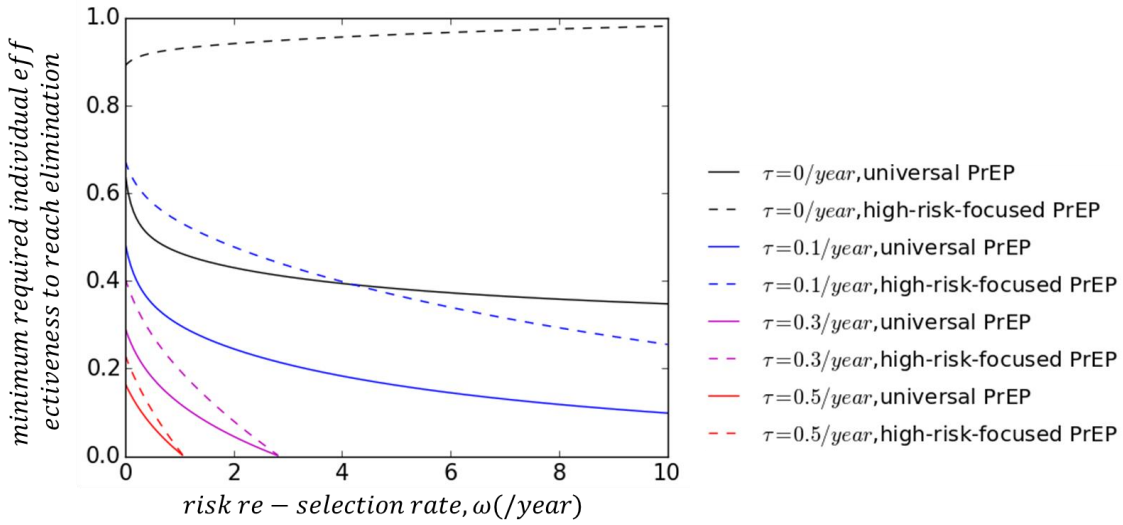

Figure IV- 5 Effect of risk re-selection rate, ω, on minimum required individual effectiveness of universal PrEP to eliminate HIV, $1-\frac{1}{R_0}$, and effect of ω on minimum required individual effectiveness of high-risk-focused PrEP, 1-1/T_H, at different effective rates universal test and treat, τ.

Fig.IV-5 shows that as the effective treatment rate, τ, increases from 0 to 0.5/year, it considerably lowers the minimum individual effectiveness needed to reach elimination of the universal PrEP, $1-\frac{1}{R_0}$, and the minimum individual effectiveness needed to reach elimination of the focused PrEP, δ_H. That is because UT&T strategy shortens the duration of infectious period and reduces overall transmission potential. Therefore, increase in τ causes both quantities to drop to zero at a lower value of ω, beyond which elimination can be achieved by UT&T alone.

In addition, Fig.IV-5 shows that the effect of episodic risk (ω) on the minimum individual effectiveness needed to reach elimination of both PrEP strategies ($1-\frac{1}{R_0}$ and δ_H) is strong at all levels of τ. However, as τ increases, the gap between $1-\frac{1}{R_0}$ and δ_H decreases. As discussed for Fig.IV-5, this gap is determined by the contribution of high risk susceptible individuals to ongoing transmission: a higher contribution leads to a smaller gap. Therefore, this indicates at higher τ episodic risk less reduces the contribution of high risk susceptible individuals to overall transmission potential. This is because UT&T acts as a competing cause for cases to leave infection period. As the effective treatment rate, τ, increases, cases are more likely to be treated before their first risk re-selection. This causes episodic risk to have smaller impact on the transmission system.

Effect of Episodic Risk on Minimum Required Individual Effectiveness of PrEP to Reach Elimination When Combined with Universal Test and Treat Given Fixed PrEP Efforts

In previous sections, our analysis has been focused on the scenarios when PrEP universally affects everyone or only affects susceptible individuals experiencing high risk. However, they are not comparable because they require very different amount of efforts. In the real world, choosing an optimal strategy is constrained by amount of available control efforts. Therefore, to incorporate the control efforts, we define two additional PrEP strategies: general PrEP and high-risk-prioritized PrEP. As mentioned, the former is PrEP strategy that randomly reach and affect susceptible individuals, and the latter is PrEP strategy that first cover all susceptible individuals experiencing high risk and randomly distribute the rest of efforts among the rest of susceptible population (who are experiencing low risk in our model). When comparing these two strategies,

we fix the coverage of PrEP strategy and compare the required individual effectiveness of either strategy to reach elimination. As mentioned in the method section, we assume that either PrEP strategy has equal proportionate effect on the acquisition risk of covered individuals. For this part of analysis, we resort to numerical simulation to find the required individual effectiveness of either PrEP strategy to reach elimination. As an exemplifying analysis, we first set the coverage at a moderate level: 50% of susceptible population. Then we vary this coverage and see how changing coverage modifies our observation.

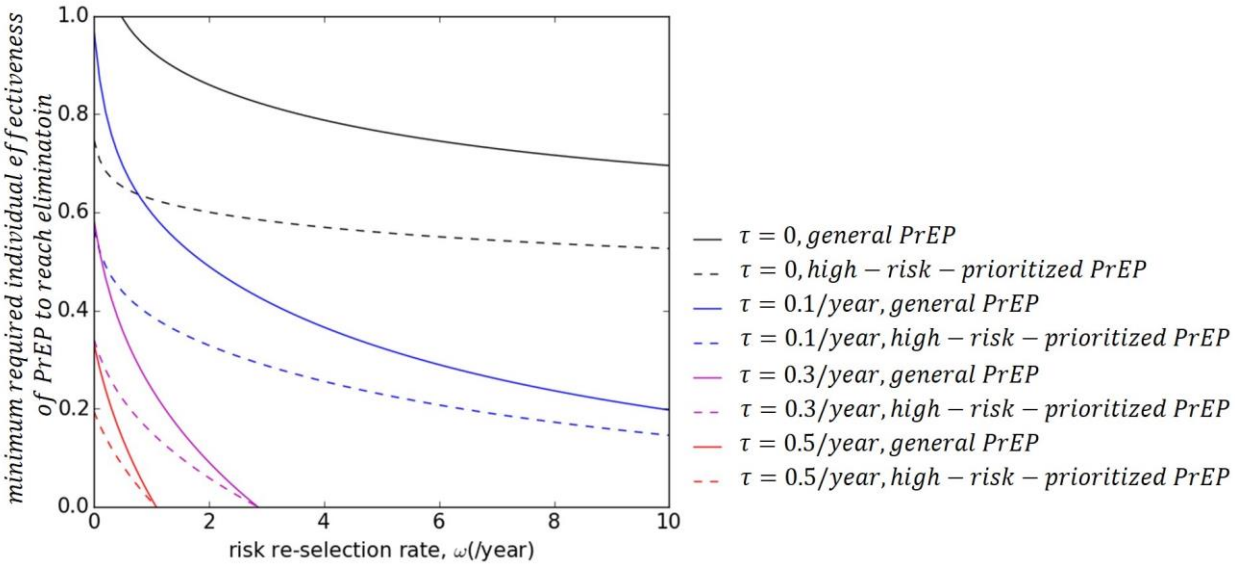

Figure IV- 6 Effect of risk re-selection rate, ω, on minimum required individual effectiveness to eliminate HIV given general PrEP (solid lines) or high-risk-prioritized PrEP (dashed lines), at different effective rates universal test and treat, τ, given that PrEP can cover 50% of the susceptible population.

Fig.IV-6 illustrates that at all levels of UT&T operation, change in risk re-selection rate, ω, can greatly change the minimum individual effectiveness to reach elimination of general PrEP and high-risk-prioritized PrEP, respectively, when PrEP can cover 50% of susceptible popoualtion. In addition, at every level of τ, increase in ω reduces the difference in the minimum required individual effectiveness to reach elimination between the two PrEP strategies (compare solid line with dashed line with the same color in each panel of Fig.IV-6). This indicates that for epidemics with higher ω prioritizing PrEP efforts to susceptible individuals experiencing high risk makes smaller difference in the minimum individual effectiveness to reach elimination. The reason was discussed for results shown in Fig.IV-5: when there is episodic risk, cases only stay at the risk level at their HIV acquisition before their first risk re-selection. Therefore, the reduces

the potential of susceptible individuals experiencing high risk to cause onward transmission. As ω increases, cases spend less time before first risk re-selection, the amount of downstream transmission susceptible individuals experiencing high risk can cause less surpass that of the rest of susceptible population. Therefore, given a fixed amount of PrEP efforts, prioritizing PrEP efforts to high-risk susceptible people or not would make a smaller difference in the required individual effectiveness to reach elimination.

Fig.IV-6 indicates that public health decisions of planning combination of UT&T and PrEP can greatly differ among population with different risk re-selection rate, ω. When ω is relatively low, it is more necessary to combine UT&T with PrEP to eliminate HIV infections (nonzero required individual effectiveness of PrEP to reach elimination). Furthermore, PrEP efforts should be prioritized to susceptible people experiencing high risk since it will considerably lower the required individual effectiveness to reach elimination (compare dashed curves with solid curves at left side of Fig.IV-6). By contrast, when ω is relatively high it is more likely that UT&T alone can eliminate HIV infections (red curves or magenta curves disappear at right side of Fig.IV-6 because at this range of ω UT&T alone can reach elimination when effective treatment rate is or above 0.3/year). In addition, when UT&T is not sufficient to eliminate HIV infections, i.e. τ=0.1/year, it can be combined with either high-risk-prioritized PrEP or general PrEP since two strategies require similar level of individual effectiveness to reach elimination (such as part of blue lines on the right side of Fig.IV-6). Choosing the best combination strategy in the real world also involve evaluation of cost and other types of input. However, the results in Fig.IV-6 indicate that impact of episodic risk should not be ignored in the early stage of planning public health intervention.

In section II-i of Appendix I, we vary the PrEP coverage from 5% to 90% and repeat the above analysis. Our results indicate that the fact that increase in ω reduces the difference in the minimum individual effectiveness to reach elimination of the two PrEP strategies does not change as we vary the coverage of PrEP (Fig.AI-1). However, such effect is less pronounced when the PrEP coverage is low or high (Fig.AI-1). This is because when PrEP can only cover a small fraction of susceptible people i.e. 5%, general PrEP miss most susceptible individuals experiencing high risk and has limited impact on overall transmission potential. Therefore, at most values of ω it is impossible to use general PrEP to reach elimination (Fig.AI-1). On the

other hand, when coverage of PrEP is high, i.e. 90%, general PrEP can also affect large portion of susceptible individuals experiencing high risk (Fig.AI-1). Therefore, the minimum individual effectiveness to reach elimination will be similar between the two PrEP strategies, regardless of individual risk behavior volatility.

Sensitivity Analysis. In section II-ii and section II-iii of Appendix I, we examine how changing behavioral or biological characteristics of HIV transmission modify our observations. Our results indicate that generally our conclusion holds as we vary the model parameter settings. However, there are several interesting phenomena worth noting.

Firstly, level of risk heterogeneity and assortative mixing may affect our observations from two perspectives. On one hand, increase in level of risk heterogeneity or degree of assortative mixing causes a stronger mixing within high-risk subpopulations. Therefore, prioritizing PrEP efforts to susceptible individuals experiencing high risk can more efficiently reduce the transmission potential and require a more lowered individual effectiveness to reach elimination than general PrEP. Such advantage is also less affected by individual risk behavior volatility (Fig.AI-8 and Fig.AI-9 in section II-iii of Appendix I). On the other hand, increase of high-to-low contact rate ratio and higher degree of assortative mixing causes risk heterogeneity to play a more important role in determining R_0. Thus, for population with higher contact rate ratio of greater degree of assortative mixing, episodic risk has greater impact on the minimum required individual effectiveness to reach elimination of universal PrEP, $1-1/R_0$, and the minimum required effective treatment rate of UT&T to reach elimination, τ_e (Fig.AI-2, Fig.AI-3, Fig.AI-4 and Fig.AI-5 in section II-ii of Appendix I).

Secondly, if a higher fraction of transmission potential is attributed to acute stage of infection episodic risk will less affect $1-1/R_0$ (Fig.AI-6 in section II-ii of Appendix I). In addition, the benefit that prioritizing PrEP efforts to susceptible individuals experiencing high risk can considerably lower individual effectiveness to eliminate HIV infections is also more robust to individual risk behavior volatility (Fig.AI-10). This is because cases are less likely to re-select risk groups at earlier stage of infection than later stage of infection. Therefore, if more transmission potential concentrate at acute HIV infection episodic risk will have smaller impact on the transmission system.

Discussion of Results in Chapter IV

Summary of Findings

This chapter takes the first step to examine how and why individual risk behavior volatility changes the level of individual effectiveness of PrEP or UT&T to eliminate HIV transmissions.

We have shown that individual risk behavior volatility reduces the effect of risk heterogeneity on a transmission system. As result, increase in individual risk behavior volatility causes R_0 and prevalence at endemic equilibrium to both converge to their values expected for a homogeneous population. Specifically, at low individual risk behavior volatility, R_0 is high while endemic prevalence is low, and this reverses at high individual risk behavior volatility. As result, for system with higher individual risk behavior volatility, it is more likely that elimination can be reached with UT&T alone. By contrast, for system with lower individual risk behavior volatility, elimination may require combination of UT&T and PrEP. We have shown that such impact is stronger when the population has higher level of risk heterogeneity or higher degree of assortative mixing.

We also demonstrated that individual risk behavior volatility can reduce the importance of susceptible individuals experiencing high risk in sustaining ongoing transmission. As result, for system with low individual behavior risk volatility, efforts of PrEP should be prioritized to susceptible individuals experiencing high risk to reduce the required individual effectiveness to reach elimination. However, for system with higher individual risk behavior volatility, prioritizing PrEP efforts to high risk susceptible individuals or not make little difference in required individual effectiveness to reach elimination. We found that such effect gets weaker for population with higher level of risk heterogeneity or higher degree of assortative mixing.

Our study demonstrates that when a greater fraction of transmission potential is attributed to acute HIV infection, individual risk behavior volatility less impacts the transmission system. This is because transmission from earlier stage of infection are less likely to be affected by individual risk behavior volatility. On the other hand, increase in intensity of UT&T operation may also reduce such impact since more cases will be treated before they change risk levels.

Rationale for the Model Simplification and How Greater Realism Might Affect the Major Findings

In order to illustrate the underlying mechanism of effect of individual risk behavior volatility on the transmission system, we use highly simplifying assumptions. First, we assume that individual risk behavior is dichotomous, which makes it possible to examine PrEP strategy focused on a single risk group. In the real world, distribution of individual risk behavior is continuous. Romero-Severson et al (2014) uses a model with continuously distributed risk behavior and found that increase in individual risk behavior volatility can decrease R_0.[48] This agrees with the observation in our model. However, there has not been studies on effect of individual risk behavior volatility on focused strategies for population with more continuously distributed individual risks. We infer that incorporating continuously distributed individual risks may quantitively change our results regarding effect of individual risk behavior volatility on focused PrEP strategies, but would not change our mechanic conclusions.

Secondly, we assume that partnerships are instantaneous events that are randomly formed between individuals instead of ongoing relationships. Such assumption enables us to algebraically analyze effect of individual risk behavior volatility on the reproduction numbers. However, it is unrealistic in two aspects. Firstly, by assuming instantaneous partnerships we exclude the possibility of concurrent partnership. However, partnership concurrency has been observed among MSM population and is found to be a major cause of high HIV prevalence in certain MSM subgroups.[43,93–96] Therefore, allowing partnership to be enduring and concurrent can modify our conclusion. For example, spread of infection may less depend on people with high partner change rate but more on people with high number of concurrent partners. Secondly, partnerships may not form randomly. Studies suggest that factors such as age, race and sero-status can also affect the probability that two individuals form a partnership.[97–99] Therefore, if people have high tendency to choose partner who share a specific characteristic, transmission will be largely confined within such subgroups. Therefore, relaxing the assumption in either aspect may cause risk heterogeneity to have less impact on the transmission system. Thus, individual risk behavior volatility may also have less impact on the transmission system. This needs to be explored in the future study.

Related Research on UT&T and PrEP Effects

Our study differs from previous studies which explore the advantage of focused PrEP from two perspectives. Firstly, previous studies suggest that targeting susceptible individuals experiencing high risk is optimal compared to allocating PrEP to the general population.[25,26,38,91,92,100] The underlying assumption is that all individuals remain at the same level of risk throughout their sexually active life. Therefore, our study shows that this is not true given individual risk behavior volatility. The population impact of high-risk focused PrEP could be overestimated if individual risk behavior volatility exits but is not considered. Secondly, the outcome of PrEP strategies is usually measured by cost-effectiveness or amount of downstream infection prevented per target individual while no single prevention goal is specified. In contrast, we set the prevention goal as elimination and examine the minimum required effectiveness to reach this goal. We argue that the minimum required effectiveness to reach elimination is important to evaluate in addition to the total control efforts because it helps single out the strategies that need a reasonable level of individual effectiveness to eliminate HIV infections.

Studies have been exploring the benefit of combining ART treatment and PrEP in preventing HIV infection.[25,36,37,101–104] Our study provides two new perspectives for this field of research. Firstly, we show that risk volatility reduces R_0. Therefore, how necessary it is to implement PrEP in addition to UT&T needs to be assessed with knowledge of individual risk behavior volatility. Secondly, studies using mathematical models indicate that combining PrEP and test and treat strategy is better than using either alone.[35,36,105] This is because they can reach two distinct subpopulations: infected people and susceptible people. However, our study indicates that the synergy between these two could be more complicated given individual risk behavior volatility. Such synergy is that increase in intensity of UT&T operation can cause more cases to be treated before they change risk levels. Therefore, scale-up of UT&T makes the population effect of PrEP strategies to be less affected by individual risk behavior volatility. Such phenomenon has not been discussed in earlier studies. It also implies that evaluation of individual risk behavior volatility may be necessary not only before UT&T implementation, but also during the operation or scale-up of UT&T.

Related Research on Individual Risk Behavior Volatility

Effect of individual risk behavior volatility on transmission of HIV has been explored in several earlier studies. In our earlier work, we used the simulation from a deterministic model to show that episodic risk can greatly change the fraction of transmission from acute infection.[23] A study by Henry and Koopman (2015) use the same model as one used in this chapter and identified that individual risk behavior volatility can considerably alter the required effort of universal test and treat to eliminate HIV infections.[24] Romero-Severson et al (2013) used a stochastic model and found that individual risk behavior volatility can increase the level of endemic prevalence and proportion of acute infectors.[106] Although each study analyzes effect of individual risk behavior volatility on HIV transmission from a unique perspective, they all support the importance of understanding the individual risk behavior volatility on guiding the HIV intervention.

The phenomena that individuals risk behavior fluctuates over time has been observed from empirical data collected among MSM population.[45,46] Factors that possibly induce change in individual risk behavior can be found in real world. For example, episodes of unprotected sexual activity induced by substance use, transition in between periods within partnership and periods without partners, or change in social context of partnerships.[107–109] Any factor or combination of them would cause variation in individual risk behavior over time. Therefore, the assumption that individuals keep their risk behavior constant over the entire sexually active life is unlikely given these factors.

Implications For Public Health Practice

Our study indicates that the ease of test and treat strategy to eliminate HIV infection among MSM population cannot be accurately estimated without investigating individual risk behavior volatility. For example, Granich et al suggested that HIV would be eliminated when everyone is tested annually and immediately treated once diagnosed HIV positive. [8] Another strategy proposed to eliminate HIV infections is 90-90-90 strategy, which suggests that 90% of infected population get tested, 90% of those confirmed HIV positive get treated and 90% of those treated have their infectiousness successfully reduced to zero.[29] This strategy indicates that HIV will be eliminated when about 73% infected people gets successfully treated. However, our study indicates that whether either strategy will work for MSM population cannot be determined unless the level of individual risk behavior volatility is well understood.

Furthermore, our study demonstrates that before planning allocation of PrEP efforts, it is crucial to understand the risk composition of the target population. If individual risk behavior volatility is neglected, heterogeneity due to individual risk behavior volatility will be erroneously attributed to risk heterogeneity at population level. This will lead to overestimation of the population-level effect of targeting susceptible people experiencing high risk. In reality, prioritizing PrEP resource to susceptible people who are experiencing high risk requires additional efforts than randomly assigning PrEP resource. For example, before prioritization it might be necessary to screen HIV negative individuals who are at high risk through risk assessment.[110,111] In addition, promoting use of PrEP during high risk period require adherence monitoring specifically around the high risk episode.[112–115] Therefore, when individual risk behavior volatility exists, the population effect of focused PrEP decreases but demands for other efforts persists. In this case, it becomes more important to evaluate the actual benefits of prioritizing PrEP efforts to people who are experiencing high risk.

It is important to note that our study focuses on population effect of PrEP strategy instead of individual level effect. Therefore, we acknowledge that using PrEP at period of high risk is always an important strategy to reduce one's risk of HIV acquisition.

Potential Effects of Our Simplifying Assumptions

In our model, we assume that if PrEP is targeted at a specific risk phase, it would not affect that individual's acquisition risk after he transits to the other risk phase. How well such assumption applies to PrEP in the real world depends on how PrEP is delivered. There are two types of PrEP based on the time frame of delivery. One is event-driven PrEP, a strategy that HIV negative people use antiretroviral medicine before and after unprotected sexual contact.[116] Another is time-driven PrEP, a strategy that HIV negative people continuously use antiretroviral medicine to maintain risk of HIV acquisition at a low level.[117] Our study can naturally apply to the event-driven PrEP. For our analysis to apply to time-driven PrEP, we need to assume that prevention effect of time-driven PrEP is limited within the period of risk behavior that PrEP is targeted at. This appears to be a valid assumption because effectiveness of PrEP is highly correlated with adherence.[118] Therefore, promoting adherence to PrEP while an individual has a specific level of risk has little impact on that individual's acquisition risk after that individual transits to another risk level. However, several conditions may make this assumption less valid. For example,

individuals may reduce their risk behavior due to increased awareness.[119] That says, once people get PrEP intervention, they may reduce risk behavior as adjustment regardless of whether they have changed risk levels or not. This would increase the population effect of focused PrEP but reduce the extent to which individual risk behavior volatility modifies population effect of PrEP.

Directions for Future Studies

There are several directions in which future research could explore impact of individual risk behavior volatility on HIV transmission in addition to our study. One possible direction would be to explore effect of individual risk behavior volatility on focused controls for population with continuously distributed behavior risk. which could be done with stochastic models. Future research could also explore effect of individual risk behavior volatility on HIV transmission by allowing partnerships to be ongoing and concurrent. In addition, future study may also integrate individual risk behavior volatility into more specific network of MSM population and explore its impact on HIV transmission.

Conclusions

The major message that we want to convey through our study is that individual risk behavior volatility, an understudied phenomenon that has been observed among MSM population, can have dramatic impact on HIV transmission. On one hand, it can alter the required individual effectiveness of UT&T or PrEP strategies to eliminate HIV infections. On the other hand, it can reduce the population effect of prioritizing PrEP efforts to susceptible individuals experiencing high risk. Therefore, it is necessary to evaluate the level of individual risk behavior volatility before planning the allocation of efforts of the two biomedical prevention strategies among the MSM population.

CHAPTER V

Detecting Signal of Individual Risk Behavior Volatility from HIV Phylogenetic Trees

Introduction

Despite increasing efforts to control transmission of HIV, the incidence of HIV infection remains stable in the U.S.[2,87] Men who have sex with men (MSM) are the risk group in which most annual new infections have occurred. This is the only risk group where incidence is still increasing.[87–89]

Making the correct inference of what drives the transmission of HIV among MSM population is the key step before planning intervention strategies. Accurate inference not only relies on sufficient data but also a good understanding of the how each characteristic revealed by the data affects HIV transmission. Given the complexity of HIV transmission, this can only be done with help of mathematical models. Studies using mathematical models have well explored many behavioral characteristics of MSM population that are potential determinant of the intense transmission and high prevalence of HIV among MSM population[43,95,113,120,121]. However, individual risk behavior volatility, a phenomenon that individuals risk fluctuates over time, has been observed among MSM population but is understudied. Studies using mathematical models indicate that individual risk behavior volatility can considerably change R_0, prevalence at endemic equilibrium and contribution of acute infection to ongoing transmission.[48,82,90,106] Furthermore, studies also suggest that individual risk behavior volatility can have strong impact on population effectiveness of HIV intervention. The analysis by Henry and Koopman (2015) indicate that episodic risk, a simplified version of individual risk behavior volatility, can reduce the required effort of universal test and treat strategy to eliminate HIV infection.[82] In our earlier study, we found that individual risk behavior volatility can reduce the population effect of PrEP that is targeted at susceptible individuals experiencing high risk (Chapter IV).

Furthermore, these studies show that risk volatility can still strongly affect HIV transmission when population has fixed heterogeneity of contact rate, same level of stage-specific transmissibility and same degree of assortative mixing.[82,90,106] This indicates that data about biological and behavior characteristic of HIV transmission cannot inform the level of individual risk behavior volatility and its population impact. Consequently, it is imperative to explore what type of data can improve the identifiability of risk volatility. As the amount of HIV sequence data grows, HIV sequence becomes an increasingly important data source for public health interventions. The advantage of HIV sequence data is that evolution of HIV virus occurs on the same timescale of HIV transmission. This makes it possible to infer the transmission dynamics of HIV from HIV sequence data.[122] Mathematical studies suggest that phylogenetic tree can reveal the contact structure of the population.[59,123,124] Phylodynamic studies, a field of study that combine behavioral data, surveillance data and phylogeny to estimate important epidemiological parameters, suggest that phylogeny of HIV sequence can inform the intensity of transmission, and tendency that the transmission occurs in a certain cluster.[50,52,53,59,125]

Motivated by these early findings, we hypothesize that the population impact of individual risk behavior volatility can be detected from HIV phylogeny. In this Chapter, we take the first step to test this hypothesis. Specifically, we aim to answer two research questions in this study: 1) what are the likely mechanisms that individual risk behavior volatility affects HIV phylogeny? 2) how likely that impact of risk volatility on HIV phylogeny can be distinguished from those of some other factors? To keep our analysis focused, we examine factors that are parameterized in our model. Answering the second research question can be seen as the first attempt of exploring the identifiability of parameter of risk volatility given HIV phylogeny. In addition, to illustrate the underlying mechanisms that risk volatility affects HIV phylogeny, we focus on stochastically simulated epidemics and simulated phylogenetic trees. Stochastic simulation enables us to track the history of transmission and the correspondence between the transmission events and process of branching of the phylogenetic trees. This enables us to relate the effect of risk volatility on HIV phylogeny to its impact on the HIV transmissions.

We use two groups of statistics to summarize a phylogenetic tree: tree imbalance statistics and the statistics that quantify the tendency that tree leaves are clustered with each other. Tree imbalance statistics measure the degree to which a phylogenetic tree is asymmetric. Studies

found that the extent to which a phylogenetic tree is imbalanced can be affected by heterogeneity in infectiousness and assortative mixing of the population.[123] On the other hand, pattern that external nodes of phylogenetic tree cluster can also be affected by the transmission dynamics. For example, Volz et al (2009) suggests that distribution of cluster size by different cluster threshold value can also reflect the transmission dynamics that shape a phylogenetic tree.[125] Robinson et al (2013) suggest that phylogenetic tree of agents from population with higher heterogeneity of degree distribution can have greater variation in cluster size and numbers.[124] Both types of statistics are gaining more interest, for that they can provide valuable information of the contact pattern of population.

Methods

In this section, we first present the structure of deterministic model that our simulation is based on. Then we give an outline of how we used its stochastic, individual-based counterpart to simulate epidemics and construct the simulated phylogenetic trees. Lastly, we describe the statistics of phylogenetic tree imbalance and clustering that will be examined in later sections.

Model Structure

This deterministic version of model in this study is firstly introduced in Zhang et al (2012). It is a model to study the impact of risk volatility on HIV transmission dynamics for population with dichotomous risk states. Individuals risk behavior is either at relatively high level or relatively low level. Risk volatility is modeled as that individuals periodically change their risk behavior between the higher level and the lower level, namely episodic risk.

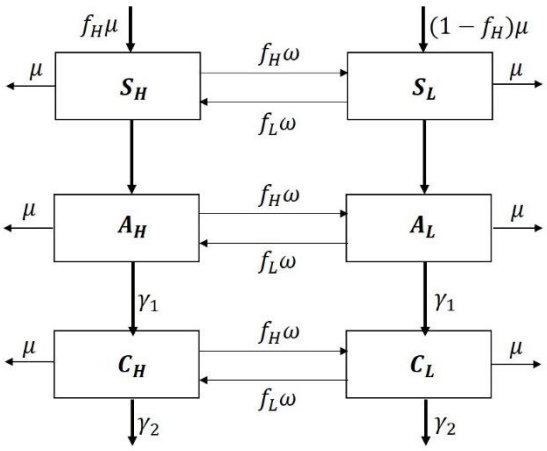

Figure V- 1. Schematic of deterministic episodic risk model

A schematic of the model is shown in Fig.V-1. The letters S, A, and C in the compartment labels denote compartments containing subpopulations that are susceptible, at the acute stage of infection, and at the chronic stage of infection, respectively. The subscripts H and L denote subpopulations with a high or low contact rate, respectively. We model assortative mixing by contact rate by allowing individuals to make certain fraction (m) of contact with people in the same risk phase, for both people who are experiencing high risk or and people with low risk.[83] In addition, individuals proportionately mix with each other and make (1-m) fraction of their contacts at a general mixing site. If not specified, m is set to zero and random mixing is assumed in default setting.

In Fig.V-1, arrows between the two patches with grey shade represent the flow of individuals between the two risk groups when they change their contact rates. We use a single risk change parameter instead of these two risk transition parameters to simplify derivation of model variables. Specifically, we adopt the way that Henry and Koopman (2015) models episodic risk.[82] It is to specify a single parameter, varying which can change the extent to which individual risk is episodic. Such parameter, as introduced by Hendy and Koopman (2015), is the rate at which individuals re-select level of risk from the population distribution, namely risk re-selection rate (denoted ω). People re-select risk group at certain rate (denoted ω), with probability of choosing risk group proportional to the fraction of average time people spend in corresponding risk group (f_H for high-risk subpopulation and f_L for low-risk subpopulation).[82] Therefore, the rate of flow from the high-risk to the low-risk subpopulation is $f_L\omega$ and $f_H\omega$ as the rate of flow in the opposite direction.

Definitions, meanings and units of model parameters are listed in Table AII-2.

Simulation and Construction of Phylogenetic Trees

We used Gillespie algorithm to simulate the stochastic, individual-based counterpart of the episodic risk model (Fig.V-1).[126] At the beginning of each simulated epidemic a single index case who was acutely infected and had high contact rate is introduced into a completely susceptible population who were assumed to be at disease-free equilibrium right before introduction of the index case.

Each round of simulation starts at year zero ends at 10th year of simulation time. Transmission trees were constructed based on the history of who transmit to whom. As the first step of exploring impact of episodic risk on HIV phylogenies, we treated the pruned binary transmission tree as the phylogenetic tree, by assuming that the evolution history perfectly coincides with the transmission history.

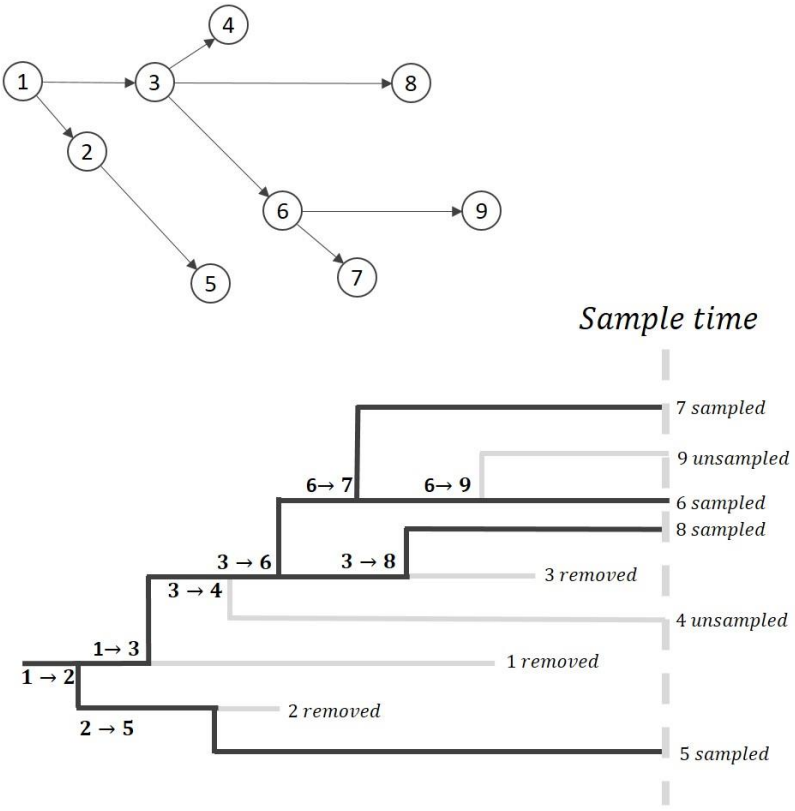

Figure V- 2. Illustration of how a simulated phylogenetic tree is constructed from a simulated transmission tree

Fig.V-2 illustrates how a simulated phylogenetic tree is constructed from a binary transmission tree. Upper part of Fig.V-2 shows an outbreak consisted of nine cases. Lower part of Fig.V-2 shows that the binary transmission tree of the nine cases where each branch corresponds to a new infection, e.g. transmission from case 1 to case 2 generates a new branch. Grey dashed vertical line on the right side of Fig.V-2 marks the time of sampling. Cases who are removed (leave sexually active life) before the sample time or not sampled due to other reasons (cases 1, 2, 3 and 9) do not consist of phylogenetic tree leaves. Therefore, when constructing the phylogenetic tree of the sampled cases, we pruned the branches corresponding to the unsampled cases (marked in color grey) from the transmission tree. The transmission events that result in a branching event in the (simulated) phylogenetic tree are those where the downstream infections (or self) of infector and the downstream infections (or self) of infectee are both sampled. For example, case 6 and case 7 are sampled, so transmission from case 6 to case 7 results in a branching event in the phylogenetic tree. In addition, the downstream infection of case 1, i.e. case 8, and the downstream infection of case 2, i.e. case 5, are both sampled. Therefore, transmission from case 1 from case 2 results in a branching event in the phylogenetic tree.

It is important to note that two necessary conditions are required for us to assume that each new branch of the phylogenetic tree is generated as result of a transmission event. Firstly, evolution of HIV virus occurs at the same time scale of HIV transmission or at a faster speed. Secondly, each host contributes only one unique HIV virus sequence. When these two conditions are met one can infer that when HIV virus is passed onto a new host, it ends up as a unique viral population in the new host, and generates a new branch in the phylogenetic tree. The first condition is met given the fast mutation of HIV virus,[127] while the second condition is satisfied when one consensus sequence is usually collected from each case.

Sampling

The structure of a phylogenetic tree can greatly depend on the sample time. To take this effect into account, we examined two sampling scenarios. One is that cases are collected at the same time, namely homochromous sample. Another is that cases are sampled at different time points, namely heterochronous sample. In the scenario of homochromous sample, cases were randomly sampled at the 10^{th} year of simulation time, while all cases who were sexually active at 10^{th} year of simulation time has equal probability of being sampled, regardless of their risk phases or

stages of infection. In the scenario of heterochronous sample, cases are randomly sampled from 6th to 10th year of simulation time. In addition, sample time points are evenly distributed within this four-year window. At each time point, one case who is sexually active is randomly sampled, regardless of the case's risk phase or stage of infection. We set the sample size to be 500. Therefore, in the scenario of heterchronous sample the sample rate is about one case per three days (sample rate=500/(4*365)=0.342/day).

Tree Imbalance

A binary phylogenetic tree is considered perfectly balanced if each branching results in two branches that lead to clades with equal size. In this study, we focus on two statistics of tree imbalance. One is Sackin index, which is calculated as the sum of number of internal nodes that each tip of a tree needs to traverse back to the tree's root.[128] Specifically, if N_i is the number of internal nodes that ith tip needs to traverse back to the tree root and n is the number of tree leaves, Sackin Index, I_s, is calculated as, $I_s = \sum_{i=1}^{n} N_i$.[128] The formulation reflects two important features of I_s: first, it is not affected by the length of tree branches and is a measure of topology of a phylogenetic tree; Secondly, it is highly dependent on the number of tree leaves. Earlier studies suggest that given a fixed number of tree leaves, a phylogenetic tree with higher I_s is more imbalanced than a tree with a smaller I_s.[129]

In order to control the effect of number of tree leaves on I_s, we normalized I_s by its value expected for a tree randomly generated with a Yule model who has the same number of tree leaves, $E(I_s)$. Such way of normalization has been introduced by Leventhal et al.[130] As suggested by Leventhal et al, the normalized Sakin's Index, I_{ns}, is calculated as,

$$I_{ns} = \frac{I_s - E(I_s)}{E(I_s)}$$

Formulation of $E(I_s)$ is given in Kirkpatrick and Slatkin (1993),[129]

$$E(I_s) = 2\sum_{i=2}^{n} \frac{1}{i}$$

Where n is the number of tree leaves. Yule model is a model of a simple birth-and-death process in which every individual reproduces randomly and has equal probability to give birth to a new individual within a time internal.[131] It is equivalent to an Susceptible-Infected (SI) model where

population has infinite size, while all infectives have equal transmissibility and all susceptible individuals have equal susceptibility.[130] Therefore, I_{ns} measures the extent to which a phylogenetic tree imbalance deviates from its level expected for a SI model for a homogeneous population.

The other statistic of tree imbalance we examined is the number of cherries. A cherry is a cluster consisted of two tree leaves. It reflects the degree of imbalance near the tree leaves: a tree with fewer cherries has more internal nodes with a terminal node as one descendent and a clade as the other group of descendants. This results in greater difference in the size of clade descending from two sibling branches, indicating an increased degree of imbalance. In this study, we adjusted the number of cherries by its maximum given the number of tree leaves. Specifically, if we denote the number of cherries as N_c, the normalized number of cherries as N_{nc}, and the number of tree leaves as n, then, $N_{nc} = N_c/(\frac{n}{2})$ when n is even and $N_{nc} = N_c/(\frac{n-1}{2})$ when n is odd.

Clustering

Cluster is a group of terminal nodes whose pairwise distances are all below a cutoff value. Pairwise distance is calculated as the sum of branch lengths from each terminal node to the most common ancestor of two terminal nodes. Since we constructed phylogenetic trees from the transmission trees, the length of tree branch was measured in calendar time and so are the pairwise distances. For homochromous sample, each pair of terminal nodes have equal distance to their most recent common ancestor, i.e. the first common ancestor that two terminals nodes find along their paths back to root. Therefore, for homochromous sample the pairwise distance of two external nodes is twice the distance from either terminal node to their most recent common ancestor. In this case, clusters can be identified by drawing a vertical line across a tree where the cutoff distance is set, and the group of terminal nodes whose most recent common ancestors all fall on the right side of the line form a cluster (Fig.V-3).

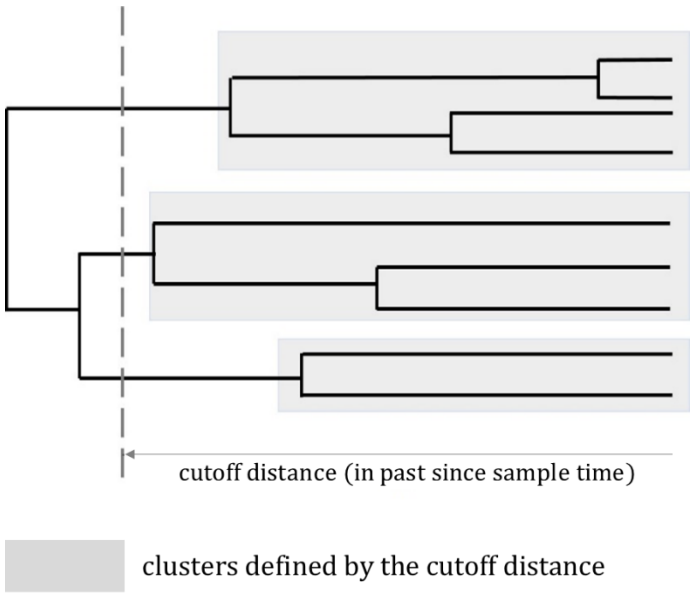

Figure V- 3. Illustration of how clusters are defined for a phylogenetic tree of homochromous sample

The process of finding clusters for heterochromous sample is more complex. For a heterochronous sample, a cluster is defined as the group of terminal nodes among whom the maximum pairwise distance is below a given cutoff distance. When analyzing heterochronous sample, we use hierarchical clustering algorithm to identify clusters at a given cutoff distance.[132]

Results

The results section is divided into four parts, and each part contributes to answering a particular research question as proposed in the introduction section. The first two sections are presented to illustrate the effect of episodic risk on the imbalance and clustering of simulated phylogenetic trees and to interpret the underlying mechanisms. The third section compare the effect of episodic risk on the simulated phylogenetic trees consisted of homochromous sample that its effect on simulated phylogenetic trees consisted of heterochronous sample. The last section presents the comparison between effect of episodic risk and effect of other model parameters on simulated phylogenetic trees. If not specified, the phylogenetic trees for analysis are constructed in the default sampling scenario, which is that 500 cases are randomly selected at 10^{th} year of simulation time.

Effect of Episodic Risk on Tree Imbalance

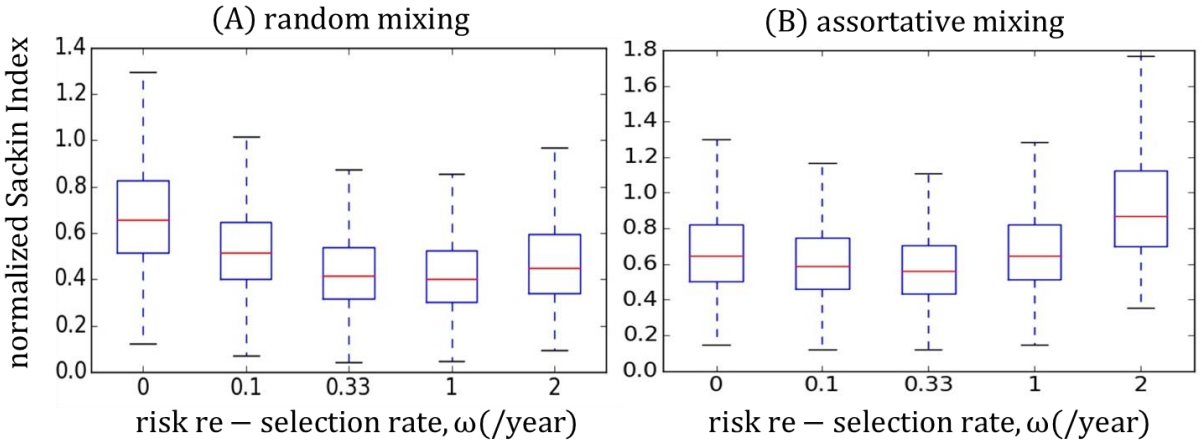

Figure V- 4. Effect of episodic risk on normalized Sackin Index when population has random mixing (panel A) and when population has assortative mixing, i.e. individuals reserve 50% of contacts for people in the same risk phase. When simulating epidemics in random mixing scenario, average transmissibility per contact β=0.008. When simulating epidemics in assortative mixing scenario, average transmissibility per contact β=0.005. Other parameters (except ω) are set at their default values as shown in Table AII-2.

Fig.V-4 shows that episodic risk affects normalized Sackin Index differently for population with random mixing and population with assortative mixing. When population has random mixing, increase in risk re-selection rate, ω, causes normalized Sackin Index to first considerably decrease and then slightly increase (panel (A) of Fig.V-4). In contrary, when

population has assortative mixing, increase in risk re-selection rate, ω, causes normalized Sackin Index to first slightly decrease and then considerably increase (panel (B) of Fig.V-4)

To understand such phenomenon, one can first consider what causes phylogenetic tree imbalance. A perfect balanced binary phylogenetic tree is one in which each internal branch divides into two branches that lead to equally-sized group of tree leaves. In other words, two branches have the same number of descendant tree leaves. Therefore, a phylogenetic tree is imbalanced when there are internal branches that split into branches who have different potency to have descendant leaves. Every time HIV virus is passed from one host to a new host, a unique virus population is built up in the new host. The transmission event would result in a new branch in the phylogenetic tree. That says, the branch that already exists corresponds to the infector while the new branch corresponds to the infectee. If the infector and the infectee cause different amount of onward transmissions after this transmission event, they would also have different downstream infections (descendants). This results in imbalanced HIV phylogenetic tree. In other words, an HIV phylogenetic tree tends to be imbalanced when HIV virus is passed among hosts with different potency to cause onward transmissions.

To understand how episodic risk affects phylogenetic tree imbalance, it is important to understand how episodic risk affects the transmission system. Our earlier mathematical studies suggest that episodic risk increases the transmission from acutely infective in high-risk phase to susceptible individuals in high-risk phase.[90] This is the consequence of the risk fluctuation of both the susceptible population and the infected population. Firstly, given episodic risk, susceptible individuals experiencing low risk can transit to high risk which increases the replenishment of susceptible individuals in high-risk phase. Thus, more high-risk susceptible individuals become infected and increases the pool of high-risk acute infections. By the time that high-risk acutely infected people progress to the less contagious chronic HIV infection, they also pass into low-risk phase. This further increases the contribution from high-risk acutely infected people to the overall transmission potential. The consequence is that there is a greater amount of transmission from high-risk acutely infected people to susceptible individuals experiencing high-risk who also become high-risk acutely infected right after infection. In other words, episodic risk increases tendency that high-risk acute infections are linked with each other in the transmission tree. Consequently, HIV virus are more likely to be passed in between high-risk

acutely infected hosts who on average have same potency of causing onward transmission. From this perspective, episodic risk tends to reduce tree imbalance. We name such effect as the transmission-aspect effect, since it changes the risk levels of hosts that HIV virus is mostly passed in between. In the supplementary materials, we show that phylogenetic trees constructed from epidemics with greater risk re-selection rate, ω, have more branches generated by the transmissions from high-risk acutely infected people to susceptible individuals experiencing high risk (Fig.AII-1). This further confirms our inference of the transmission-aspect effect.

Another perspective in which episodic risk affects phylogenetic tree imbalance relates to risk transition itself. We name such effect as the transition-aspect effect. It is that a case who changes risk phases have a changed potency to cause onward transmissions. If this case's descendants are sampled, this case's risk transition would change a branch's potency to split to new branches. When both infector and infectee are in the same risk phase at the time of transmission, their corresponding branches (sibling branches) have similar potency to split. Under this condition, risk transition of one of the them would make one branch's tendency of splitting into new branches to be different from the other. Thus, risk transition tends to increase the tree imbalance. The transition-aspect effect is more pronounced for population with greater degree of assortative mixing because infector and infectee are more likely to be in the same risk phase at the time of transmission. By contrast, when population has random mixing, there are relatively frequent inter-risk-group transmissions, which limits the transition-aspect effect.

Impact of episodic risk on the Sackin Index is the consequence of the synergy of the transmission-aspect effect and the transition-aspect effect. When risk re-selection rate, ω, is relatively small, risk transition is slow. Cases do not change risk phases until they progress to the less contagious chronic infection, so risk transition has limited impact on a case's potency of transmitting HIV infection. Thus, transition-aspect effect is limited and the transmission-aspect effect dominates, which reduces normalized Sackin Index (left side of panel A and left side of panel B). When ω is relatively large, risk transition is frequent enough to occur within the short period of the highly contagious acute HIV infection. Therefore, transition-aspect effect becomes stronger and increases the normalized Sackin Index (right side of panel A and right side of panel B). As discussed earlier, transition-aspect effect increases as the degree of assortative mixing increases, due to the reduced frequency of inter-risk-group transmissions. Therefore, episodic

risk increases Sackin Index more for population with assortative mixing than for population with random mixing (compare right side of panel B with right side of panel A of Fig.V-4).

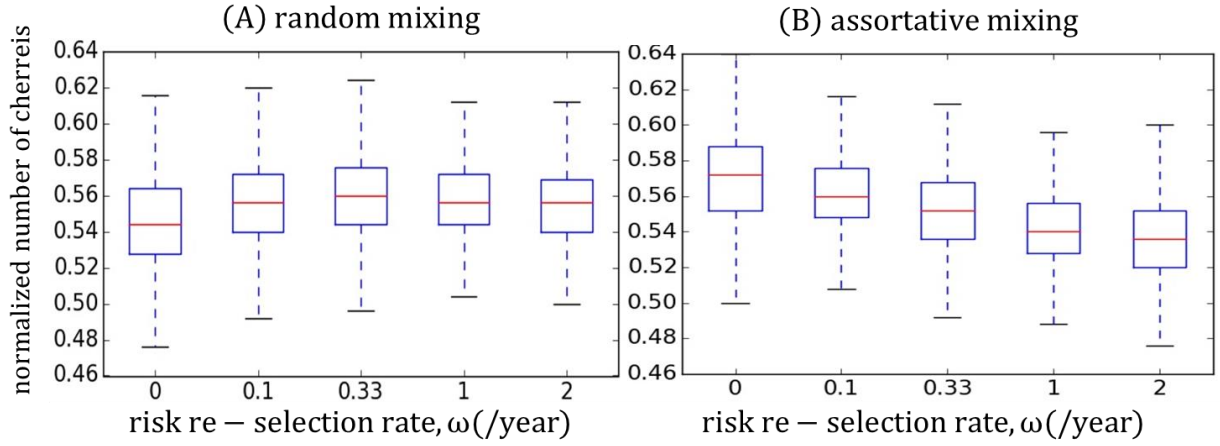

Figure V- 5. Effect of episodic risk on normalized number of cherries when population has random mixing (panel A) and when population has assortative mixing, i.e. individuals reserve 50% of contacts for people in the same risk phase (panel B). For the random mixing scenario, average transmissibility per contact β=0.008. For the assortative mixing scenario, average transmissibility per contact β=0.005. Other parameters (except ω) are set at their default values as shown in Table AII-2.

Fig.V-5 shows that episodic risk affects the normalized number of cherries differently for phylogenetic trees simulated assuming random mixing and phylogenetic trees simulated assuming assortative mixing. When population is randomly mixed, increase in risk re-selection rate, ω, causes normalized number of cherries to first increase and then decrease (panel A of Fig.V-5). By contrast, when population has assortative mixing, increase in ω causes normalized number of cherries to monotonically decrease (panel B of Fig.V-5). As discussed in the method section, trees with larger normalized number of cherries are less imbalanced near tree leaves. Therefore, this suggests that when population has random mixing increase in ω first reduces and then increases the phylogenetic trees imbalance near tree leaves (panel A of Fig.V-5). In contrary, when population has assortative mixing increase in ω monotonically increases tree imbalance near leaves (panel B of Fig.V-5). Effect of episodic risk on the normalized number of cherries can be largely explained by the mechanism through which episodic risk affects the Sackin Index, which measures the overall tree imbalance

However, the comparison between the Fig.V-5 and Fig.V-4 suggests that effect of episodic risk on the two statistics are not totally the same. For example, when population has random mixing, increase in ω considerably reduces normalized Sackin Index (panel A of Fig.V-4), but only slightly increases normalized number of cherries (panel A of Fig.V-5). This suggests

that the effect of episodic risk in reducing tree imbalance, i.e. transmission-aspect effect, is weaker near the tree leaves than on the whole tree. On the other hand, for population with assortative mixing, increase in ω does not increase Sackin Index until it reaches above 0.33/year (panel B of Fig.V-4), but monotonically reduces number of cherries (panel B of Fig.V-5). This suggests that the effect of episodic risk in increasing tree imbalance, i.e. transition-aspect effect, is stronger near tree leaves than on the whole tree. Both indicate that the strength of transition-aspect effect relative to that of transmission-aspect effect increases near tree leaves. This is due to the difference in the nature of the two effects. As discussed earlier, transmission-aspect effect is that episodic risk increases the amount of transmissions linking high-risk acutely infected cases. Thus, such effect can be cumulated over the simulated phylogenetic tree. By contrast, transition-aspect effect is that cases change risk phases. Such effect does not accumulate over the tree. Rather, a case's risk transition can be counterbalanced if this case later transits back. Therefore, transition-aspect effect is stronger near tree leaves than throughout the entire tree.

Effect of Episodic Risk on Clustering

Unlike imbalance statistics, clustering can be considerably affected by the temporal distribution of tree internal nodes. This is because the number of internal nodes increases when branches or clades merge into clusters. The more internal nodes a tree has under a given cutoff distance the more likely tree leaves are clustered within this cutoff distance. Therefore, to control for such confounding effect, we matched phylogenetic trees simulated under each value of risk re-selection rate, ω, by their cumulative number of internal nodes at different cutoff distances. The criteria of matching is the sum of squared distance between the cumulative number of internal nodes of each simulated tree and the average cumulative number of branches at a given cutoff distance of all the simulated trees. We set the threshold of sum of square distance arbitrarily to select trees that meet two requirements: 1) the cumulative number of branches of selected trees are matched as close as possible (sum of square distance as small as possible); 2) there is large enough sample of simulated trees per scenario (for each value of ω). The details of matching simulated phylogenetic trees by their cumulative number of branches at a given cutoff distance are described in section I-ii of Appendix II.

Same as the section "effect of episodic risk on tree imbalance", we examine effect of episodic risk on clustering assuming that population either has random mixing or assortative

mixing. In each contact mixing scenario, we simulated epidemics assuming zero to high values of risk re-selection rate, ω. Phylogenetic trees (simulated) were constructed from each set of epidemics. Within each contact mixing scenario, we selected simulated phylogenetic trees whose cumulative number of internal nodes are matched as close as possible for clustering analysis (Fig.V-6).

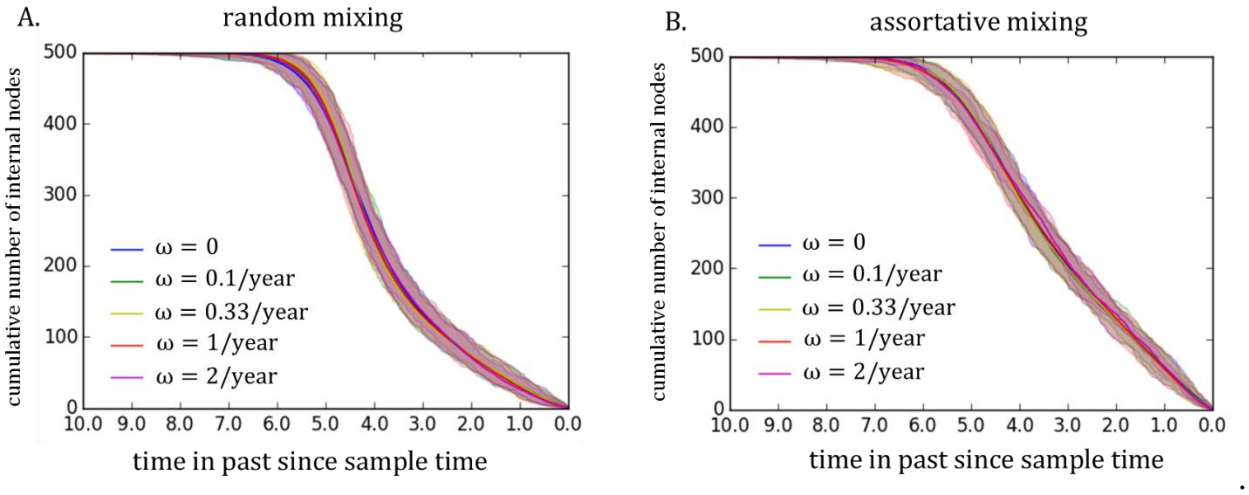

Figure V- 6. Cumulative number of internal nodes of the phylogenetic trees simulated with different risk re-selection rate, ω, that are matched as close as possible. Each shaded area represents the collection of cumulative number of internal nodes of the phylogenetic trees simulated assuming a specific value of ω. Solid line represents the average cumulative number of internal nodes of selected phylogenetic trees simulated assuming a specific value of ω. Panel A: when population has random mixing, simulations are done by assuming that average transmissibility per contact, β=0.008. Panel B: when population has assortative mixing, and average transmissibility per contact, β=0.005. Other parameters are set at their default values as shown in Table AII-2.

The first row of Fig.V-7 illustrates that when population has random mixing episodic risk has slight but observable impact on clustering (panel A and B of Fig.V-7). At cutoff distances greater than 4 years in past, increase in ω increases the average cluster size (left side of panel B of Fig.V-7). At cutoff distances smaller than 4 years in past, increase in ω reduces the average cluster number (right side of panel B of Fig.V-7). Since trees have been matched by the cumulative number of branching within each cutoff distance (panel A of Fig.V-6), both cluster size and number of clusters reflect the tendency that tree leaves are clustered with each other. This is because that when the number of branches under a cutoff distance is fixed, smaller cluster number and large cluster size indicate that branches more occur within the same cluster. Therefore, the phenomenon as shown in upper rows of Fig.V-7 suggests that when population

has random mixing, episodic risk slightly increases the average tendency of clustering and such effect is observable at both small and large cutoff distances.

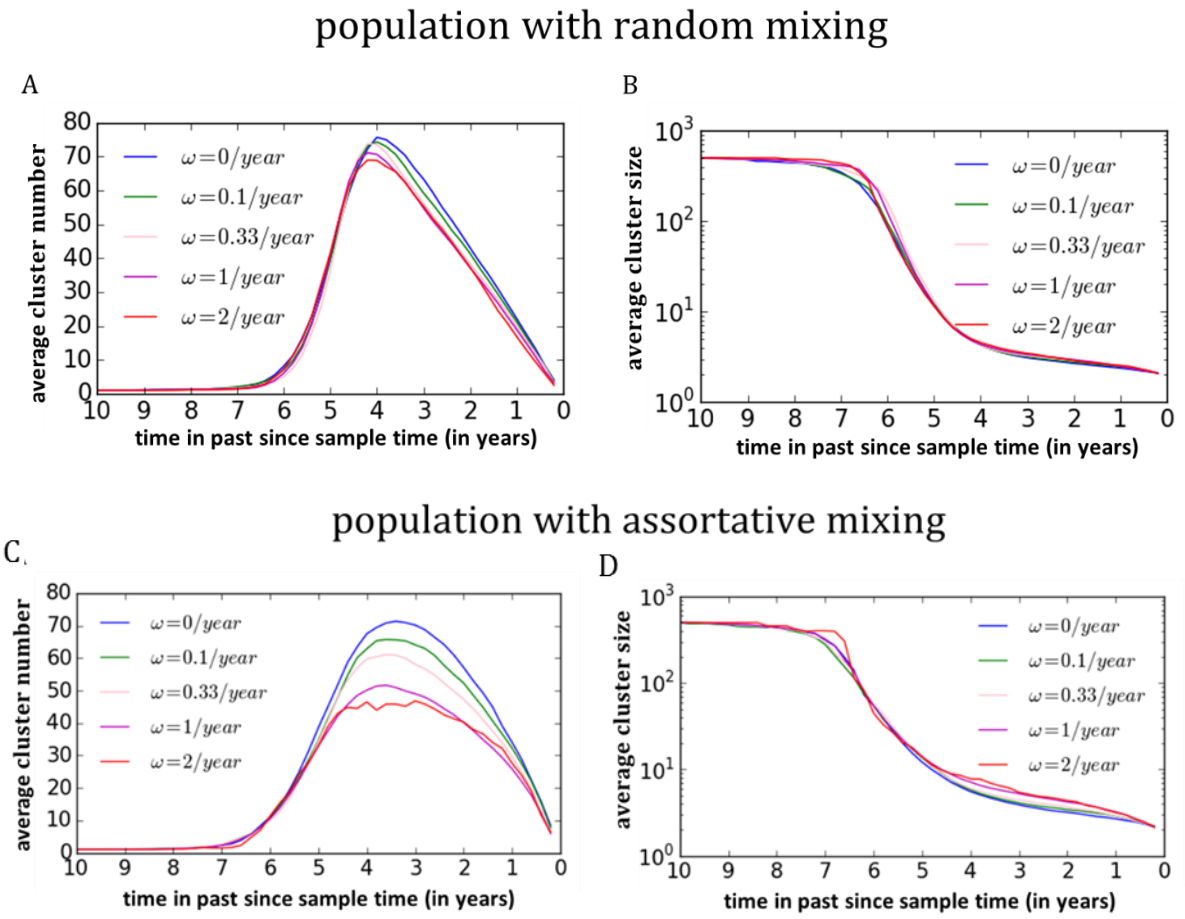

Figure V- 7. Effect of episodic risk on average size of cluster (panel A) and average number of cluster (panel B) when population has random mixing, and effect of episodic risk on average size of cluster (panel C) and average number of cluster (panel D) when population has assortative mixing, i.e. individuals reserve 50% of contacts for people at the same level of risk behavior. For epidemics that are simulated assuming random mixing of population, average transmissibility per contact, β=0.008. For epidemics that are simulated assuming assortative mixing, average transmissibility per contact, β=0.005. Other parameters are set at their default values as shown in Table AII-2. Single branches are not counted as clusters in this analysis.

By contrast, lower row of Fig.V-7 shows that when population has assortative mixing, episodic risk considerably reduces the average clustering number and increases the average cluster size at the cutoff distances smaller than 5 years since sample time. As trees are matched by the cumulative number of branches at each cutoff distance (panel B of Fig.V-6), this indicates that episodic risk considerably increases the tendency that tree leaves are clustered together at relatively small cutoff distances, i.e. near tree leaves. In addition, comparison between upper row

of Fig.V-7 and lower row of Fig.V-7 suggest that episodic risk more greatly affects clustering tendency when population has assortative mixing than when there is random mixing.

To understand these observations, it is again helpful to consider the mechanism through which episodic risk affects the phylogenetic tree. As discussed in section “effect of episodic risk on tree imbalance”, the effect of episodic risk on the simulated phylogenetic trees can be categorized into two aspects: transmission-aspect effect and transition-aspect effect. Transmission-aspect effect is attributed to that episodic risk increases the amount of transmissions linking high-risk acute infected individuals. Consequently, high-risk acutely infected individuals more likely form large clusters on the transmission tree. Such effect would be observable on the simulate phylogenetic trees, since they are the pruned transmission trees. That says, tree leaves are more likely to be members or descendants of the members of the high-risk acute infection cluster.

As discussed in earlier sections, the transition-aspect effect of episodic risk on the phylogenetic trees is that as a case change risk levels, the corresponding branch would also have changed potency to split into new branches. If a case transit from high risk to low risk, corresponding branch would have lowered potency to split into new branches. This limits the chance that a branch with a high potency to have descendants maintain this high potency and generate large clusters. Therefore, it reduces the chance that a cluster grows over time on a phylogenetic tree. Rather, clusters are more likely to be initiated locally by branches whose corresponding case transits from low risk to high risk. Briefly speaking, the transition-aspect effect of episodic risk on clustering is to reduce the extent to which large clusters grow over time but increases the tendency that small clusters grow locally. We found the transition-aspect effect is more pronounced for population with assortative mixing than for population with random mixing. This is because that for population with random mixing inter-risk-group transmissions are frequent, which also change branches’ state. This limits the impact of transition-aspect effect. This has also been observed in section “effect of episodic risk on tree imbalance”, where we showed that transition-aspect effect mainly affects phylogenetic tree imbalance when population has assortative mixing.

The impact of episodic risk on the clustering is the consequence of synergy of the transmission-aspect effect and transition-aspect effect. When population has random mixing, the

frequent inter-risk-group transmissions limit the impact of transition-aspect effects. Therefore, the transmission-aspect effect dominates, which increases the clustering tendency (upper row of Fig.V-7). When population has assortative mixing, both effects play important roles. At large cutoff distances, i.e. near the tree root, episodic risk increases clustering tendency through the transmission-aspect effect. Meanwhile episodic risk limits the extent that large clusters grow over the entire tree, which we refer as the transition-aspect effect. The two effects counterbalance each other, so at large cutoff distances we hardly observe impact of episodic risk on clustering (left side of panel C and panel D of Fig.V-7). At smaller cutoff distances, i.e. cutoff distance smaller than 5 years, both effects increase the tendency that clusters grow near tree leaves. In addition, transmission-aspect effect increases due to assortative mixing: high-risk acute HIV infection are more likely to be clustered due to assortative mixing. Therefore, episodic risk considerably increases the clustering tendency at relatively small cutoff distances (right side of panel C and panel D of Fig.V-7).

In addition to the average cluster size and average cluster number, we also examined the cluster size distribution (CSD) and skewness of CSD for each simulated phylogenetic tree. The relevant results are included in section II of Appendix II. The results largely agree with the results as shown in Fig.V-7. Firstly, when population has random mixing, episodic risk considerably increases the probability of large cluster size at large cutoff distance, i.e. near the tree root (Fig.AII-2). In addition, episodic risk reduces the skewness of cluster size distribution, which suggest that cluster size distribution is less skewed towards right, or in other words, increase in the mass probability of large clusters (Fig.AII-3). Such change of skewness is observed from the cutoff distances of 6 year in past to 1 year in past, which covers a large part of the tree (Fig.AII-3), indicating that episodic risk affects the clustering of the entire phylogenetic tree. When population has assortative mixing, episodic risk reduces the probability of large cluster at relatively large cutoff distance, but increases the probability of large cluster near tree leaves (Fig.AII-4). This agrees with earlier inference: when population has assortative mixing, episodic risk causes cluster to more likely grow near the tree leaves but less likely near the tree root.

Effect of Episodic Risk on Phylogenetic Tree of Heterochronous Sample

In this section, we further investigate how effect of episodic risk on a phylogenetic tree differs between a homochromous sample and a heterochronous sample. For the illustrative purpose, we constrain our analysis to the assortative mixing scenario, where individuals reserve 50% of their contacts for people experiencing the same risk phase. In addition, we observe the effect of episodic risk on simulated phylogenetic trees by comparing trees simulated assuming no episodic risk, i.e. risk re-selection rate $\omega=0$, and trees simulated assuming a moderate level of episodic risk, $\omega=1$/year.

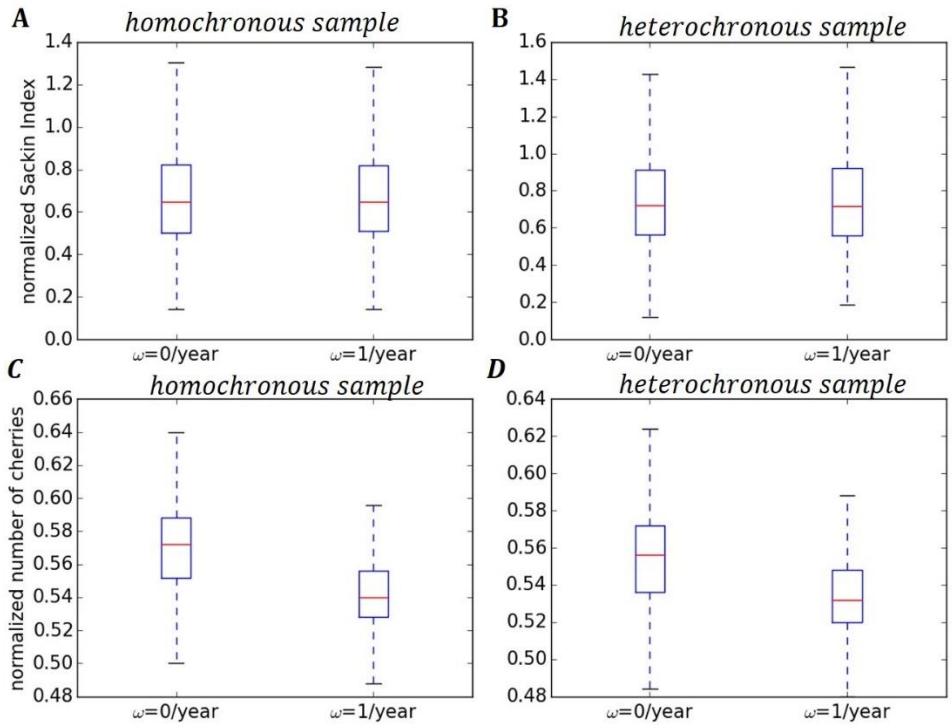

Figure V- 8. Effect of episodic risk on normalized Sackin Index (upper rows) and normalized number of cherries (lower rows) when sample is collected homochromously (left column) or heterochronously (right column). Sample size is 500 for both sample scenarios. For both samping scenarios, we assume that population has assortative mixing, i.e. 50% of contacts reserved for people at the same level of risk. Average transmissibility per contact, β=0.005. Parameters (not including risk re-selection rate, ω, or β) are set at their default values as shown in Table AII-2.

Fig.V-8 shows that effect of episodic risk on the phylogenetic tree imbalance are consistent between the homochromous sample and heterochronous sample. Firstly, episodic risk has minor impact on the normalized Sackin's Index in both sample scenarios (panel A and panel B of Fig.V-8). Secondly, episodic risk reduces the number of normalized cherries in both sample scenarios (panel C and panel D of Fig.V-8). However, episodic risk less reduces number of cherries for heterochronous sample than homochromous sample. This is because that cases sampled at different time are less likely to be clustered and also less likely to form cherries. This reduces the impact of episodic risk on cherry. An extreme example is that if cases are sequentially collected with a large time gap in between each case then cases are most likely to form ladders. Such effect will drive the pattern of clustering regardless of episodic risk.

We further explore whether and how episodic risk differently affects clustering of the phylogenetic tree for homochromous sample and heterochronous sample. As mentioned in the method section, tree leaves that belong to the same cluster is the group of leaves whose maximum pairwise distance is below the cutoff value. For homochromous sample, the pairwise

distance of two leaves is simply twice the distance from either leaf to the most common ancestor of the two leaves. For heterochronous sample, clusters at a cutoff pairwise distance are identified using hierarchical clustering function in R.[132]

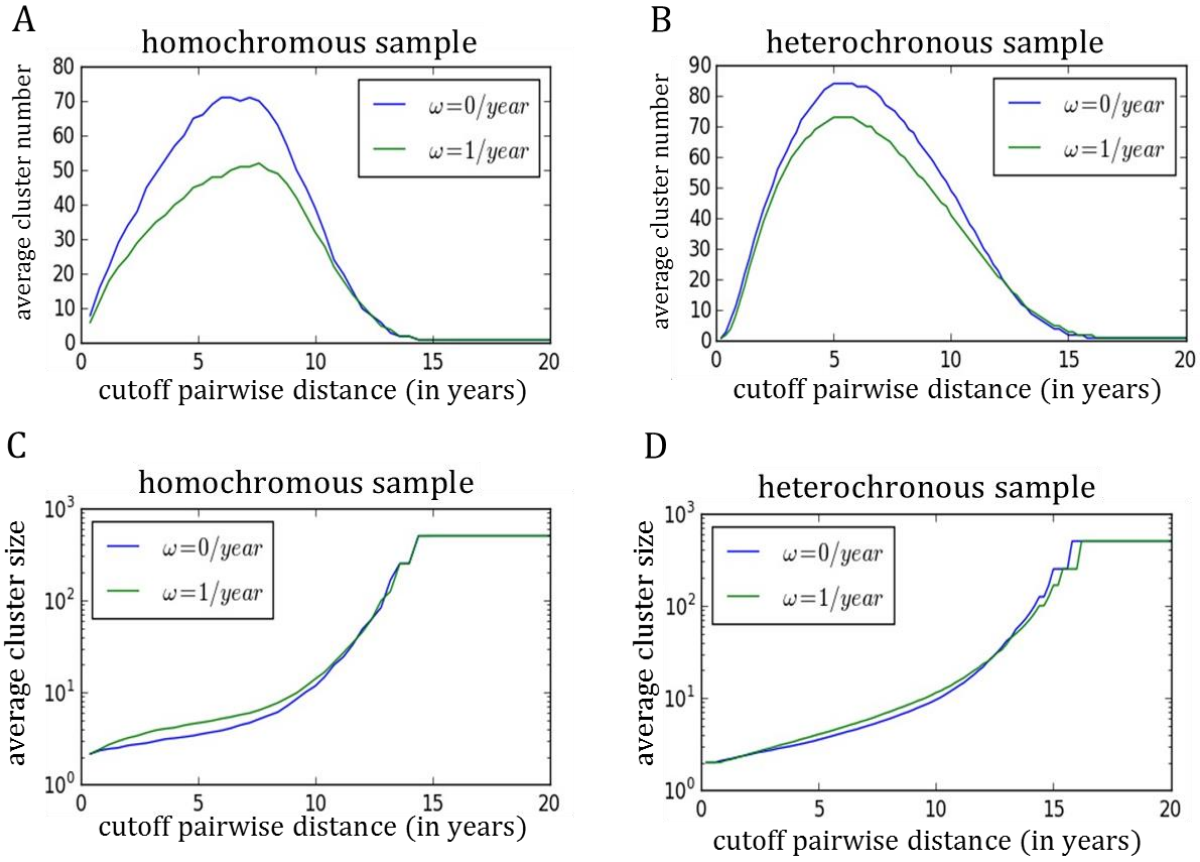

Figure V- 9 Effect of episodic risk on the average number of clusters (upper rows) and average cluster size (lower rows) at a given cutoff pairwise distance when sample is collected homochromously (left column) or heterochronously (right column). Sample size is 500 for both sample scenarios. For both sample scenarios, we assume that population has assortative mixing, i.e. 50% of contacts reserved for people at the same level of risk. Average transmissibility per contact, β=0.005. Parameters (not including risk re-selection rate, ω, or β) are set at their default values as shown in Table AII-2.

Fig.V-9 shows that effect of episodic risk on clustering at a given cutoff pairwise distance is still noticeable for heterochronous sample (panel B and panel D of Fig.V-9). However, such effect is smaller than when sample is collected at different time than when sample is collected at the same time (compare right column with left column). As before, this is because the tree leaves tend to be less clustered if they are collected at different times. This also reduces the impact of episodic risk on clustering.

Comparing Effect of Episodic Risk and Other Model Parameters on Simulated Phylogenetic Trees

As indicated by earlier results, episodic risk can considerably affect imbalance and clustering of the simulated phylogenetic trees. In this section, we further explore that whether such impact can be distinguished from those of some other model parameters. The purpose of this section is not to systematically examine the identifiability of risk volatility parameter from HIV phylogeny. Rather, it is to explore the uniqueness of effect of risk volatility on the HIV phylogeny within the scope of our model. In earlier sections, we have demonstrated that effect of episodic risk on the dynamics of HIV transmission can be summarized into two aspects. One is transmission-aspect effect, which is that episodic risk increases the amount of transmissions linking high-risk acute infections. The other is transition-aspect effect, which is that episodic risk changes the state of a branch by inducing cases' risk transition. Therefore, we choose two model parameters whose change may affect HIV transmissions in a similar way. The first is the relative transmissibility of acute infection, ζ, increase in which can also increase the contribution of acute infection to ongoing transmissions. The second is fraction of contacts reserved for people with the same risk level, m, reducing which can increase the possibility of inter-risk-group transmissions and change the state of a branch.

When examining the uniqueness of one aspect of effect of episodic risk, we set the model parameters at the level that largely eliminate the other aspect of effect. That says, when we compare the transmission-aspect effect of episodic risk and the effect of increasing ζ on phylogenetic trees, we restrict this comparison under the condition that population has random mixing, i.e. m=0, so that the transition-aspect effect is largely controlled. Likewise, when comparing the transition-aspect effect of episodic risk and the effect of reducing m on the simulated phylogenetic trees, we set equal transmissibility between acute infection and chronic infection, i.e. ζ=1, so that episodic risk's transmission-aspect effect is eliminated.

Comparing Transmission-Aspect Effect of Episodic Risk and Effect of Heterogeneity of Infectiousness on Simulated Phylogenetic Trees. Fig.V-10 shows that although increasing risk re-selection rate, ω, changes tree imbalance statistics and clustering in a way similar as increasing relative transmissibility of acute infection, ζ, does, they change these statistics to different extents. Firstly, compared to increasing ζ, increasing ω causes a greater reduction in the

normalized Sackin's Index (panel A of Fig.V-10) and a slightly greater increase in the normalized number of cherries (panel B of Fig.V-10). Both indicate that increasing ω can alter tree imbalance more than increasing ζ does. This is because that increasing ω and increasing ζ affect HIV transmissions and HIV phylogeny in different ways. In section "Effect of Episodic Risk on Tree Imbalance", we discussed that episodic risk can affect transmission from acute infection in two ways: firstly, because chronic infection comes after acute infection, people who were at high risk phase at HIV acquisition more likely transit to low risk during chronic infection than during acute infection. This increases the contribution of acute infection to ongoing transmissions. Secondly, episodic risk increases replenishment of susceptible individuals experiencing high risk. The synergy of these two results in greater amount of transmissions linking high-risk acute infections. Consequently, phylogenetic trees have higher fraction of branching events which generates two sibling branches that both correspond to high-risk acute infections. Such effect reduces both the Sackin Index and number of cherries. By contrast, increasing ζ only increases the contribution from acute infection but does not increase replenishment of high risk susceptible individuals. As result, increasing ζ does not considerably increase the fraction of tree branching caused by transmissions linking high-risk acute cases (Fig.AII-8). Consequently, increasing ζ change the tree imbalance statistics less than increasing ω does.

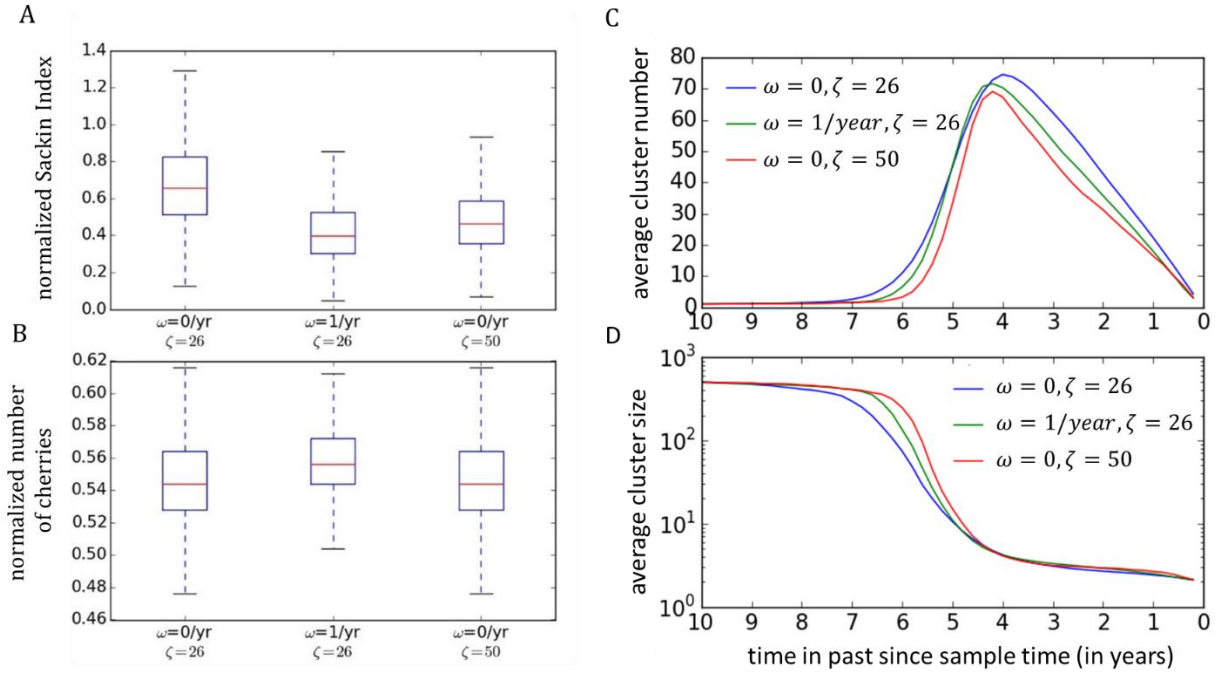

Figure V- 10. Effect of increasing relative transmissibility of acute infection, ζ, and effect of increasing risk re-selection rate, ω, on normalized Sackin's Index (panel A), normalized number of cherries (panel B) and pattern of clustering at different cutoff distance (panel C and panel D). For the three sets of simulations, population has random mixing. For all three sets of epidemics, average transmissibility per contact, β=0.008.

However, panel C and D of Fig.V-10 shows that increasing ω changes clustering less than increasing ζ does (compare the difference between blue curve and green curve with the difference between blue curve with red curve). This is also because these two parameters affect HIV transmission differently. Increasing ω results in increased amount of transmission linking high-risk acute infections. The consequence is that prevalence of high-risk acute infections increases. On the phylogenetic tree, there are more branches which correspond to high-risk acute HIV infections. These branches have equally high potency to have descendants. Given the fixed number of tree leaves, this limits the extent that a single branch has large group of descendants。

By contrast, increasing ζ increases the transmission potential from acute HIV infection, regardless of the risk phases of acutely infected people. In addition, increasing ζ does not increase the replenishment of susceptible individuals experiencing high risk. Therefore, increasing ζ does not considerably increase the prevalence of high-risk acute HIV infections. Consequently, there are fewer branches that correspond to the high-risk acute HIV infection. In other words, fewer branches compete with each other to have descendants . Therefore, it is more likely that few branches lead to large group of descendants, resulting in a greater clustering tendency. This is further confirmed by Fig.AII-9 in Appendix II, which shows that the average number of branching caused by a single infector is considerably greater among phylogenetic trees with high ζ than among phylogeny trees simulated with nonzero ω.

Comparing Transition-aspect Effect of Episodic Risk and Effect of Reduced Degree of Assortative Mixing on Phylogenetic Trees.

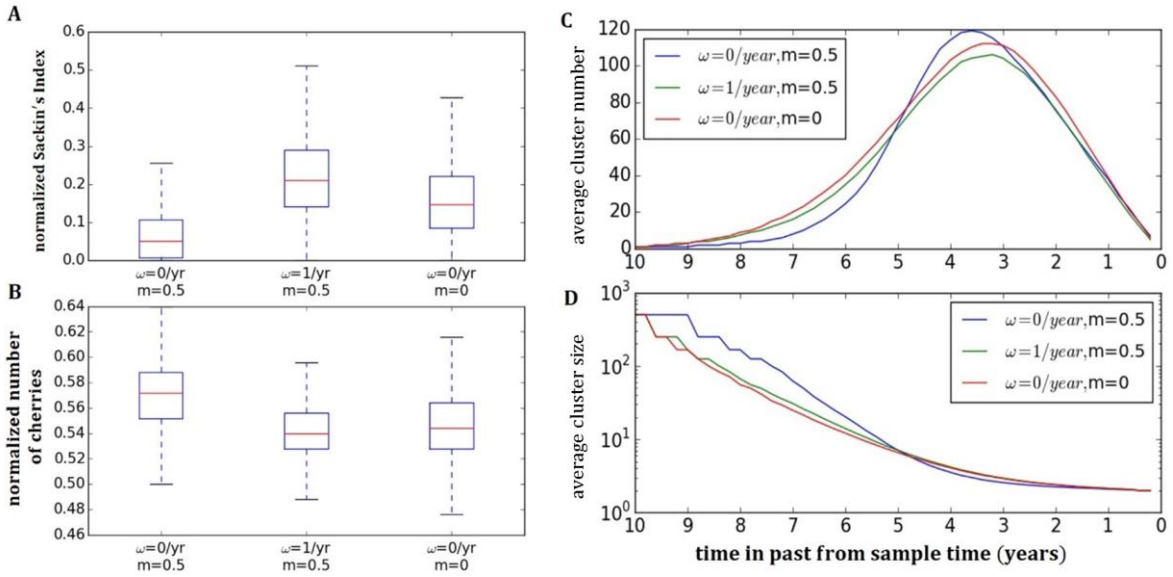

Figure V- 11. Effect of increasing risk re-selection rate, ω, and reducing fraction of transmission reserved for individuals with same level of risk, m, on normalized Sackin Index (panel A), normalized number of cherries (panel B) and average cluster size (panel C) or average number of cluster at different cutoff distances (panel D). For the three sets of simulated epidemics, transmissibility is set equal between acute infection and chronic infection, i.e. relative transmissibility of acute HIV infection, ζ=1. Average transmissibility per contact, β=0.01. Other model parameters are set at their default values as shown in Table AII-2.

Fig.V-11 shows that compared to reducing the degree of assortative mixing, m, increasing risk re-selection rate, ω, changes normalized Sackin Index (panel A of Fig.V-11) and clustering (panel C, D of Fig.V-11) to a greater extent but have similar impact on the normalized number of cherries (panel B of Fig.V-11). This indicates that compared to reducing m, increasing ω can more impact the overall tree imbalance and the pattern of cluster than the tree imbalance near tree leaves (cherry). To understand this, it is helpful to again consider how changing either parameter affects the phylogeny. As discussed in section "Effect of Episodic Risk on Tree Imbalance", increasing ω increases the frequency of risk transitions. If a case whose descendants (infections on the downstream of the transmission pathway) are in the sample, risk transition of this case results in change in state of a branch of the phylogenetic tree. Similarly, reducing degree of assortative mixing, i.e. m, increases inter-risk group transmission. This increases the chance that an inter risk group transmission occurs along a branch and can also change the branch's state. From this perspective, increasing ω and reducing m have similar impacts on the phylogeny. However, the difference between them is that risk transition only requires a single case but inter-risk group transmission requires interaction of two persons. This makes risk transition to be less dependent on the transmission dynamics of HIV, especially the prevalence of HIV infection. As result, increasing ω tends to affect a phylogeny tree more than

reducing m. In addition, such difference needs to cumulate at multiple branches for it to be seen. Therefore, effect of increasing ω and effect of reducing m on the phylogeny tree can be distinguished by examining the overall tree imbalance statistics and clustering throughout the entire tree.

Discussion of Results in Chapter V

Summary of findings. Our study used stochastic simulations to understand that how individual risk behavior volatility, a phenomenon that individual risk behavior fluctuates over time, affect HIV phylogeny. This study is also the first step of exploring the possibility of detecting population impact of risk volatility from HIV phylogeny. Individual risk behavior volatility is an important but understudied behavior characteristic among men who have sex with men (MSM) population. As the first attempt at this topic, we simplify the risk volatility as individuals risk altering between high and low level over time.

Our results demonstrate that individual risk behavior volatility considerably affects the simulated phylogenetic trees. Specifically, the effect of individual risk behavior on HIV phylogeny can be summarized in two aspects: transmission-aspect effect and transition-aspect effect. The transmission-aspect effect is that risk volatility increases the amount of branches that are generated by transmissions linking high-risk acute HIV infections. It tends to reduce tree imbalance and increase the tendency that tree leaves are clustered together. The transition-aspect effect is that as cases change risk phases their corresponding branches have changed potency to have descendants. It tends to increase tree imbalance, and causes clusters to more likely to grow further from tree root. The synergy of the two effects change both the tree imbalance and tendency that tree leaves are clustered together.

We illustrated that how risk volatility affects the imbalance and clustering of the simulated phylogenetic trees depends on the degree of assortative mixing. This is because that as degree of assortative mixing increases, within-risk-group transmissions become more frequent, so that HIV virus are more likely to be passed among cases experiencing the same risk phase. This makes phylogenetic trees more sensitive to cases' risk transitions. That says, the transition-aspect effect of risk volatility increases as the degree of assortative mixing increases.

In addition, our results demonstrate that the effect of risk volatility on the simulated HIV phylogeny cannot be reproduced by varying two other model parameters: the relative transmissibility of acute HIV infection or the fraction of contacts reserved for people experiencing the same risk phase. This is because that mechanism through which risk volatility affects HIV transmissions is unique and cannot be reproduced by varying either model parameter.

Lastly, our results suggest that given a fixed sample size, heterochromous sample is less sensitive to risk volatility than homochromous sample. This is because as the gap between sampling time increases, tree leaves are less likely to be clustered. This particularly limits the extent to which risk volatility affects clustering.

Related Research on Effect of Population Contact Pattern on Phylogenies

Our study takes the first step to explore the effect of an understudied behavior characteristic of MSM population, individual risk behavior volatility, on HIV phylogeny. It can be related to earlier studies that examine how population contact structure affects phylogenies. For example, Leventhal et al (2012) suggest that contact structure greatly impact the phylogenetic tree imbalance.[130] However, our results suggest that Sackin Index is not be a good indicator of the contact structure if the population has individual risk behavior volatility. The strong interaction among the risk volatility, high infectiousness of acute HIV infection and assortative mixing makes it less likely to rely on one single imbalance statistic to infer the population contact structure. Rather, our study indicates that Sackin Index needs to be combined with number of cherry, which measures the imbalance near tree leaves to better reveal the effect of individual risk behavior volatility. This agrees with conclusion of several earlier studies. For example, Dearlove and Frost (2013) suggest how tree asymmetry distribute along the phylogeny is important to assess in addition to the overall tree imbalance.[133] Our study agrees with their conclusion and suggest it is more likely to detect signal of risk volatility by looking at the clustering pattern and tree imbalance at different parts of a phylogenetic tree. Colijn et al (2014) suggest that the local tree imbalance and overall tree imbalance can be combined to inform the type of population where an outbreak occurs.[134] Our study agrees with such conclusion: the effect of individual risk behavior volatility on the combination of tree imbalance and clustering is more unique than its effect on any single statistic.

Similar topics have been explored by previous studies. For example, Robinson et al (2013) suggest that when contact network of hosts becomes dynamic, phylogenies can less reflect the network structure.[124] To some extent our study agrees with this finding: individual risk behavior volatility, a specific type of dynamic individual risk behavior, can reduce the extent to which assortative mixing affects a phylogenetic tree and cause a tree to show a clustering pattern that resemble one from population with random mixing. However, we found effect of risk volatility

on the simulated phylogenies is not exactly as reducing the degree of assortative mixing because they affect HIV transmissions through different mechanisms.

Implication for public Health Interventions

The motivation behind our study is that empirical data suggest the existence of individual risk behavior volatility[45,46], and that studies indicate that risk volatility may considerably change the population impact of HIV interventions. Henry and Koopman (2015) found that individual behavior volatility can reduce R_0 and the minimum effort that universal test and treat needs to eliminate HIV infections. [82] The mechanism is that risk volatility reduces the correlation between a case's risk level at HIV acquisition and that case's risk level when he transmits HIV infection. As result, risk volatility reduces the effect of contact heterogeneity on the transmission system.[82] The consequence is that cases who were at high risk at HIV acquisition cause smaller amount of onward transmission than when there is no risk volatility. Therefore, as our earlier study indicated, individual risk behavior volatility reduces the population impact of PrEP that is targeted at susceptible individuals experiencing high risk. In addition, study by Alam et al (2012), suggest that acute HIV infection are more likely to form large cluster for epidemics with higher individual risk behavior volatility.[56] Our study again confirms this conclusion. All these studies support the importance of understanding risk volatility in guiding public health prevention of HIV transmission.

Furthermore, our study demonstrates that investigating individual risk behavior volatility may help understand what is driving acute HIV infections to cluster in a phylogenetic tree. Earlier studies use real HIV sequence data to understand what generates clusters of acute infection. For example, Volz et al (2011) suggested that acute infections form large clusters, and this is largely attributed to the short duration of acute HIV infection.[50] Brenner et al (2007) finds out that the pattern that primary HIV infection (within one year since HIV infection) reflect the underlying contact structure.[52] Lewis et al (2014) finds clusters of acute infection from phylogenetic tree of HIV virus that suggest episodic sexual transmission.[53] These studies suggest that acute infection cluster is a result of intense transmission during early stage of HIV infection and the short period of acute infection. However, our study found that individual risk behavior volatility can further alter the tendency of acute infection to cluster despite the transmissibility

from acute infection, level of contact heterogeneity and degree of assortative mixing all remains the same.

Model Assumption and Realistic Relaxing Assumptions

Our study uses individual based model. The advantage is that we can track the transmission history and find the clear correspondence between the transmission events and generation of new branches on the phylogenetic tree. In addition, we are able track the history of risk transition and disease progression of each case. These information facilitate our understanding of impact of individual risk behavior volatility on transmission dynamics of HIV and how it relates to HIV phylogeny.

To make our analysis more tractable we make several assumptions. Firstly, we assume dichotomous individual risk behavior. This ensures the efficiency of stochastic simulation risk transitions. However, this is not realistic. In reality, individual risk behavior follows a continuous distribution. Romero-Severson et al (2014) explored effect of individual risk behavior volatility on R_0 by assuming continuous distribution of individual risk.[48] Henry and Koopman (2015) research same topic by assuming dichotomous risks.[82] The two studies reach the same conclusion: increase in individual risk volatility reduces R_0.[48,82] This indicates that change in distribution of individual risk from continuous to dichotomous does not change the mechanism that risk volatility affects transmission dynamics of HIV. As result, we expect that incorporating more continuously distributed individua risk will not fundamentally change our observations. However, it may change scale of the effect that we observed with dichotomous risk states. Secondly, we assume that individuals randomly pick partners and each partnership dissolves instantly. This excludes the possibility of concurrent partnership. However, partnership concurrency is common among MSM population.[96] We expect that high degree of concurrency will increase the tendency that cases cluster even without individual risk behavior volatility. This may reduce the extent to which phylogenetic tree clustering pattern is affected by individual risk behavior volatility.

Furthermore, our analysis is based on stochastically simulated phylogenetic trees. Our understanding of effect of risk volatility on simulated HIV phylogeny is based on our understanding of impact of risk volatility on transmission dynamics of HIV. For this to be valid, there are two underlying assumptions. First is that each new branch of the phylogenetic tree is generated due to a transmission event. The second is that the position of internal nodes of the

phylogenetic tree is highly correlated with the time that corresponding transmission event occurs. These assumptions help simplify our analysis. However, phylogenetic trees built from HIV sequences could considerably differ from the transmission tree.[135] Within host evolution, stochasticity of evolution all makes phylogenetic tree to less resemble transmission tree.[136,137] Therefore, for real phylogenetic trees, uncertainty due to stochastic evolution and bias introduced during reconstruction need to be both considered when inferring the parameters related with individual risk behavior volatility. This suggest that although our observations apply to the simulated transmission trees, whether they apply to HIV phylogeny in the real world remains explored.

In addition, an important research question that we explored is whether population impact of risk volatility can be distinguished from those of other model parameters. Our analysis is based on the episodic risk model. Therefore, the model parameters do not include all the possible factors that may affect HIV phylogenies in a similar way as risk volatility does. This implies that it remains unknown that whether effect of risk volatility on HIV phylogeny can be distinguished from other behavioral or biological factors of HIV transmission or their combination.

In our results, we found that if cases are sampled at different times rather being collected at the same time, they are less likely to form cluster. This limits our ability to observe impact of individual risk behavior volatility on clustering pattern of HIV phylogenetic tree. However, such issue can be potentially handled by using method such as one used in Gray et al (2011), which analyzes temporal clustering pattern by taking into account the sequence sample time.

Future Directions

There are several directions that may be valuable to explore in future. One direction is to further test the observations in our paper with simulated sequence or real sequence data. Another is to better understand how different factors alone and in combination affect tree shapes in a manner that might improve our understanding about what information we can get out of tree shapes with regard to different factors. Ultimately, future studies can also develop the algorithm that estimate different factors such as individual risk behavior volatility from the HIV sequence data, using information such as HIV phylogenetic tree imbalance, clustering pattern and branching times. For example, using Sequential Monte Carlo methods suggested by Smith et al or the new

artificial intelligence analytic methods to discover and detect tree shape patterns emerging as a result of different factors.[138]

Conclusions

Our findings indicate that the population impact of individual risk behavior volatility can be potentially detected from HIV phylogenetic tree data. However, it requires good knowledge of the biological and behavioral characteristics of HIV transmissions because effect of risk volatility on HIV phylogeny may highly depend on these factors. Given HIV phylogeny the population impact of risk volatility is likely to be distinguished from that of some other model parameters. HIV phylogenetic tree should be examined from multiple perspectives not to miss any valuable information.

CHAPTER VI

Discussion

Summary of findings

This thesis takes the first step to explore two important issues people usually face when using mathematical model to guide public health decision-making to control HIV transmission: robustness inference and identifiability of model parameters. Specifically, this thesis focused on one commonly adopted assumptions for model of HIV transmission among men who have sex with men (MSM): individual risk behavior remains constant over time. Research presented in this thesis can be summarized in three sections.

Firstly, we examined the robustness of inference of a simple model of HIV transmission among MSM by relaxing one commonly adopted model assumption: individual risk behavior remains constant throughout sexually active life (Chapter II and Chapter IV). To do so, we build a simple deterministic model with two stages of HIV infection, and incorporate individual risk behavior volatility in the model. We define a parameter that takes the rate of changing individual risk between low and high levels through time. This relaxes the assumption of constant individual risk behavior over time. Increasing this parameter considerably alters several epidemiological quantities that are essential to guide the public health decision of HIV intervention. These quantities include transmission from acute stage of HIV infection, prevalence at endemic equilibrium and minimum required individual effectiveness of HIV intervention strategy to eliminate HIV transmission. Therefore, a model assuming constant individual risk behavior would make erroneous inference of these important quantities.

Secondly, we derived type reproduction numbers for model with risk volatility both with next generation matrix approach and an approach based on their epidemiological meanings (Chapter III). The second approach derives the probabilities that a model case is in a risk phase during a stage of infection. With the help of state probabilities, we get clear understanding of

how risk volatility changes the type reproduction numbers, which cannot be revealed by next generation matrix approach.

Thirdly, we explore whether signal of individual risk behavior volatility can be detected from HIV phylogenies and whether such signal can be distinguished from that of two other model parameters (Chapter V). This is also the first step of exploring the identifiability of parameter that quantifies individual risk behavior volatility from HIV genetic data. To do so, we simulate the stochastic, individual-based version of the risk volatility deterministic model, and observe how imbalance and clustering pattern of the simulated phylogenetic tree (pruned transmission tree) changes as we vary the parameter that quantifies the individual risk behavior volatility. We found that varying this parameter can change tree imbalance and considerably alter the tendency that tree leaves cluster at a given cutoff distance. In addition, the findings also suggest that individual risk behavior volatility has a unique impact on the combination of phylogenetic tree imbalance and clustering pattern, which cannot be reproduced by varying two other model parameters.

Implication for Public Health Intervention

Findings in Chapter II and Chapter IV suggest that it is important to assess individual risk behavior volatility among the target population before making the decision of how to allocate control efforts of HIV prevention. Such decisions include but may not be limited to: how much effort needs to be invested into detecting and treating early HIV infections, how effectiveness a HIV intervention needs to be at individual level to eliminate HIV infections and whether control efforts should be prioritized to people who are experiencing high risk. Earlier studies show that individual risk behavior volatility cannot be detected unless individual risk behavior is evaluated through multiple periods and long term.[45] However, most risk behavior assessment collect risk behavior in past several months or one year and only record the average level of risk behavior. This makes it impossible to detect the variation of individual risk behavior over time. Therefore, it is imperative to improve the behavior data collection so that individual risk behavior can be more thoroughly assessed.

Individual risk volatility may be assessed from multiple perspectives, such as number of unprotected sex within a partnership, number of partners, possible risk factors that induce change in individual risk behavior. Studies indicate that episodes of unprotected sexual activity can

induced by substance use, transition in between periods within partnership and periods without partners, or change in social context of partnerships.[107–109] Any factor or combination of them would cause variation in individual risk behavior over time. Therefore, collecting individual behavior data through longer period and relevant context in which sexual contact happens may help better detect risk volatility. The ultimate goal of assessing individual risk behavior volatility is to distinguish risk heterogeneity attributed to individual risk behavior volatility and risk heterogeneity at population level. To make such inference one may not only need to collect relevant behavior data but also utilize statistical models to estimate how much individual risk behavior volatility contributes to the overall risk heterogeneity.

Findings in Chapter V indicate that HIV genetic data is potentially valuable to assess the individual risk behavior volatility. This chapter focused on two features of phylogenetic trees: tree imbalance and tendency that tree leaves cluster at a given cutoff distance. They are less commonly examined than measures such as distribution of branching time and branch lengths in phylodynamic study of HIV epidemic. This is because compared to the commonly assessed measures, these two features can less inform the demographic history of the population. However, findings in Chapter V suggest that individual risk behavior volatility has a unique effect on the contact structure of population and such effect can be reflected from tree imbalance and tendency that tree leaves cluster. This implies that phylodynamic studies of HIV epidemic may combine every characteristic of a phylogenetic tree to improve the identifiability of individual risk behavior volatility. A recent study by Smith et al (2017) proposed a Sequential Monte Carlo (SMC) method to estimate epidemiological parameters directly from genetic sequence data.[138] This suggests the possibility of estimating individual risk behavior volatility from HIV sequence data. Results in Chapter V imply that phylogenetic tree imbalance and clustering pattern can be potentially integrated into a framework such as SMC methods to improve identifiability of parameter that quantify risk volatility.

Future directions

In order to illustrate the mechanism through which individual risk behavior volatility shapes HIV transmission dynamics, we build a simple model with several unrealistic assumptions. Future studies may make realistic variation of the model structure and study that how effect of individual risk behavior volatility on HIV transmission change accordingly. For example, one

may incorporate directional sex role, concurrent partnership or continuous distribution of individual risk behaviors into the model and examine individual risk behavior volatility interacts with either behavior characteristics to affect HIV transmission. In addition, in Chapter III, we found the rules by which state probabilities are formulated for each risk phase and each stage of infection. Such rule does not change as more risk levels or stages of infections are added. Therefore, future study may use the framework of this method for parameter estimation using models with more risk states or more stages of infections. Furthermore, to more directly observe the mechanism that individual risk behavior volatility affects phylogenetic tree through changing the HIV transmission dynamics, we treat part of the transmission tree as phylogenetic tree. Future studies can relax this assumption by stochastically simulating HIV sequences and examine that how the randomness of virus evaluation and bias introduced during phylogenetic tree reconstruction affect the observations made in this thesis. Lastly, future studies can also develop the algorithm that estimate different factors such as individual risk behavior volatility from the HIV sequence data, using information such as HIV phylogenetic tree imbalance, clustering pattern and branching times. For example, using SMC methods suggested by Smith et al (2017) or the new artificial intelligence analytic methods to discover and detect tree shape patterns emerging as a result of different factors.[133]

Conclusion

There are two major messages that we want to convey in this thesis. Firstly, individual risk behavior volatility among MSM population greatly change fraction of transmission from acute infection, endemic prevalence and individual-effectiveness-population-effect relationship. This suggests that when there is risk volatility, model assuming constant individual risk over time may make erroneous inference of these quantities. Secondly, population impact of individual risk behavior volatility can be potentially detected from HIV phylogenetic tree. Therefore, study in this thesis address the importance of collecting behavior data that can reveal individual risk behavior volatility and the value of HIV sequence data from a new perspective.

APPENDICES

Appendix I-Supplementary Materials for Chapter IV

This document is divided into two parts: supplementary methods for Chapter IV and supplementary results for Chapter IV. In the supplementary methods section, we list the model equations of episodic risk model with universal test and treat, default values of model parameters used for simulation in Chapter IV. In the supplementary results section, we present the sensitivity analysis of the results we present in Chapter IV.

Supplementary Methods for Chapter IV

Section-I-i Model Equations. Structure of the model has been introduced in the main text. Here we present the equations of the deterministic compartment model.

Table AI- 1 Equations of Episodic Risk Model With Universal Test and Treat Simulated in Chapter IV

$$\frac{dS_H}{dt} = f_H\mu - \lambda_H S_H + f_H\omega S_L - f_L\omega S_H - S_H\mu$$

$$\frac{dS_L}{dt} = f_L\mu - \lambda_L S_L + f_L\omega S_H - f_H\omega S_L - S_L\mu$$

$$\frac{dA_H}{dt} = \lambda_H S_H + f_H\omega A_L - f_L\omega A_H - A_H\tau - A_H\gamma_1 - A_H\mu$$

$$\frac{dA_L}{dt} = \lambda_L S_L + f_L\omega A_H - f_H\omega A_L - A_L\tau - A_L\gamma_1 - A_L\mu$$

$$\frac{dA_{H,t}}{dt} = A_H\tau + f_H\omega A_{L,t} - f_L\omega A_{H,t} - A_{H,t}\gamma_1 - A_{H,t}\mu$$

$$\frac{dA_{L,t}}{dt} = A_L\tau + f_L\omega A_{H,t} - f_L\omega A_{H,t} - A_{L,t}\gamma_1 - A_{L,t}\mu$$

$$\frac{dC_H}{dt} = A_H\gamma_1 + f_H\omega C_L - f_L\omega C_H - C_H\tau - C_H\gamma_2 - C_H\mu$$

$$\frac{dC_L}{dt} = A_L\gamma_1 + f_L\omega C_H - f_H\omega C_L - C_L\tau - C_L\gamma_2 - C_L\mu$$

$$\frac{dC_{H,t}}{dt} = C_H\tau + f_H\omega C_{L,t} - f_L\omega C_{H,t} - C_{H,t}\gamma_2 - C_{H,t}\mu$$

$$\frac{dC_{L,t}}{dt} = C_L\tau + f_L\omega C_{H,t} - f_H\omega C_{L,t} - C_{L,t}\gamma_2 - C_{L,t}\mu$$

Meaning, value, unit and definition of the parameters or derived variables used in model equations are listed in Table AI-2 and Table AI-3.

Table AI- 2 Model parameter symbols, default values, units and definitions

Parameter	Default Value	Unit	Definition
μ	1/40	/year	Rate of removal from the sexually-active population unrelated to HIV. Because we set the equilibrium population in the absence of disease to 1, this is also the (absolute) rate of entry of new individuals into the sexually active population
γ_1	4	/year	Rate of progressing from acute to chronic infection
γ_2	1/10	/year	Rate of death from AIDS during chronic infection
β	0.0047	/contact	Average per-contact transmission probability across both stages of infection
ζ	17.5	-	Ratio of per-contact transmissibility during acute infection over per-contact transmissibility during chronic infection
f_H	0.05	-	Fraction of average time spent at high-risk phase
χ	36	/year	Average contact rate in the entire population at disease free equilibrium

r_{HL}	7	-	Ratio of high contact rate over low contact rate
m	0	-	Fraction of contacts reserved for people experiencing the same level of risk

Table AI- 3 Derived variables symbols, default values, units and definitions for calculation of derivatives of ODE model simulated in Chapter IV.

Variable	Formulation	Meaning
X_h	$(S_H + A_H + C_H)\chi_H m$	Total contacts made at the site where only high-risk subpopulation make fraction m of contacts with each other
X_g	$(S_H + A_H + C_H)\chi_H(1-m) + (S_L + A_L + C_L)\chi_L(1-m)$	Total contacts made at the general mixing site where high-risk subpopulation make contacts with low-risk subpopulation proportionately
X_l	$(S_L + A_L + C_L)\chi_L m$	Total contacts made at the site where only low-risk subpopulation make fraction m of contacts with each other
λ_{Hh}	$\frac{(A_H\beta_1\chi_H m + C_H\beta_2\chi_H m)\chi_H m}{X_h}$	Force of infection that causes new high-risk infections at the site where only high-risk subpopulation make fraction m of contacts with each other
λ_{Hgh}	$\frac{(A_H\beta_1\chi_H(1-m)+C_H\beta_2\chi_H(1-m))\chi_H(1-m)}{X_g}$	Force of infection from infected people in high risk phase that causes new infections in high risk phase at the general mixing site where subpopulation in high risk phase make contacts with subpopulation in low risk phase proportionately
λ_{Hgl}	$\frac{(A_L\beta_1\chi_L(1-m) + C_L\beta_2\chi_L(1-m))\chi_H(1-m)}{X_g}$	Force of infection from infected people in high risk phase that causes new infections in high risk phase at the general mixing site
λ_{Ll}	$\frac{(A_L\beta_1\chi_L m + C_L\beta_2\chi_L m)\chi_L m}{X_l}$	Force of infection that causes new low-risk infections at the site where only high-risk subpopulation make fraction m of contacts with each other
λ_H	$\lambda_{Hh} + \lambda_{Hgh} + \lambda_{Hgl}$	Total force of infection that causes new infections among susceptible people experiencing high risk

λ_L	$\lambda_{Ll}+\lambda_{Hgh}+\lambda_{Hgl}$	Total force of infection that causes new infections among susceptible people experiencing low risk

Section I-ii Formulation of Reproduction Numbers given Episodic Risk and Assortative Mixing

Formulation of Type Reproduction Numbers given Episodic Risk. In Chapter III, we have formulated type reproduction numbers for episodic risk model. We found that the type reproduction number for cases who were in high risk phase at HIV acquisition, $T_{H_{Inf}}$ and type reproduction number for cases who were in low risk phase at HIV acquisition, $T_{L_{Inf}}$, can be expressed as,

$$T_{H_{Inf}} = R_{HH}\psi + (1-\psi)Hg_H + \frac{\left(R_{LH}\psi + (1-\psi)Hg_L\right)\left(R_{HL}\psi + (1-\psi)Hg_H\right)}{1-\left(R_{LL}\psi + (1-\psi)Hg_L\right)}$$

$$T_{L_{Inf}} = R_{LL}\psi + (1-\psi)Hg_L + \frac{\left(R_{HL}\psi + (1-\psi)Hg_H\right)\left(R_{LH}\psi + (1-\psi)Hg_L\right)}{1-\left(R_{HH}\psi + (1-\psi)Hg_H\right)}$$

Where R_{pq} denotes the expected number of new infections in risk phase p a model case would cause if this model case spends the whole infection time in q risk phase. Formulations of R_{pq}s are given in Table AI-4. Variable ψ denotes the probability that a model case transmits HIV infection given that this case has not changed risk level since HIV acquisition. In Chapter III, we found,

$$\psi = \frac{\left(\frac{\beta_1}{\upsilon_1+\omega}\right)+\beta_2\gamma_1\frac{\upsilon_1}{(\upsilon_1+\omega)}\frac{\upsilon_2}{(\upsilon_2+\omega)}}{\left(\frac{\beta_1}{\upsilon_1}+\left(\frac{\gamma_1}{\upsilon_2}\right)\left(\frac{\beta_2}{\upsilon_2}\right)\right)}$$

In Chapter III, we denote υ_1 as the total rate of leaving acute HIV infection and υ_2 as the total rate of leaving chronic HIV infection. In the scenario with UT&T treatment,

$$\upsilon_1 = \gamma_1 + \mu + \tau$$

$$\upsilon_2 = \gamma_2 + \mu + \tau$$

In expressions of both type reproduction numbers, H denotes the expected number of new infections a case would cause during entire infection period if the expected contact rate of this case equals population average contact rate, χ.

$$H = \chi\frac{\beta_1}{\upsilon_1} + \chi\left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{\beta_2}{\upsilon_2}\right)$$

Table AI- 4 Formulations of variables used in calculation of type reproduction numbers

Symbol	formulation	meaning
υ_1	$\gamma_1 + \mu + \tau$	Total rate of leaving acute HIV infection
υ_2	$\gamma_2 + \mu + \tau$	Total rate of leaving chronic HIV infection
R_{HH_1}	$\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H(g_H + mg_L)$	Expected number of new infections in high risk phase caused by a case during acute infection if this case spends the whole acute infection in high risk phase (no episodic risk)
R_{HH_2}	$\left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_H(g_H + mg_L)$	Expected number of new infections in high risk phase caused by a case during chronic infection if this case spends the whole chronic infection in high risk phase (no episodic risk)
R_{HH}	$R_{HH_1} + R_{HH_2}$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case spends the whole infection time in high risk phase (no episodic risk)
R_{HL_1}	$\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_L(1-m)g_H$	Expected number of new infections in high risk phase caused by a case during acute infection if this case spends the whole acute infection in low risk phase (no episodic risk)
R_{HL_2}	$\left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_L(1-m)g_H$	Expected number of new infections in high risk phase caused by a case during chronic infection if this case spends the whole acute infection in low risk phase (no episodic risk)
R_{HL}	$R_{HL_1} + R_{HL_2}$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case spends the whole infection time in low risk phase (no episodic risk)
R_{LH_1}	$\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_H(1-m)g_L$	Expected number of new infections in high risk phase caused by a case during acute infection if this case spends the whole acute infection in high risk phase (no episodic risk)
R_{LH_2}	$\left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_H(1-m)g_L$	Expected number of new infections in high risk phase caused by a case during chronic infection if this case spends the whole chronic infection in high risk phase (no episodic risk)
R_{LH}	$R_{LH_1} + R_{LH_2}$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case spends the whole infection time in high risk phase (no episodic risk)
R_{LL_1}	$\left(\frac{1}{\upsilon_1}\right)\beta_1\chi_L(g_L + mg_H)$	Expected number of new infections in low risk phase caused by a case during acute infection if this case spends the whole acute infection in low risk phase (no episodic risk)
R_{LL_2}	$\left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{1}{\upsilon_2}\right)\beta_2\chi_L(g_L + mg_H)$	Expected number of new infections in low risk phase caused by a case during chronic infection if this case spends the whole chronic infection in low risk phase (no episodic risk)
R_{LL}	$R_{LL_1} + R_{LL_2}$	Expected number of new infections in low risk phase caused by a case during entire infection time if this case spends the whole infection time in low risk phase (no episodic risk)
ψ_1	$\frac{\upsilon_1}{\upsilon_1 + \omega}$	Probability that a case does not re-select contact rate during acute infection
ψ_2	$\left(\frac{\upsilon_1}{\upsilon_1 + \omega}\right)\left(\frac{\upsilon_2}{\upsilon_2 + \omega}\right)$	Probability that a case has not re-selected contact rate at chronic infection

ψ	$\frac{\left(\frac{\beta_1}{\upsilon_1}\right)\psi_1 + \left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{\beta_2}{\upsilon_2}\right)\psi_2}{\left(\frac{\beta_1}{\upsilon_1} + \left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{\beta_2}{\upsilon_2}\right)\right)}$	Probability that a case transmits HIV infection when this case has not re-selected contact rate
H	$\chi\frac{\beta_1}{\upsilon_1} + \chi\left(\frac{\gamma_1}{\upsilon_1}\right)\left(\frac{\beta_2}{\upsilon_2}\right)$	expected number of new infections a case would cause during entire infection period if the expected contact rate of this case equals population average contact rate, χ
R_{HHe}	$R_{HH}\psi + (1-\psi)Hg_H$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case was in high risk phase at HIV acquisition, given episodic risk
R_{HLe}	$R_{HL}\psi + (1-\psi)Hg_H$	Expected number of new infections in high risk phase caused by a case during entire infection time if this case was in low risk phase at HIV acquisition, given episodic risk
R_{LHe}	$R_{LH}\psi + (1-\psi)Hg_L$	Expected number of new infections in low risk phase caused by a case during entire infection time if this case was in high risk phase at HIV acquisition, given episodic risk
R_{LLe}	$R_{LL}\psi + (1-\psi)Hg_L$	Expected number of new infections in low risk phase caused by a case during entire infection time if this case was in low risk phase at HIV acquisition, given episodic risk

Supplementary Results for Chapter IV

Section II-i. Minimum Required Individual Effectiveness to Eliminate HIV Infections of PrEP Given Fixed Control Efforts At Different Levels of Coverage and Effective Treatment Rate of UT&T.

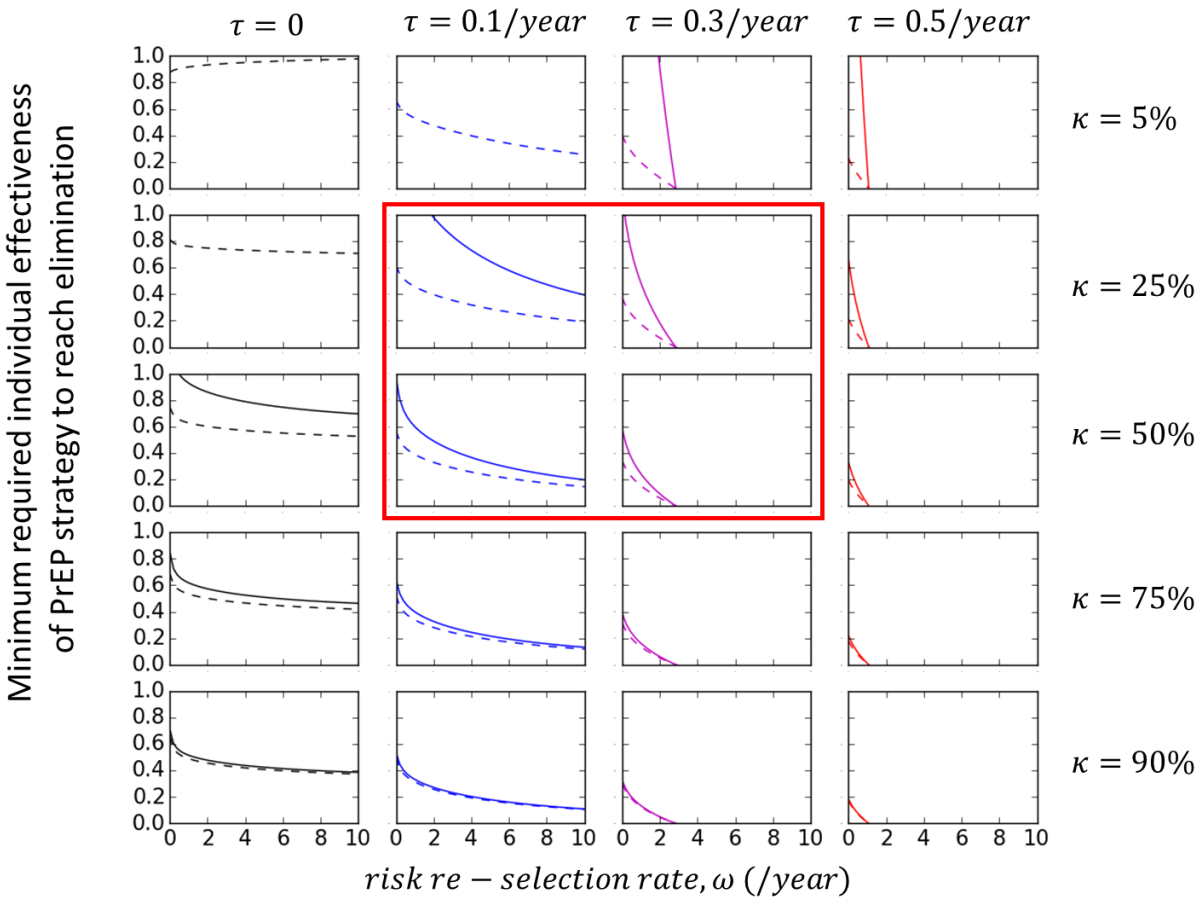

Figure AI- 1 Effect of risk re-selection rate, ω, on the minimum required individual effectiveness of general PrEP (solid) and high-risk-prioritized PrEP to reach elimination when coverage of PrEP efforts. Columns from left to right: effective treatment rate of UT&T, τ, is 0, 0.1/year, 0.3/year, or 0.5/year. Rows from top to bottom: coverage of PrEP efforts, κ, is 5%, 25%, 50%, 75% or 90%.

In Fig.AI-1 the solid curve and dashed curve represent the minimum required individual effectiveness of general PrEP or high-risk-prioritized PrEP to reach elimination, respectively. As shown in Fig.AI-1, in most panels the gap between these two curves decreases as risk re-selection rate, ω, increases. This indicates that for epidemics with greater individual risk behavior volatility, prioritizing PrEP efforts to susceptible individuals experiencing high risk makes smaller difference in the required individual effectiveness to eliminate HIV infections. In addition, such effect is most prominent when there is moderate coverage of PrEP, κ, and UT&T

is operated with moderate level of effective treatment rate, τ, as shown by panels in the middle of Fig.AI-1 (panels bounded by red frame). This is because when PrEP can only cover a small part of population, i.e. 5%, general PrEP can reach very few susceptible individuals experiencing high risk, which limits its ability to reduce transmission potential. In this case, at most values of ω general PrEP cannot eliminate HIV infections (first row of Fig.AI-1). Similarly, when effective treatment rate of UT&T is low general PrEP cannot eliminate HIV infections with small coverage (first panel in second row of Fig.AI-1). On the other hand, when coverage of PrEP is high, general PrEP can also cover most susceptible people who are experiencing high risk. Therefore, the minimum required individual effectiveness of two PrEP strategies remain similar regardless of the risk re-selection rate, ω (two lower rows of Fig.AI-1).

Section II-ii. Impact of Model Parameter Setting on Effect of Episodic Risk on Minimum Individual Effectiveness of Universal PrEP and Minimum Effective Treatment Rate to Eliminate HIV Infections.

Our results indicate that episodic risk tends to reduce $1-1/R_0$ and τ_e more for system with higher level of risk heterogeneity or higher degree of assortativity (Fig.AI-2, Fig.AI-3, Fig.AI-4 and Fig.AI-5). This is because increase in either parameter results in greater contribution of risk heterogeneity to R_0, which makes R_0 more sensitive to change in ω, and so does $1-1/R_0$ and τ_e. However, note that increase in ω reduces $1-1/R_0$ most when system has a moderately high r_{HL} instead of highest examined value of r_{HL} (green curve in lower panel of Fig.AI-2). This is because how much $1-1/R_0$ decreases per unit increase of ω depends on level of R_0. Generally, $1-1/R_0$ tends to be more robust to change in ω when R_0 is relatively high. For example, R_0 decreases from 10 to 5 results in 10% reduction of the minimum individual effectiveness of PrEP to reach elimination, $1-1/R_0$, i.e. 1-1/10-(1-1/5)=0.1, R_0 decreases from 8 to 4 results in 0.125 reduction of $1-1/R_0$. However, such counteracting effect is not prominent since overall R_0 for the examined values of r_{HL} is at low to moderate level (within the range of 1.4 to 7).

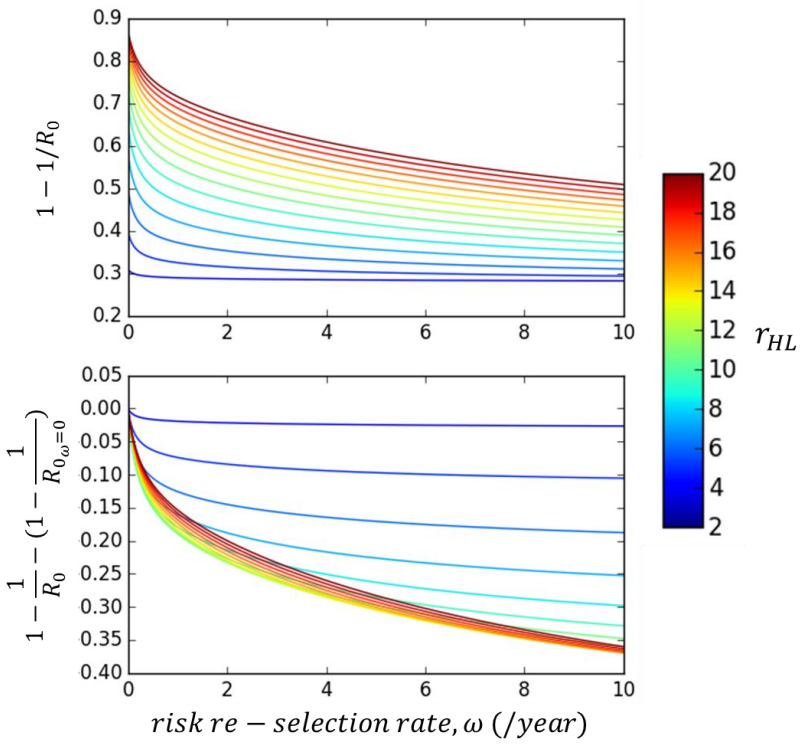

Figure AI- 2 Effect of risk re-selection rate, ω, on the minimum individual effectiveness of universal PrEP to reach elimination, 1-1/R_0 (upper panel) and the difference between 1-1/R_0 and its level when ω=0 1-1/$R_{0\_\omega=0}$(lower panel) when ratio of high contact rate over low contact rate, r_{HL} increases from 2 to 20.

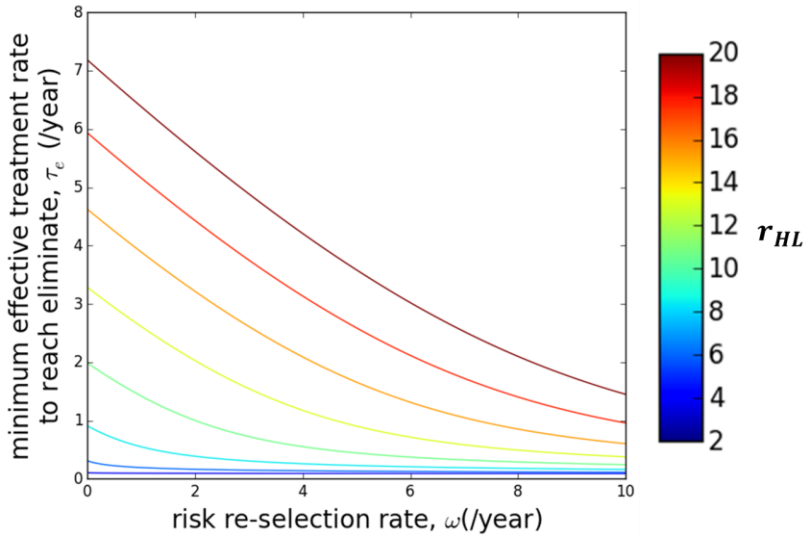

Figure AI- 3 Effect of risk re-selection rate, ω, on the minimum effective treatment rate of UT&T to reach elimination, τ_e as high-to-low contact rate ratio, r_{HL} increases from 2 to 20.

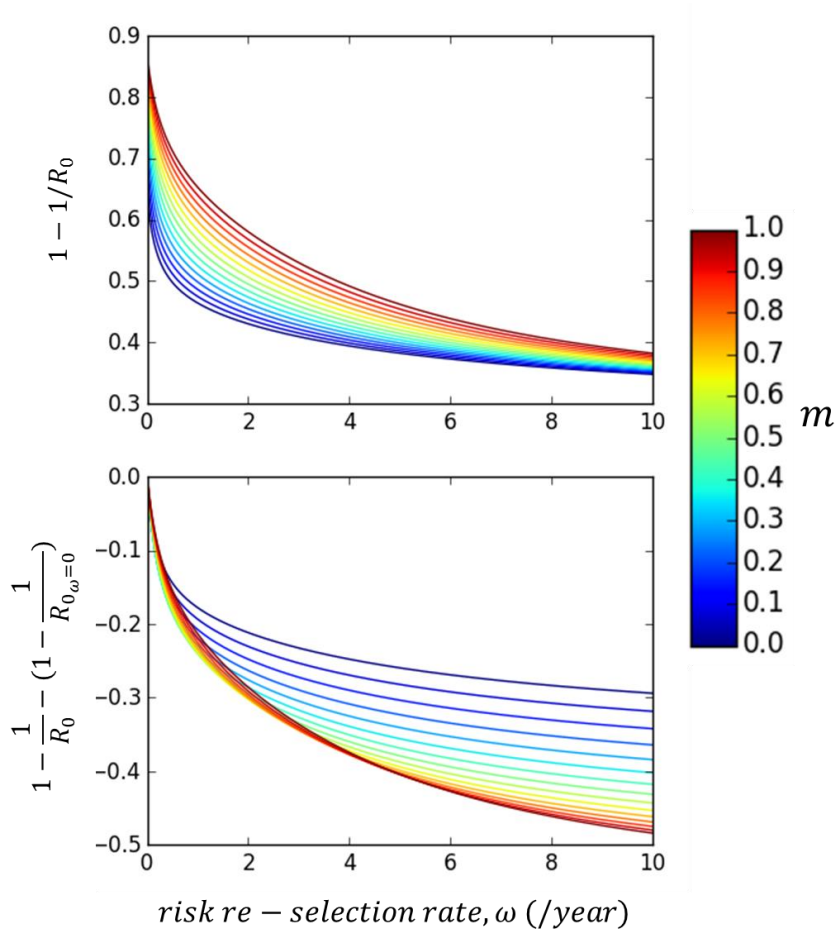

Figure AI- 4 Effect of increasing risk re-selection rate, ω, on minimum individual effectiveness of universal PrEP to reach elimination, 1-1/R_0 (upper panel) and the difference between 1-1/R_0 and its level when ω=0 1-1/$R_{0\_\omega=0}$(lower panel) when fraction of contacts reserved for people with the same level of risk, m, increases from 0 to 1.

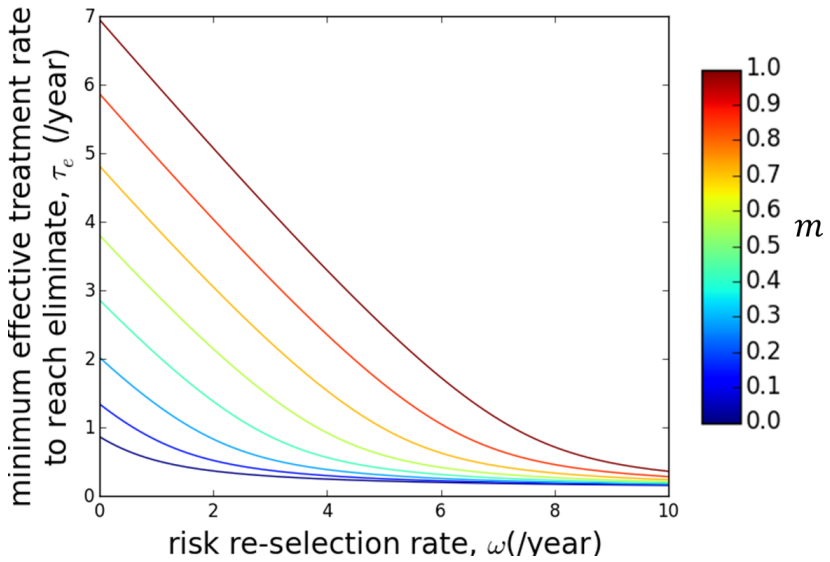

Figure AI- 5 Effect of risk re-selection rate, ω, on minimum individual effectiveness of universal PrEP to reach elimination, τ_e, when fraction of contacts reserved for people with the same level of risk, m, increases from 0 to 1.

Fig.AI-6 shows that when system has a greater relative transmissibility of acute HIV infection, ζ increase in ω less reduces the minimum required individual effectiveness of universal PrEP to reach elimination, 1-1/R_0. This is because that the longer a case stays in infection time, the more likely that this case changes risk level. Therefore, episodic risk has greater impact on later stage of infection than on early stage of infection. If a greater fraction of transmission potential is attributed to acute HIV infection, R_0 will be less sensitive to episodic risk. Consequently, increase in ω less reduces 1-1/R_0 when there is a greater relative transmissibility of acute HIV infection, ζ.

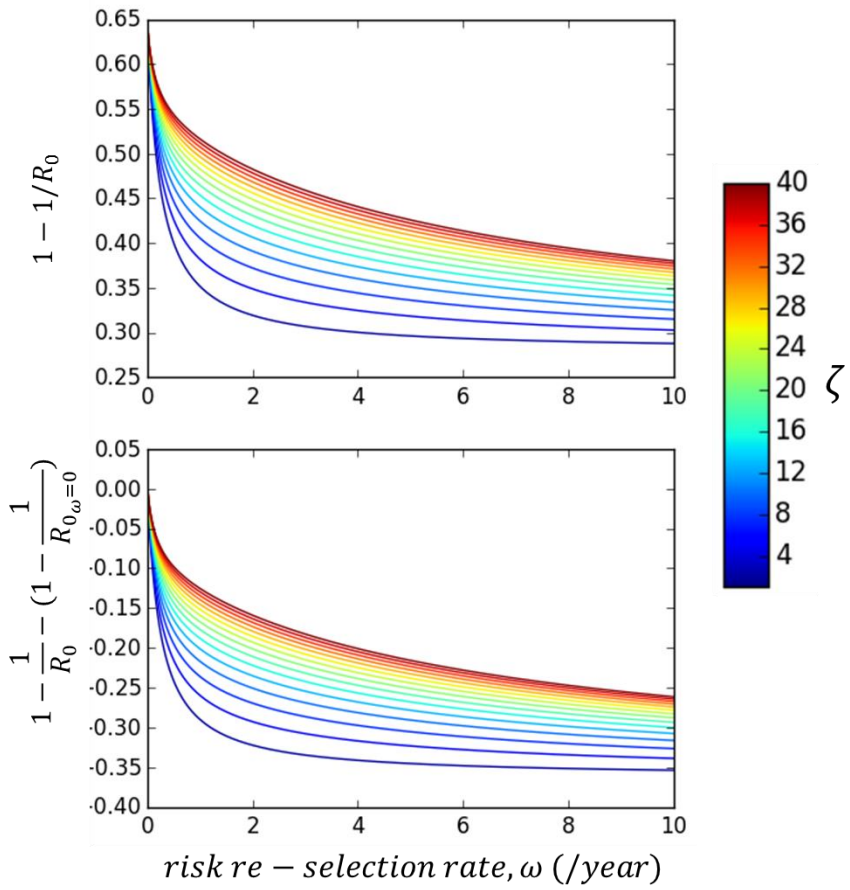

Figure AI- 6 Effect of risk re-selection rate, ω, on minimum individual effectiveness of universal PrEP to reach elimination, 1-1/R_0 (upper panel) and the difference between 1-1/R_0 and its level when ω=0 1-1/$R_{0\_\omega=0}$(lower panel) when relative transmissibility of acute infection, ζ, increases from 1 to 40.

Fig.AI-7 shows that increase in risk re-selection rate, ω, more considerably reduces the minimum required effective treatment rate of UT&T, τ_e, when acute infection has a higher relative transmissibility, ζ. This seems to disagree with the observation for universal PrEP (Fig.AI-6). The reason is that the mechanism through which UT&T reduces R_0 is different from that of PrEP: UT&T is targeted at infected population and is operated at certain rate instead of taking effects instantly. Therefore, effect of UT&T on R_0 is sensitive to the heterogeneity in infectiousness by stage. Specifically, cases are more likely to be treated as they spend more time at infectious period. That says, UT&T tends to miss more acutely infected cases than chronically infected cases. Increase in relative transmissibility of acute infection will thus reduce impact of UT&T on R_0 and increase τ_e. Furthermore, the elevated infectiousness during acute infection also interacts with high contact rate of high risk individuals. Due to such synergy risk heterogeneity can more greatly increase τ_e: a greater fraction of transmission potential is attributed to high risk acutely

infected people. Therefore, as episodic risk reduces effect of risk heterogeneity on τ_e, it will reduce τ_e more when there is a greater heterogeneity of infectiousness by stage.

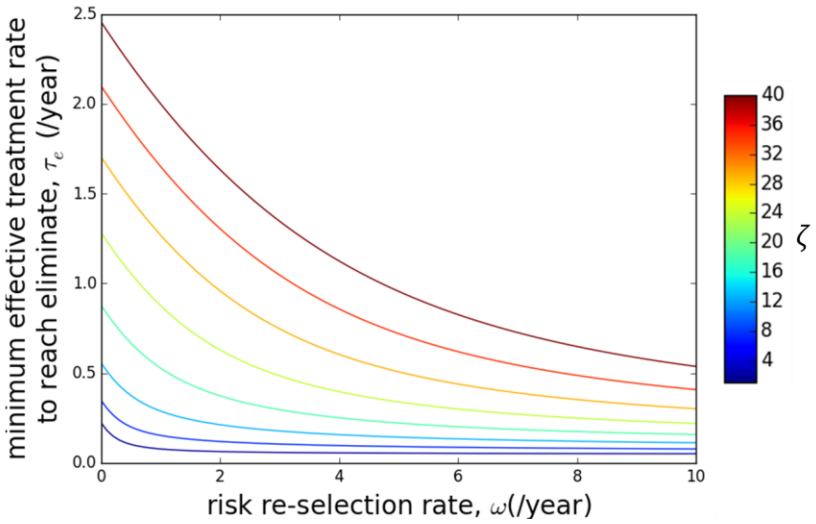

Figure AI- 7 Effect of risk re-selection rate, ω, on minimum effective treatment rate of UT&T to reach elimination, τ_e when relative transmissibility of acute infection, ζ, increases from 1 to 40.

Section II-iii. Effect of Episodic Risk on Difference Between Minimum Required Individual Effectiveness Given General PrEP or High-Risk-Prioritized PrEP to Reach Elimination Under Different Model Parameter Settings. In this section, we aim to find the conditions where episodic risk can considerably reduce the advantage of prioritizing PrEP efforts to high risk susceptible individuals in terms of minimum required individual effectiveness to reach elimination and the conditions where high-risk-prioritized PrEP needs considerably lower individual effectiveness to reach elimination than general PrEP despite of episodic risk. In section II-i of this document, we demonstrated that effect of episodic risk on the required individual effectiveness of general PrEP and high-risk-prioritized PrEP to reach elimination is most pronounced when there is moderate coverage of PrEP and moderate level of UT&T. Therefore, our analysis will focus on these two situations. We examine the situation when PrEP can cover 25% or 50% of the susceptible population while the effective treatment rate of UT&T, τ, is 0.1/year, 0.3/year, or 0.5/year. We focus on three model parameters that each determines a behavioral or biological characteristic of HIV transmission. They are high-to-low contact rate

ratio, r_{HL}, fraction of contacts reserved for people with the same level of risk, m, and relative transmissibility of acute HIV infection, ζ.

To show the extent to which prioritizing PrEP efforts to high risk susceptible individuals is more optimal than general PrEP, we examine the difference in the minimum required individual effectiveness of general PrEP to reach elimination, δg, and minimum required individual effectiveness of high-risk-prioritized PrEP to reach elimination, δ_{hp}, i.e. δg- δhp. A large δ_g- δ_{hp} indicates that prioritizing PrEP to susceptible individuals experiencing high risk can cause a large reduction in the minimum required individual effectiveness to eliminate HIV infections, and vice versa.

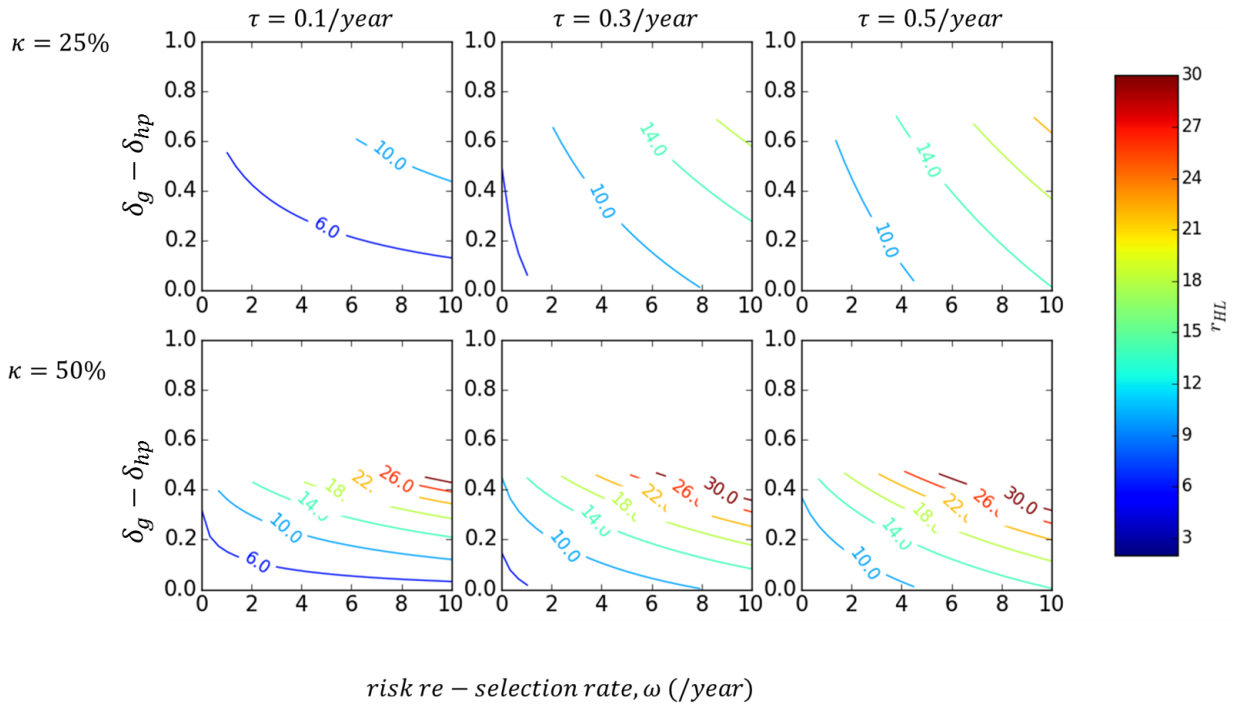

Figure AI- 8 Difference in minimum individual effectiveness to reach elimination of general PrEP and that of high-risk prioritized PrEP when risk re-selection rates, ω, increases from zero to 10/year at different levels of r_{HL}. Coverage of PrEP, κ, is set at either 25% (upper row) or 50% (lower row) and effective treatment rate, τ, is set at 0.1/year (left column), 0.3/year (middle column) or 0.5/year (right column).

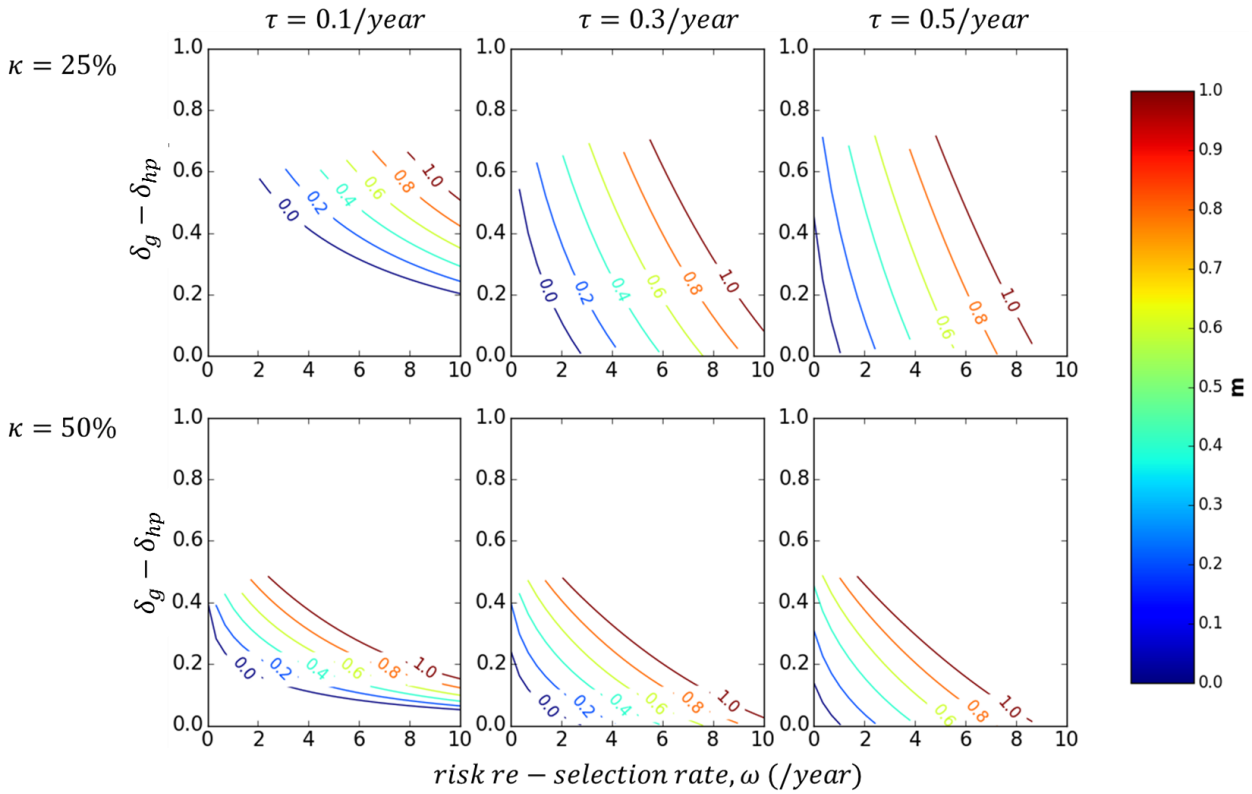

Figure AI- 9 Difference in minimum individual effectiveness to reach elimination of general PrEP and that of high-risk prioritized PrEP, δ_g- δ_{hp}, when risk re-selection rates, ω, increases from zero to 10/year at different levels of m (fraction of contacts reserved for people with same level of risk). Coverage of PrEP, κ, is set at either 25% (upper row) or 50% (lower row) and effective treatment rate, τ, is set at 0.1/year (left column), 0.3/year (middle column) or 0.5/year (right column).

Fig.AI-8 and Fig.AI-9 show that as risk re-selection rate, ω, increases δ_g- δ_{hp} is more likely to remain high when there is a higher level of risk heterogeneity, i.e. high-to-low contact rate ratio $r_{HL,}$ or a higher degree of assortativity, i.e. fraction of contacts reserved for people with same level of risk, m. This is largely because that given a higher r_{HL} or m, there is a wider range of ω where general PrEP cannot eliminate HIV infections (curves start at a higher value of ω, below which δ_g- δ_{hp} is undefined). As mentioned, the difference δ_g- δ_{hp} is the reduction of the required individual effectiveness to reach elimination due to prioritizing PrEP to high risk susceptible individuals. A high δ_g- δ_{hp} means that high-risk-prioritized PrEP has a large advantage as compared to general PrEP. Therefore, the trend shown in Fig.AI-8 and Fig.AI-9 indicates that individual risk behavior volatility less reduces the advantage of high-risk-prioritized PrEP when there is a greater level of risk heterogeneity or higher degree of assortativity. The reason is that increase in either parameter causes stronger transmission within high risk subpopulation. Therefore, the potential contribution of susceptible individuals experiencing high risk to ongoing

transmission increases. As result, high-risk-prioritized PrEP can more efficiently prevent HIV transmission and such advantage would also more be robust to individual risk behavior volatility.

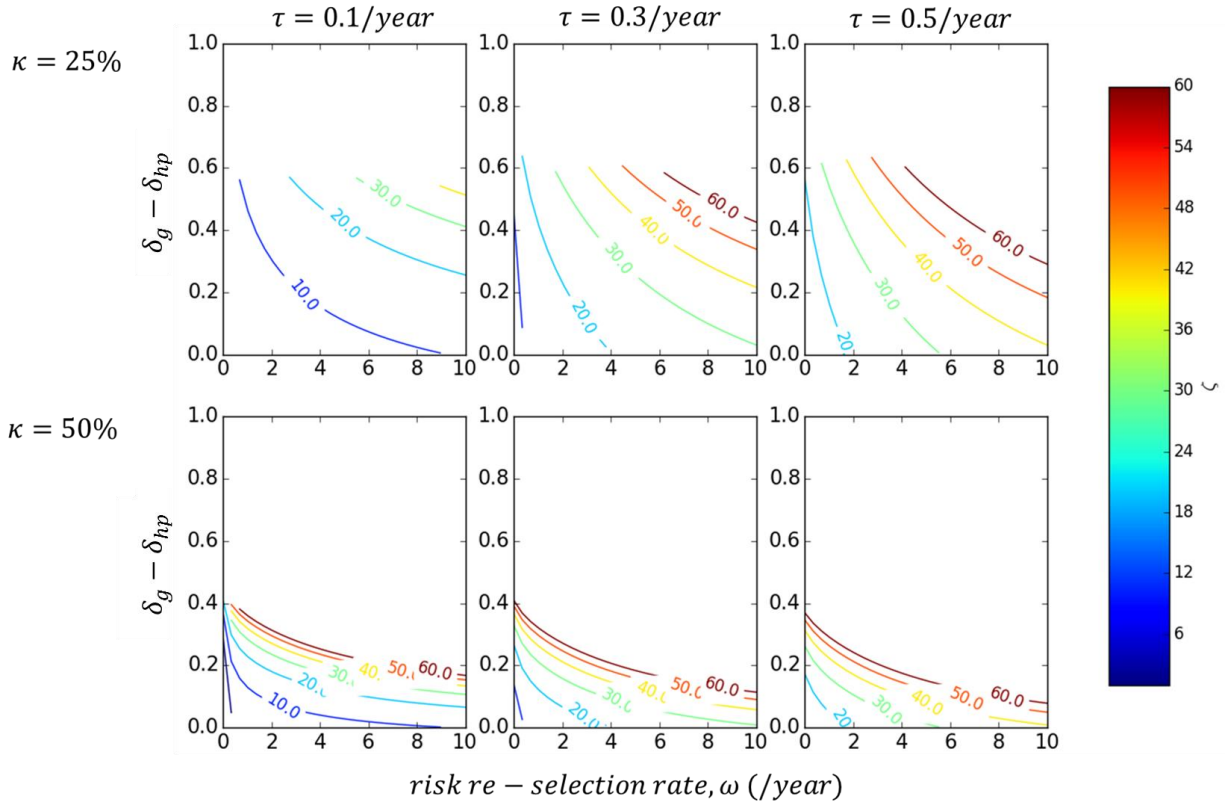

Figure AI- 10 Difference in minimum individual effectiveness to reach elimination of general PrEP, δ_g, and that of high-risk-prioritized PrEP, δ_{hp}, i.e. δ_g-δ_{hp}, when risk re-selection rates, ω, increases from zero to 10/year at different levels of relative transmissibility of acute infection, ζ.

Fig.AI-10 shows that increase in ω less reduces difference δ_g- δ_{hp} when there is a higher relative transmissibility of acute HIV infection, ζ. The reason has been discussed in section II-ii of this document: cases are more likely to change risk levels at later stage of infection than at earlier stage of infection. Therefore when a higher fraction of transmission potential concentrates at acute stage of infection, individual risk behavior volatility will less affect the transmission system.

Appendix II-Supplementary Materials for Chapter V

Section I. Supplementary Methods

Section I-i. Model equations and parameters

Table AII- 1. Equations of deterministic version of model

$$\frac{dS_H}{dt} = \mu f_H - S_H\lambda_H - f_H\omega S_H + f_L\omega S_L - \mu S_H$$

$$\frac{dS_L}{dt} = \mu(1 - f_H) - S_L\lambda_L + f_H\omega S_H - f_L\omega S_L - \mu S_L$$

$$\frac{dA_H}{dt} = S_H\lambda_H - A_H\gamma_1 - f_H\omega A_H + f_L\omega A_L - \mu A_H$$

$$\frac{dA_L}{dt} = S_L\lambda_H - A_L\gamma_1 + f_H\omega A_H - f_L\omega A_L - \mu A_L$$

$$\frac{dC_H}{dt} = A_H\gamma_1 - f_H\omega C_H + f_L\omega C_L - \gamma_2 C_H - \mu C_H$$

$$\frac{dC_L}{dt} = A_L\gamma_1 + f_H\omega C_H - f_L\omega C_L - \gamma_2 C_L - \mu C_L$$

Table AII- 2. Parameter symbols, default values, units and definitions

Parameter	Default Value	Unit	Definition
μ	1/40	/year	Rate of removal from the sexually-active population unrelated to HIV. Because we set the equilibrium population in the absence of disease to 1, this is also the (absolute) rate of entry of new individuals into the sexually active population
γ_1	4	/year	Rate of transitioning from acute to chronic infection
γ_2	1/10	/year	Rate of death from AIDS during chronic infection
β	variable	/contact	Average per-contact transmissibility across both stages of infection
ζ	26	-	Relative transmissibility of acute stage of infection, ratio of transmissibility during acute infection over transmissibility during

$\boldsymbol{\beta_1}$	$\frac{\beta\zeta(\gamma_1+\gamma_2+\mu)}{\gamma_1+\zeta(\gamma_2+\mu)}$	/contact	Per-contact transmissibility during acute stage
$\boldsymbol{\beta_2}$	$\frac{\beta(\gamma_1+\gamma_2+\mu)}{\gamma_1+\zeta(\gamma_2+\mu)}$	/contact	Per-contact transmissibility during chronic stage
f_H	0.1	-	Fraction of population that is at high-risk phase at disease-free equilibrium
f_L	$1-f_H$	-	Fraction of population that is at low-risk phase at disease-free equilibrium
r_{HL}	8	-	Ratio of average contact rate at high-risk phase over average contact rate at low-risk phase
χ	36	/year	Average contact rate of the overall population expected for population at disease-free equilibrium
χ_H	$\frac{\chi r_{HL}}{f_H r_{HL}+f_L}$	/year	Average contact rate of the high-risk population
χ_L	$\frac{\chi}{f_H r_{HL}+f_L}$	/year	Average contact rate of the low-risk population
m	0	-	Fraction of contacts reserved for people experiencing the same level of risk
ω	variable	/year	Rate of reselecting risk group

Section I-ii. Selecting Phylogenetic Trees for Analysis on Clustering

In order to control for the effect of number of internal nodes within a given cutoff distance on clustering, we select phylogenetic trees with similar cumulative number of internal nodes at a given cutoff distance (time in past since sample time). In order to select trees which have similar cumulative number of internal nodes, we calculated the sum of square of distance between the cumulative number of internal nodes of a simulated phylogenetic tree and the average of the cumulative number of internal nodes of all simulated phylogenetic trees to be matched. Trees with sum of squared distance under a certain threshold value are selected.

For example, suppose that there are N sets of simulated trees, with each set simulated under a specific parameter setting, i.e. a specific risk re-selection rate ω. Suppose there are R sets of trees per set. We name rth tree in the nth set of trees as Tree(n_r) (order of trees are arbitrarily assigned for the purpose of illustration) (0<n<=N, 0<r<=R). Suppose there are in total *M* sampled cases for each tree, and we denote the time since sample time in past that mth branching occurs on Tree(n_r) as $t_{n\_r\_m}$ (0<m<=M-1). In order to obtain a reference cumulative number of internal nodes for each simulated tree to be compared to, we get average of $t_{n\_r\_m}$ across all trees

in all sets. That says, for this reference curve, the time for mth branching to occur is to occur, t_m, is,

$$t_m = \sum_{n=1}^{N} \sum_{r=1}^{R} t_{n\_r\_m} / (N * R).$$

For rth tree in nth set, Tree(n_r), we calculate the square of distance between the time of mth internal node since sample time on Tree(n_r) and t_m for m from 1 to total number of branching, M-1,

$$Sum\ of\ square\ distance = \sum_{m=1}^{M-1} \left(t_{n\_r\_m} - t_m\right)^2$$

We calculated the sum of squared distance as above for each simulated phylogenetic tree. We choose trees which have relatively small distance from the average. We set the threshold sum of square distance arbitrarily so that trees have visually close enough cumulate number of branching by cutoff distance and there are enough trees per set for analysis.

Section II. Supplementary Results

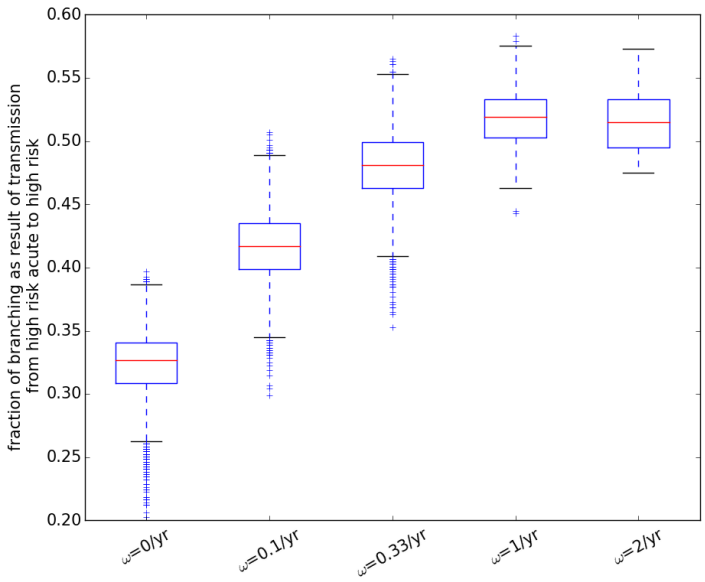

Figure AII- 1. Effect of episodic risk on fraction of branching events caused by transmission linking high-risk acute infections. Simulation is done assuming random mixing. Average transmissibility per contact, β=0.008. Other parameters (other than ω) are set at their default values as shown in Table AII-2.

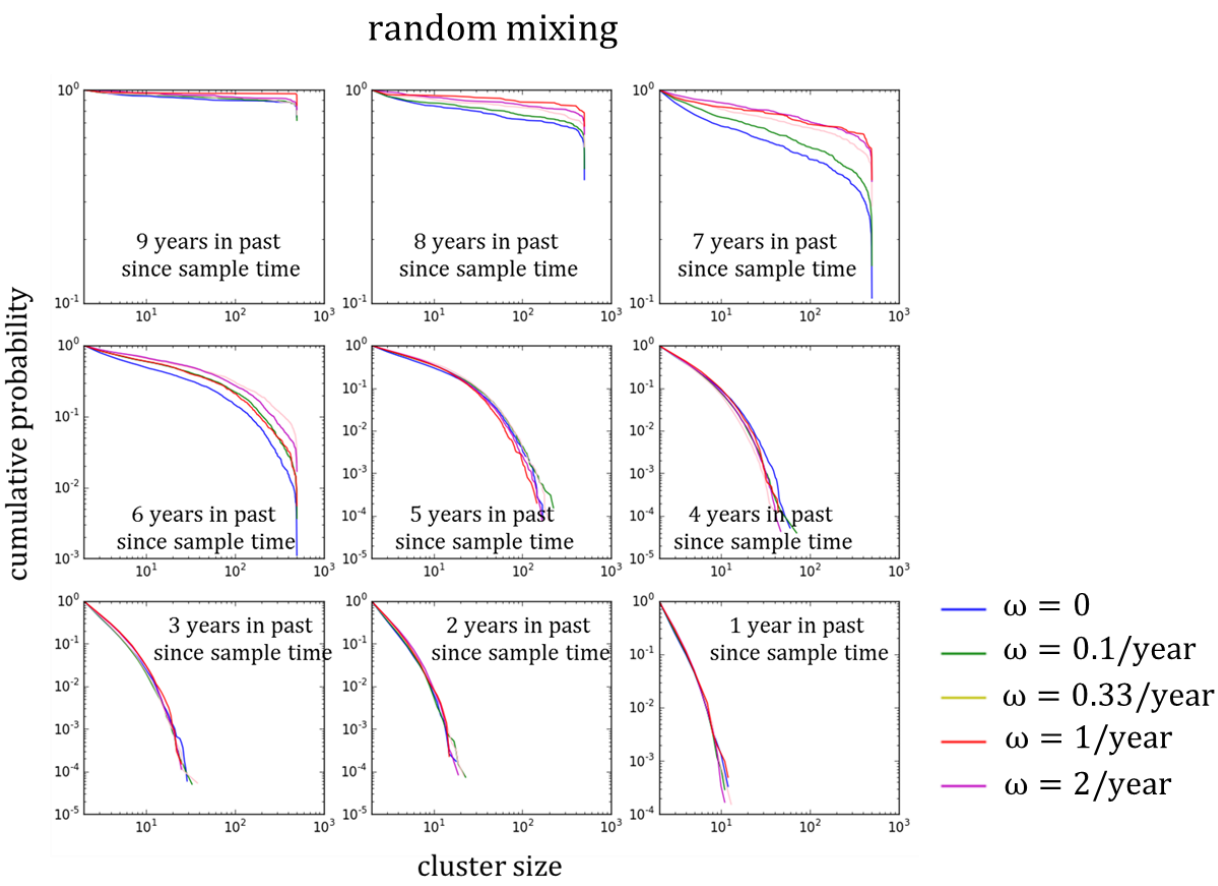

Figure AII- 2. Cumulative probability of cluster size at different cutoff distance (time in past since sample time) of phylogenetic trees simulated with different values of risk re-selection rate, ω. The phylogenetic trees have been matched by their cumulative number of internal nodes by cutoff distances. Simulations are done assuming random mixing. Average transmissibility per contact, β=0.008. Other parameters (except ω) are set at their default values as shown in Table AII-2.

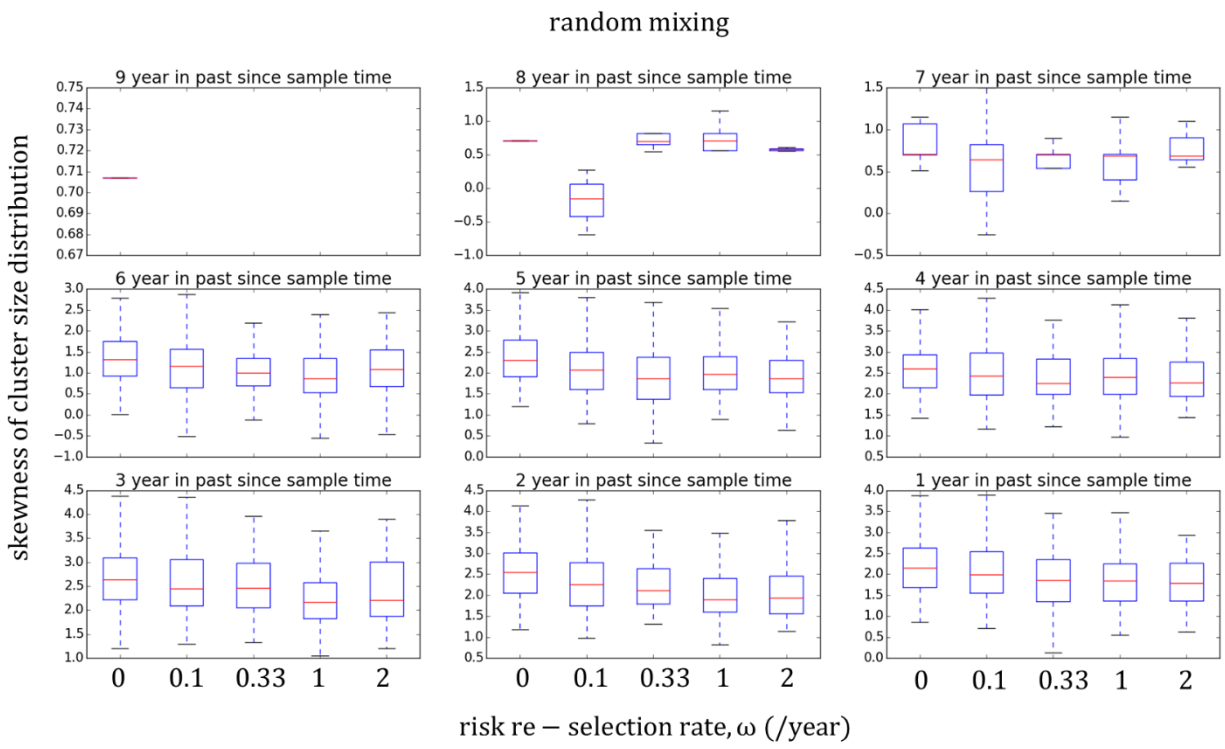

Figure AII- 3. Skewness of cluster size distribution at each cutoff distance (time in past since sample time) for phylogenetic trees simulated assuming different values of risk re-selection rate, ω. Simulations are done assuming random mixing. Average transmissibility per contact, β=0.008. Other parameters (except ω) are set at their default values as shown in Table AII-2.

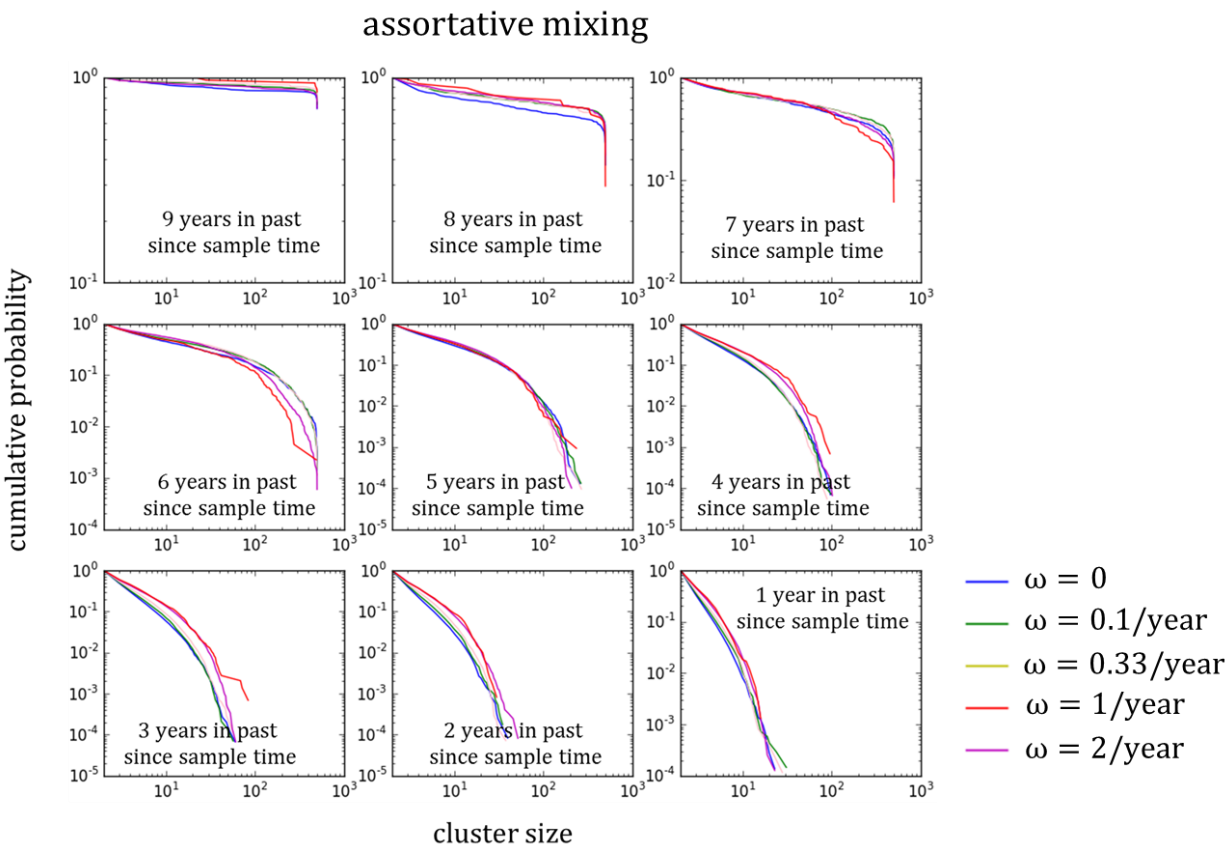

Figure AII- 4. Cumulative probability of cluster size at different cutoff distance (time in past since sample time) of phylogenetic trees simulated with different values of risk re-selection rate, ω. The phylogenetic trees have been matched by their cumulative number of internal nodes by cutoff distances. Simulations are done assuming assortative mixing, i.e. individuals reserve 50% of their contacts for people experiencing the same risk phase. Average transmissibility per contact, β=0.005. Other parameters (except ω) are set at their default values as shown in Table AII-2.

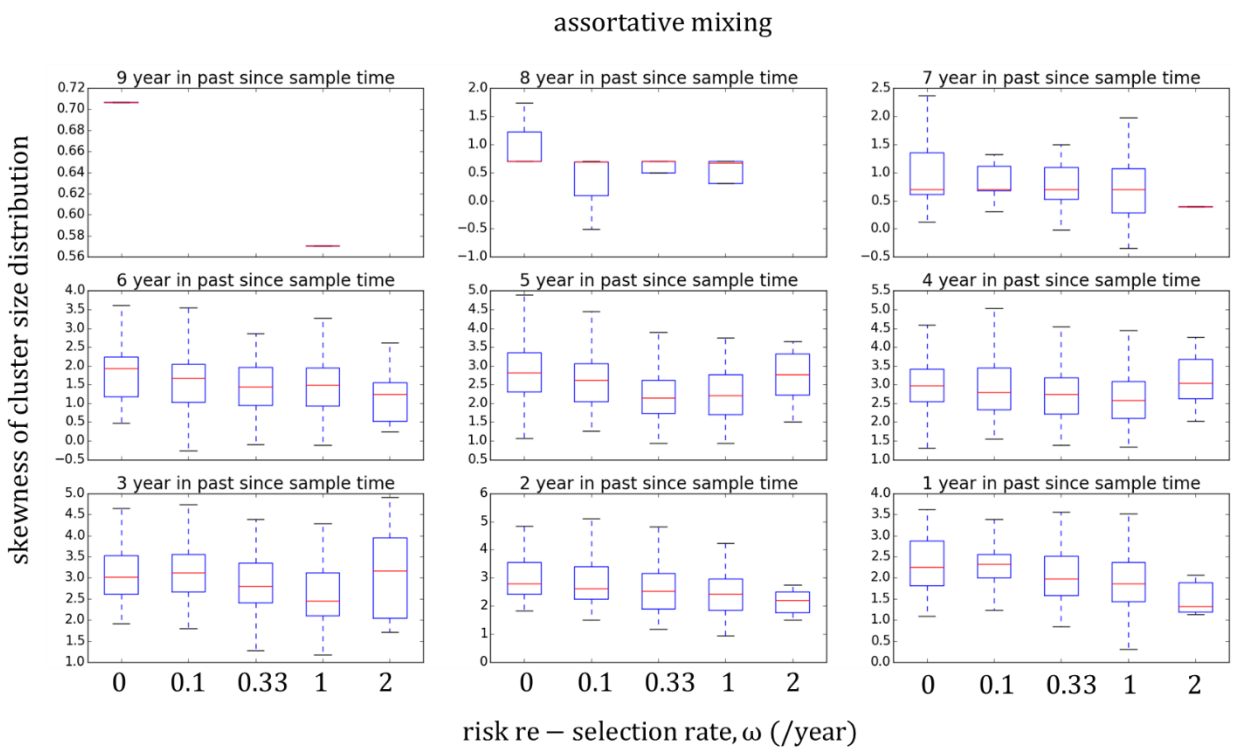

Figure AII- 5. Skewness of cluster size distribution at each cutoff distance (time in past since sample time) for phylogenetic trees simulated assuming different values of risk re-selection rate, ω. Simulations are done assuming assortative mixing. Average transmissibility per contact, β=0.005. Other parameters (except ω) are set at their default values as shown in Table AII-2.

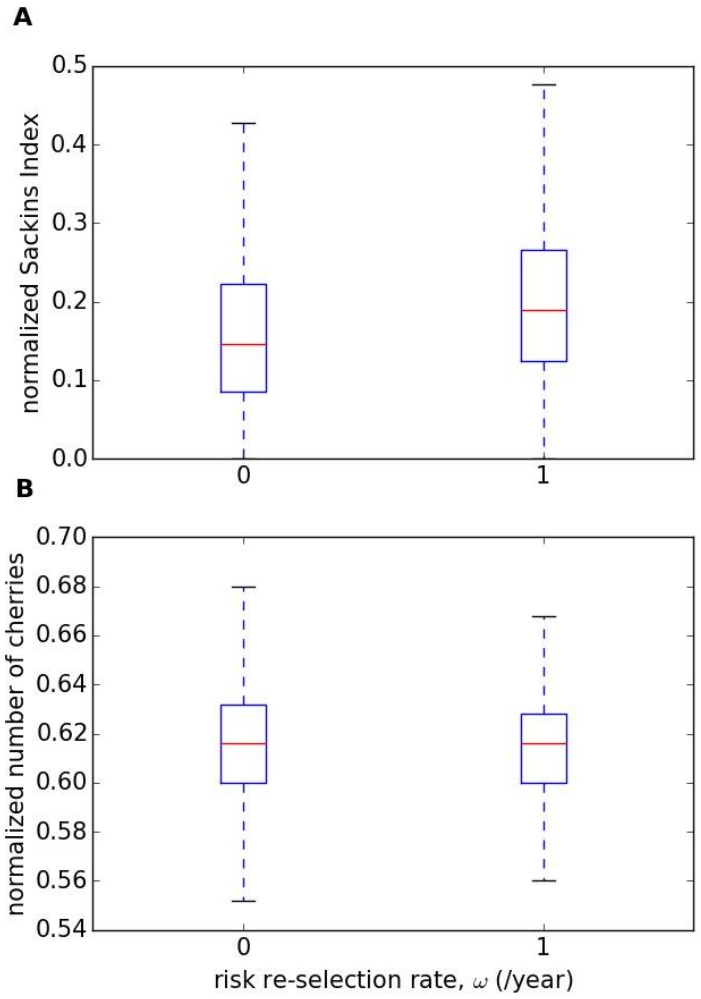

Figure AII- 6. Effect of episodic risk on normalized Sackin Index (panel A) and normalized number of cherries (panel B) for population with random mixing and homogeneous stage- specific transmissibility, i.e. relative transmissibility of acute infection ζ=1.

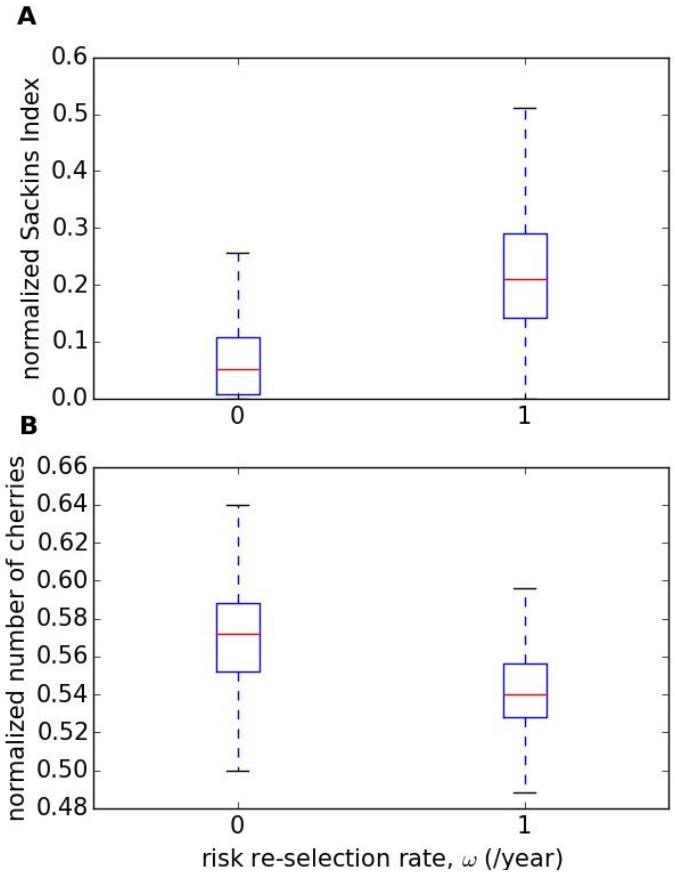

Figure AII- 7. Effect of episodic risk on normalized Sackin Index (panel A) and normalized number of cherries (panel B) for population with assortative mixing and homogeneous stage- specific transmissibility, i.e. relative transmissibility of acute infection ζ=1.

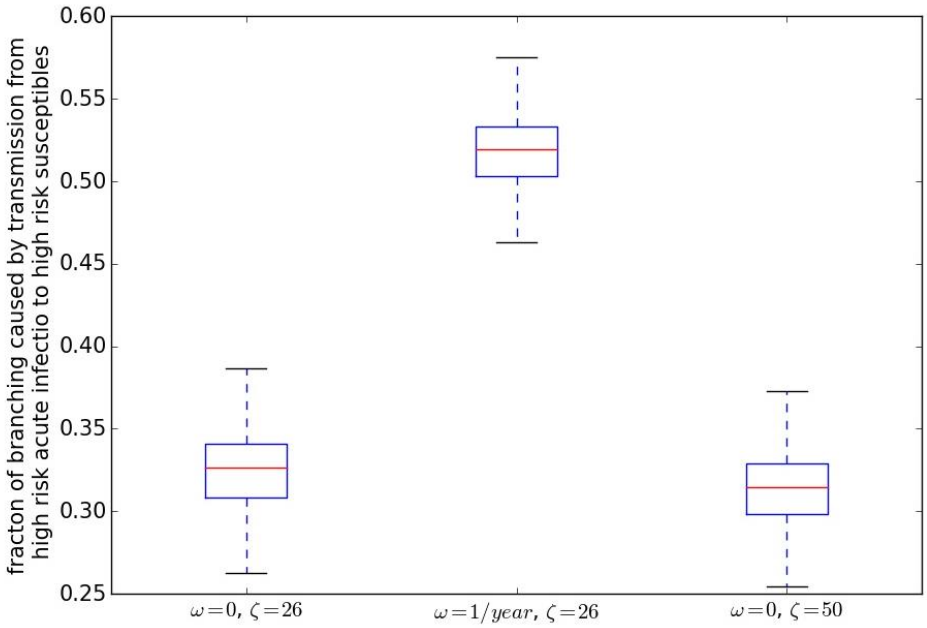

Figure AII- 8. Boxplots of fraction of branching events caused by transmission linking high risk acute infections of phylogenetic trees for population where risk re-selection rate, ω=0, relative transmissibility of acute infection, ζ=26, and for population where ω=1/year and ζ=26, and for population where ω=0 and ζ=50.

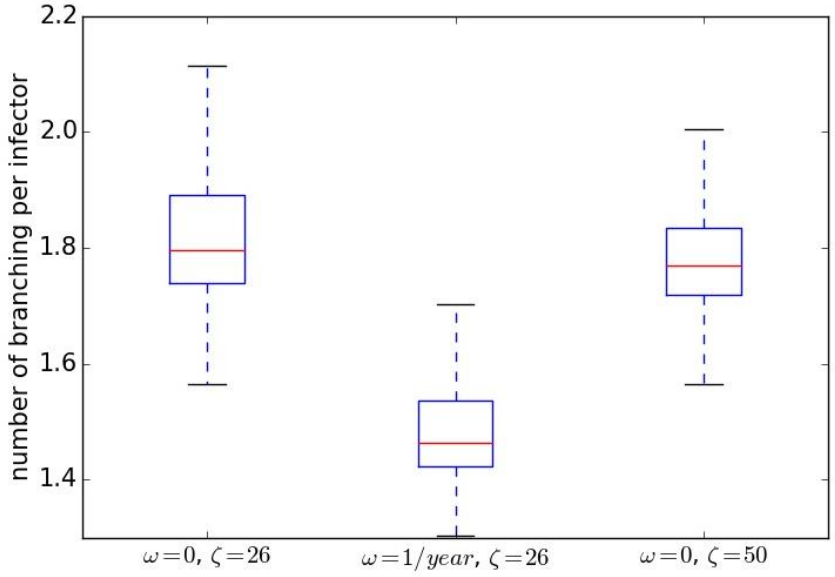

Figure AII- 9. Boxplots of average number of branching events caused by transmission from each infector for population where risk re-selection rate, ω=0, relative transmissibility of acute infection, ζ=26, and for population where ω=1/year and ζ=26, and for population where ω=0 and ζ=50.

Reference

1. Hall HI, Song R, Rhodes P, et al. Estimation of HIV Incidence in the United States. *JAMA*. 2008;300(5):520-529. doi:10.1001/jama.300.5.520.

2. Prejean J, Song R, Hernandez A, et al. Estimated HIV incidence in the United States, 2006-2009. *PloS One*. 2011;6(8):e17502. doi:10.1371/journal.pone.0017502.

3. Blower SM, Samuel MC, Wiley JA. Sex, Power Laws, and HIV Transmission. [Letter]. *J Acquir Immune Defic Syndr*. 1992;5(6):633-634.

4. Goodreau SM, Golden MR. Biological and demographic causes of high HIV and sexually transmitted disease prevalence in men who have sex with men. *Sex Transm Infect*. 2007;83(6):458-462. doi:10.1136/sti.2007.025627.

5. Blower SM, van Griensven GJ, Kaplan EH. An analysis of the process of human immunodeficiency virus sexual risk behavior change. *Epidemiol Camb Mass*. 1995;6(3):238-242.

6. Vittinghoff E, Douglas J, Judson F, McKirnan D, MacQueen K, Buchbinder SP. Per-contact risk of human immunodeficiency virus transmission between male sexual partners. *Am J Epidemiol*. 1999;150(3):306-311.

7. Cohen MS, Shaw GM, McMichael AJ, Haynes BF. Acute HIV-1 Infection. *N Engl J Med*. 2011;364(20):1943-1954. doi:10.1056/NEJMra1011874.

8. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *Lancet*. 2009;373(9657):48-57. doi:10.1016/S0140-6736(08)61697-9.

9. Lange JMA. "Test and Treat": Is It Enough? *Clin Infect Dis*. 2011;52(6):801-802. doi:10.1093/cid/ciq254.

10. Dieffenbach CW, Fauci AS. Universal voluntary testing and treatment for prevention of HIV transmission. *JAMA J Am Med Assoc*. 2009;301(22):2380-2382. doi:10.1001/jama.2009.828.

11. Dodd PJ, Garnett GP, Hallett TB. Examining the Promise of HIV Elimination by "Test and Treat" in Hyper-Endemic Settings. *AIDS Lond Engl*. 2010;24(5):729-735. doi:10.1097/QAD.0b013e32833433fe.

12. Wagner BG, Blower S. Voluntary universal testing and treatment is unlikely to lead to HIV elimination: a modeling analysis. *Nat Preced*. 2009;(713). doi:10.1038/npre.2009.3917.1.

13. Brown A, Gill O, Delpech V. HIV treatment as prevention among men who have sex with men in the UK: is transmission controlled by universal access to HIV treatment and care? *HIV Med*. 2013;14(9):563-570. doi:10.1111/hiv.12066.

14. Kretzschmar ME, Schim van der Loeff MF, Birrell PJ, De Angelis D, Coutinho RA. Prospects of elimination of HIV with test-and-treat strategy. *Proc Natl Acad Sci U S A*. 2013;110(39):15538-15543. doi:10.1073/pnas.1301801110.

15. Hayes R, Sabapathy K, Fidler S. Universal Testing and Treatment as an HIV Prevention Strategy: Research Questions and Methods. *Curr HIV Res*. 2011;9(6):429-445. doi:10.2174/157016211798038515.

16. Okano JT, Robbins D, Palk L, Gerstoft J, Obel N, Blower S. Testing the hypothesis that treatment can eliminate HIV: a nationwide, population-based study of the Danish HIV epidemic in men who have sex with men. *Lancet Infect Dis*. 2016;16(7):789-796. doi:10.1016/S1473-3099(16)30022-6.

17. De Cock KM, Gilks CF, Lo Y-R, Guerma T. Can antiretroviral therapy eliminate HIV transmission? *Lancet*. 2009;373(9657):7-9. doi:10.1016/S0140-6736(08)61732-8.

18. Garnett GP, Baggaley RF. Treating our way out of the HIV pandemic: could we, would we, should we? *Lancet*. 2009;373(9657):9-11. doi:10.1016/S0140-6736(08)61698-0.

19. Naswa S, Marfatia YS. Pre-exposure prophylaxis of HIV. *Indian J Sex Transm Dis*. 2011;32(1):1-8. doi:10.4103/0253-7184.81246.

20. Prevention of rectal SHIV transmission in macaques by daily or intermittent prophylaxis with emtricitabine and tenofovir. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/18254653. Accessed June 30, 2017.

21. Antiretroviral Pre-exposure Prophylaxis Prevents Vaginal Transmission of HIV-1 in Humanized BLT Mice. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0050016. Accessed June 30, 2017.

22. Grant RM, Lama JR, Anderson PL, et al. Preexposure Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. *N Engl J Med*. 2010;363(27):2587-2599. doi:10.1056/NEJMoa1011205.

23. Golub SA, Gamarel KE, Rendina HJ, Surace A, Lelutiu-Weinberger CL. From Efficacy to Effectiveness: Facilitators and Barriers to PrEP Acceptability and Motivations for Adherence Among MSM and Transgender Women in New York City. *AIDS Patient Care STDs*. 2013;27(4):248-254. doi:10.1089/apc.2012.0419.

24. Grant RM, Anderson PL, McMahan V, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. *Lancet Infect Dis*. 2014;14(9):820-829. doi:10.1016/S1473-3099(14)70847-3.

25. Punyacharoensin N, Edmunds WJ, Angelis DD, et al. Effect of pre-exposure prophylaxis and combination HIV prevention for men who have sex with men in the UK: a mathematical modelling study. *Lancet HIV*. 2016;3(2):e94-e104. doi:10.1016/S2352-3018(15)00056-9.

26. Gomez GB, Borquez A, Case KK, Wheelock A, Vassall A, Hankins C. The cost and impact of scaling up pre-exposure prophylaxis for HIV prevention: a systematic review of cost-effectiveness modelling studies. *PLoS Med*. 2013;10(3):e1001401. doi:10.1371/journal.pmed.1001401.

27. Ying R, Barnabas RV, Williams BG. Modeling the implementation of universal coverage for HIV treatment as prevention and its impact on the HIV epidemic. *Curr HIV/AIDS Rep*. 2014;11(4):459-467. doi:10.1007/s11904-014-0232-x.

28. Long EF, Brandeau ML, Owens DK. The Cost-Effectiveness and Population Outcomes of Expanded HIV Screening and Antiretroviral Treatment in the United States. *Ann Intern Med*. 2010;153(12):778-789. doi:10.7326/0003-4819-153-12-201012210-00004.

29. UNAIDS. 90-90-90 An ambitious treatment target to help end the AIDS epidemic. *JC2684 Geneva Switz UNAIDS*. 2014.

30. Rozhnova G, Loeff MFS van der, Heijne JCM, Kretzschmar ME. Impact of Heterogeneity in Sexual Behavior on Effectiveness in Reducing HIV Transmission with Test-and-Treat Strategy. *PLOS Comput Biol*. 2016;12(8):e1005012. doi:10.1371/journal.pcbi.1005012.

31. Wagner BG, Kahn JS, Blower S. Should we try to eliminate HIV epidemics by using a "Test and Treat" strategy? *AIDS Lond Engl*. 2010;24(5):775-776. doi:10.1097/QAD.0b013e3283366782.

32. Levi J, Raymond A, Pozniak A, Vernazza P, Kohler P, Hill A. Can the UNAIDS 90-90-90 target be achieved? A systematic analysis of national HIV treatment cascades. *BMJ Glob Health*. 2016;1(2):e000010. doi:10.1136/bmjgh-2015-000010.

33. Williams BG, Gouws E. R0 and the elimination of HIV in Africa: Will 90-90-90 be sufficient? *ArXiv13043720 Q-Bio*. April 2013. http://arxiv.org/abs/1304.3720. Accessed August 8, 2016.

34. Iwuji C, Newell M-L. Towards control of the global HIV epidemic: Addressing the middle-90 challenge in the UNAIDS 90–90–90 target. *PLOS Med*. 2017;14(5):e1002293. doi:10.1371/journal.pmed.1002293.

35. Abbas UL, Glaubius R, Mubayi A, Hood G, Mellors JW. Antiretroviral Therapy and Pre-exposure Prophylaxis: Combined Impact on HIV Transmission and Drug Resistance in South Africa. *J Infect Dis*. 2013;208(2):224-234. doi:10.1093/infdis/jit150.

36. Cremin I, Alsallaq R, Dybul M, Piot P, Garnett G, Hallett TB. The new role of antiretrovirals in combination HIV prevention: a mathematical modelling analysis. *AIDS*. 2013;27(3):447-458. doi:10.1097/QAD.0b013e32835ca2dd.

37. Celum C, Baeten JM, Hughes JP, et al. Integrated Strategies for Combination HIV Prevention: Principles and examples for men who have sex with men in the Americas and heterosexual African populations. *J Acquir Immune Defic Syndr 1999*. 2013;63(0 2):S213-S220. doi:10.1097/QAI.0b013e3182986f3a.

38. Desai K, Sansom SL, Ackers ML, et al. Modeling the impact of HIV chemoprophylaxis strategies among men who have sex with men in the United States: HIV infections prevented and cost-effectiveness: *AIDS*. 2008;22(14):1829-1839. doi:10.1097/QAD.0b013e32830e00f5.

39. Alistar SS, Grant PM, Bendavid E. Comparative effectiveness and cost-effectiveness of antiretroviral therapy and pre-exposure prophylaxis for HIV prevention in South Africa. *BMC Med*. 2014;12:46. doi:10.1186/1741-7015-12-46.

40. Freedberg KA, Losina E, Weinstein MC, et al. The Cost Effectiveness of Combination Antiretroviral Therapy for HIV Disease. *N Engl J Med*. 2001;344(11):824-831. doi:10.1056/NEJM200103153441108.

41. Koopman J. Modeling Infection Transmission. http://dx.doi.org/10.1146/annurev.publhealth.25.102802.124353. http://www.annualreviews.org/doi/10.1146/annurev.publhealth.25.102802.124353. Published March 11, 2004. Accessed June 20, 2017.

42. Cobelli C, DiStefano JJ. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am J Physiol - Regul Integr Comp Physiol*. 1980;239(1):R7-R24.

43. Bohl DD, Raymond HF, Arnold M, McFarland W. Concurrent sexual partnerships and racial disparities in HIV infection among men who have sex with men. *Sex Transm Infect*. 2009;85(5):367-369. doi:10.1136/sti.2009.036723.

44. Goodreau SM, Goicochea LP, Sanchez J. Sexual Role and Transmission of HIV Type 1 among Men Who Have Sex with Men, in Peru. *J Infect Dis*. 2005;191(0 1):S147-S158. doi:10.1086/425268.

45. Romero-Severson EO, Alam SJ, Volz EM, Koopman JS. Heterogeneity in Number and Type of Sexual Contacts in a Gay Urban Cohort. *Stat Commun Infect Dis*. 2012;4(1). doi:10.1515/1948-4690.1042.

46. Romero-Severson EO, Volz E, Koopman JS, Leitner T, Ionides EL. Dynamic Variation in Sexual Contact Rates in a Cohort of HIV-Negative Gay Men. *Am J Epidemiol*. 2015;182(3):255-262. doi:10.1093/aje/kwv044.

47. Romero-Severson EO, Alam SJ, Volz E, Koopman J. Acute-stage transmission of HIV: effect of volatile contact rates. *Epidemiol Camb Mass*. 2013;24(4):516-521. doi:10.1097/EDE.0b013e318294802e.

48. Romero-Severson EO, Meadors GD, Volz EM. A generating function approach to HIV transmission with dynamic contact rates. *Math Model Nat Phenom*. 2014;9(2):121-135. doi:10.1051/mmnp/20149208.

49. Holmes EC, Grenfell BT. Discovering the Phylodynamics of RNA Viruses. *PLoS Comput Biol*. 2009;5(10):e1000505. doi:10.1371/journal.pcbi.1000505.

50. Volz EM, Ionides E, Romero-Severson EO, Brandt M-G, Mokotoff E, Koopman JS. HIV-1 transmission during early infection in men who have sex with men: a phylodynamic analysis. *PLoS Med*. 2013;10(12):e1001568; discussion e1001568. doi:10.1371/journal.pmed.1001568.

51. Rasmussen DA, Ratmann O, Koelle K. Inference for Nonlinear Epidemiological Models Using Genealogies and Time Series. Lemey P, ed. *PLoS Comput Biol*. 2011;7:e1002136. doi:10.1371/journal.pcbi.1002136.

52. Brenner BG, Roger M, Stephens D, et al. Transmission Clustering Drives the Onward Spread of the HIV Epidemic Among Men Who Have Sex With Men in Quebec. *J Infect Dis*. 2011;204(7):1115-1119. doi:10.1093/infdis/jir468.

53. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Brown AJL. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med*. 2008;5(3):e50.

54. Pybus OG, Rambaut A, Harvey PH. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*. 2000;155(3):1429.

55. Volz EM, Koelle K, Bedford T. Viral Phylodynamics. *PLoS Comput Biol*. 2013;9(3):e1002947. doi:10.1371/journal.pcbi.1002947.

56. Alam SJ, Zhang X, Romero-Severson EO, et al. Detectable signals of episodic risk effects on acute HIV transmission: strategies for analyzing transmission systems using genetic data. *Epidemics*. 2013;5(1):44-55. doi:10.1016/j.epidem.2012.11.003.

57. Brenner BG, Roger M, Routy J-P, et al. High Rates of Forward Transmission Events after Acute/Early HIV-1 Infection. *J Infect Dis*. 2007;195(7). http://www.jstor.org.proxy.lib.umich.edu/stable/30086025.

58. Hughes GJ, Fearnhill E, Dunn D, et al. Molecular Phylodynamics of the Heterosexual HIV Epidemic in the United Kingdom. *PLOS Pathog*. 2009;5(9):e1000590. doi:10.1371/journal.ppat.1000590.

59. Volz EM, Koopman JS, Ward MJ, Brown AL, Frost SDW. Simple Epidemiological Dynamics Explain Phylogenetic Clustering of HIV from Patients with Recent Infection. *PLoS Comput Biol*. 2012;8(6):e1002552. doi:10.1371/journal.pcbi.1002552.

60. Volz EM, Pond SLK, Ward MJ, Brown AJL, Frost SDW. Phylodynamics of Infectious Disease Epidemics. *Genetics*. 2009;183(4):1421-1430. doi:10.1534/genetics.109.106021.

61. Volz EM. Complex Population Dynamics and the Coalescent Under Neutrality. *Genetics*. 2012;190(1):187-201. doi:10.1534/genetics.111.134627.

62. Koopman JS. Infection transmission science and models. *Jpn J Infect Dis*. 2005;58(6):S3-8.

63. Kim J-H, Riolo RL, Koopman JS. HIV Transmission by Stage of Infection and Pattern of Sexual Partnerships. *Epidemiology*. 2010;21(5):676-684.

64. Alam SJ, Romero-Severson E, Kim J-H, Emond G, Koopman JS. Dynamic Sex Roles Among Men Who Have Sex With Men and Transmissions From Primary HIV Infection. *Epidemiology*. 2010;21(5):669-675.

65. Koopman JS, Simon CP, Riolo CP. When to Control Endemic Infections by Focusing on High-Risk Groups. *Epidemiology*. 2005;16:621-627. doi:10.1097/01.ede.0000172133.46385.18.

66. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. OUP Oxford; 1992.

67. Hethcote HW, Van Ark JW. Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs. *Math Biosci*. 1987;84(1):85-118. doi:10.1016/0025-5564(87)90044-7.

68. Koopman JS, Jacquez JA, Welch GW, et al. The Role of Early HIV Infection in the Spread of HIV Through Populations. *JAIDS J Acquir Immune Defic Syndr*. 1997;14(3). http://journals.lww.com/jaids/Fulltext/1997/03010/The_Role_of_Early_HIV_Infection_in_the_Spread_of.9.aspx.

69. Jacquez JA, Koopman JS, Simon CP, Longini IM. Role of the primary infection in epidemics of HIV infection in gay cohorts. *J Acquir Immune Defic Syndr*. 1994;7(11):1169-1184.

70. Rapatski BL, Suppe F, Yorke JA. HIV epidemics driven by late disease stage transmission. *J Acquir Immune Defic Syndr 1999*. 2005;38(3):241-253.

71. Abu-raddad LJ, Longini IM. No Hiv stage is dominant in driving the Hiv epidemic in sub-saharan Africa. *Aids*. 2008;22(9):1055-1061. doi:10.1097/QAD.0b013e3282f8af84.

72. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. *J Infect Dis*. 2008;198(5):687-693. doi:10.1086/590501.

73. Pilcher CD, Joaki G, Hoffman IF, et al. Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection: *AIDS*. 2007;21(13):1723-1730. doi:10.1097/QAD.0b013e3281532c82.

74. Pinkerton SD. Probability of HIV transmission during acute infection in Rakai, Uganda. *AIDS Behav*. 2008;12(5):677-684. doi:10.1007/s10461-007-9329-1.

75. Romero-Severson EO, Alam SJ, Volz EM, Koopman JS. Heterogeneity in Number and Type of Sexual Contacts in a Gay Urban Cohort. *Stat Commun Infect Dis*. 2012;4(1). doi:10.1515/1948-4690.1042.

76. Eaton JW, Hallett TB, Garnett GP. Concurrent Sexual Partnerships and Primary HIV Infection: A Critical Interaction. *AIDS Behav*. 2011;15(4):687-692. doi:10.1007/s10461-010-9787-8.

77. Powers KA, Ghani AC, Miller WC, et al. The Role of Acute and Early HIV Infection in the Spread of HIV-1 in Lilongwe, Malawi: Implications for “Test and Treat” and Other Transmission Prevention Strategies. *Lancet*. 2011;378(9787):256-268. doi:10.1016/S0140-6736(11)60842-8.

78. Akpa OM, Oyejola BA. Modeling the Transmission Dynamics of HIV/AIDS epidemics: an introduction and a review. *J Infect Dev Ctries*. 2010;4(10):597-608. doi:10.3855/jidc.542.

79. Dietz K. The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res*. 1993;2(1):23-41. doi:10.1177/096228029300200103.

80. Roberts MG, Heesterbeek JAP. A new method for estimating the effort required to control an infectious disease. *Proc R Soc B Biol Sci*. 2003;270(1522):1359-1364. doi:10.1098/rspb.2003.2339.

81. Heesterbeek JAP, Roberts MG. The type-reproduction number T in models for infectious disease control. *Math Biosci*. 2007;206(1):3-10. doi:10.1016/j.mbs.2004.10.013.

82. Henry CJ, Koopman JS. *Sci Rep*. 2015;5:9467. doi:10.1038/srep09467.

83. Nold A. Heterogeneity in disease-transmission modeling. *Math Biosci*. 1980;52(3):227-240. doi:10.1016/0025-5564(80)90069-3.

84. Diekmann O, Heesterbeek J a. P, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface*. November 2009:rsif20090386. doi:10.1098/rsif.2009.0386.

85. de-Camino-Beck T, Lewis MA, Driessche P van den. A graph-theoretic method for the basic reproduction number in continuous time epidemiological models. *J Math Biol*. 2009;59(4):503-516. doi:10.1007/s00285-008-0240-9.

86. de-Camino-Beck T, Lewis MA. A new method for calculating net reproductive rate from graph reduction with applications to the control of invasive species. *Bull Math Biol*. 2007;69(4):1341-1354. doi:10.1007/s11538-006-9162-0.

87. Hall HI, Song R, Rhodes P, et al. Estimation of HIV incidence in the United States. *JAMA J Am Med Assoc*. 2008;300(5):520-529. doi:10.1001/jama.300.5.520.

88. Estimated HIV Incidence in the United States , 2007 – 2010. 2010;17(4):2007-2010.

89. Jaffe HW, Valdiserri RO, De Cock KM. The reemerging HIV/AIDS epidemic in men who have sex with men. *JAMA J Am Med Assoc*. 2007;298(20):2412-2414. doi:10.1001/jama.298.20.2412.

90. Zhang X, Zhong L, Romero-Severson E, et al. Episodic HIV Risk Behavior Can Greatly Amplify HIV Prevalence and the Fraction of Transmissions from Acute HIV Infection. *Stat Commun Infect Dis*. 2012;4(1). doi:10.1515/1948-4690.1041.

91. Gomez GB, Borquez A, Caceres CF, et al. The Potential Impact of Pre-Exposure Prophylaxis for HIV Prevention among Men Who Have Sex with Men and Transwomen in Lima, Peru: A Mathematical Modelling Study. *PLOS Med*. 2012;9(10):e1001323. doi:10.1371/journal.pmed.1001323.

92. MacFadden DR, Tan DH, Mishra S. Optimizing HIV pre-exposure prophylaxis implementation among men who have sex with men in a large urban centre: a dynamic modelling study. *J Int AIDS Soc*. 2016;19(1). doi:10.7448/IAS.19.1.20791.

93. Eaton JW, Hallett TB, Garnett GP. Concurrent Sexual Partnerships and Primary HIV Infection: A Critical Interaction. *AIDS Behav*. 2011;15(4):687. doi:10.1007/s10461-010-9787-8.

94. Piot P, Bartos M, Larson H, Zewdie D, Mane P. Coming to terms with complexity: a call to action for HIV prevention. *The Lancet*. 2008;372(9641):845-859. doi:10.1016/S0140-6736(08)60888-0.

95. Glick SN, Morris M, Foxman B, et al. A comparison of sexual behavior patterns among men who have sex with men and heterosexual men and women. *J Acquir Immune Defic Syndr 1999*. 2012;60(1):83-90. doi:10.1097/QAI.0b013e318247925e.

96. Rosenberg ES, Khosropour CM, Sullivan PS. High prevalence of sexual concurrency and concurrent unprotected anal intercourse across racial/ethnic groups among a national, web-based study of men who have sex with men in the United States. *Sex Transm Dis*. 2012;39(10):741-746. doi:10.1097/OLQ.0b013e31825ec09b.

97. Same race and older partner selection may explain higher HIV... : AIDS. LWW. http://journals.lww.com/aidsonline/Fulltext/2007/11120/Same_race_and_older_partner_selection_may_explain.15.aspx. Accessed April 25, 2017.

98. Service SK, Blower SM. HIV Transmission in Sexual Networks: An Empirical Analysis. *Proc R Soc Lond B Biol Sci*. 1995;260(1359):237-244. doi:10.1098/rspb.1995.0086.

99. Eaton LA, Kalichman SC, O'Connell DA, Karchner WD. A strategy for selecting sexual partners believed to pose little/no risks for HIV: serosorting and its implications for HIV transmission. *AIDS Care*. 2009;21(10):1279-1288. doi:10.1080/09540120902803208.

100. Paltiel AD, Freedberg KA, Scott CA, et al. HIV Preexposure Prophylaxis in the United States: Impact on Lifetime Infection Risk, Clinical Outcomes, and Cost-Effectiveness. *Clin Infect Dis*. 2009;48(6):806-815. doi:10.1086/597095.

101. Kurth AE, Celum C, Baeten JM, Vermund SH, Wasserheit JN. Combination HIV Prevention: Significance, Challenges, and Opportunities. *Curr HIV/AIDS Rep*. 2010;8(1):62-72. doi:10.1007/s11904-010-0063-3.

102. Velasco-Hernandez JX, Gershengorn HB, Blower SM. Could widespread use of combination antiretroviral therapy eradicate HIV epidemics? *Lancet Infect Dis*. 2002;2(8):487-493.

103. Khademi A, Anand S, Potts D. Measuring the Potential Impact of Combination HIV Prevention in Sub-Saharan Africa. *Medicine (Baltimore)*. 2015;94(37):e1453. doi:10.1097/MD.0000000000001453.

104. Buchbinder SP, Liu A. Pre-Exposure Prophylaxis and the Promise of Combination Prevention Approaches. *AIDS Behav*. 2011;15(1):72-79. doi:10.1007/s10461-011-9894-1.

105. Brookmeyer R, Boren D, Baral SD, et al. Combination HIV Prevention among MSM in South Africa: Results from Agent-based Modeling. *PLoS ONE*. 2014;9(11):e112668. doi:10.1371/journal.pone.0112668.

106. Romero-Severson EO, Alam SJ, Volz E, Koopman J. Acute-stage transmission of HIV: effect of volatile contact rates. *Epidemiol Camb Mass*. 2013;24(4):516-521. doi:10.1097/EDE.0b013e318294802e.

107. Colfax G, Coates TJ, Husnik MJ, et al. Longitudinal patterns of methamphetamine, popper (amyl nitrite), and cocaine use and high-risk sexual behavior among a cohort of San Francisco men who have sex with men. *J Urban Health*. 2005;82(1):i62. doi:10.1093/jurban/jti025.

108. Kelly JA, DiFranceisco WJ, Lawrence JSS, Amirkhanian YA, Anderson-Lamb M. Situational, Partner, and Contextual Factors Associated with Level of Risk at Most Recent Intercourse Among Black Men Who Have Sex with Men. *AIDS Behav*. 2013;18(1):26-35. doi:10.1007/s10461-013-0532-y.

109. Koblin BA, Husnik MJ, Colfax G, et al. Risk factors for HIV infection among men who have sex with men. 2006;(September 2005).

110. Krakower D, Mayer KH. Engaging Healthcare Providers to Implement HIV Pre-Exposure Prophylaxis. *Curr Opin HIV AIDS*. 2012;7(6):593-599. doi:10.1097/COH.0b013e3283590446.

111. Baeten JM. Making an Impact With Preexposure Prophylaxis for Prevention of HIV Infection. *J Infect Dis*. 2016;214(12):1787-1789. doi:10.1093/infdis/jiw224.

112. Haberer JE, Bangsberg DR, Baeten JM, et al. Defining success with HIV pre-exposure prophylaxis: A prevention-effective adherence paradigm. *AIDS Lond Engl*. 2015;29(11):1277-1285. doi:10.1097/QAD.0000000000000647.

113. Pines HA, Gorbach PM, Weiss RE, et al. Sexual risk trajectories among MSM in the United States: implications for pre-exposure prophylaxis delivery. *J Acquir Immune Defic Syndr 1999*. 2014;65(5):579-586. doi:10.1097/QAI.0000000000000101.

114. Mansergh G, Koblin BA, Sullivan PS. Challenges for HIV Pre-Exposure Prophylaxis among Men Who Have Sex with Men in the United States. *PLOS Med*. 2012;9(8):e1001286. doi:10.1371/journal.pmed.1001286.

115. Myers GM, Mayer KH. Oral Preexposure Anti-HIV Prophylaxis for High-Risk U.S. Populations: Current Considerations in Light of New Findings. *AIDS Patient Care STDs*. 2011;25(2):63-71. doi:10.1089/apc.2010.0222.

116. Molina J-M, Capitant C, Spire B, et al. On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. *N Engl J Med*. 2015;373(23):2237-2246. doi:10.1056/NEJMoa1506273.

117. McCormack S, Dunn DT, Desai M, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *The Lancet*. September 2015. doi:10.1016/S0140-6736(15)00056-2.

118. Fonner VA, Dalglish SL, Kennedy CE, et al. Effectiveness and safety of oral HIV preexposure prophylaxis for all populations. *AIDS Lond Engl*. 2016;30(12):1973-1983. doi:10.1097/QAD.0000000000001145.

119. Changes in Sexual Risk Behavior Among Participants in a PrEP... : Sexually Transmitted Diseases. LWW. http://journals.lww.com/stdjournal/Fulltext/2008/12000/Changes_in_Sexual_Risk_Behavior_Among_Participants.11.aspx. Accessed June 3, 2017.

120. Wilson PA, Cook S, McGaskey J, Rowe M, Dennis N. Situational predictors of sexual risk episodes among men with HIV who have sex with men. *Sex Transm Infect*. 2008;84:506-508. doi:10.1136/sti.2008.031583.

121. Levy I, Mor Z, Anis E, et al. Men Who Have Sex With Men, Risk Behavior, and HIV Infection: Integrative Analysis of Clinical, Epidemiological, and Laboratory Databases. *Clin Infect Dis*. 2011;52:1363-1370. doi:10.1093/cid/cir244.

122. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet*. 2009;10(8):540-550. doi:10.1038/nrg2583.

123. Frost SDW, Volz EM. Modelling tree shape and structure in viral phylodynamics. *Philos Trans R Soc B Biol Sci*. 2013;368(1614). doi:10.1098/rstb.2012.0208.

124. Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. How the Dynamics and Structure of Sexual Contact Networks Shape Pathogen Phylogenies. *PLoS Comput Biol*. 2013;9(6):e1003105. doi:10.1371/journal.pcbi.1003105.

125. Volz EM, Pond SLK, Ward MJ, Brown AJL, Frost SDW. Phylodynamics of Infectious Disease Epidemics. *Genetics*. 2009;183(4):1421-1430. doi:10.1534/genetics.109.106021.

126. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*. 1977;81(25):2340-2361. doi:10.1021/j100540a008.

127. Moya A, Elena SF, Bracho A, Miralles R, Barrio E. The evolution of RNA viruses: A population genetics view. *Proc Natl Acad Sci*. 2000;97(13):6967-6973. doi:10.1073/pnas.97.13.6967.

128. Sackin MJ. "Good" and "Bad" Phenograms. *Syst Biol*. 1972;21(2):225-226. doi:10.1093/sysbio/21.2.225.

129. Kirkpatrick M, Slatkin M. Searching for Evolutionary Patterns in the Shape of a Phylogenetic Tree. *Evolution*. 1993;47(4):1171-1181. doi:10.2307/2409983.

130. Leventhal GE, Kouyos R, Stadler T, et al. Inferring Epidemic Contact Structure from Phylogenetic Trees. *PLoS Comput Biol*. 2012;8(3):e1002413. doi:10.1371/journal.pcbi.1002413.

131. Aldous DJ. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat Sci*. 2001;16(1):23-34. doi:10.1214/ss/998929474.

132. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016. https://www.R-project.org/.

133. Dearlove BL, Frost SDW. Measuring Asymmetry in Time-Stamped Phylogenies. *PLoS Comput Biol*. 2015;11(7). doi:10.1371/journal.pcbi.1004312.

134. Colijn C, Gardy J. Phylogenetic tree shapes resolve disease transmission patterns. *Evol Med Public Health*. 2014;2014(1):96-108. doi:10.1093/emph/eou018.

135. Jombart T, Eggo RM, Dodd PJ, Balloux F. Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity*. 2011;106(2):383-390. doi:10.1038/hdy.2010.78.

136. Ypma RJF, Ballegooijen WM van, Wallinga J. Relating Phylogenetic Trees to Transmission Trees of Infectious Disease Outbreaks. *Genetics*. September 2013:genetics.113.154856. doi:10.1534/genetics.113.154856.

137. Giardina F, Romero-Severson EO, Albert J, Britton T, Leitner T. Inference of Transmission Network Structure from HIV Phylogenetic Trees. *PLOS Comput Biol*. 2017;13(1):e1005316. doi:10.1371/journal.pcbi.1005316.

138. Smith RA, Ionides EL, King AA. Infectious Disease Dynamics Inferred from Genetic Data via Sequential Monte Carlo. *Mol Biol Evol*. April 2017. doi:10.1093/molbev/msx124.