The Effects of Demographic Processes on Dynamic Networks and The Role of Sexual Behavior and Acquired Immunity on Chlamydia Transmission in Young Adults

Emily Pollock

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Steven M. Goodreau, Chair

Wanda (Martina) Morris

Darryl Holman

Christine Khosropour

Program Authorized to Offer Degree:
Anthropology

University of Washington

Abstract

The Effects of Demographic Processes on Dynamic Networks and The Role of Sexual Behavior and Acquired Immunity on Chlamydia Transmission in Young Adults

Emily Pollock

Chair of the Supervisory Committee:
Professor Steven M. Goodreau
Anthropology

The complex social networks which we maintain and navigate every day make our uniquely elaborate life possible, but also put us at risk of exposure to infectious diseases. This dissertation explores the effects of different demographic processes on dynamic networks and examines the intersection of sexual behavior, biomedical interventions, and acquired immunity in the reinfection of chlamydia trachomatis. Chapter 1 demonstrates that simulating dynamic networks (separable, temporal, exponential random graph models, or STERGMs) with open populations can have unexpected effects on key network metrics and implements several corrections. Chapter 2 uses survival analysis to conclude that de-coupling marriages and cohabitations in the dissolution components of STERGMs can improve the representation of the full distribution of relationship lengths in populations with a large age range. Finally, Chapter 3 finds that while concurrency can increase the rate of reinfection of chlamydia, the greater effect is the treatment of chlamydia without concurrent treatment of a patient's local network. Reinfection rates among those treated with low levels of expedited partner therapy (EPT) are higher than among those whose infections clear naturally, but high levels of EPT treatment can be protective.

Table of Contents

List of Figures

List of Tables

Introduction

Anthropologists have long recognized the importance of social connections and behavioral variation among humans and our nonhuman primate relatives. Indeed, the ability for us to participate in distinct but potentially interlocking complex social networks has likely fueled our evolution as a species and made our uniquely elaborate life possible. Network analysis has often been utilized as a way to visually and quantitatively represent these ties in order to understand their effects on those connected to each other, from kinship, social support and social capital, to the diffusion of information and transmission of disease. These latter networks are crucially important to our understanding of how human biosocial variation influences our health, where these often beneficial complex social networks we maintain and navigate every day can also put us at risk of exposure to infection. The work in this dissertation is guided by the theoretical framework of the human ecology of infectious disease: the investigation of how human behavior, social patterns, and built environments interact with the broader pathogen environment to influence our health. Of particular interest is not just aggregate behavior, but also how variation in individual behavior influences social patterns and alters the landscape through which diseases spread, particularly as this variation relates to biological age.

0.1 Mathematical Modeling

For those interested in human behavior and disease transmission, particularly sexually transmitted diseases, one particular approach to investigation is mathematical modeling. Mathematical models are quan-

titative approximations of real-world systems. This form of inquiry is particularly useful when classic scientific experiments or epidemiological studies to understand disease spread or intervention efficacy cannot be conducted for either practical or ethical reasons, or when specific processes or parameter values in a system are uncertain. In these situations, we use mathematical modeling as an in-silico laboratory to explore ideas and test hypotheses. The level of detail often required to articulate and test specific hypotheses within these artificial constructions has increased over time as we learn more about the world, generate more specific questions, and develop the tools necessary to handle increasingly complex structures. We do not always need a highly complex model, but the form of the model should reflect the mechanism one is interested in addressing. For example, several different types of mathematical models exist that could address the question "how do partner treatment interventions reduce the prevalence of chlamydia," but models that do not explicitly represent sexual partnerships can only attempt to answer this question indirectly. However, all models, regardless of their complexity, have implicit assumptions in their structure that can influence the outcomes of interest in unexpected ways. It is not my goal here to provide a thorough history of epidemic modeling, but below I give a brief overview of the various model forms to highlight the most common implicit assumption: the constant hazard of events as operationalized by the exponential distribution.

The majority of historical mathematical models designed to understand epidemic dynamics are deterministic and compartmental in nature. They do not represent people individually. Instead, they group them into homogeneous compartments representing specific states of interest, a portion of which transition between compartments at each time step based on a rate using differential equations. In the most basic models, the compartments are usually "susceptible" and "infected" and the rate of transition from susceptible to infected depends on the rate of contact between the groups and the size of the infected group relative to the whole population (Kermack & Mckendrick, 1927). Additional complexity can be represented by adding more compartments or states, like breaking down the state of susceptible and infected into demographic groups like race or age groups, adding compartments for vector populations like mosquitoes, or by representing a more complex natural history of the pathogen by including states for groups such as "exposed but not infectious," "recovered," "infected and symptomatic," or "infected and

asymptomatic" to name a few. These models are deterministic in nature because the transitions between compartments rely on unchanging rates and if you run a deterministic compartmental model (DCM) multiple times you will always have the same result.

Stochastic models grew out of this original framework as a way to capture variability and uncertainty in the systems we wish to study. They expanded the framework to include individual actors that could have variation in certain attributes, and most importantly, had the capability to probabilistically evaluate contacts or transition rates rather than assuming a constant rate of change. For example, if the estimated mean contact rate from empirical data was two contacts per day, it is unlikely that a real individual person has *exactly* two contacts per day. Some days they may have more, some days fewer. In this stochastic, agent-based scenario, some or all transitions between states were based on a *probability* of transitioning rather than a set rate, meaning that not the same proportion of individuals in a certain state transitioned at every time step, but the *average* rate of change was preserved. Note however, that both strategies rely on a memoryless assumption that generates an exponential distribution (or geometric if working in discrete time).

Separable, Temporal, Exponential Family Random Graph Models (STERGMs) are a subset of these stochastic, agent-based models that specifically govern the dynamic formation and dissolution of relationships between individual actors. The static form of these models, ERGMs, were originally developed to quantify the likelihood of ties between actors based on a variety of possible concerns, including the individual attributes of the actors and the presence of ties elsewhere in the network. These models have a wide variety of applications but their dynamic form in particular is of special interest to social scientists interested in the transmission of sexually transmitted infections (STIs) like HIV, syphilis, gonorrhea and chlamydia. The strength of the STERGM approach to modeling these sexual transmission networks is driven by three observations. First, our sexual contacts are not random. Our preference for partners can be influenced by a wide range of attributes including age, race, whether or not someone is currently in a relationship, disease status, etc. Second, that there is some level of dependence between time steps - the formation of a new relationship is likely related to the immediately previous state of the network. Third, the factors that contribute to the formation of a relationship may be different than those that contribute

to the dissolution of a relationship. STERGMs provide a statistically principled framework that in the model building phase allows for inference and hypothesis testing of a wide variety of generative processes for contact structure and in the simulation phase dynamically maintains the cross-sectional distribution of relationships of the empirical data from which the models were estimated.

0.2 Goals of this Dissertation

This focus on maintaining the integrity of the sexual contact network makes these methods especially attractive to social scientists because these tools allow us to explicitly represent the aspects of sexual partnership dynamics like overlapping partnerships (concurrency) that increase the connectivity of a network, contribute to disease spread and demonstrate how small variation in individual behavior can have large outcomes for transmission across the full network. All models are sets of assumptions made by researchers about what they believe to be important relative to the outcome. Where many epidemiological models have assumed behavior to be secondary to the question of interest or not relevant at all, we say here that in order to understand the complex patterns by which STIs are transmitted throughout populations, we first need to understand the behavior of human relationships and how these behaviors generate the dynamic sexual network across which these types of infections spread.

We will discuss more thoroughly in the following chapters why particular sexual partnership dynamics including concurrency and relationship duration are important factors in the transmission of STIs, but much of this work will be focused on improving the ways in which we represent these networks in specific contexts. First, I will outline the ways in which including formation model terms depend on a time-varying covariate (nodal age) influence key network metrics in the simulation phase, and explore a variety of corrections. Second, I will evaluate how well the exponential assumption in the dissolution model captures the full distribution of relationship lengths and will attempt to identify simple stratifications under which a constant hazard is appropriate. Lastly, I will use the corrections tested in the first chapter to calibrate a STERGM and explore how both variation in concurrency and immunological processes can influence the rate of reinfection among males and females infected with chlamydia.

As with many projects, the original plan was not ultimately realized. While the survival analysis piece that evaluates the exponential assumption of relationships lengths comes second in this document, it was actually completed first. The conclusions that we outline there have clear next steps which we hoped to implement and test in the next project. However, this first required the development of a set of networks based on the current structure of published models in order to compare the adjustment. During this construction we observed some unexpected network effects that led us to the work presented in Chapter 1. While the goal was always to develop a network model for chlamydia transmission as the final project, and the underlying networks use calibrations developed in Chapter 1, it was beyond the scope of the project to also implement any of the recommendations from Chapter 2. As such, Chapter 2 can largely be read as a standalone project with implications for future epidemic network models.

0.3 Primary Data Source

The empirical behavioral data used in this dissertation are drawn from the 2006-2010 and 2011-2015 waves of the National Survey of Family Growth (NSFG). The NSFG surveys men and women aged 15-44 on many aspects of family life, including but not limited to marriage and divorce, pregnancy, contraception use, infertility, and other aspects of sexual and reproductive behavior. In addition to the demographic information recorded for each respondent and their sampling weights, in this study we use the data collected in section C of the public use files on each respondent's recent sexual partnerships with opposite-sex partners in the last year, with a maximum of three partnerships reported. These data include demographic information about each partner, the century-month of first sexual contact with a particular partner, the century-month of last sexual contact, whether the respondent considers this sexual partnership to be ongoing, and the partnership status (marriage, cohabitation, or other). In Chapters 1 and 3 we use data from all respondents to estimate models for the sexual networks and other epidemiological parameters. In Chapter 2 we focus on the reported duration of relationships, so in that study we limit combined data set to those respondents who report at least one partnership in the last year. Out of the original 43,303 respondents, this subset contains 32,516 respondents who report on 40,443 sexual partnerships.

1

Demography and Dynamic Network Simulations

The choice of model terms in mathematical models depends on the question of interest and the underlying patterns in the data and this is no less true for network models of sexual partnerships developed to understand disease transmission. One of the primary key decisions about model structure is whether to use a "closed" population, where no nodes exit or enter, or an "open" population, where nodes enter and exit, often to represent demographic processes like births, deaths, and aging. Most of the recently published models using using ERGMs and EpiModel to simulate epidemics use this open framework and have modeled adult men who have sex with men (MSM) populations aged 18-39 (Goodreau et al., 2017; Jenness et al., 2017). A second key decision relates to which model terms are selected to generate network structure, and as we will see, these decisions will interact with the first. The previously men-

tioned models focused on terms related to the average number of relationships of certain types, mixing patterns by age, the likelihood of concurrent partnerships. Because prevalence of both main and casual relationships remained relatively stable over the small age range, the models did not include terms that used age as a predictor of relationship formation. However, in this project, we focus on heterosexual relationships over a larger age range (15-45). Unlike the MSM models, there are large differences in the prevalence of different relationship types over this age range, so we will need to include terms that involve age in our model (see Figure 1.1). However, while individual age is straightforward to represent in this modeling framework, using age-dependent relational formation terms introduces several complicating factors for network models largely related to the boundaries imposed by age range and the aging process. In this chapter, first I will demonstrate that the dynamic networks estimated from empirical data (separable temporal exponential random graph models, STERGMs) reproduce key network statistics in the absence of dynamic vital processes. Then I will document how simulations that incorporate individual births, deaths, aging, and sexual debut lead the network to deviate from these key statistics. Third, I will explore the possible underlying causes for these deviations, implement new and/or extend existing corrections for demographic effects, evaluate the effectiveness of each, and outline some possible future directions for improved simulations.

1.1 Some Definitions

Node / Ego: An individual in the network.
Nodal Attribute: A trait of an individual (sex, age, sexual debut status, etc).

1.2 Base Model Overview

The dynamic networks developed here represent two types of relationships: main partnerships defined as marriages and cohabitations, and casual partnerships, defined as any relationship with duration greater than one week that is not a marriage or cohabitation. These relationships are governed by two separate but potentially overlapping networks among the same set of nodes. The models used in this chapter focus

largely on the age-related effects of relationship formation, with several additional terms and structural offsets. One-off relationships ("instantaneous" or "one night stands") are not included in the current model, but will be included in later chapters when we model the transmission of disease. Both the marriage/cohabitation network and the casual networks have terms for:

1. the overall density of the network,
2. the prevalence of relationships by age and age-squared of each node,
3. a mixing term for the difference in the square root of each node's age and that of their partner's age,
4. an offset term prohibiting the formation of relationships among specific nodes to mimic the sexual debut process, and
5. an additional offset term to prohibit the formation of relationships among same-sex nodes.

In an applied example at the end of this chapter, we will also add terms that govern how many relationships overlap between the networks (the number of relationships that can form among individuals who have a casual relationship as well as a marriage), but we do not explicitly represent this process in these initial networks. The models are estimated from egocentrically collected data from the NSFG using the ergm.ego package in R, which first estimates target statistics for the specified model terms from the data and then fits the ERGMs based on these targets (Krivitsky & Morris, 2017). Table 1.1 displays the formation models and their estimated coefficients.

1.3 Closed-System Dynamics

First we demonstrate that the estimated models dynamically reproduce statistics we are interested in, particularly the mean degree by age in each network and the expected duration of relationships, in a closed system (i.e. without aging, births, or deaths). This is one of the several steps to check model performance prior to epidemic simulations. In this diagnostic, we simulate the STERGMs for 5 repetitions of 7500 time steps (representing almost 150 years) and evaluate the cross-sectional network statistics over time. At each time step, ties can form and ties can dissolve based on the model coefficients. If the model converges properly (MCMC diagnostics can be used for assessing this), the network formation statistics should hover around their estimated targets. In this diagnostic we also evaluate the duration of ties and the rate of tie

Table 1.1: Summary of Formation Model Fits

Model Term	Estimate	SE	Z Value	Pr(>|z|)
Casual Network				
offset(netsize.adj)	-10.8197783			
edges	2.8953778	0.4736468	6.112947	0.0000000
nodecov.age	0.0470762	0.0167334	2.813301	0.0049036
nodecov.agesquared	-0.0015307	0.0002800	-5.466910	0.0000000
absdiff.sqrtage	-2.8356952	0.0650198	-43.612827	0.0000000
concurrent	-2.3870360	0.1002761	-23.804643	0.0000000
offset(nodematch.male)	-Inf			
offset(nodefactor.debuted.0)	-Inf			
Marriage/Cohabitation Network				
offset(netsize.adj)	-10.8197783			
edges	-10.8388692	0.8782038	-12.342089	0.0000000
nodecov.age	0.5014342	0.0301891	16.609768	0.0000000
nodecov.agesquared	-0.0079037	0.0004850	-16.294620	0.0000000
absdiff.sqrtage	-3.0938425	0.0446581	-69.278395	0.0000000
offset(nodematch.male)	-Inf			
offset(nodefactor.debuted.0)	-Inf			
offset(concurrent)	-Inf			

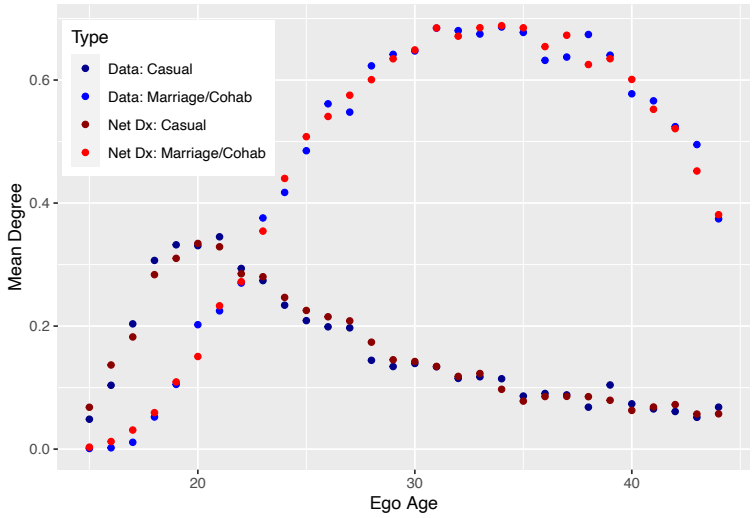

Figure 1.1: Comparison: Egodata vs Mean Degree in Closed-Pop Simulation

dissolution to ensure the dissolution targets are met. At this step, the node set is static - all nodal attributes including age are fixed, no nodes exit, and no new nodes enter the population. Figures 1.1 and 1.2 show that in this closed system simulation, the models perform exceptionally well. Not only is the overall mean degree of each network reproduced, but mean degree by age in both networks is also reproduced, which is not necessarily guaranteed by the parsimonious parameterization. Additionally, both models reach the target mean cross-sectional relationship duration after sufficient time (once the initial censoring artifacts disappear). Any deviation from these targets as we move to the simulation then, should be related to the introduction of vital dynamics and other processes like sexual debut.

1.4 Existing Demographic corrections

This section provides an overview of existing corrections for dynamic networks and demography and outlines some of the boundary effect issues that we will explore in more detail below.

1. Formation Approximation

In many cases, a full STERGM cannot be estimated directly due to the computational burden when networks are large, sparse, and have relatively long relational durations (a near perfect description of mar-

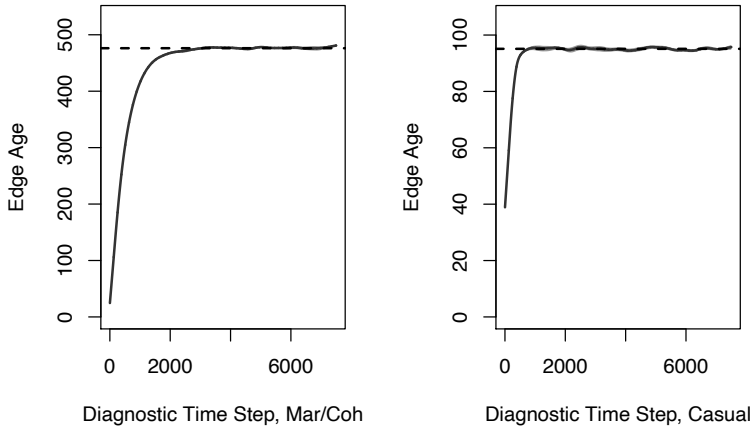

Figure 1.2: Mean Relationship Lengths in Diagnostic

riage/cohabitation networks). Carnegie, Krivitsky, Hunter, & Goodreau (2015) introduced an approximation to full STERGM estimation for such cases that uses the same data: cross-sectional egocentric network data and information on tie duration. In this approximation, the static ERGM is first estimated on the cross-sectional network. Then the edges formation term (which represents the base propensity for ties to form between any two individuals in the network) is decreased by the log odds of the probability of edge persistence, in effect transforming the formation term from *prevalence* of ties in the network to the *incidence* of ties. The explorations below do not attempt to modify this approach. Instead we will explore the relationship between this adjustment of the formation coefficient and the probability of tie persistence as estimated from the long relationship duration expected in the marriage network relative to the length of time we actually observe each individual in the simulation.

2. Departure Correction

The node departure correction used in the model estimation-to-simulation workflow is necessary due to the observation that when nodes were removed from the simulation to mimic, for example, background age-specific mortality, the mean degree of the network became lower than expected, as does the mean duration of relationships. The logic is relatively straightforward: the statistical model underlying these network simulations balances the probability of tie formation with the probability of tie dissolution in order to maintain a target number of ties in the network. However, the empirical data from which we esti-

mate our duration targets captures two sets of dissolution: relationships that end due to a "break-up" and relationships that end due to node death. Because demographic processes are handled separately from the network formation/dissolution, we use a correction to separate the STERGMs dissolution ("break-ups"), from dissolutions due to node death or departure. To counter the lowering of relationship duration (and subsequently mean degree) related to node death, the expected (endogenous) duration of ties is increased such that the *average* duration in the presence of both forms of dissolution is maintained. The departure correction implemented in previous models has two components: 1) the average mortality rate per time step across the entire population (often weighted by age and/or race) and 2), the rate at which individuals depart the simulation due to the age boundary, calculated as 1/(time steps each node is expected to be observed in the simulation in the absence of early death). For example, if the age range of the model was 18:39, then the weekly rate at which each individual is expected to exit the simulation is 1/(52*(39-18)). In the past this approach has successfully balanced the additional unexpected dissolution of relationships due to node departure. Below we will explore situations where this correction is not sufficient in its current form.

3. Population Size Correction

Occasionally it is of interest to model a population that is growing, declining, or stochastically varying around some mean size. The correction outlined in Krivitsky, Handcock, & Morris (2011) makes small adjustments at each time step based on the difference in population size between time t and t-1 to the coefficient on the edges (density) term. This correction is designed to maintain the target mean degree of the network in the presence of changes in the size of the population by while preserving the odds of forming ties based on nodal attributes as specified by the other model terms (e.g. matching by age, race, classroom, etc). This correction adjusts for both changes in the population size and the composition, so we will not further modify it here.

1.5 Overview of Open-Population Demographic Dynamics

The open-population simulations run using the EpiModel API that handles demographic dynamics and disease transmission separately from the STERGM dynamics described above, but in a sequence with

each other. In these models, at every time step, a series of modules is run that govern important demographic dynamics: node departure, node entry, aging, and sexual debut, followed by relationship formation and dissoluton. We describe the demographic dynamics here.

Nodes automatically depart the model at age 45. This boundary was selected for two main reasons: 1) According to the Centers for Disease Control, in their 2018 surveillance report, 97.4% of all chlamydia infections were diagnosed in the 15-44 age range and 2) the National Survey of Family Growth, the empirical data from which we estimate our model, only surveys adults aged 15-44. There are likely other sources of information that we could use to increase the age range, but it did not seem necessary to our questions of interest. Note that implicit in this decision is the elimination of all reported relationships among egos aged 15-45 whose *partners* are outside of this age range. The degree distribution that we actually use to estimate the model (and are trying to maintain during simulation) looks rather different than the original distribution shown above, particularly in the marriage/cohabitation network (see 1.4). We will consider the consequences of this in a later section.

In addition to the age boundary at 45, all individuals experience the possibility of mortality at each time step. I will refer to this as their age-specific mortality rate, or ASMR. The per-time-step-ASMR is based on data from published in recent U.S. Vital Statistics documents adn stratified by age category and sex. Given that our age range is relatively young, departures due to background mortality are uncommon relative to the effect of the age boundary on which nodes depart the model. Nodes enter at age 15 at a rate based on the expected number of departures per time step in order to keep the population size relatively stable. Like the number of deaths due to ASMR, the actual number of entries per time step is stochastic but maintains a population size within 1-2% of the starting size of 50,000 nodes. Each time step in the simulation represents one week, so nodes age by 1/52 per time step. Nodes enter the population at age 15 and are evaluated for sexual debut at each time step, with probability that increases until age 29 to match the age-at-debut distribution as reported in the NSFG. In accordance with the data where a small proportion of the population never reports intercourse with a member of the opposite sex, some individuals will never "debut" and will therefore never form a tie in these networks.

Table 1.2: Mean Degree and Duration Comparison, Targets and Base Simulation

	Mean Degree			Mean Duration		
	Target	Sim Result	Pct Off	Target	Sim Result	Pct Off
Marriage/Cohab	0.455	0.431	-5.27	476	365	-23.32
Casual	0.159	0.152	-4.40	95	103	8.42

1.6 Initial Simulation Results

Unlike closed-population scenario above, when we run these simulations with demographic processes, several metrics stray from their target values. First, while the mean degree in the initial networks match the targets, the equilibrium mean degrees, or average number of relationships per person across each network, are lower than expected in both the marriage/cohabitation network and in the casual network (by roughly five and three percent respectively). These deviations are not large overall, but they are especially concerning when considering the equilibrium distribution of relationships by age. The mean degree by age is underrepresented in both networks for the youngest ages but overrepresented in the mid-30s. Finally, the mean relationship length is 24% too short in the marriage network but 7% too long in the casual network. In the next few sections, we describe possible explanations for these deviations and explore several corrections.

1.7 Considering the effect of older partners

The empirical data show that the prevalence of marriages and cohabitations is higher at older ages than at younger ages, and the model coefficients support this observation (as demonstrated by the closed-system results). However, we observe that in simulations when the characteristics of the node set are largely in equilibrium but each individual node enters, ages, and eventually exits, there are too many relationships among the older egos. We theorize that this may be due to the age boundary imposed by the model. When nodes leave the simulation age at 45, they will dissolve any relationship that they were in at the

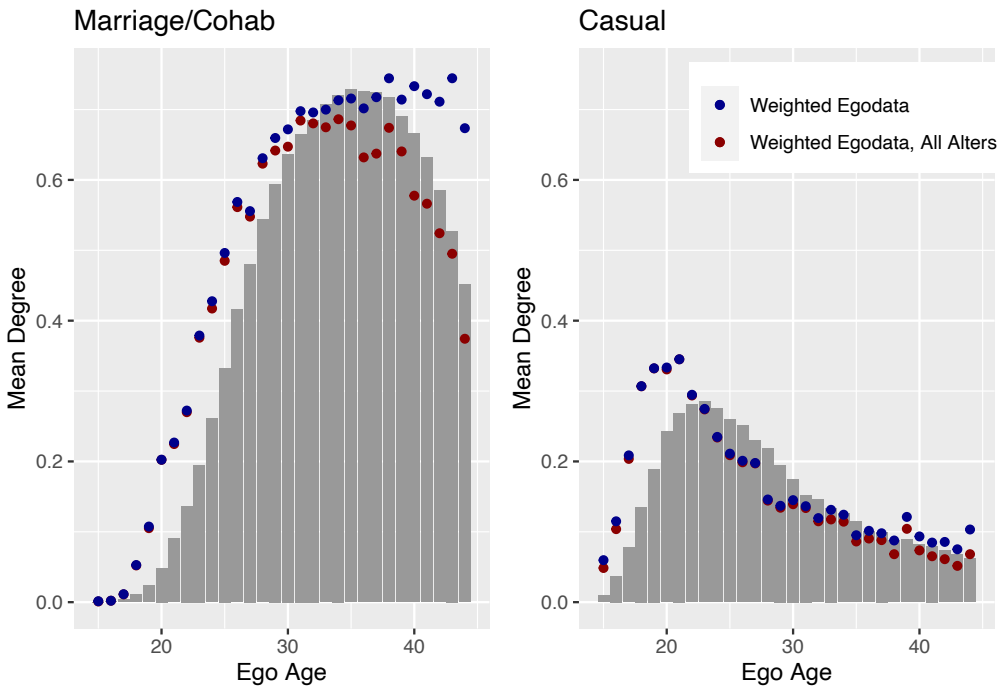

Figure 1.3: Base Simulation: Mean Degree by Age.

previous time step. While these additional dissolutions are theoretically corrected for using the Departure Correction, the positive model coefficient on age and prohibition on concurrency suggest that the now-unpartnered node who remains in the simulation after their partner departs has a high likelihood of forming a new relationship. And because the age-mixing term increases the odds of forming relationships with nodes of similar ages, the incidence of relationships at older ages increases. It is possible that these new, short relationships in older ages contribute both to the lower than expected mean relationship duration in the marriage network and the lower than expected mean degree at younger ages. The problem is that the tie that dissolved as a result of one partner leaving the simulation due to this age boundary is not a true dissolution, and these newly formed relationships should not actually exist because the remaining partner should not actually be eligible to form a new relationship in the network yet. That is, they should still be in their original relationship, even if we no longer observe it. Figure 1.4 highlights this age boundary effect in the empirical data. The light blue and light red dots reflect the mean degree by age among egos and their partners aged 15-44. Their darker counterparts reflect egos reporting on their partners of

all ages. The effect is particularly pronounced among the oldest ages in the marriage network, while there is a much smaller effect in the casual network. The casual network also displays some small differences in the youngest ages, where a few 15 and 16 year-olds report relationships with partners younger than 15, although the below corrections focus only on correcting for the effect of partners outside the upper end of the age boundary.

We consider two ways to address the effect of partners outside the age boundary. First, we prevent egos whose partners have aged out from immediately forming new relationships by adding an offset term for egos who meet this condition. In this scenario we hope that by preventing new relationships from forming among egos whose previous relationships were terminated artificially by the age boundary, the simulation will better match the data with the restricted alter set and increase the mean relationship length by generating new relationships at earlier ages. In the second scenario, we increase the age at which egos depart the simulation to age 65. While we may not be interested in modeling individuals older than 44 for epidemiological reasons, it may be worthwhile to keep them in the simulation over a longer period of time to avoid the artificial ending of relationships. In this case we hope to match the empirical mean degree distribution among egos with the age-unrestricted alter set. However, because we would be simulating individuals outside the age range in the data we used for estimation, we may run into additional issues.

1.7.1 Offset for Partner Age-Out

This scenario adds an offset term to the formation model ("olderpartner") for egos whose alters are outside of the 15-44 age range modeled in the simulation. We have a target count for this offset during estimation because as the figure above demonstrates, there are nodes that exist in the empirical data who have a partner older than 44. During the simulation, if a node ages out while they are in a relationship, the remaining partner gets flagged by the "olderpartner" attribute and are prohibited from forming a new relationship. The probability of becoming available for a relationship on any future time step (i.e. removing the "olderpartner" flag by resetting that attribute to 0) is equal to 1/expected duration of the relationship type, although in the case of the marriage/cohabitation network relationships last so long that it is un-

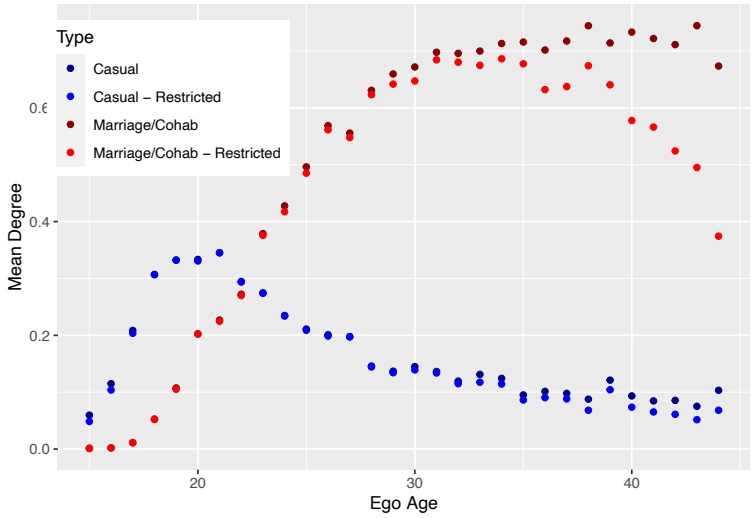

Figure 1.4: Observed Data: Mean Degree by Ego Age and Relationship Type, Restricted and Unrestricted Alters

Table 1.3: Mean Degree and Duration Comparison, Targets vs Older Partner Offset

	Mean Degree			Mean Duration		
	Target	Sim Result	Pct Off	Target	Sim Result	Pct Off
Marriage/Cohab	0.455	0.434	-4.62	476	364	-23.53
Casual	0.159	0.152	-4.40	95	102	7.37

likely that a node becomes available for the rest of their simulation life-course (unless the age difference between partners was exceptionally large, which is not impossible). Figure 1.5 plots the mean degree by age in the simulation with the older partner offset included compared to both the base model simulation and the egodata. The first thing we note is that this offset did not largely influence the overall mean degree in either network, nor did it increase the mean relationship duration in the marriage/cohabitation network (mean relationship length was also unchanged in the casual network, but we did not necessarily expect it to). When comparing mean degree by age between scenarios, the offset did not correct the general trend of the overrepresentation of relationships at the older ages, but it did slightly increase the mean degree in nodes ages roughly 30-35. The casual network was largely uninfluenced by this offset.

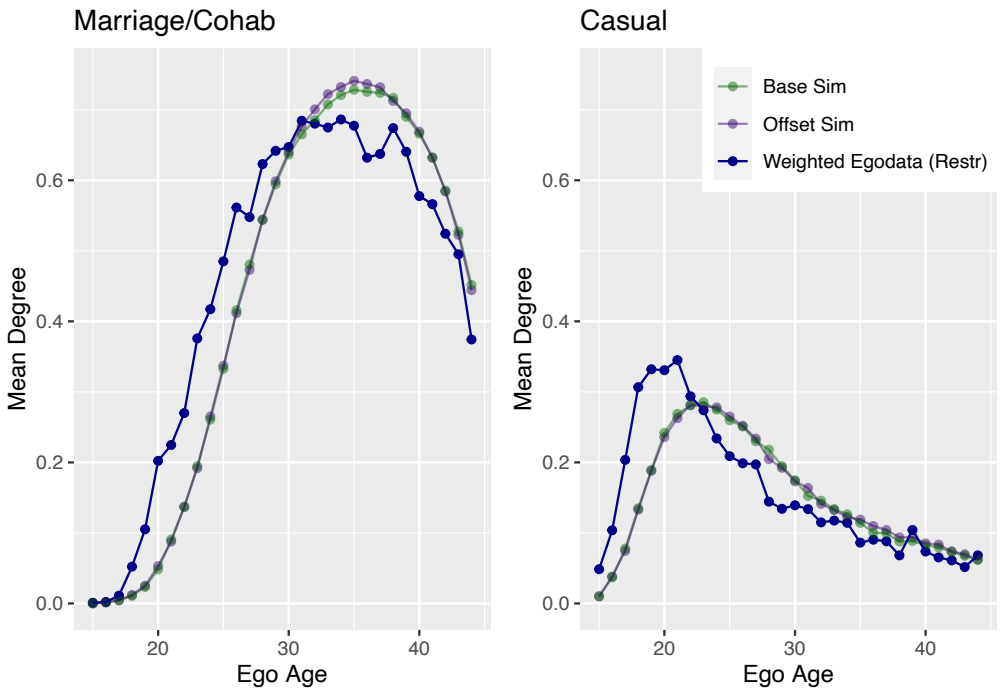

Figure 1.5: Mean Degree Comparison: Base vs Offset.

Increase Age Boundary

In this scenario, we hope to move the degree distribution closer to the egodata distribution with the age-unrestricted alters, the distribution that better represents reality when surveying egos aged 15-44 about their relationships. This scenario also includes the offset for "older partners" but employs it in a slightly different fashion. In the previous scenario, edges dissolved artificially when one of the partners left the model at age 45. We now allow all individuals to remain in the simulation until age 65 and allow those relationships to continue as they would normally. We use the offset to prevent any nodes older than 45 from forming new relationships. This means that only relationships that began prior both partners turning 45 exist in this simulation.

Table 1.4 and Figure 1.6 present the results from this scenario. It is clear that in the marriage/cohabitation network, we can easily represent the partnerships lost to the upper age boundary simply by keeping their older partners in the model, even if the data used to estimate the model did not include these partners.

However, this approach has some unintended consequences. The edges coefficient in the formation model is a density term, and its target is based on a mean degree estimated from the restricted partner data. When we prevent relationships from dissolving when one partner turns 45, we increase the mean degree of those at older ages, but this same logic this would imply then a decrease in the mean degree at younger ages. And indeed this is what we observe: the increased age boundary reduces the mean degree of those below 30 in the casual network and those roughly 25-35 in the marriage/cohabitation network. So while the overall mean degree now slightly exceeds the target and the mean relationship length has increased, this approach on its own fails to substantially improve the fit of the mean degree distribution overall.

We find that these corrections that focus on the effet of older partners have limited utility on their own. That we could capture the distribution of relationships with unrestricted partner ages by a simple extension of the age boundary is heartening and may have a role to play in other contexts, but the consequences in the casual network in particular are too strong to continue down this path. The small improvements we see in mean relationship duration and mean degree in the marriage network suggest that while the older age boundary may not be the primary factor governing the misrepresentation of relationships by age, it did contribute. The fact that very little effect at all on the casual network is somewhat expected given that older ages are actually less likely to form casual partnerships than younger ages. Additionally, whereas each node is allowed a maximum of one relationship in the marriage network, no such limit exists for the casual network, so the prohibition on relationship formation in the casual network is not strictly necessary. We continue to include the older partner offset in the following scenarios because it makes intuitive sense in the marriage network to discourage partner turnover at the oldest ages due to artificial relationship dissolution. However, more work is needed to address the broader issues in these simulations.

1.8 Relationship Length & The Simulation Window

We now turn our focus to the issue of relationship length. So far, our attempts to represent relationships at older ages in a more accurate way has not corrected the issues with relationship duration in these networks.

Table 1.4: Mean Degree and Duration Comparison, Targets vs Increased Age Boundary

	Mean Degree			Mean Duration		
	Target	Sim Result	Pct Off	Target	Sim Result	Pct Off
Marriage/Cohab	0.455	0.472	3.74	476	414	-13.03
Casual	0.159	0.141	-11.32	95	104	9.47

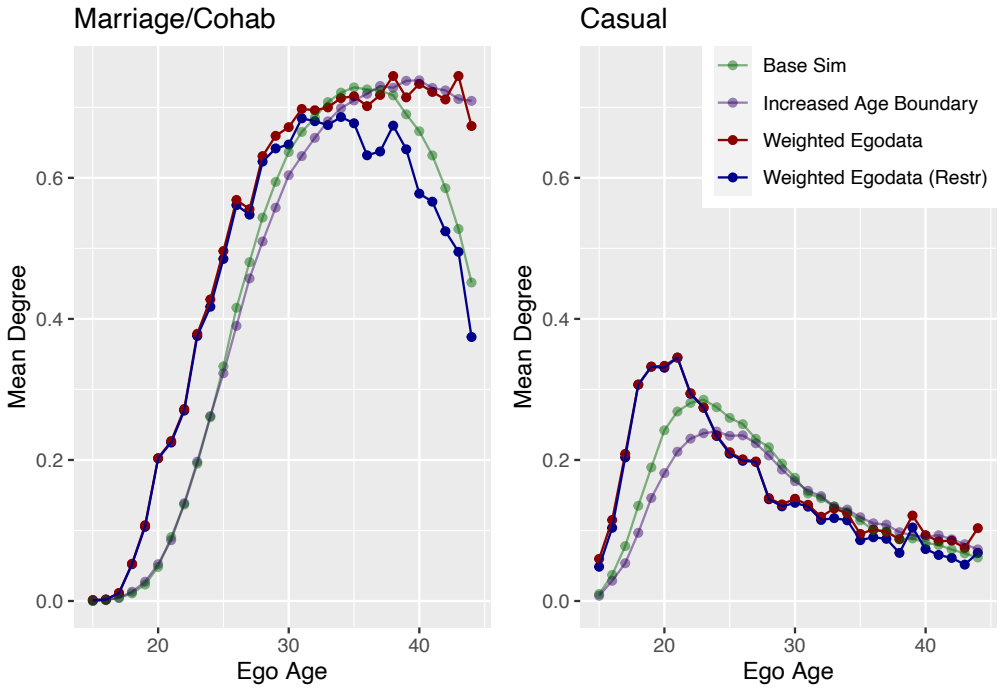

Figure 1.6: Mean Degree Comparison: Increased Age Boundary.

In the marriage/cohabitation network, the mean relationship length falls nearly 2 years short of the target length and the length among casual relationships is roughly 10% too long. This may not seem substantial, but these shorter, occasionally overlapping, relationships are important for the transmission of STIs and an increase in the mean length (and poentially decreasing the rate of new partner acquisition) could have consequences for our understanding of epidemics across these networks (Jolly, Muth, Wylie, & Potterat, 2001; Morris, Kurth, Hamilton, Moody, & Wakefield, 2009; Niccolai, Rowhani-Rahbar, Jenkins, Green, & Dunne, 2005). Conversely, marriages that are too short may decrease the time certain portions of the network are isolated and protected from exposure. There are a few possible reasons that there may be a mismatch between the formation and dissolution coefficients in-simulation that may contribute to these outcomes. Here we explore a possible issue related to the window of observation for each node in the simulation and how that influences the observable mean relationship duration.

The dissolution component of the STERGM in these models assumes a homogenous (exponential) hazard of dissolution within each network (i.e. marriages and casual relationships have different expected duration, but *each* marriage has the same expected length). The model then evaluates each relationship at every time step for stochastic dissolution, and this generates a distribution of simulated relationship lengths within each network that is exponential. There are consequences of this assumption of a constant hazard that we explore in greater detail in Chapter 2, but for now we will address the relationship between the range of possible relationship lengths predicted by the exponential and the length of the simulation window that we are actually able to observe in the simulation. An exponential distribution with a mean of roughly 476 weeks (the mean cross-sectional length of marriages in this data) has a very long right tail extending to 77 years. Clearly this tail is not possible to observe when you consider that the window of observation in the simulation is equal to the age range of the population, 15-44 (30 years). 1.7 shows the density plot of 1000 relationships lengths that are randomly generated from an exponential distribution with a mean of 476 weeks based on the data for marriages and cohabitations, and 95 weeks for the casual. While 97.3% of randomly generated marriages lay within the simulation window, the removal of the tail lowers the mean observable relationship length based on this distribution (the mean of relationship lengths if you remove the observations that are impossible to occur in the simulation) from

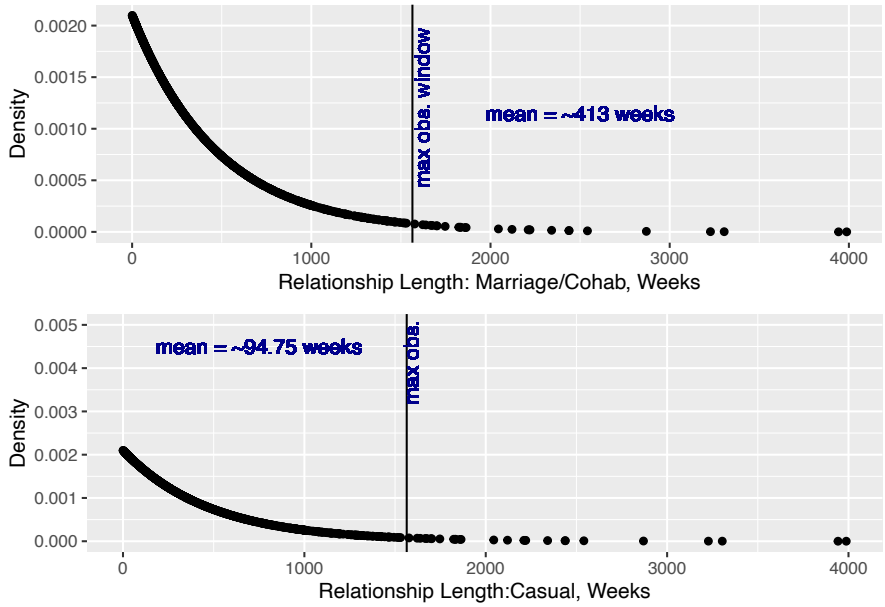

Figure 1.7: Predicted Distribution of Relationship Lengths and Simulation Window

476 weeks to 426 weeks. The mean relationship duration in the casual network is also shown to demonstrate that the simulation window of each node is unlikely to contribute to the variation we see in the mean simulated relationship duration in the same way and as such we will only implement a correction for the marriage network.

Recall the previous description of the formation approximation that used log of the expected relationship duration to convert the edges coefficient estimated by a static ERGM from a prevalence term to an incidence term for a dynamic network. In the present scenario, we modify this approximation for the marriage/cohabitation. Instead of using the log of the target duration, we instead use use the log of the mean relationship length that we estimate from the distribution that was truncated by the simulation window length. This will slightly increase the underlying rate of formation in the network and hopefully will both help us recover missing edges across the network but specifically increase the number of edges that form earlier in the life-course, improving our fit of the full mean degree distribution and increasing the mean observed relationship length. The boost in the edges coefficient successfully increased the overall mean degree of the network to within 2% of the target mean degree. However, very little of the

Table 1.5: Mean Degree and Duration Comparison, Targets vs Edapprox Correction

	Mean Degree			Mean Duration		
	Target	Sim Result	Pct Off	Target	Sim Result	Pct Off
Marriage/Cohab	0.455	0.447	-1.76	476	369	-22.48
Casual	0.159	0.152	-4.40	95	103	8.42

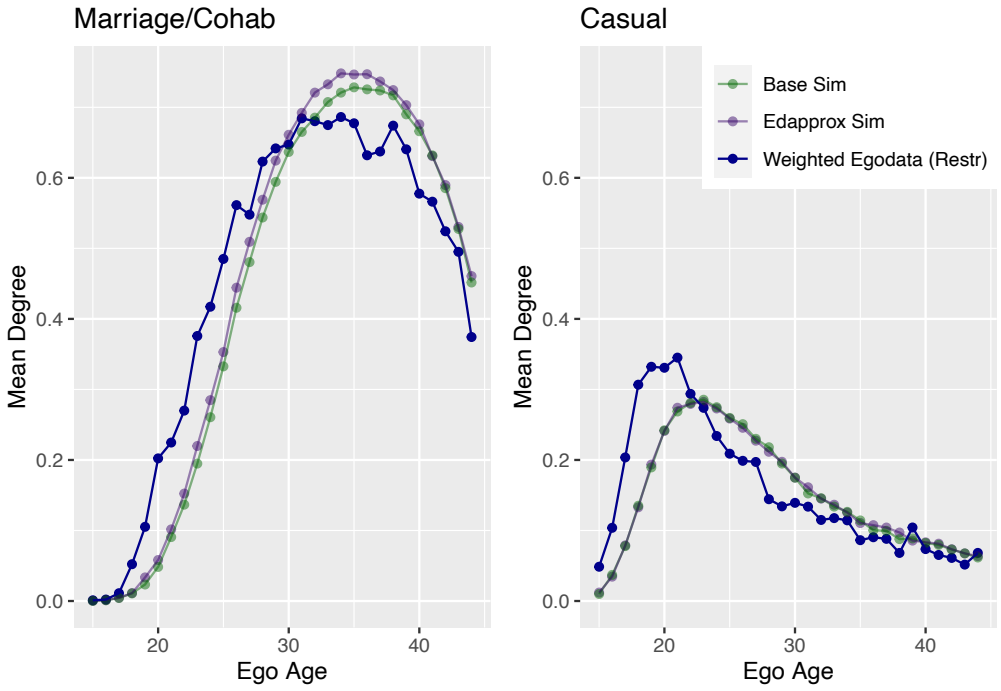

Figure 1.8: Mean Degree Comparison: Edapprox Correction

increase in mean degree came from an increase in relationship prevalence at younger ages. Instead, the boost largely only increased the degree at the peak, which was already over-representing relationships in the mid-to-late 30s. Additionally, we only gained about two months in mean relationship duration. This is again likely due to the increase in mean degree in individuals in their 30s rather than across the network more evenly. If more relationships had formed among younger individuals, more relationships would have the opportunity to last longer, improving mean length. Unfortunately while this is a step in the right direction, it is becoming clear that in order to meet the target mean degree in both the marriage network and the casual network we will have to address the issue of formation at younger ages more directly. In the next section we will address one more facet of the age-boundary related issues before moving our focus to the formation of edges among the younger ages.

1.9 Departure

The departure correction seeks to balance out the "unexpected" edge dissolutions due to nodes departing the simulation due to aging out or age-specific mortality by increasing the underlying expected length of relationships in each network (or rather, decreasing the log-odds of dissolution). The current departure correction generates a single estimate that is applied to the dissolution coefficient of both networks. When considering these marriage and casual networks among heterosexuals across a wide age range however, this assumption may not hold given the extreme variation in the prevalence of relationships by age. Most nodal departures in the simulation are due to nodes departing at age 45, but nodes of this age are far more likely to dissolve a marriage or cohabitation than a casual relationship upon departure. In this scenario, I re-consider the standard implemented departure correction by incorporating information about the prevalence of ties across likely departures.

The current correction is calculated by adding the probability that any one node departs the network due to aging out multiplied by the mean weighted age category-specific mortality rate per time step to estimate the average probability of departure for any given node in the network per time step:

$$drate = \frac{1}{w} + ASMR_{weighted}$$

Table 1.6: Original and Updated Mortality Rates

Original Mortality Rate	Updated Marriage Rate	Updated Casual Rate
0.0006645	0.000749	0.0002807

where w is the number of weeks we observe each node from entry to exit in the absence of death, 1560.

The new correction represents the likelihood that if a node departs, it also dissolves a edge.

$$1 - \sum_{a=1}^{6} S_a * D_a * P_a$$

where

a = each 5-year age category labeled 1-6, representing ages 15-19...40-44

S = the survival probability for a node in a given age category due to aging out per week

D = the probability of death for a node in a given age category per week

P = mean degree of a given age category relative to the cumulative mean degree

Table 1.6 displays two sets of departure correction: the original departure correction (which is the same for both networks), and a second using the new formula. Notice that new estimate for the marriage network that is slightly higher than the original departure correction, this formula produces a significantly smaller departure correction for the casual network. This makes some intuitive sense given the a node departing at age 45 has roughly a 50% change of dissolving a marriage, but less than a 10% chance of dissolving a casual relationship. The updated departure corrections improved key metrics in each network in different ways. First, the marriage network reached its target mean degree and slightly increased the mean duration of relationships. This appears to be largely due to the slight increase in the prevalence of relationships among nodes in their late-20s and a larger increase in relationships among nodes in their 30s. Once again, because this correction is not age-specific, the largest effect is seen at ages where the mean degree peaks. In the casual network, the mean relationship length has reached its target. Unfortunately, because relationships are now slightly shorter than in previous simulations, without a corresponding increase in the rate of

Table 1.7: Mean Degree and Duration Comparison, Targets vs Edapprox + Mortality Corrections

	Mean Degree			Mean Duration		
	Target	Sim Result	Pct Off	Target	Sim Result	Pct Off
Marriage/Cohab	0.455	0.455	0.00	476	387	-18.7
Casual	0.159	0.144	-9.43	95	95	0.0

relationship formation, the mean degree is slightly lower than before this correction. While this departure correction has strong theoretical support, it on its own is not sufficient to address all of the observed issues. We now finally turn to the issues relating to the left side of the distribution: the under-formation of ties between the ages of 15-25.

1.10 Arrival & Sexual Debut

The failure of these networks to adequately form relationships among the youngest ages is yet another form of a boundary problem. The big-picture problem is that when 15-year-olds enter the population, they do not bring in any existing relationships. This creates a problem for the model because the formation coefficients that govern the incidence of relationships at each age 15 are not estimated with the need to form all of the *prevalent* relationships among 15-year olds almost immediately upon entry. Additionally, unlike the diagnostics that occur in the closed system with a static node set, the age of each node is now a time-varying attribute (aging). This makes large jumps in the expected mean degree by age challenging because there is a limited time frame for nodes of a certain age to form sufficient new relationships like in the marriage network between age 18 and 25 or in the casual network between ages 15-20. Essentially, the seemingly straightforward change from a static nodal attribute to a time-varying attribute means we need slightly different network conditions in order to meet the expected age-specific mean degree targets.

In this section we test two possible approaches to this problem. The first involves manipulating the number of individuals eligible for relationships based on the sexual debut process, and the second takes a more direct approach to manually calibrate the formation coefficients at certain ages to boost the rate of for-

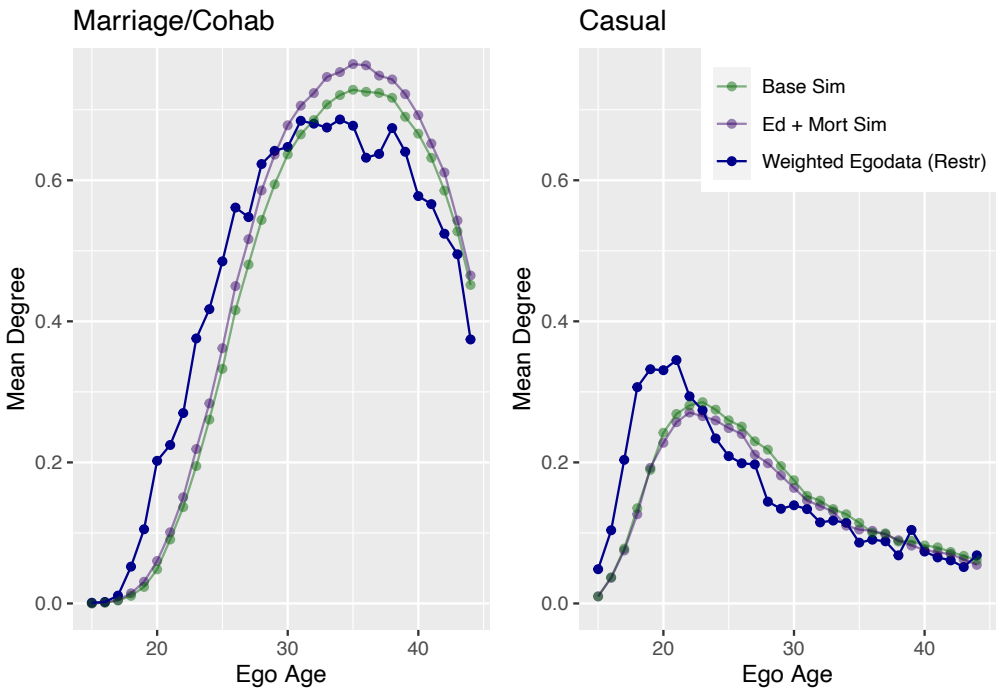

Figure 1.9: Mean Degree Comparison: Departure Corrections.

mation to account for the insufficient incidence rate. In order to evaluate the efficacy of each scenario, we will consider both the effect on the cross-sectional prevalence of relationships by age as in previous sections, and additionally the proportion of individuals who ever form a relationship (sexually debuted) by age while in the simulation.

1.10.1 Sexual Debut vs Readiness

Representing the sexual debut process is both complex and highly important if we wish to model sexually transmitted diseases in adolescents and young adults. In the U.S., more than 50% of all sexually transmitted bacterial infections such as chlamydia and gonorrhea diagnosed yearly occur among individuals aged 15-24, but not everyone in the age group is sexually active. This concentrates the transmissions into a subset of the population and increases the probability of exposure to an STI for those sexually active more so than at older ages. It is important then, to approximate this process in simulation as faithfully as possible.

If too many individuals are able to form sexual partnerships in the model, we may under-represent the risk of exposure for those sexually active and conversely over-represent the risk of exposure if too few are sexually active.

Here we take a moment to outline a few key definitions before describing the various possible implementations of this process.

1. Sexual Debut occurs when an individual first has a sexual intercourse of any kind with a member of the opposite sex (remember, we only represent opposite-sex contacts in this project). The NSFG explicitly asks if an individual has had sexual intercourse with a member of the opposite sex, and if so, what month/year did they first have sex.

2. Readiness is the state of an individual who has not yet had sexual intercourse (or "debuted"), but feels ready to do so. This state is not captured in our empirical data, but is the parameter we would ideally want to use in our dynamic networks to signal that a particular node is eligible to form a partnership. In the following scenarios we will use the nomenclature of sexual debut to model different ways we can represent readiness in these models.

For our baseline model, we follow common practice and assume that sexual debut and readiness are the same metric. Individuals enter the model with a 10.6% probability of debut, based on the proportion of 15 year olds in the NSFG who reported having sexual intercourse with a member for the opposite sex prior to age 15. For the rest of the age distribution, we used the responses to "have you ever had sexual intercourse with a member of the opposite sex" to generate a cross-sectional distribution of sexual debut status. From this data we estimated the weekly probability of debut among those who have not already debuted. The empirical data and the in-simulation distribution of the sexual debut attribute from the base model scenario before any adjustment are shown in Figure 1.10. Unfortunately, while this approach is straightforward it also creates somewhat of a catch-22. Because an individual in our model cannot form a sexual partnership *until* they have been labeled as "debuted" by the attribute adjustment process described above, there is a lag between receiving the attribute flag of "debuted" and actually forming a partnership. So while we can match the distribution of this attribute to the empirical data, the number of individuals who have truly formed a relationship for the first time in our networks is lower than the

observed data, and could contribute to the ongoing issues surrounding matching the target mean degree, particularly at younger ages. In the following sections, we refer to sexual debut as the act of actually forming a relationship in the simulation, and readiness as the attribute that governs whether a node is available for this debut.

In our next scenario, we to use the "debuted" nodal attribute as a signal of readiness to form a tie. Unfortunately our survey data do not allow us estimate the average time-to-debut directly (i.e. at what age did you decide you were ready for sex vs at what age did you actually start having sex), and the literature has largely focused more on individual characteristics and within-partnership dynamics that predict sexual debut rather than quantifying the time to readiness or the time from readiness to debut (Cavazos-Rehg et al., 2009; Kaestle, Morisky, & Wiley, 2002; Lara & Abdo, 2016). In the absence of additional information, we instead alter only the probability of having sexually debuted at entry at age 15 such that the rate of relationship formation *in-simulation* matches the proportion of 15 year olds who report having had sexual intercourse. This calibration results in a 90% probability of sexual readiness upon entry into the simulation. We then assume that after age 15 readiness to form relationships increases at the same rate we used earlier for sexual debut. We hope that increasing the number of individuals who are available to form partnerships while maintaining the originally estimated coefficients for the rate of relationship formation will help us better match the total number of relationships expected within the younger population.
The switch to this readiness metric had some dramatic effects on the casual network and moderate effects on the marriage/cohabitation network. In the marriage network, the overall mean degree has increased to about 9% greater than the target, and although we do see increases in the mean degree at younger ages that almost matches the targets, once again, the majority of the degree increase is seen between ages 30-40. The increase in the number of relationships that begin at earlier ages has increased the mean relationship length by roughly one year, but we still fall far short of the target. In the casual network, the increase in available egos for casual relationships has led to a very large increase both the overall mean degree and in the mean degree in the under-30 population. The mean relationship duration in this network has stayed within 1% of the target, but over-represents the total number of relationship at most ages. This scenario will come the closest to reproducing the actual debut distribution of the data (Figure 1.13), but largely at

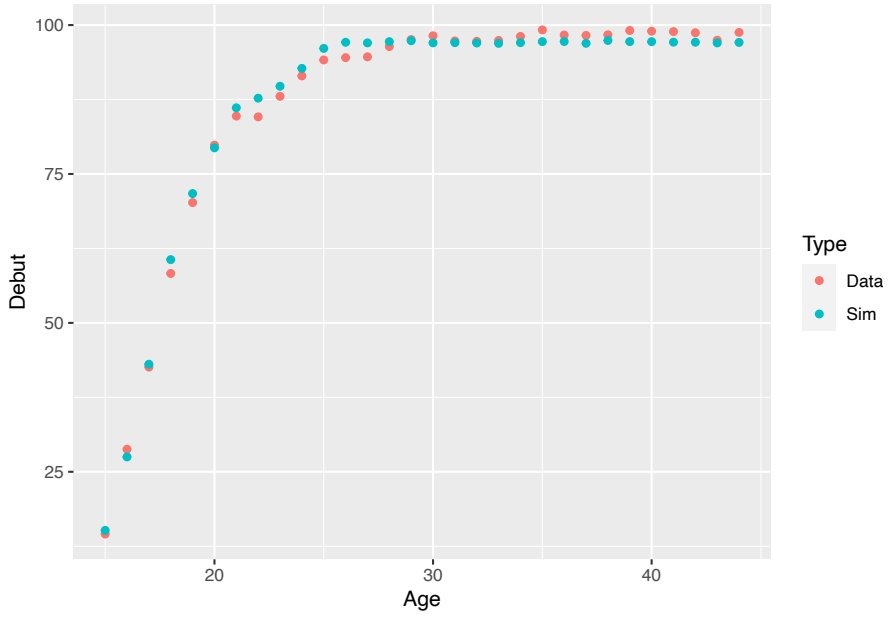

Figure 1.10: Sexual Debut Status: NSFG vs Simulation

Table 1.8: Mean Degree and Duration Comparison, Targets vs Expanded Eligibility

	Mean Degree			Mean Duration		
	Target	Sim Result	Pct Off	Target	Sim Result	Pct Off
Marriage/Cohab	0.455	0.488	7.25	476	401	-15.76
Casual	0.159	0.227	42.77	95	96	1.05

the expense of the casual network's degree distribution.

1.10.2 Young Age Formation Boost

In this scenario, our goal is to increase the rate of relationship formation among certain younger ages to increase the number of prevalent relationships at these ages by reducing the lag between becoming available for a sexual relationships and actually forming one. We revert the likelihood of sexual debut at entry to the baseline parameter and instead add an additional term to the network formation model in order to boost the log-odds of forming a tie among certain ages. I do so using a manual calibration process using

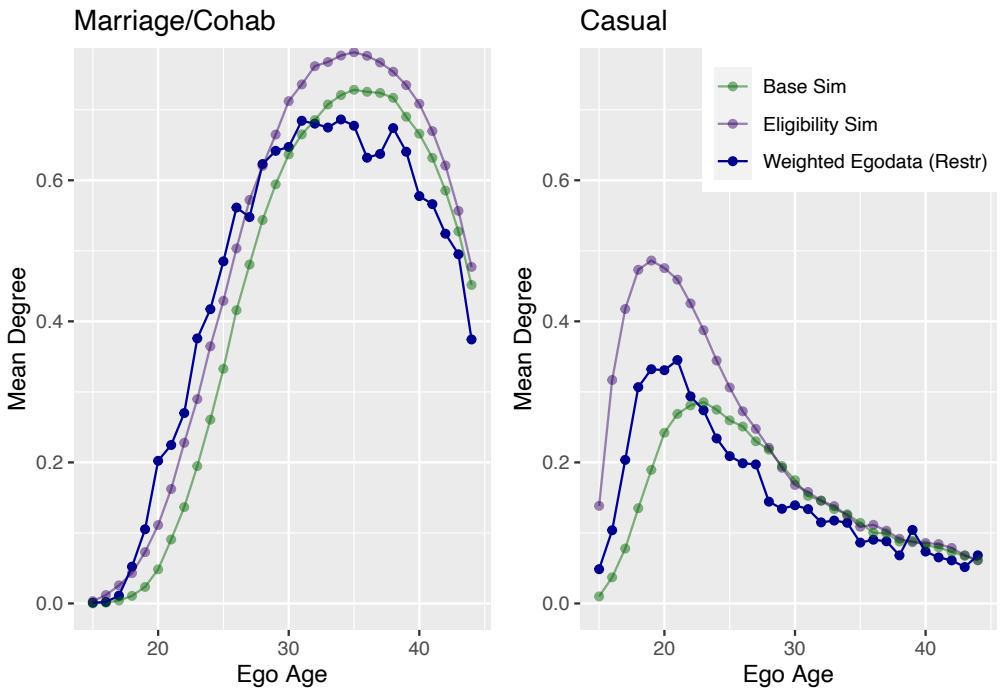

Figure 1.11: Mean Degree Comparison: Eligibility.

several parameters: a formation coefficient to boost certain nodes at specific ages, and the proportion of nodes at these ages that required the boost in order to match the prevalence of relationships in the 15-25 year-old age range. The original plan for this scenario did not include the second set of parameters, we intended only to boost formation at entry to correct for the boundary issues discussed above. However, in order to match the expected degree distribution in these younger ages, we found that additional boosts among nodes who had just aged into the next integer year (i.e. recently turned 17) were necessary at certain ages where the mean degree increased rapidly year-over-year. In the casual network we applied to this additional formation probability to all nodes at age 15 for the full year they are 15, and at age 17 and 18 for the first three months they are 17 and 18. In the marriage network, this additional rate of formation was applied to all nodes at ages 18 and then again for the first three months that nodes are aged 20 and 23. (At the end of this chapter, in the applied example, we will use a similar correction but will include separate formation coefficients at each age for ease of calibration which increases the number of coefficients needed to calibrate, but is more intuitive than applying the same boost to different proportions of

Table 1.9: Mean Degree and Duration Comparison, Targets vs Young Formation Boost

	Mean Degree			Mean Duration		
	Target	Sim Result	Pct Off	Target	Sim Result	Pct Off
Marriage/Cohab	0.455	0.497	9.23	476	413	-13.24
Casual	0.159	0.183	15.09	95	96	1.05

those nodes of a certain age). This formation boost had an interesting effect on the networks. First off, we were able to finally match the distribution of relationships at younger ages. However, the increase in incidence at younger ages seemed to have increased the prevalence at older ages, so we still over-represent those relationships and overshoot our network-wide mean degree in both networks. The increase in relationships increased the average relationship duration in the marriage network, although it still falls short of the target by roughly one year. In terms of sexual debut (Figure 1.13), the rate of effective debut using the "eligibility" framework came closest to matching the empirical data, but the scenario that boost formation with the default debut framework failed to boost the effective debut enough (although it was an improvement over previous scenarios without any boosting of young-age formation). Interestingly, we ran an additional with all baseline parameters but without a simulation-governed debut process and found that the proportion of nodes who have ever formed a relationship by age in the model was term was almost identical to the model that contained a simulation-governed debut flagging process plus boosted formation at younger ages.

1.11 Summary of Individual Corrections

It is clear that there are no one-size-fits-all corrections that addresses all of the effects that emerge when demographic dynamics are simulated. There are however some key conclusions and recommendations regarding how to implement future adjustments. First, while most adjustments influenced many of the key metrics at the same time, we need to focus on three primary issues: network-specific departure correction, boosting the formation at young ages, and making additional adjustments for the duration of

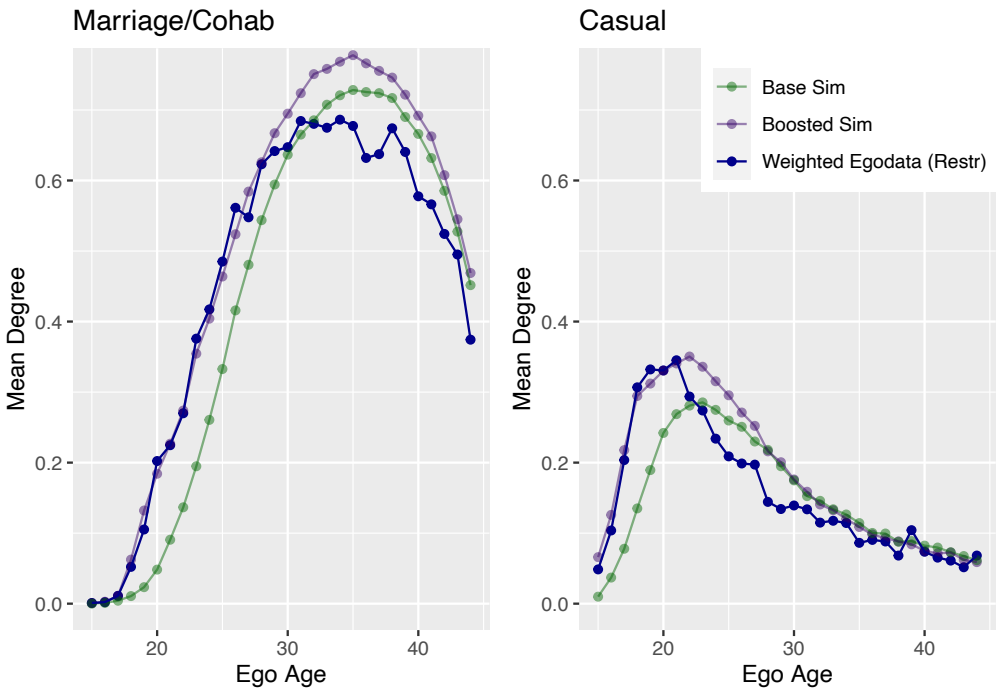

Figure 1.12: Mean Degree Comparison: Young Age Formation Boost.

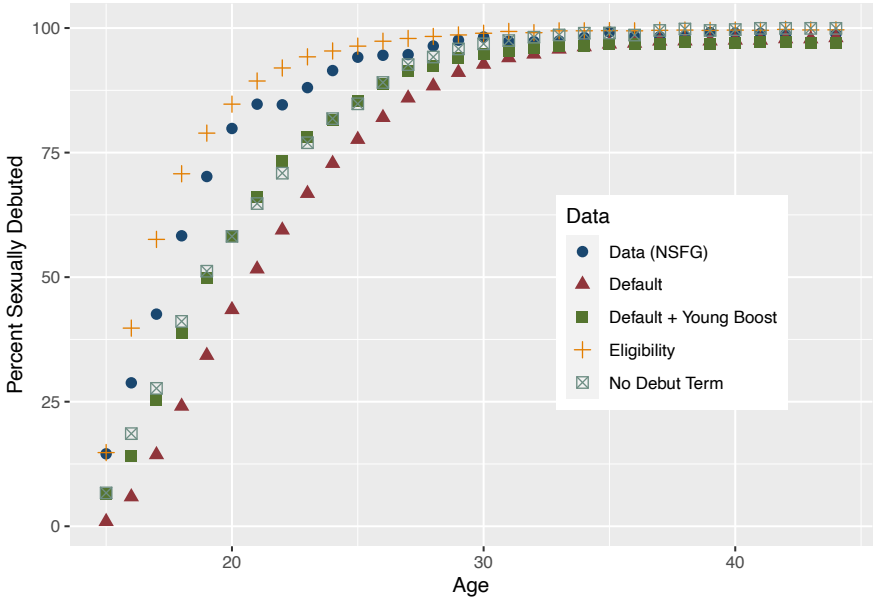

Figure 1.13: Percent Debuted In-Sim vs Data, Various Scenarios

Table 1.10: Mean Degree Comparison Summary Table

	Target	Base	Older Partner Offset	Increased Age Boundary	Sim Window Correction	Sim Window + Departure	Increased Eligibility	Young Age Boost
Marriage/Cohab	0.455	0.431	0.434	0.472	0.447	0.455	0.488	0.497
Casual	0.159	0.152	0.152	0.141	0.152	0.144	0.227	0.183

Table 1.11: Mean Relationship Duration Comparison Summary Table

	Target	Base	Older Partner Offset	Increased Age Boundary	Sim Window Correction	Sim Window + Departure	Increased Eligibility	Young Age Boost
Marriage/Cohab	476	365	364	414	369	387	401	413
Casual	95	103	102	104	103	95	96	96

very-long relationships. The second takeaway is that sexual debut may not be as large of a problem as originally expected, and it is worth exploring what the distribution of all nodes who ever form a relationship looks like when there is no specific debut term *and* boosted formation. The next section will use these conclusions to implement several of these corrections into a set of networks estimated from the same data and will eventually be used in Chapter 3.

It is possible that some of the issues we have fitting these networks is related to the assumption that marriages and casual relationships both have a constant hazard of dissolution over time. If we are not capturing the cross-sectional distribution of relationship lengths, perhaps that would also influence our ability to capture the distribution of relationship prevalence. In the below section I show that we can recapture several important targets using several of the corrections tested above, but Chapter 2 will explore this line of thought further and attempt to describe the pattern of relationships lengths over the life-course and under what stratification the constant hazard may or may not be a reasonable assumption.

Table 1.12: Key Network Targets and Pre-Calibration Results

Mean Degree			Mean Relationship Length		
Target	Simulation	Pct Off	Target	Simulation	Pct Off
Marriage/Cohabitation Network					
0.486	0.352	-27.574%	476	304.128	-36.108%
Casual Network					
0.161	0.232	44.441%	55	52.063	-5.34%

1.12 Example: A synthesis of approaches for an applied project

In this example, I construct the marriage/cohabitation and casual networks that will be used in Chapter 3 to model chlamydia transmission. The networks are similar in structure to the networks previously described in this chapter, and the model terms are described in more detail in Chapter 3 and the Appendix. I calibrate specific parameters among these two networks based on several insights we gained in Chapter 1: that we need a boost in the formation coefficients at younger ages to account for left-boundary issues and rapidly changing targets in this narrow age range, and that the the same departure correction should not be applied to both the marriage/cohabitation and casual networks. In the marriage network, we boost the log-odds of forming a tie among nodes aged 15-19 and 20-24 by 25% and 12% respectively, and decreased the log-odds of tie formation for all nodes older than 30 by 12% to reduce the excess number of relationships we observed in various approaches to demographic corrections earlier. In the casual network, we boost the log-odds of tie formation for males at age 15 (at entry) by 24% and at age 18 by 9%. For the departure corrections in this project, instead of estimating the correction needed to account for relationship dissolution due to node departure, we simply calibrated the dissolution adjustment necessary for each network in order to match the target mean relationship length, i.e. the endogenous relationship length needed to match the target mean relationship length in the presence of these excess dissolution. Calibrated parameters generated an increase in the underlying length of relationships by a factor of five in the marriage network, but an increase in the length of casual relationships by only 3%. All formation coefficients for the baseline networks and relationship length targets are outlined in the appendix. Tables

Table 1.13: Key Network Targets and Post-Calibration Results

Mean Degree			Mean Relationship Length		
Target	Simulation	Pct Off	Target	Simulation	Pct Off
Marriage/Cohabitation Network					
0.486	0.491	1.109%	476	487.123	2.337%
Casual Network					
0.161	0.156	-2.931%	55	54.918	-0.15%

1.12 and 1.13 show key overall metrics for the baseline network pre-and-post calibration. With only several small changes, we are able to reduce the difference in mean degree in the marriage/cohabitation network between the target and the simulation mean from 28% to 1.1%, and the error in relationship duration from 36% to 2%. In the casual network, we reduce the error in mean degree from 44% to 3%, and relationship length from 5% to almost 0. We also are able to represent the mean degree by age much closer to the empirical data, although for the sake of space the pre-calibration and post-calibration plots are located in the appendix. While I manually calibrated these adjustments, in the future it may be advisable to use a method like Approximate Bayesian Computation in order to minimize the errors even further.

2

A Survival Analysis Perspective on Relationship Duration for ERGM-Based Epidemic Simulations

In the previous chapter, we stratified relationship types into two main categories: marriages/cohabitations, and casual relationships, with a single term for the probability of dissolution among each type. This is the standard in the recent literature, alongside a third network for one-off relationships that is important for the transmission of disease but that I set aside in the previous chapter in order to focus on models with relationship durations greater than one time step (I will continue to ignore these one-off relationships in the work below). Some recent models have also added terms to the dissolution models to stratify by race-dyad characteristics among each network type. However these networks, even with additional

terms for race/ethnicity, assume an exponential process within each stratification. This reliance on a memoryless process makes the estimation and simulation of the underlying temporal ERGM (TERGM) more tractable, but may have some unintended consequences. While we can reproduce the mean relationship length estimated from empirical data in the network simulations, it is currently unknown how well the exponential framework reproduces the full distribution of empirical relationship lengths. In this chapter I will first begin with an explanation of the importance of relationship length on the transmission of STIs, highlight some key issues related to demographic trends and constraints of current epidemic models, and set up the scope of the analysis. Then I will use parametric and non-parametric tools from survival analysis to compare models of relationship duration and some simple extensions to the exponential framework, with the goal of exploring how well this memoryless processes captures the empirical distributions of relationship length in the National Survey of Family Growth overall and among various stratifications.

2.1 Relationship Length Overview

The duration of sexual relationships across a population is a key component of the network structure responsible for either exposing individuals to or protecting individuals from sexually transmitted infections (STIs). Relationship duration determines the length of exposure to pathogens, or in the case of a disease-free monogamous partnership, protection from pathogens. In addition to dictating this period of possible exposure, relationship durations relative to the pathogen-specific duration of infection are an important driver of how quickly STIs can spread throughout a population. Transmission beyond a pair of actors for infections with short durations relative to relationship lengths is challenging and slow, and it is more likely that an infection will be detected and treated or resolved naturally prior to the dissolution of the relationship. If the duration of infection and duration of relationships are more equal, there is a greater chance that the infection can spread to future partners and throughout the network. When partnerships overlap, transmission pathways increase even among those individuals with few lifetime partners, and this effect is even greater when the duration of overlap is large (Armbruster, Wang, & Morris, 2017; Morris & Kretzschmar, 1997).

The pattern of relationship durations across the life-course is also important because STIs often have distinct age patterns in terms of prevalence. Individual age is often used as a predictor for risky sexual behavior, but there is additional complexity when considering the effect of age on the duration of relationships across the life-course. Young age likely influences the immediate intentions for relationships (i.e. serious or casual), and the frequency at which individuals form new relationships, but it is also true that the only people who can report extremely long relationships are those who started them at young ages. This also introduces complex sampling issues because most data on relationship durations is collected cross-sectionally or retrospectively – not longitudinally (see description of methods below for more on this). Given the importance of relationship duration to features of STI epidemiology discussed above, there is growing interest in improving the representation of relational durations in dynamic network models used to study epidemics.

As we used in the first chapter and will continue to use throughout this dissertation, one common class of models used to understand network influences on patterns of STI transmission is known as separable temporal exponential-family random graph models (STERGMs). These models are governed by two expressions: one that represents the set of processes that influence the formation of relationships, and a comparable one for dissolution (Krivitsky & Handcock, 2014). We have previously explored some corrections to these expressions related to unexpected effects of certain demographic processes, but here we explore assumptions inherent in the dissolution component in more detail. The current standard practice for the dissolution models in this modeling framework assumes that once a relationship begins, its persistence is governed by a constant hazard. As previously alluded to, this memoryless process is a convenient simplifying assumption that makes TERGM estimation easier, but it seems unlikely that this assumption faithfully represents the distribution all relationship durations we observe across a wide range of ages.

Epidemic models in the recent literature have addressed this simplification by splitting out relationships into two categories: the first, marriages and cohabitations or main partnerships, and the second, persistent or casual partnerships. These are then modeled as separate networks simultaneously. This strategy is what we employed in the first chapter. By structuring the model in this fashion, each network has a hazard of dissolution specific to its type. (These models often have a third network for one-time sex-

ual contacts which last only one time-step, but this network is not the focus of our study). While these models are indeed able to reproduce the mean relationship lengths drawn from empirical data, it remains unknown how well these strategies reproduce the full distribution of lengths observed. In particular, the memoryless assumption means that the modal length of main partnerships remains near zero across all ages, which basic intuition says is not true and descriptive data analysis confirms. Other work has considered disaggregating relational durations by a single demographic attribute of their members related to a hypothesis or prevention modality being explored (Goodreau et al., 2017; Jenness et al., 2017).

2.2 Data

The combined 2006-2015 waves of the National Survey of Family Growth once again provide the empirical data for this investigation. In these analyses, however, instead of using only the relationship active at the time of interview (the cross-sectional distribution), we now use all of the data on current and past relationships (except for one-time partners). As mentioned briefly in the introduction, the survey design makes the information on relationship duration somewhat more complex to analyze than the other questions of interest. Each participant, if they have indicated they have had sexual intercourse, is asked about their three most recent relationships that are either ongoing or have ended within the last year. We then define relationship age as the difference between the month the ego reported first having sex with this partner and either the last month they reported sexual intercourse with that partner or the day of the interview if the relationship is ongoing. All relationships active on the day of the interview have right-censored duration since we do not know if or when they will end. Additionally, because we calculate duration from retrospective information, we also introduce left truncation that biases mean duration estimates upwards. For example, if someone reports having one monogamous 15-year relationship, we essentially know their 15-year relationship history. However, if someone has serial short relationships or long time intervals between relationships, we do not see these relationships going back 15 years, so we actually gather different amounts of information from each participant. Figure 2.1 highlights these phenomena. Blue relationships are those lengths that we observe via the NSFG questionnaire. Red extensions to the blue lines represent the theoretical true duration among the right-censored relationships. Green lines

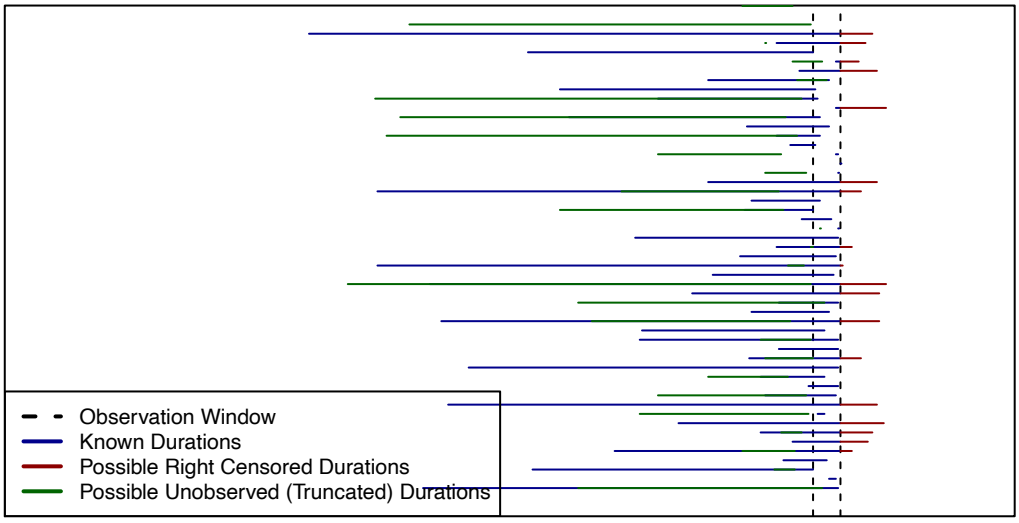

Figure 2.1: Known and Censored Relationships in NSFG

are those hypothetical relationships that could have occurred in the intervals that we do not observe for each participant. Many methods in survival analysis have corrections for these types of censoring and are employed in the relevant analyses.

2.3 Methods

First, the relational duration data is displayed using histograms (overall, by relationship type, and by age category). These histograms are not corrected for any censoring, therefore are solely used to get a visual sense of the underlying patterns. In the main analysis we explore several parametric survival models to gain insights into the underlying heterogeneity in hazard of dissolution. The goal here is not to find the most perfect fitting model, but to explore some simple extensions of the exponential that may be implemented within the constraints of current epidemic network models to better capture the full distribution of relationship lengths. Unless otherwise specified, the parametric models are fit using the R package 'flexsurv' adjusting for the right-censoring and left-truncation (Jackson, 2016). All models use the survey weights provided by the NSFG, which weight the observations to the age, sex, and race composition of the United States. Model fit is evaluated by the Akaike Information Criterion (AIC) and visually by using a

Modified Kaplan-Meier (following Burington et al. (2010) and fit using the R package 'survival') as reference curves to compare the survival of the empirical relationships to that of the fitted models. (BIC is not presented because the number of model parameters is so small between fits that the BIC and AIC provide almost identical outputs and any conclusions about model fit are unaltered). Additional visual comparison will be done using a PP-Plot, which plots two survival distributions against one another in order to visually evaluate the divergence between them (Cox, 2014). Models with ego attribute covariates will be fit twice: once using the all relationships and once stratified by two-category relationship type (marriages and cohabitations, casual) to reflect the current standard practice in the literature and to explore under what conditions the exponential process with stratifications may be a reasonable approximation of the data.

2.4 Descriptive Histograms

At first glance, the histogram of all relationships looks like something we would expect from an exponential distribution: a high decay right at the beginning, and a long right tail. However, it is clear from Figure 2.3 that this shape is primarily driven by the casual relationships rather than the marriages and cohabitations. The marriages and cohabitations are not uniformly distributed, but have a slower, linear-looking decay after an initial peak around three months. These trends are largely maintained if we break these types down further by age category of the reporting ego. Among casual relationships the primary age differences are 1) the number of relationships, 2) the frequency of short casual relationships versus longer casual relationships, with more frequent short relationships in younger age categories and a wide range of longer casual relationships maintained at older ego ages. Interestingly, the marriage and cohabitations look increasingly uniform with age. This may indicate that for certain relationship types, and for certain age groups, a simple constant hazard of dissolution may not accurately capture the distribution overall or over the life-course. Appendix Figure A.1 further breaks down these histograms by censoring status (i.e. ended or ongoing). The overall shape of the distribution is similar between relationships that are ongoing versus those that are ended, although there are far more short casual relationships that have ended than are ongoing and far fewer ended marriages and cohabitations than are ongoing.

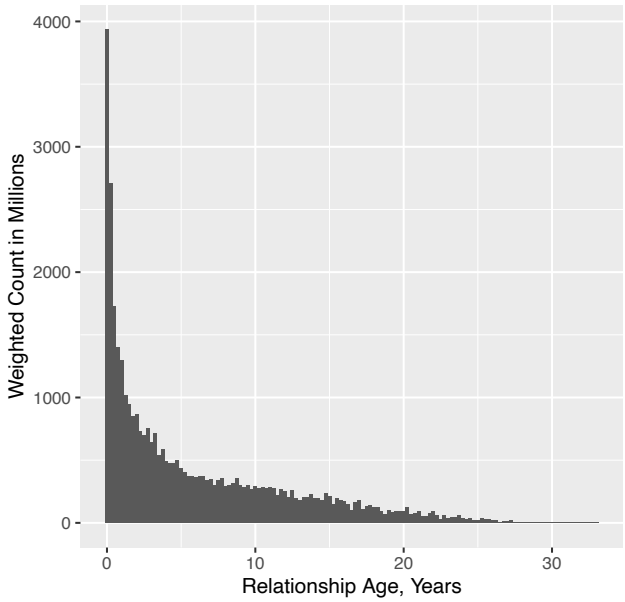

Figure 2.2: All Relationships either Current or Ended in the Last Year

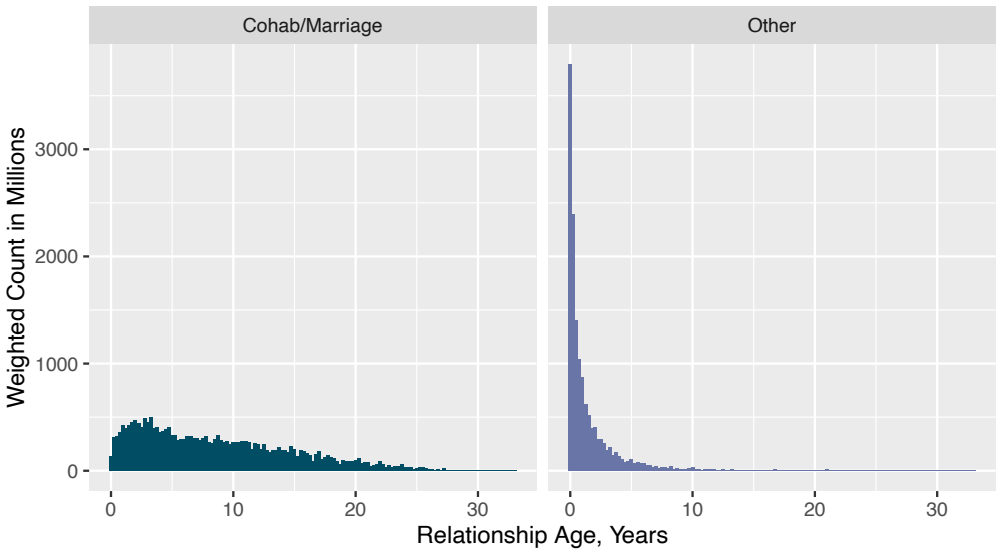

Figure 2.3: All Relationships either Current or Ended in the Last Year, By Type

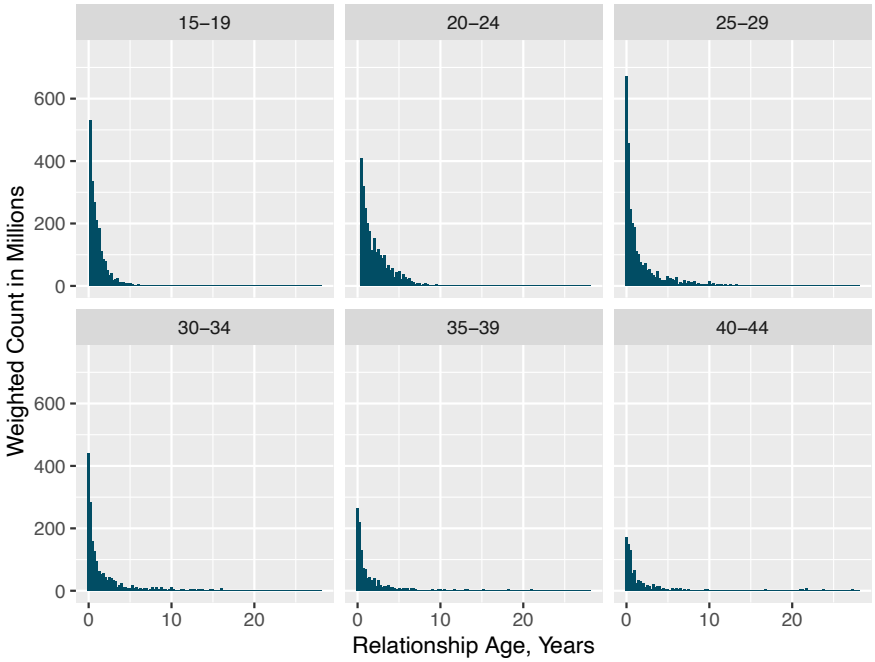

Figure 2.4: Casual Relationships, by Age Category

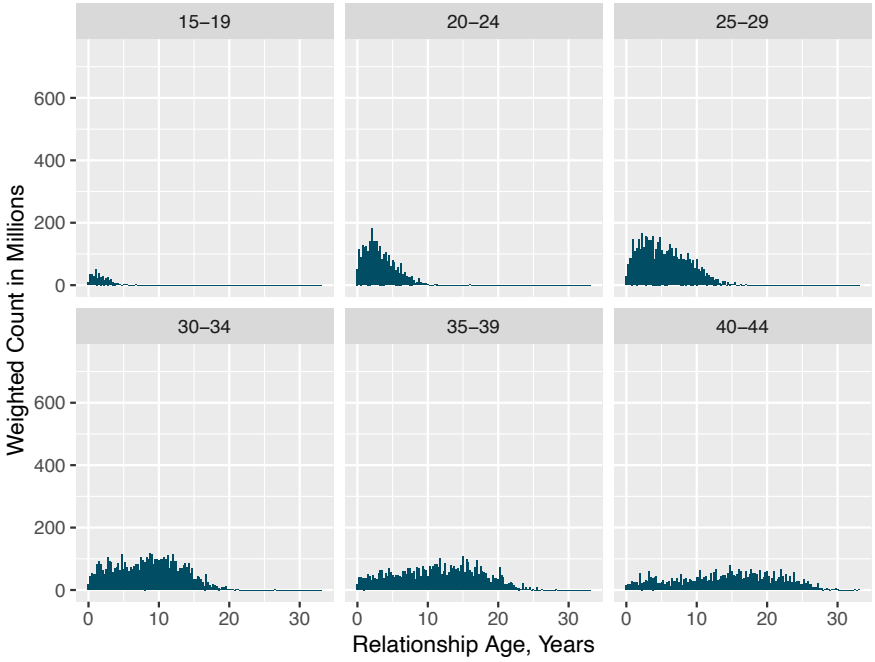

Figure 2.5: Mariages and Cohabitations, by Age Category

2.5 Duration-Only Models

As a first pass, we fit several duration-only (covariate free) models using 3 different but related distributions: the exponential, the Weibull, and the gamma. It would not be ideal to use either the Weibull or gamma as a dissolution model in our STERGMs because the parameterization would be dependent on the current duration of each relationship. This is not impossible, but it is computationally intensive. However, we choose to look at them because they are related to the exponential and a better fit using these distributions would represent that there is a heterogeneity in the data that the single-parameter exponential doesn't capture.

Figures 2.6 and 2.7 display these results. In 2.6, the data are represented by the black Kaplan-Meier curve and the fitted models with various distributions are in color. In Figure 2.7, the parametric models are plotted against the Kaplan-Meier survival estimates at each time step. Lines that fall above the (0,1) reference line represent places in the curve where the parametric model overestimates the survival of relationships and lines below the curve are where the parametric models underestimate the survival relative to the data. As we expected based on the histograms, a single exponential does not capture the full distribution of relationships well, but while the weibull and gamma capture the survival of shorter relationships better, all of the distributions fail to capture the very long, almost flat right tail of the data. While none of the covariate-free models capture the overall distribution of relationships, it is clear that there is important heterogeneity in the data that the exponential cannot capture alone. It is worth noting at this point that the parametric models assume that even though much of the data is right-censored, eventually all relationships dissolve and the survival curve will go to zero. This would be true if our data was gathered among egos across the full distribution of the human lifespan. But this is not the case, and because the Kaplan-Meier curves do not have this assumption, we will likely have poor fit in the tails of the the long relationship distributions across all models.

2.6 Simple Extensions to the Exponential

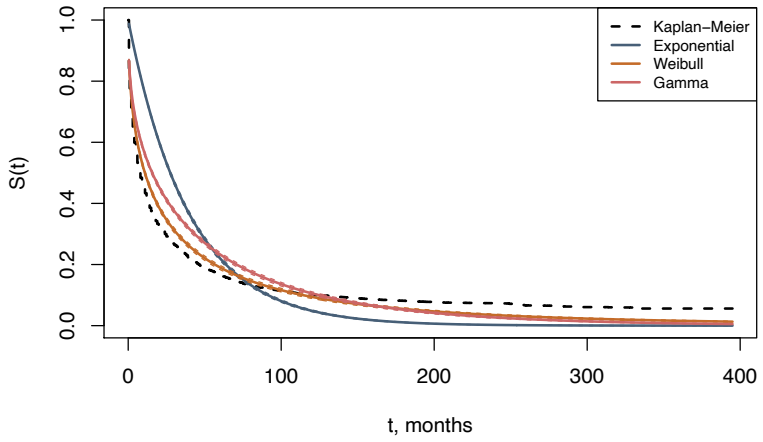

Figure 2.6: Various Duration-Only Survival Models, All Relationships

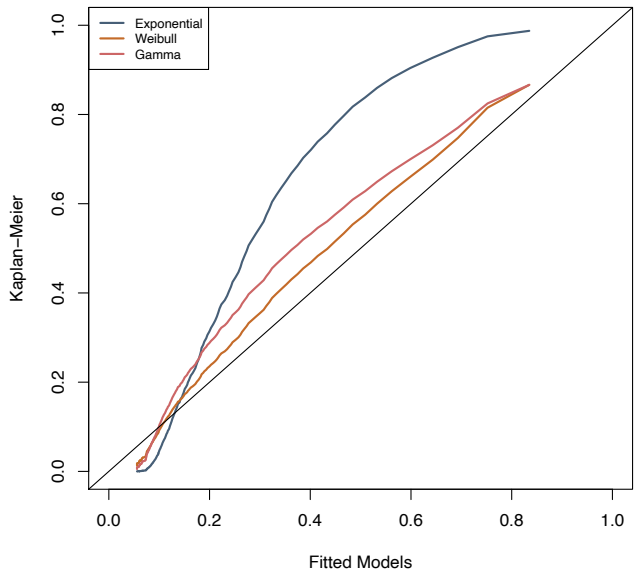

Figure 2.7: P-P Plot Comparison, K-M vs Various Probability Distributions

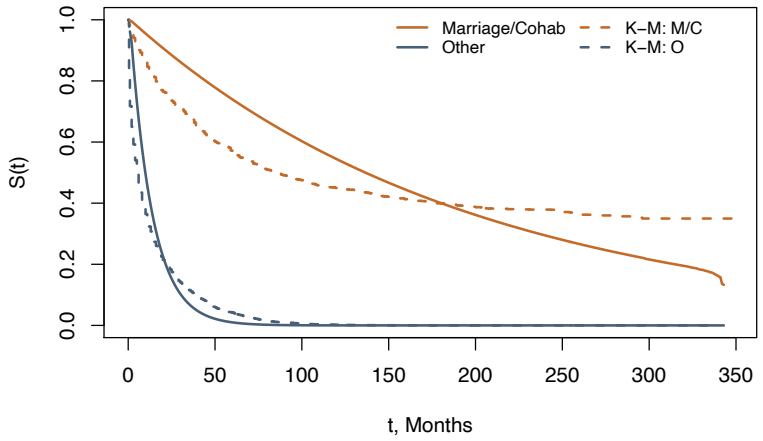

Figure 2.8: Kaplan-Meier vs Exponential - By Relationship Type

2.6.1 Relationship Type: Current Standard

Figure 2.8 stratifies the relationships into their Marriage/Cohabitation or Casual designations. Each relationship type then has its own hazard of dissolution, but within each relationship type the hazard is constant. The curve for the casual partnerships fits remarkably well to the reference Kaplan-Meier at first glance, but the p-p-plot highlights the difference in dissolution rate at the start of casual relationships. The Kaplan-Meier tells us that almost 20% of casual relationships fail within the first month but the exponential estimate is rather more conservative. Conversely, the exponential somewhat underestimates the survival of the longest casual relationships - likely a reflection of what we saw in the histograms of casual relationships at older ages that had much greater variation in length. The curve for marriage/cohabitations demonstrates similar issues. This model over-represents the survival of relationships that last less than four years, but under-estimates the survival of longer relationships. Clearly there is more heterogeneity here that we will try to tease out in the next examples.

2.6.2 Age Category

Here we break down the relationships by age category of reporting ego (Figure 2.10), and then further by relationship type and age category (Figure 2.11. The Kaplan-Meier reference curve by age category alone

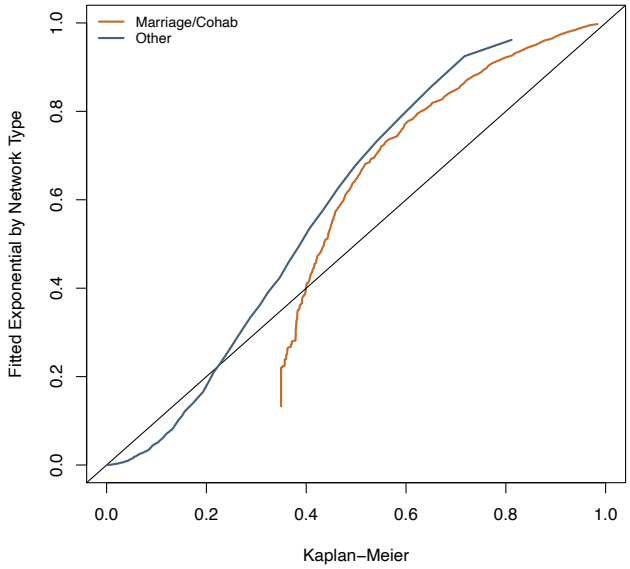

Figure 2.9: PP Plot, Kaplan-Meier (Obs) vs Exponential - By Relationship Type

shows very little difference in the survivorship of relationship among egos aged 25 and above, although the maximum length observed increases with age category (as expected). These difference in the maximum length however likely explain why the exponential curves predict large differences in survivorship by age category and overestimate the survival relative to the reference curves. Interestingly, relationships seem to dissolve at very similar rates in the first few months regardless of age category - a property that this model certainly does not reflect. The youngest ages come somewhat closer to their reference curves, but as the p-p-plot demonstrates, all strata suffer from overestimating the survival of young relationships and underestimating the long relationships. The model of age category among casual relationships (Figure 2.11, left) reveals very small differences in the curve between age categories, but all curves follow the now well-established deviation patterns relative to their Kaplan-Meier references. It seems we gain very little by adding age category as a covariate among casual relationships. While the *frequency* of casual relationships decreases across the life course, it seems that the *length* of these relationships follows a relatively similar pattern. Conversely, age category among marriages and cohabitations does seem to add to our overall fit (Figure 2.11, right). In particular, the curves within three youngest age groups (15-19, 20-24, and 25-30) fits their K-M references remarkably well. These observations are confirmed in the p-p-plots, Figure 2.12.

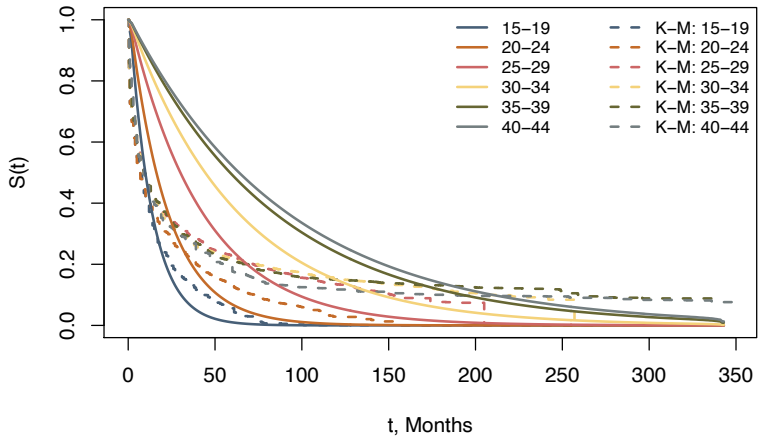

Figure 2.10: Exponential and K-M by Current Age Category

We might expect this lack of fit at older ages in the marriage/cohabitation network pattern based on the increasingly uniform distribution in the histograms among older ages. This is perhaps not surprising, in that the length of relationships is at least partly an emergent property rather than a causal one. That is, no individual can have a relationship that has lasted longer than they have been sexually active, so the range of relationship lengths for young age categories is relatively small and easier to represent. Meanwhile, the older age categories are challenging to represent because the possible range of relationships is so much larger, and are likely influenced not only by dissolution probabilities but also by the changing formation probabilities over the life-course – that is, older people in long-term relationships do not start new relationships at the same rate as others, and thus have relatively few relationships that are short.

2.6.3 Race/Ethnicity

Here we add a covariate for race/ethnicity of respondents. The results here are somewhat analogous to the age category covariate results. In the pooled relationship model (Figure (2.13), the Kaplan-Meier curves are almost identical between race/ethnicity groups until roughly 40% of relationships remain, around 20 weeks. The differences lie in the survival of the longest relationships. The exponential fits here are worse than using age category as a covariate. Among casual relationships (Figure 2.14, left)), we similarly gain little in adding race/ethnicity as a covariate. As above, the story for marriages and cohabitations is

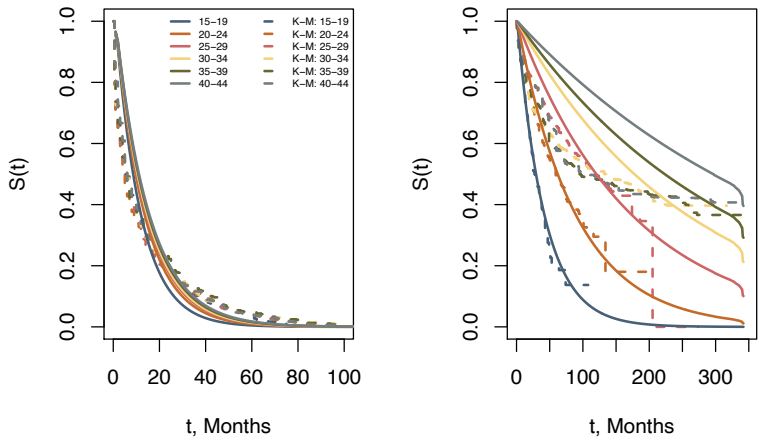

Figure 2.11: Exponential and K-M, By Age Category and Relationship Type

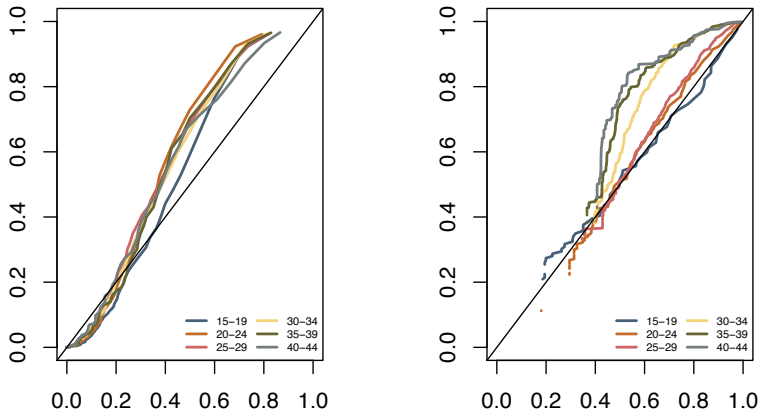

Figure 2.12: PP Plot, Exponential vs K-M, By Age Category and Relationship Type

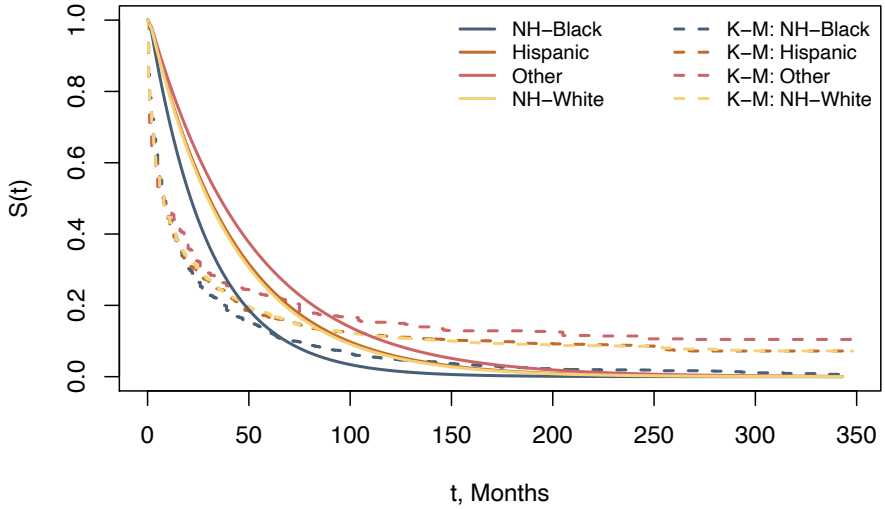

Figure 2.13: Kaplan-Meier vs. Constant Hazard by Race/Ethnicity

different. Here, the Kaplan-Meier shows clear differences in surviorship of relationships among Non-Hispanic Black respondents relative to all others. While this observation is reflected in the race covariate model as well, this stratified exponential models still fail to capture these relationships (particularly at the longest relationships). Figure 2.15 shows that the deviance between the Kaplan-Meier and the exponential curves are very similar across race/ethnicity groups, highlighting the poor fit.

2.7 Additional Relationship Types

Here we try to address the two phenomena that have been through-lines in the above results: first, that neither age category nor race adds meaningful value to the casual models and consistently underestimates the rate of dissolution in the first few weeks of beginning a relationships and second, that these covariates when applied to the marriage/cohabitation only seem to help fit the relationships that are somewhat shorter: relationships among younger and/or Non-Hispanic Black respondents. It is my hypothesis that these latter observations are due to a false assumption that cohabitations and marriages have similar properties and dissolution rates. Recent work in the field of family demography that has shown that there are significant differences in the risk of dissolution between cohabitations and marriages and that these dif-

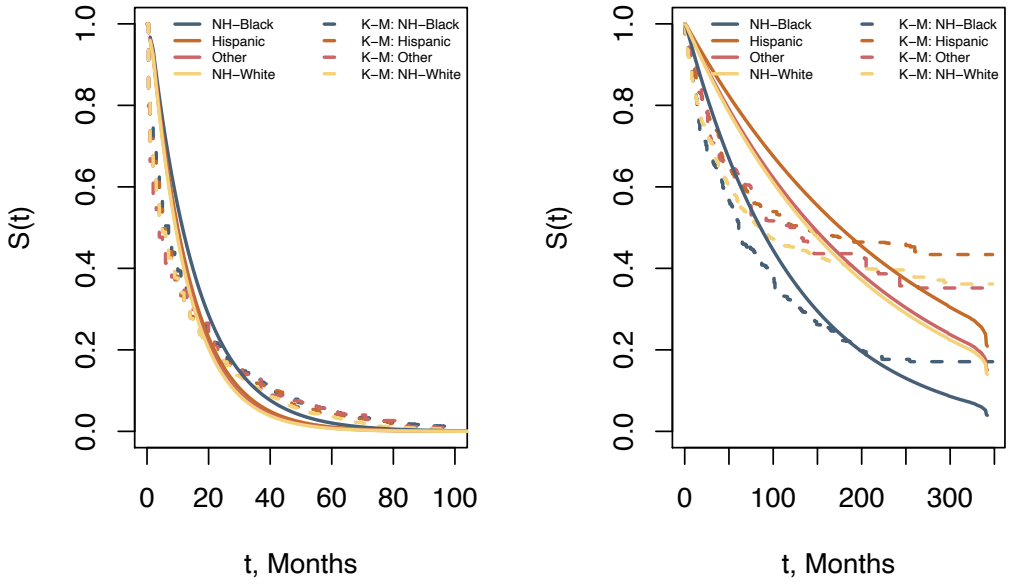

Figure 2.14: Kaplan-Meier vs. Exponential by Race/Ethnicity and Relationship

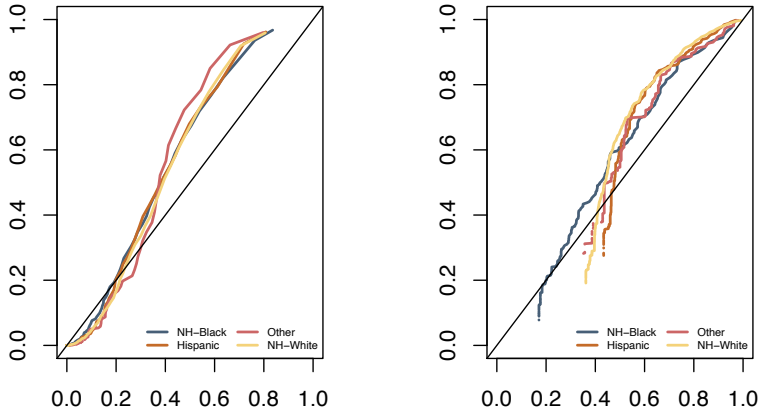

Figure 2.15: PP Plot, Kaplan-Meier vs Exponential - By Race/Ethnicity and Relationship

ferences are due to variation in joint lifestyles (van Houdt and Poortman 2018). Additionally, the role of cohabitation is complex: some couples use cohabitation as a trial prior to marriage, some prefer to cohabitate with no intent to marry, and some skip cohabitation and get married prior to living together. This suggests that while cohabitation itself is a heterogeneous category, it is distinct from marriage and we could improve the overall accuracy of our models if we had a separate dissolution risk for those in this cohabitation phase.

In the case of casual relationships, I return to the observation that the Kaplan-Meier curve shows that roughly one quarter of all casual relationships will fail within the first month, whereas the exponential models overestimate this survival. And indeed, the first quartile of observations in the empirical data are one month or less. This is likely, at least partially, an artifact of the way that relationships are reported and described in the NSFG. If a respondent reports a partner that they only had sex with once and do not expect to have sex with them again in the future, this relationship is labeled as "ended" and "once." These are the relationships that provide the data for our instantaneous networks when simulating for epidemics. If a respondent reports a relationship beginning in the same month of their interview (one month being the smallest unit of time in the NSFG) but expects to continue to have intercourse with this partner, this relationship is labeled as ongoing and we label these relationships as having duration of 0.5 months. Some of these relationships may be true brand-new relationships, but some of them also may be instantaneous relationships reported on by optimistic respondents.

In this next scenario, we make two changes. First, we split out cohabitations and marriages and model them separately. Second, we re-define casual relationships as those relationships that have lasted at least one month. Figures 2.16 and 2.17 show that these changes improve the fit by a remarkable amount.

2.8 Summary of Model Fits and Discussion

While age category and race/ethnicity of individuals may be important factors in determining the rate of relationship formation and partner selection, these covariates add little to the overall model fit of relationship *dissolution*. Both overall and within each of the two original relationship types, age category and

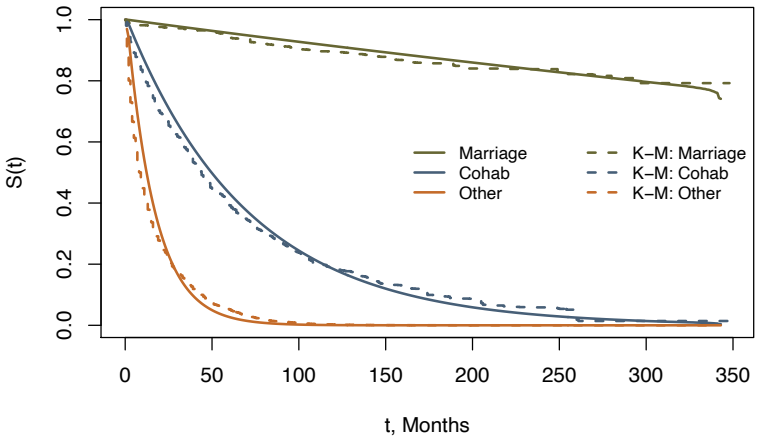

Figure 2.16: Kaplan-Meier vs. Exponetial with Three Relationship Categories

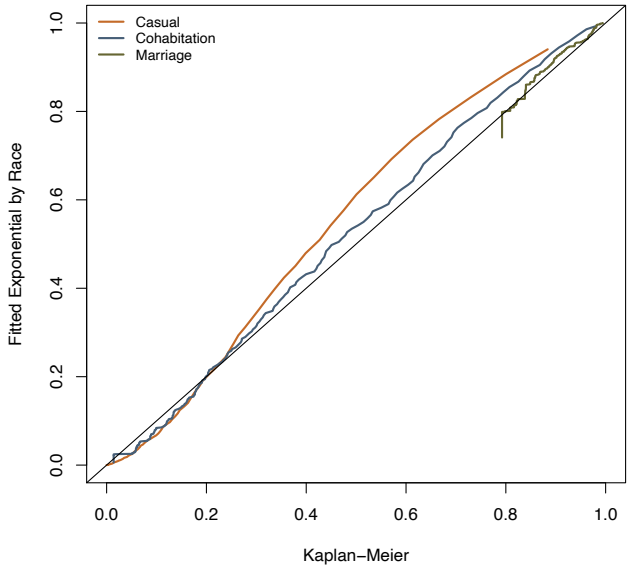

Figure 2.17: PP Plot, Kaplan-Meier vs Exponential - By Three Relationship Types

Table 2.1: AIC

	All Relationships	Casual	Marriage/Cohab
Duration-Only			
Exponential	84512	54633	16676
Weibull	74356		
Gamma	76027		
Ego Attributes			
Age Category	80720	54583	16183
Race	84326	54571	16622
Relationship Type			
Reltype-2	71309		
Reltype-3	59929		

race/ethnicity of the reporting ego have very little effect on the AIC, particularly among the casual relationships. This suggests a somewhat more universal experience for casual relationships, although from the p-p-plots and survival curves we can see that there is a consistent lack of fit at the start, suggesting that there is still heterogeneity in dissolution risk not explained by age or race: across the board, if relationships are going to fail, they do so very quickly. Model fit is most dramatically improved when we exclude casual relationships that have an ambiguous classification (those that started in the most recent time unit, and their true status as ongoing or ended is not known), and when we stratify the long relationships into marriages and cohabitation. Indeed, the small improvement in model fit among young egos and Non-Hispanic Black egos can likely be explained by this separation. Figure 2.18 shows the proportion of the types of relationships by race/ethnicity and age category. Relative to the other groups, Non-Hispanic Blacks have a higher proportion of cohabitations relative to marriages than the other groups, and the same is true for the two youngest age categories. So much of the difference observed between groups in the grouped long-duration network can likely be attributed to the higher proportion of cohabitations in these groups, which we capture better when we split by relationship type rather than ego attribute. It appears that within these three final categories (casual relationships one month or longer, cohabitations, and marriages) relationship lengths across the lifecourse *can* be reasonably approximated by an exponen-

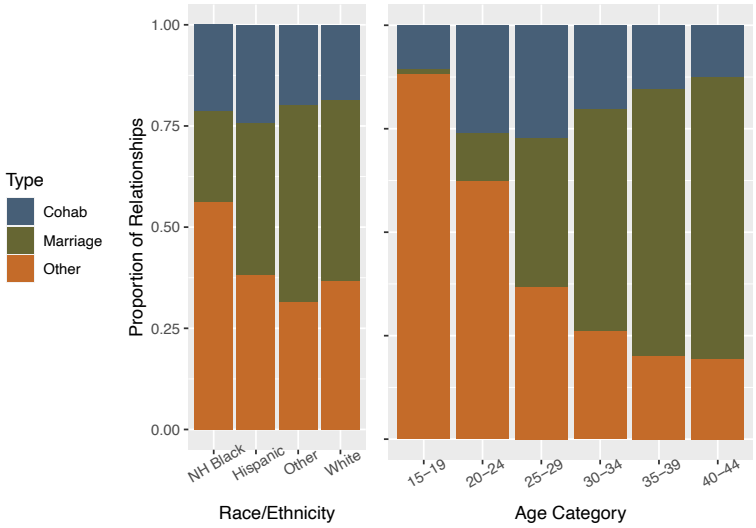

Figure 2.18: Proportion of Relationships Types by Race and Age Category

tial process.

These results have some implications for STERGMs developed for epidemic models moving forward, and two obvious strategies come to mind. The first and most straightforward given the current workflow would be to simply add an additional dynamic network based on the subsets in Figure 2.16 and add the relationships that we re-classified as instantaneous (but reported on by optimistic respondents) to the data that informs the instantaneous network. However given that the time unit in the simulations is one week and time unit in the data is one month, we may have a mismatch if we continue to assume that these relationships only last one time step or that only one sex act occurs (the instantaneous network at that point may not be truly separable, that is too many relationships start and end in between time steps to properly estimate the network). Of course, some of these suggestions, particularly relating to the instantaneous network, may not necessarily be relevant when using survey data sources that use a finer-grained time scale and were designed with mathematical modeling applications in mind, like the ARTnet survey for MSM (Weiss et al., 2020).

A more parsimonious solution however, would be to have a single dynamic network that captures all relationships (or at least all relationships longer than one month) with time-varying dissolution rates by

relationship type. In the current framework and ignoring instantaneous relationships for a moment, marriages and cohabitations are labeled as such from day one, and casual relationships are not allowed to transition to the longer class of relationships, which, while we *are* able to accurately model the prevalence of each relationship type, is still somewhat awkward. With this new approach, each relationships would begin as casual and transition type over time, with the constant hazard assumption maintained *within* each relationship type, but could vary over the full duration of each individual relationship. This representation of relationships as having decreasing dissolution probabilities over time provides a more intuitive framework for the structure of relationship in these models. As of writing this chapter, the statistical theory and software support for valued TERGMs, the class of models needed to label each relationship as a specific type with distinct dissolution probabilities, are not currently available (although valued ERGMs do exist). Krivitsky (2012) also outlines several possible methods for inducing non-constant dissolution hazards over the age of a relationship in STERGMs including a piece-wise function that could prove useful to this type of combined model, but it has not been implemented in the software and is outside the scope of this project.

3

The Role of Expedited Partner Therapy in Reinfection of Chlamydia

We now finally turn to an application of STERGMs to understand the role of partner therapy and concurrency on the reinfection of chlamydia. C. trachomatis is an obligate intracellular bacterium transmitted through sexual contact among humans. Chlamydial infections are most often asymptomatic, and untreated infections in women are an additional public health concern because they can lead to a variety of sequalae including pelvic inflammatory disease, scarring of ovaries and fallopian tubes, ectopic pregnancies, chronic pain, and infertility. Repeat infections are common and are an additional risk factor for the development of the above sequelae (Brunham & Rey-Ladino, 2005). There is a great deal of uncer-

tainty regarding the natural history of chlamydia, but the duration of infection for untreated individuals is generally thought to be up to six months for men and a year or more for women, although recent mathematical modeling studies suggests that the asymptomatic and undiagnosed infections may last up to three years (Davies, Anderson, Turner, & Ward, 2014; Geisler, 2010; Satterwhite et al., 2013). Chlamydia is usually treated with azithromycin or doxycycline, and unlike other common STIs like syphilis and gonorrhea, true antibiotic resistance is rare although treatment failure does occur (Kong & Hocking, 2015). In this project, we will use the demographically calibrated networks from the end of Chapter 1 sets of calibrated networks to explore several hypotheses about how reinfection of chlamydia can be influenced by biomedical and immunological processes.

3.1 Concurrency and Partner Therapy

Chlamydia is the most common reportable disease in the United States, particularly adolescents and young adults. The Centers for Disease Control and Prevention (CDC) estimates that half of all new STI infections (including gonorrhea, syphilis, and others) occur in those aged 15-24 despite them making up only a quarter of the sexually active population. The United States has some of the highest STI rates in the industrialized world, and despite this, funding for public health programs dedicated to these issues has largely declined (CDC 2016 STD Report). As a consequence, few health departments are able to offer traditional partner notification services, where a patient who tests positive for an STI gives the contact information of their recent sex partners to the health department, and the department then contacts their partners with the hope that these partners will then get tested and, if necessary, treated. Expedited partner therapy (EPT) was developed with this scenario in mind.

Under an EPT regime, a patient who tests positive, upon receipt of their own treatment, receives either additional antibiotic pills for their recent sexual partners or prescriptions for treatment that their partners can fill. The patient then is expected to hand-deliver either the treatment or prescription to their current or previous partner(s), who take the medicine at their own discretion and without the need for a visit to a provider or positive lab test. Both EPT and traditional partner services work by leveraging the sexual network, but this intervention eliminates the contact tracing burden on the health department and aims

both to decrease the time to treatment for all possible infected partners and to increase the total number of partners treated. It also has the potential to reduce re-infection among the index patients if the partnerships are ongoing.

There have been several clinical trials of EPT across the US (and Europe), including Washington State. These trails demonstrated that relative to traditional referral practices, EPT provision increased the proportion of partners who were ultimately treated, reduced the number of individuals who were re-infected at follow-up, and was less costly if at least 30% of partners were treated via EPT (Gift et al., 2011). However, there are several concerns about the real-world feasibility of this type of intervention. The primary concern concerns the overuse of antibiotics and lack of testing of sexual partners (Van Aar, Van Benthem, Van Den Broek, & Götz, 2018; Weiss et al., 2019). The over-use of antibiotics in general has been documented, and treating partners regardless of infection status could be highly wasteful if a large proportion of partners are not actually infected. Although the rise of antibiotic resistance is less of a concern for chlamydia compared to gonorrhea and syphilis, because infection with one STI is often associated infection with other STIs, many physicians are concerned that the lack of screening among sexual partners is a lost opportunity to test for these additional STIs and also monitor the various infections for antibiotic-resistant strains.

Additionally, community trials and implementation of expedited partner therapy at local clinics has provided mixed results regarding how well this intervention reduces reinfection of the initially diagnosed patient (Kerns et al., 2011; Taylor et al., 2013). Even in clinical trials, expedited partner therapy only reduced reinfection slightly compared to standard methods of partner referral (Golden et al., 2005; Shiely et al., 2010). This suggests that there are still barriers to reducing rates of reinfection that expedited partner therapy does not address.

In order for an individual to become reinfected following treatment one of several things must have happened. First, their own treatment could have failed, leading to another positive test at follow-up. Second, their infected partner(s) could have reinfected them either due to treatment failure or failing to get treated at all. Third, they could form a relationship with a new partner who is chlamydia-positive and become infected. There is a fourth possible option, where an individual who gets treated has a partner

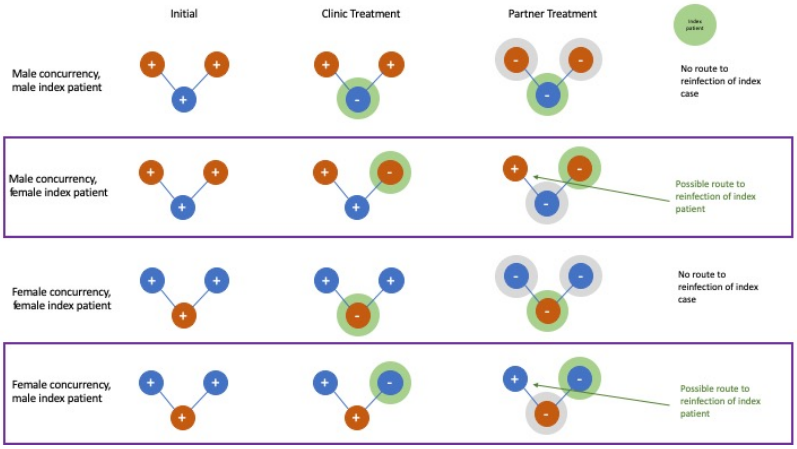

Figure 3.1: Pathways to reinfection following treatment due to concurrency

with other partner. This partner-of-a-partner could be infected, and if the index patient only treats their own partner, there is path to reinfection from the 3rd partner to the index patient (see Figure 3.1). Note that possibilities 2 and 4 in this list both rely on concurrency: first concurrency of the index patient, and second the concurrency of the non-index patient.

Hypothesis 1. The presence of concurrency exacerbates the rate of reinfection among individuals who provide treatment for their partners.

Concurrency creates a direct pathway from an infected third partner in a chain to reinfect the index node after the index and the immediate partner have been treated. We will test this effect by comparing the rate of reinfection following treatment between networks with rates of concurrency as reported in the National Survey of Family Growth, and a counterfactual scenario with a prohibition on concurrency. If we find that concurrency plays a large role in reinfection following treatment, this would provide more evidence for the need to have partners of diagnosed patients to seek in-person testing and refer any additional partners to the health department.

3.2 Arrested Immunity Hypothesis

One of the paradoxes in the era of modern public health is that chlamydia incidence has actually increased overall in the presence of mass control programs. In Sweden, Norway, Finland and Canada the rates initially decreased but then resumed increasing, and in Australia, United States, and the United Kingdom the rates never stopped increasing even after program initiation, although this second pattern has been attributed to the challenges of implementing control programs consistently throughout a large population (Brunham & Rekart, 2008). These areas now experience incidence rates higher than rates prior to introduction of control programs. Additionally, a regression analysis using data from family planning clinics in Region X of the United States (Alaska, Washington, Idaho, and Oregon) found that, after controlling for any changes in demographics, sexual behaviors, and increased sensitivity of clinical tests, there was a remaining 5% 'true' and unexplained annual increase in chlamydia positivity from 1997-2004 (Fine, Dicker, Mosure, & Berman, 2008). In response to these and other examples of unabated chlamydia infection in the presence of control programs, Brunham and Reckart have proposed the arrested immunity hypothesis. Under this hypothesis, early detection and treatment of chlamydia interrupts the development of acquired immunity, making treated individuals particularly vulnerable to reinfection almost immediately after treatment.

In the last decade researchers have demonstrated that this immunity to chlamydia is likely to develop in several possible ways. Among women who returned to a clinic in Birmingham, Alabama for treatment following a positive chlamydia screening test, those who had spontaneously cleared their infection before receiving treatment were 4 times less likely to get reinfected in the following 1-12 months than women who had persisting infections and received treatment (Geisler, Lensing, Press, & Hook, 2013). Additionally, a 2010 review article acknowledged that in several studies of infection status among couples, the rates of discordance (i.e. one partner is infected while the other is not), are higher for chlamydia than for gonorrhea and that this discordance increases with age, providing indirect evidence for some level of protective immunity to chlamydia that increases with age, likely due to exposure over time. There is little immunity that develops to gonorrheal infection due to high levels of antigenic variation (Batteiger et al., 2010). Recent modeling using data from both the UK and United States has demonstrated that at least some im-

munity to chlamydia following natural clearance is necessary to generate observed patterns in incidence (Omori, Chemaitelly, Althaus, & Abu-Raddad, 2019). These questions are particularly relevant in the context of expedited partner therapy, where the goal is to interrupt transmission by treated individuals and their partners as quickly as possible. However, due to the arrested immunity of those treated quickly, if the timing of delivery and uptake of partners is not sufficient, the initially treated is likely at higher risk of reinfection than under the standard referral scenario. If sufficient numbers of partners are treated effectively and quickly and transmission throughout the network is greatly diminished, then EPT may be able to overcome the effects of this arrested immunity.

Hypothesis 2. Treatment increases the rate of reinfection among diagnosed individuals compared to those whose infections clear naturally.

To test this theory, we will compare the difference in reinfection rates among individuals whose infections clear naturally (and are temporarily immune to reinfection) versus those individuals who clear their infections following treatment and are immediately susceptible to new infections. We will also compare these rates between concurrency scenarios to evaluate the interacting role of multiple overlapping partnership on reinfection between these types of recovery.

3.3 Sexual Networks

The baseline networks (with concurrency) are similar in structure to the networks initially described in Chapter 1, with several exceptions (the demographic calibration of these networks was described at the end of that chapter). First, we now stratify both the nodecov and concurrent terms that influence relationship formation by age and the number of individuals who have multiple ongoing partnerships within the casual network into separate terms by sex in order to capture small variations in the age-wise distribution of relationship frequency over the life course. Second, we make the age mixing term asymmetric by sex to reflect that on average females form relationships with slightly older males. Third, we add several nodefactor terms for a range of ages to both the casual and marriage/cohabitation networks for ease of calibration. Fourth, we add "cross-network" terms to specify the rate at which those in a casual network

also form a cohabitation (and vice versa). Lastly, we add a network for relationships that only last for one time step. We call this the instantaneous network and it is meant to reflect the frequency of one-offs, or more colloquially, "one night stands" by age category. We do not at this time attempt to implement the suggestions surrounding relationship duration in Chapter 2, which would be an entire project on its own and outside the scope of this dissertation.

The second set of three networks is identical to this initial set with one major exception: no individual is allowed to be in more than one active relationship at a time. This means that there is no concurrency within the casual network and also no cross-network concurrency (i.e. a marriage/cohabitation partnership where one partner forms an additional casual relationship). We do this by manually assigning the formation terms that would normally govern these processes a coefficient of negative infinity, thereby prohibiting them. This set of networks will provide the counterfactual to our baseline scenario to test our theories about the role of concurrency in reinfection following treatment.

3.3.1 The Epidemic

The dynamic demographics (births, aging, and deaths) are handled almost identically to the simulations in chapter one, with two exceptions. First, the sexual debut process is no longer explicitly modeled. This exclusion is based on the issues with implementation as described in the first chapter of this dissertation, and also the observation that once the degree-by-age targets are met post-calibration, the rate at which individuals form partnerships in the dynamic network simulation mirrors the distribution of individuals every having had sexual intercourse as reported in the National Survey of Family Growth. Second, to facilitate faster and more consistent simulations, the population size is reduced to 20,000 and is maintained at this size by setting the number at births at each time step to the number of deaths in the previous step. These small adjustments do not meaningfully change the proportional distribution of age and sex from the initial networks in chapter one.

In addition to the demographic modules, we now also include modules that govern the number of sexual acts per time step, the probability of condom use, the disease transmission process, testing, treatment, recovery, and the expedited partner treatment intervention. I will discuss each of these modules briefly,

but they are based on previously published work using an MSM population and publicly available code (Weiss et al., 2019).

At the beginning of each time step, those who turned 45 at the previous time step and any other deaths based on the age-category specific mortality rates are removed from the population. All existing nodes age by one week, and new nodes are added to the population to counter the deaths. Relationships then form and dissolve based on model coefficients. Then, all infected nodes who were treated in the previous time step recover, and asymptomatically infected nodes are probabilistically evaluated for spontaneous recovery. Nodes that spontaneously recover begin a period of immunity to subsequent infection, and all nodes that are currently immune are probabilistically evaluated to return to susceptible status. Next, we model transmission of chlamydia across existing partnerships. The number of sexual acts per week and the probability of condom use per week is based on the combined age of each partner and the type of relationship (i.e. marriage/cohabitation or casual). Once the number of acts with and without condoms has been determined, we focus on those relationship pairs who have a discordant infection status, where one member is infected with chlamydia and is eligible to transmit, and their partner is susceptible. Then each sexual act is evaluated for transmission, based on the probability of transmission per act and the protective effect of condom use.

Then we focus on testing and treatment of infections that are prevalent at the beginning of this time step (new infections are not eligible for testing or treatment). If any male or female has an infection that is symptomatic, the individual will seek testing within a month of infection. Females additionally screen for asymptomatic infections with probability according to their age category based on reports in NSFG. Diagnosed males and females are offered treatment for their current partners, and if they accept, their partners are then offered treatment at the following time step, and they accept, are flagged for treatment. In these scenarios, if a diagnosed individual accepts treatment for a current partner, we assume that partner also accepts the treatment. If any person is diagnosed, or flagged for treatment via EPT, they receive treatment and recover at the next time step (there is a present but low probability of treatment failure). Nobody who is flagged for treatment is allowed to have sex in the time step before their recovery. Finally, statistics about prevalence, incidence, and reinfection are generated before we move on to the next time

step.

3.3.2 Scenarios and Epidemic Calibration

Both sets of networks (concurrency and no concurrency) are modeled under five partner treatment scenarios: no partner treatment, and with 25%, 50%, 75% and 100% of diagnosed egos accepting expedited partner therapy for their current partners. Each of these ten sets are run for 80 years with five repeat simulations within each set. The first 40 years are used as a burn-in period to allow the system to reach its dynamic equilibrium. The results below reflect data from the last 40 years of the simulation sets for each scenario. Mean statistics represent the average within all simulations, and the standard errors represent the variation between the five simulations. We calibrate the transmission parameters in the model to match observed CT prevalence in our baseline model under 25% EPT (Torrone, Papp, Weinstock, & Centers for Disease Control and Prevention (CDC), 2014). Two types of parameters were chosen to be included in this calibration based on different types of uncertainty in the starting values. We included the probability of transmission per condomless sexual act and the duration of asymptomatic, undiagnosed infection, two commonly used calibration parameters due to uncertainty in the natural history of chlamydia (Davies et al., 2014). We also used parameters that modify the average number of sexual acts per week, the probability of condom use, and the rate of testing among women based on uncertainty in the reliability of reporting (desirability bias) and variation of reports between different sources (Broad et al., 2013; Khosropour et al., 2014).

3.4 Results

3.4.1 Prevalence and Incidence

We first begin with a high-level perspective on how concurrency and partner treatment influence population prevalence. We present average prevalence between males and females because they are so similar to each to each other across all scenarios (this is not true for incidence, as we shall discuss below). In the absence of partner therapy, when only those symptomatic individual and women who get diagnosed

Table 3.1: Overall Prevalence by Behavior and Partner Treatment Scenario

Treatment	M.F.Concurrency	No.Concurrency
None	4.3% (4.2 - 4.4)	2.9% (2.8 - 3)
25%	3.4% (3.3 - 3.5)	2.4% (2.3 - 2.6)
50%	2.7% (2.6 - 2.9)	2.1% (2 - 2.2)
75%	2.4% (2.3 - 2.6)	1.9% (1.9 - 2)
100%	2.1% (2 - 2.2)	1.7% (1.6 - 1.8)

via screening receive treatment, concurrency increases prevalence in the population by 1.4%. Increasing levels of partner treatment for partners of diagnosed individuals decreases population prevalence in the concurrency and no concurrency networks, by 51% and 41%, respectively. This is by no means the first study to show that partner treatment interventions can reduce prevalence in a population (Golden et al., 2015; Kretzschmar, Satterwhite, Leichliter, & Berman, 2012). Figure 3.2 displays incidence by age category and sex across all partner treatment scenarios and behavioral networks. Much like how concurrency increases the rate of prevalence in the population, we see that concurrency also drives incidence, particularly in among the 25-35 year range, where concurrency is high. Concurrency also seems to have a greater effect on incidence among females than on males, which makes sense given that males are more likely to form relationships with multiple females, creating additional opportunities for exposure and transmission to these females. It is notable that the absolute reductions in prevalence and incidence are higher in the concurrency network. A small proportion of people having multiple partnerships increases prevalence relative to a network with exclusively monogamous couplings, but this behavior also gives us more access to partners via partner treatment interventions and having high rates of partner treatment goes a long way towards mitigating the effect of concurrency.

3.4.2 Concurrency & Reinfection Among Treatment

Table 3.2 shows the proportion of males and females who became infected with a new infection following recovery due to treatment in the previous three months. The rate of reinfection at three months post-recovery for those individuals who got diagnosed and received treatment is exceptionally high at

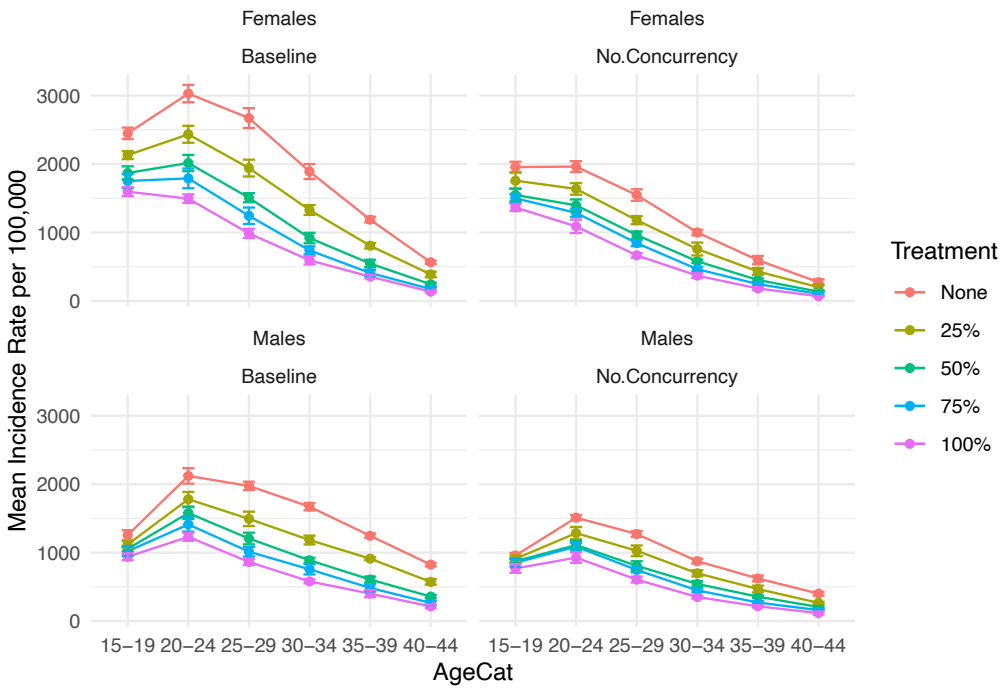

Figure 3.2: Incidence per 100,000 By Sex and Scenario

low levels of partner therapy between both the concurrency and no-concurrency populations. Concurrency increases the rate of reinfection among those people who get diagnosed by 2-3% among females and by 4-5% among males at low levels of partner treatment but again, the effect of concurrency seems to be greatly mitigated at high levels of partner treatment. Interestingly, at low levels of partner therapy males and females have similar rates of reinfection, but at the highest levels of partner therapy, females become reinfected at much higher rates than males do. This is likely a reflection of testing behavior among females, where the younger age categories have higher rates of testing, so diagnosed cases are more likely to come from higher incidence groups. Males only get diagnosed if they are symptomatic, and their incidence rates are not as variable across the life course as the rates among females, so their diagnosed cases are at much lower risk of reinfection. Table 3.3 shows the proportion of reinfected individuals at one year following treatment and recovery. The results are similar to three months but larger in magnitude, likely showing the role of new partnerships in additional reinfections in the longer term.

Table 3.2: Rate of Reinfection at Three Months Post-Treatment and Recovery

Treatment	Baseline	No.Concurrency	Difference
Females			
None	39% (38.2 - 39.7)	36.2% (35.7 - 36.6)	2.79%
25%	31.9% (30.7 - 33)	30.5% (29.1 - 31.9)	1.38%
50%	26.2% (25.4 - 27)	23.9% (23.4 - 24.5)	2.29%
75%	20% (19.3 - 20.7)	18.8% (17.8 - 19.8)	1.24%
100%	15.8% (15.1 - 16.4)	14.6% (13.3 - 15.9)	1.16%
Males			
None	41.9% (40.5 - 43.3)	37.2% (34.8 - 39.7)	4.64%
25%	33.9% (32.7 - 35.1)	30.3% (28.4 - 32.1)	3.66%
50%	22.7% (20.9 - 24.6)	18.9% (18 - 19.8)	3.84%
75%	13.6% (12.3 - 14.9)	12.8% (10.4 - 15.2)	0.78%
100%	5.1% (3.8 - 6.4)	3.7% (2.4 - 5)	1.39%

Table 3.3: Rate of Reinfection at 1 Year Post-Treatment and Recovery

Treatment	Baseline	No.Concurrency	Difference
Females			
None	59.1% (58.1 - 60.1)	54.7% (53.6 - 55.9)	4.42%
25%	48.8% (47.2 - 50.4)	45.2% (44 - 46.5)	3.59%
50%	40.2% (39.5 - 41)	35.9% (34.7 - 37.1)	4.32%
75%	31.9% (30.1 - 33.8)	28.2% (27.2 - 29.1)	3.74%
100%	24.5% (23.8 - 25.1)	21.1% (20 - 22.3)	3.31%
Males			
None	67.1% (64.9 - 69.4)	61% (59.5 - 62.6)	6.09%
25%	53.6% (50.4 - 56.9)	48.4% (45.4 - 51.3)	5.28%
50%	38.7% (36.5 - 40.9)	34.3% (32.1 - 36.4)	4.46%
75%	23.3% (21.9 - 24.6)	22.8% (19.7 - 25.9)	0.47%
100%	10.7% (9.2 - 12.3)	6% (4 - 8.1)	4.71%

Table 3.4: Rate of Reinfection at Three Months Post-Natural Clearance

Treatment	Baseline	No.Concurrency	Difference
Females			
None	10.4% (10 - 10.7)	9.8% (9.6 - 10.1)	0.53%
25%	9.8% (9.5 - 10.2)	9.6% (9.1 - 10.2)	0.22%
50%	9.8% (9.6 - 10)	9.1% (8.8 - 9.4)	0.72%
75%	9.3% (8.9 - 9.7)	8.7% (8.1 - 9.2)	0.63%
100%	8.9% (8.6 - 9.1)	8.6% (8.3 - 8.9)	0.29%
Males			
None	9.3% (9 - 9.6)	8.6% (8.2 - 9)	0.67%
25%	9% (8.5 - 9.5)	8.1% (7.6 - 8.6)	0.93%
50%	8.7% (8.4 - 9)	7.5% (6.9 - 8.1)	1.21%
75%	8.4% (8.3 - 8.6)	7.3% (6.9 - 7.7)	1.13%
100%	8.1% (7.9 - 8.3)	6.4% (5.8 - 6.9)	1.75%

3.4.3 Reinfection Among Those Who Clear Infections Naturally

Tables 3.4 and 3.5 show the proportion of males and females who become infected with a new infection following recovery due to natural clearance at three months and at one year, respectively. We show these tables to demonstrate how partner treatment of the diagnosed can influence the rate of reinfection among the undiagnosed as well, albeit indirectly. Here we see the effect of the short-term immunity following a natural clearance. The rate of reinfection varies very little between behavior scenarios and partner treatment levels for both males and females within three months, although males in the no concurrency setting seem to benefit slightly more at high levels of partner treatment relative to their counterparts when concurrency is allowed. In the short term, the immunity to chlamydia mitigates the effect of concurrency but once the immunity wears off, concurrency increases the rate of reinfection by 2-3.5%.

3.4.4 Comparison of Reinfection Between Treatment and Natural Recovery

A high proportion of those who cleared an infection naturally became re-infected within one year, but overall those who clear their infections naturally still have lower rates of reinfection than do those in-

Table 3.5: Rate of Reinfection at 1 Year Post-Natural Clearance

Treatment	Baseline	No.Concurrency	Difference
Females			
None	37.7% (37.2 - 38.3)	35.4% (35 - 35.7)	2.37%
25%	36.1% (35.7 - 36.5)	33.6% (31.7 - 35.6)	2.48%
50%	33.7% (33.2 - 34.3)	31.8% (31.4 - 32.3)	1.92%
75%	32.4% (31.9 - 32.9)	30.5% (29.6 - 31.4)	1.9%
100%	30.6% (30.4 - 30.8)	28.5% (27.2 - 29.8)	2.1%
Males			
None	34.5% (33.8 - 35.2)	31.4% (30.6 - 32.3)	3.03%
25%	32.1% (31.2 - 33)	29.2% (28.6 - 29.8)	2.9%
50%	30.2% (29.9 - 30.5)	26.6% (25.8 - 27.4)	3.59%
75%	29% (28.5 - 29.5)	25.6% (25.1 - 26.2)	3.35%
100%	27.3% (26.7 - 27.9)	23.5% (22.6 - 24.3)	3.85%

dividuals who received treatment. Table 3.6 highlights the difference in reinfection rates between the treatment group and natural recovery groups at three months following clearance. Even at 100% treatment of current partners, females still experience a 6-7% higher probability of reinfection relative to those who cleared their infections naturally. Males however, actually experience a lower probability of reinfection relative to their natural clearance counterparts in the first three months post-treatment. At one year post-clearance, the two groups experience rather different rates of reinfection. Reinfection decreases by 7% in both behavioral scenarios from no treatment to 100% partner treatment in the natural clearance group, suggesting that the reduction in overall prevalence due to partner treatment reduces reinfection from new partnerships. This effect is far more dramatic among the group who gets diagnosed.

3.5 Discussion

We find that while concurrency does seem to increase the proportion of individuals who become reinfection with chlamydia relative to our counterfactual population with no overlapping partnerships, the effect of concurrency is small, and can be almost entirely mitigated at high levels of treatment of current

Table 3.6: Rate of Reinfection at Three Months Post-Treatment and Recovery

	Females			Males		
Treatment	Following Treatment	Following Natural Recovery	Difference	Following Treatment	Following Natural Recovery	Difference
Baseline						
None	39% (38.2 - 39.7)	10.4% (10 - 10.7)	28.6%	41.9% (40.5 - 43.3)	9.3% (9 - 9.6)	32.6%
25%	31.9% (30.7 - 33)	9.8% (9.5 - 10.2)	22.1%	33.9% (32.7 - 35.1)	9% (8.5 - 9.5)	24.9%
50%	26.2% (25.4 - 27)	9.8% (9.6 - 10)	16.4%	22.7% (20.9 - 24.6)	8.7% (8.4 - 9)	14%
75%	20% (19.3 - 20.7)	9.3% (8.9 - 9.7)	10.7%	13.6% (12.3 - 14.9)	8.4% (8.3 - 8.6)	5.2%
100%	15.8% (15.1 - 16.4)	8.9% (8.6 - 9.1)	6.9%	5.1% (3.8 - 6.4)	8.1% (7.9 - 8.3)	-3%
No Concurrency						
None	36.2% (35.7 - 36.6)	9.8% (9.6 - 10.1)	26.4%	37.2% (34.8 - 39.7)	8.6% (8.2 - 9)	28.6%
25%	30.5% (29.1 - 31.9)	9.6% (9.1 - 10.2)	20.9%	30.3% (28.4 - 32.1)	8.1% (7.6 - 8.6)	22.2%
50%	23.9% (23.4 - 24.5)	9.1% (8.8 - 9.4)	14.8%	18.9% (18 - 19.8)	7.5% (6.9 - 8.1)	11.4%
75%	18.8% (17.8 - 19.8)	8.7% (8.1 - 9.2)	10.1%	12.8% (10.4 - 15.2)	7.3% (6.9 - 7.7)	5.5%
100%	14.6% (13.3 - 15.9)	8.6% (8.3 - 8.9)	6%	3.7% (2.4 - 5)	6.4% (5.8 - 6.9)	-2.7%

Table 3.7: Rate of Reinfection at One Year Post-Treatment and Recovery

	Females			Males		
Treatment	Following Treatment	Following Natural Recovery	Difference	Following Treatment	Following Natural Recovery	Difference
Baseline						
None	59.1% (58.1 - 60.1)	37.7% (37.2 - 38.3)	21.4%	67.1% (64.9 - 69.4)	34.5% (33.8 - 35.2)	32.6%
25%	48.8% (47.2 - 50.4)	36.1% (35.7 - 36.5)	12.7%	53.6% (50.4 - 56.9)	32.1% (31.2 - 33)	21.5%
50%	40.2% (39.5 - 41)	33.7% (33.2 - 34.3)	6.5%	38.7% (36.5 - 40.9)	30.2% (29.9 - 30.5)	8.5%
75%	31.9% (30.1 - 33.8)	32.4% (31.9 - 32.9)	-0.5%	23.3% (21.9 - 24.6)	29% (28.5 - 29.5)	-5.7%
100%	24.5% (23.8 - 25.1)	30.6% (30.4 - 30.8)	-6.1%	10.7% (9.2 - 12.3)	27.3% (26.7 - 27.9)	-16.6%
No Concurrency						
None	54.7% (53.6 - 55.9)	35.4% (35 - 35.7)	19.3%	61% (59.5 - 62.6)	31.4% (30.6 - 32.3)	29.6%
25%	45.2% (44 - 46.5)	33.6% (31.7 - 35.6)	11.6%	48.4% (45.4 - 51.3)	29.2% (28.6 - 29.8)	19.2%
50%	35.9% (34.7 - 37.1)	31.8% (31.4 - 32.3)	4.1%	34.3% (32.1 - 36.4)	26.6% (25.8 - 27.4)	7.7%
75%	28.2% (27.2 - 29.1)	30.5% (29.6 - 31.4)	-2.3%	22.8% (19.7 - 25.9)	25.6% (25.1 - 26.2)	-2.8%
100%	21.1% (20 - 22.3)	28.5% (27.2 - 29.8)	-7.4%	6% (4 - 8.1)	23.5% (22.6 - 24.3)	-17.5%

partners. Additionally, when all current partners of diagnosed individuals are treated, males experience lower rates of reinfection than their undiagnosed counterparts in the short term, suggesting that this extreme scenario can also mitigate the loss of short-term protective immunity following natural clearance. This finding is underscored by the rates of reinfection at one year, where increasing treatment reduces the relative proportion of reinfection so much that both males and females who receive treatment for their partners in the long run actually have lower rates of reinfection than those whose infections cleared naturally.

A 2009 meta-analysis of published chlamydia reinfection rates estimated that the median reinfection rate among females across studies was 13.9% with a range from 0-30% (Hosenfeld et al., 2009). There was a wide range of study types and length of follow-up window, but in general the rates are somewhat lower than we observe in these simulations. There are a few possible explanations for this discrepancy. First, while we included the development of immunity following spontaneous resolution of asymptomatic infections in our model, we did not additionally include immunity following multiple exposures to chlamydia. This could inflate the rate of reinfection if partners are trading asymptomatic infections back and forth, or for people who have multiple exposures over the life course. Second, we also assumed that no immunity to chlamydia developed if the patient was treated with antibiotics, regardless of the duration of infection prior to diagnosis. If some partial immunity did occasionally develop, our reinfection rates could be higher than what we would see in reality. Finally, many of the studies included in the meta-analysis are several decades old, and likely used diagnostic tests that have since been replaced with more sensitive tools. This could have underestimated the number of positive test and rate of reinfection in the studies. Nevertheless, it would be worthwhile in the future to conduct a sensitivity analysis of the immunity assumptions in the model to understand their effect on reinfection.

However, even if these simulations have somewhat higher reinfection rates due to the omission of these forms of immunity, it is still concerning that the rate of reinfection at low and medium levels of partner treatment among those who are treated is so high. The mechanisms of partner treatment in the real world are of course more varied than the narrow scope of this study, but this work underscores both the need to re-test individuals who have previously been diagnosed and to reach a broader population

for periodic screening of both females and males (Peterman et al., 2006). EPT certainly has the capability under extreme scenarios to dramatically reduce the population prevalence, but for those individuals who get diagnosed with chlamydia, their chances of reinfection are still high, which could lead to pelvic inflammatory disease and infertility.

Finally, the structure of the baseline model could have underestimated the effect of concurrency. We built the counterfactual by assuming the same number of relationships across the network by age, but prohibited any overlapping partnerships. This means that while the mean degree is preserved between behavior scenarios, the prohibition on concurrency actually increases the total number of egos in relationships at any given point in time. We could have instead altered the NSFG dataset to only include the first reported ongoing relationship of any ego, which would have reduced the overall mean degree of the network. It is unknown what effect this choice would have had on the results, but it is possible that this strategy underestimated the effect of concurrency in this study. Figure 3.3 displays the mean cross-sectional proportion of all individuals who ever become infected with chlamydia. Despite having the same mean degree, the concurrency network drives infection through a greater proportion of individuals, and lowering the mean degree in the no concurrency network would likely have increased this difference.

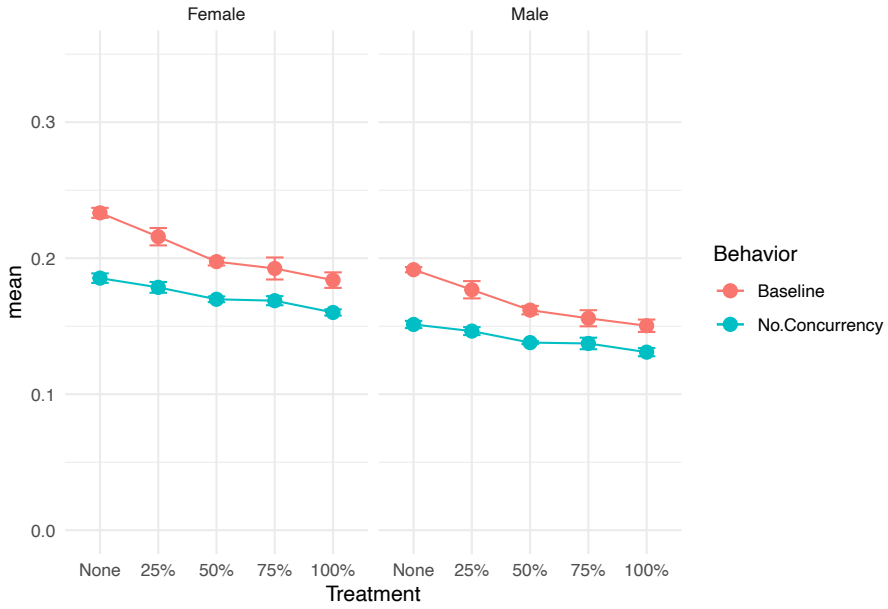

Figure 3.3: Proportion of all individuals who ever get infected

Conclusion

In this dissertation we explored many issues relating to demography, dynamic network modeling, and how behavior and biomedical interventions interact to influence our health.

Chapter 1 highlighted the need to run network diagnostics on open populations prior to epidemic simulation, particularly when terms in the formation and/or dissolution model reflect patterns in relationship prevalence that vary over the life course and may be influenced by the the boundaries of the modeled population. These boundary issues, in particular non-partnered entry and partnered dissolution, can generate equilibrium network statistics far from the expected statistics, even when these metrics are met during traditional network diagnostics using a closed population. However, we also show that several simple adjustments can be made to counter these effects and allow our networks to reproduce desired metrics. It may be worthwhile in the future to develop more standardized methods for network diagnostics in an open population (similar to the current "netdx" functionality), and for calibration using common STERGM terms.

Chapter 2 demonstrated that under many circumstances, the assumption of a constant hazard among relationships lengths is not a reasonable approximation of empirical data. However, among heterosexuals in the National Survey of Family Growth, we show this exponential assumption *is* reasonable when we use stricter definition of casual relationships and stratify longer relationships into cohabitations and marriages. We have a number of tasks moving forward. First, we must develop a method that allows us to use these new relationship categories more effectively. As mentioned previously, the current framework

of multiple and potentially overlapping relationship-specific networks is somewhat awkward because we must assume that we know a relationship's final state when it begins. It would make more sense for relationships to transition type over time, with corresponding dissolution probabilities by type in order to induce a non-constant hazard over the course of each relationship. While the necessary software to implement different dissolution rates by dyad-type (valued TERGMs) is not currently available, it may be possible to handle the transition and labeling of relationships within the EpiModel API as a work-around. It may also be possible to specify different dissolution probabilities by relationship length (edge age), which may be able to capture much of the variation in dissolution hazards over the course of a relationship without needing to specify the type of relationship represented by each edge. Second, it would also behoove us to compare epidemics between models with this new structure against those using the current framework in order to evaluate the effect of better capturing the distribution of relationship duration on epidemic outcomes.

In Chapter 3, we explored the role of concurrency and acquired immunity on chlamydia reinfection. We found that while concurrency does drive some reinfections, the overall rate of reinfection following treatment is concerningly high relative to those infected individuals who were able to clear their infections naturally. These effects were particularly striking at low levels of Expedited Partner Therapy (EPT). However, receiving treatment when EPT had high rates of uptake was protective against reinfection relative to those who did not get diagnosed. Given that the availability and usage of EPT varies widely across the United States, it may be worth exploring some of the threshold effects in greater detail to understand the return-on-investment of increasing resources for EPT programs.

In simulation modeling, there is often a trade-off between model complexity and clarity of results. However, much of this work demonstrates that occasionally only small increases in complexity are needed in order to greatly improve what are models are capable of representing. In order to capture the relationship dynamics of a certain age group within our model, we found that it was very necessary to also consider the behavior outside of that group. Approximating relationship durations with broad categories meant that the simulated distributions were far from the empirical data, but small changes to these relationship definitions resulted in remarkable fit improvements. There is still much work do be done to understand

the dynamics of immunity and partner therapy on reinfection, but even in a relatively simple model epidemic model with the addition of acquired immunity, we were able to demonstrate that interrupting the immunity process by treating infections without also treating a patient's local network can actually put them at greater risk of reinfection. As highly sophisticated modeling tools become more useable, the types of scientific questions were are able to ask will also increase in detail and specificity. As such it will be more important than ever to focus on *which* processes are important and *how* they are important to epidemic transmission - the simplest solutions are often the best, but the devil is always in the details.

A

Appendix

A.1 Survival Analysis

Figure A.1 shows the histograms of relationship length in the NSFG broken out by relationship type and censored status.

A.2 Chlamydia & EPT

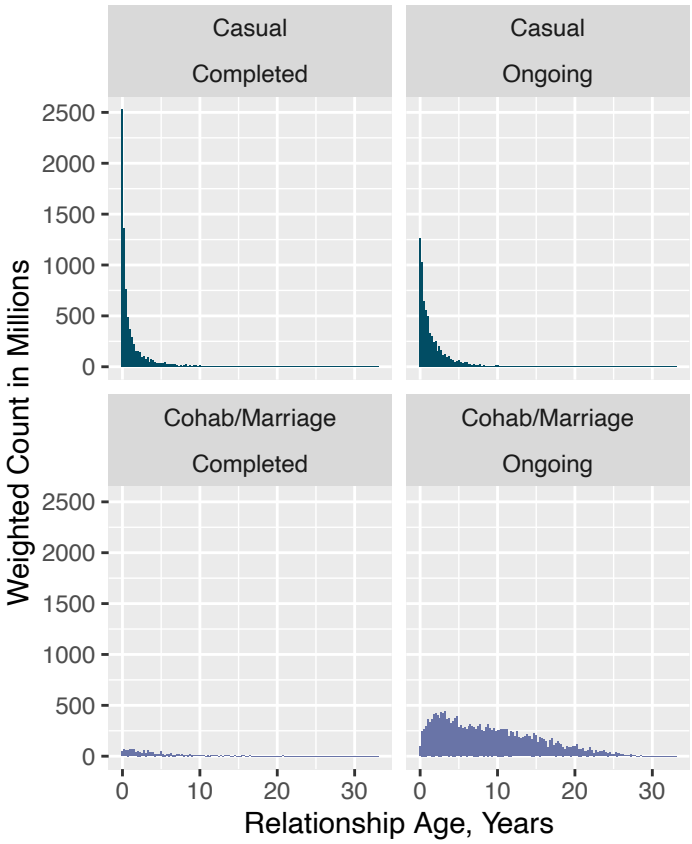

Figure A.1: Histograms of Relationship Duration by Censored Status & Type

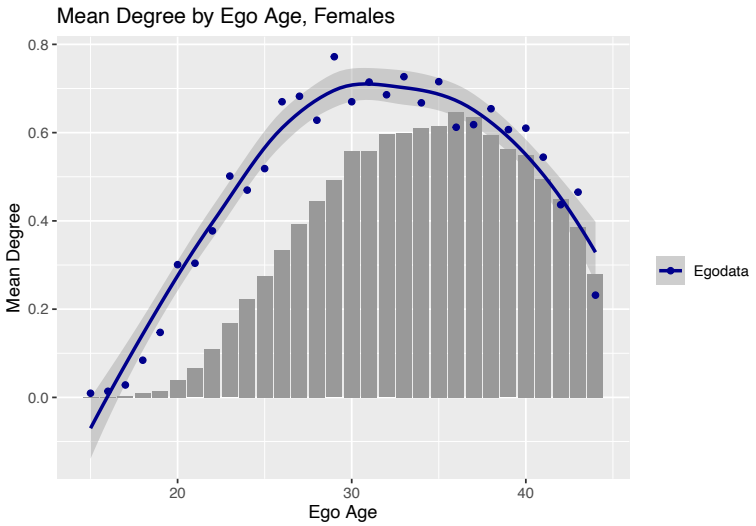

Figure A.2: Mean Degree by Age in Uncalibrated Simulations - Marriage/Cohabitation Network

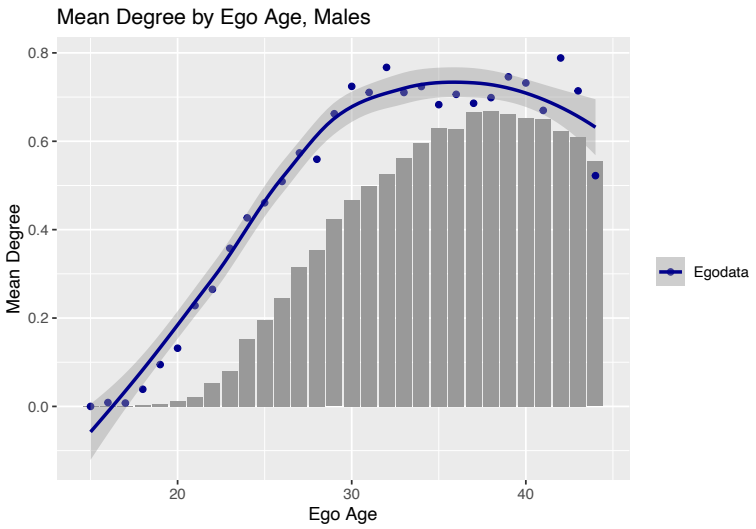

Figure A.3: Mean Degree by Age in Uncalibrated Simulations - Marriage/Cohabitation Network

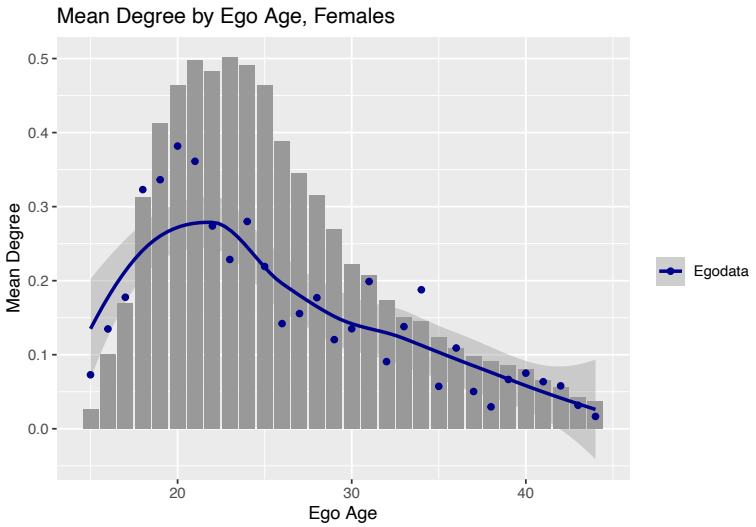

Figure A.4: Mean Degree by Age in Uncalibrated Simulations - Casual Network

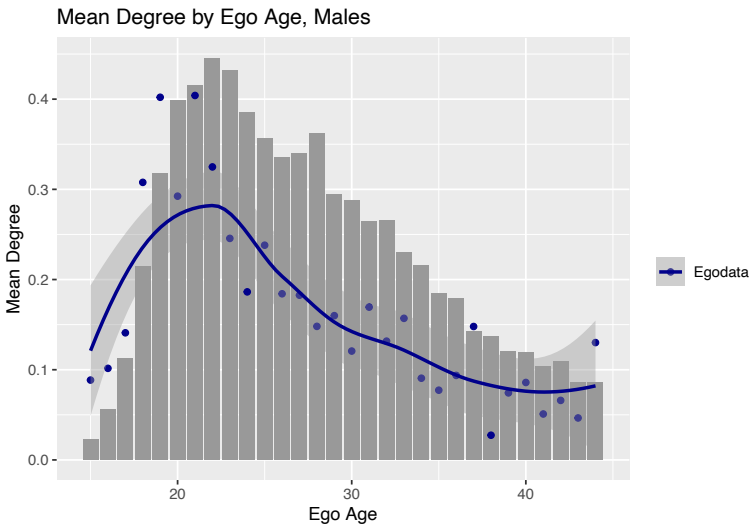

Figure A.5: Mean Degree by Age in Uncalibrated Simulations - Casual Network

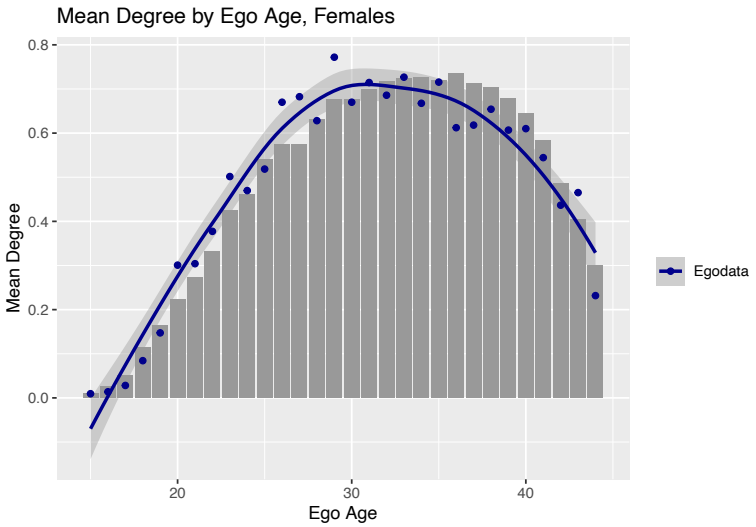

Figure A.6: Mean Degree by Age in Calibrated Simulations - Marriage/Cohabitation Network

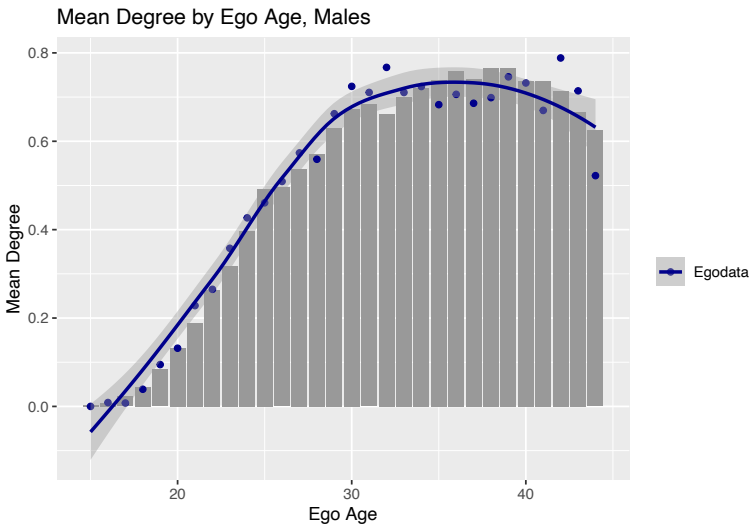

Figure A.7: Mean Degree by Age in Calibrated Simulations - Marriage/Cohabitation Network

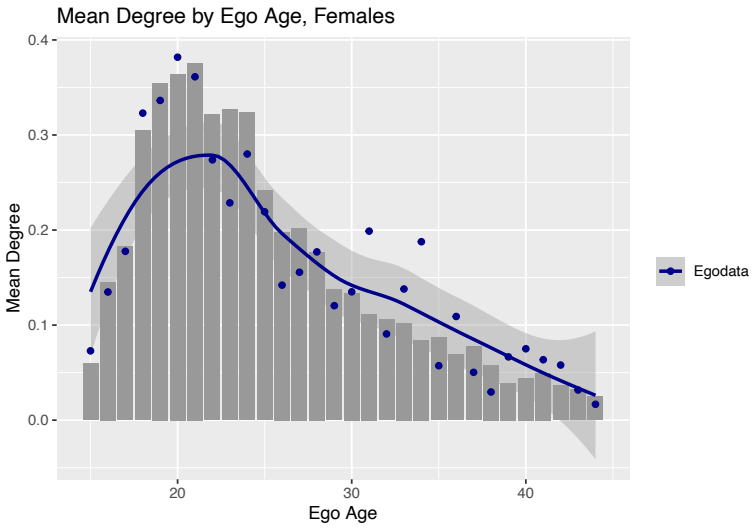

Figure A.8: Mean Degree by Age in Calibrated Simulations - Casual Network

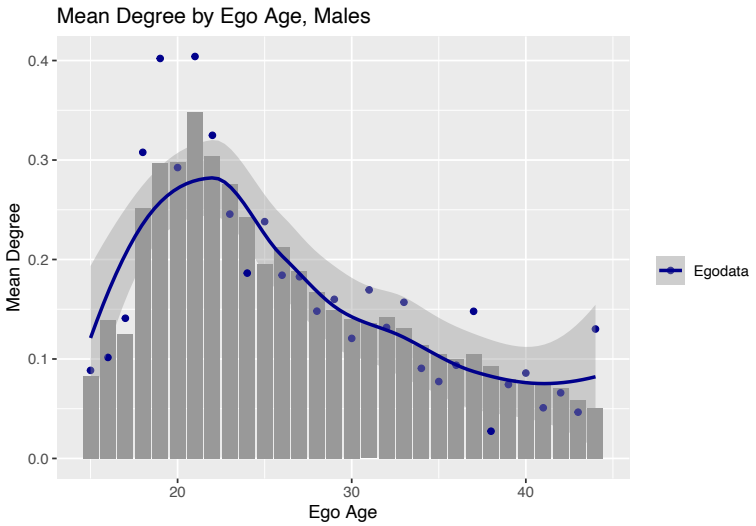

Figure A.9: Mean Degree by Age in Calibrated Simulations - Casual Network

Table A.1: Marriage/Cohabitation Network: Summary of Formation Model Fit

Model Term	Estimate	SE	Z Value	Pr(>|z|)
offset(netsize.adj)	-9.9034876	NA	NA	NA
edges	-21.5043377	5.8325350	-3.686962	0.0002269
nodecov.ageF	0.7701599	0.1810178	4.254608	0.0000209
nodecov.ageF^2	-0.0109692	0.0026536	-4.133714	0.0000357
nodecov.ageM	0.7275492	0.1873103	3.884192	0.0001027
nodecov.ageM^2	-0.0092757	0.0027760	-3.341427	0.0008335
nodefactor.agecat.1	2.0476375	0.7599847	2.694314	0.0070534
nodefactor.agecat.2	1.6343049	0.5162024	3.166016	0.0015454
nodefactor.agecat.3	0.4114626	0.4011854	1.025617	0.3050721
nodefactor.age>30.TRUE	-0.6151742	0.4342964	-1.416485	0.1566336
absdiff.sqrtage+0.1511228*(male==0)	-3.5461269	0.1237456	-28.656582	0.0000000
nodefactor.deg.other.binary.1	-4.6938120	0.2285503	-20.537328	0.0000000
offset(nodematch.male)	-Inf	NA	NA	NA
offset(nodefactor.olderpartnerMC.1)	-Inf	NA	NA	NA
offset(concurrent)	-Inf	NA	NA	NA

Table A.2: Casual Network: Summary of Formation Model Fit

Model Term	Estimate	SE	Z Value	Pr(>|z|)
offset(netsize.adj)	-9.9034876	NA	NA	NA
edges	0.8189628	3.5668123	0.2296064	0.8183977
nodecov.ageF	-0.1701293	0.2036730	-0.8353058	0.4035456
nodecov.ageF^2	0.0013856	0.0029842	0.4643166	0.6424209
nodecov.ageM	0.3213165	0.1373386	2.3395943	0.0193047
nodecov.ageM^2	-0.0038923	0.0020529	-1.8959858	0.0579619
absdiff.sqrtage+0.14505*(male==0)	-3.2824563	0.1512915	-21.6962358	0.0000000
nodefactor.deg.marcoh.1	-4.4311755	0.2050212	-21.6132504	0.0000000
nodefactor.olderpartnerMC.1	-3.4033511	0.4357694	-7.8099819	0.0000000
nodefactor.ageM.15	2.1559900	0.7280975	2.9611279	0.0030651
nodefactor.ageM.16	0.6155908	0.6116980	1.0063640	0.3142405
nodefactor.ageM.17	0.7564320	0.5196495	1.4556581	0.1454872
nodefactor.ageM.18	1.3939751	0.4435948	3.1424515	0.0016754
nodefactor.ageM.19	0.9902745	0.3852571	2.5704252	0.0101574
nodefactor.ageM.20	0.7109074	0.3673281	1.9353472	0.0529477
nodefactor.ageM.21	0.7143414	0.3236543	2.2071122	0.0273062
nodefactor.ageM.22	0.4639386	0.3020156	1.5361409	0.1245038
nodefactor.ageM.23	0.2019091	0.2472843	0.8165057	0.4142110
nodefactor.ageF.15	-2.9801110	1.0131202	-2.9415178	0.0032661
nodefactor.ageF.16	-2.4691210	0.8707324	-2.8356830	0.0045728
nodefactor.ageF.17	-2.2888585	0.7573093	-3.0223562	0.0025082
nodefactor.ageF.18	-1.1576376	0.6584646	-1.7580863	0.0787328
nodefactor.ageF.19	-1.2780281	0.5820540	-2.1957208	0.0281119
nodefactor.ageF.20	-0.6023302	0.5083085	-1.1849698	0.2360294
nodefactor.ageF.21	-0.5216052	0.4107761	-1.2698040	0.2041544
nodefactor.ageF.22	-0.7205107	0.3593705	-2.0049244	0.0449711
nodefactor.ageF.23	0.2238072	0.3554722	0.6296055	0.5289527
concurrent.male0	-2.9836341	0.4020156	-7.4216881	0.0000000
concurrent.male1	-1.5323965	0.2483277	-6.1708642	0.0000000
offset(nodematch.male)	-Inf	NA	NA	NA

Table A.3: Instantaneous: Summary of Formation Model Fit

Model Term	Estimate	SE	Z Value	Pr(>|z|)
offset(netsize.adj)	-9.9034876	NA	NA	NA
edges	-0.7489813	0.5032150	-1.4883925	0.1366474
nodefactor.agecat.1	0.4376583	0.2742326	1.5959383	0.1105026
nodefactor.agecat.2	0.7037547	0.2930926	2.4011345	0.0163443
nodefactor.agecat.3	0.6293726	0.2978318	2.1131816	0.0345852
nodefactor.agecat.4	0.6011840	0.3515956	1.7098737	0.0872892
nodefactor.agecat.5	-0.0696064	0.3458209	-0.2012788	0.8404806
absdiff.sqrtage	-3.0456947	0.2702213	-11.2711138	0.0000000
nodefactor.deg.marcoh.1	-2.5453253	0.3249382	-7.8332585	0.0000000
nodefactor.deg.other.binary.1	-0.0445363	0.1819218	-0.2448103	0.8066033
offset(nodematch.male)	-Inf	NA	NA	NA

References

10 Armbruster, B., Wang, L., & Morris, M. (2017). Forward reachable sets: Analytically derived properties of connected components for dynamic networks. *Netw. Sci.*, *5*(3), 328–354. http://doi.org/10.1017/nws.2017.10

Batteiger, B. E., Tu, W., Ofner, S., Van Der Pol, B., Stothard, D. R., Orr, D. P., ... Fortenberry, J. D. (2010). Repeated Chlamydia trachomatis genital infections in adolescent women. *J. Infect. Dis.*, *201*(1), 42–51. http://doi.org/10.1086/648734

Broad, J. M., Manhart, L. E., Kerani, R. P., Scholes, D., Hughes, J. P., & Golden, M. R. (2013). Chlamydia screening coverage estimates derived using healthcare effectiveness data and information system procedures and indirect estimation vary substantially. *Sex. Transm. Dis.*, *40*(4), 292–297. http://doi.org/10.1097/OLQ.0b013e3182809776

Brunham, R. C., & Rekart, M. L. (2008, January). The arrested immunity hypothesis and the epidemiology of chlamydia control. http://doi.org/10.1097/OLQ.0b013e31815e41a3

Brunham, R. C., & Rey-Ladino, J. (2005, February). Immunology of Chlamydia infection: Implications for a Chlamydia trachomatis vaccine. Nature Publishing Group. http://doi.org/10.1038/nri1551

Burington, B., Hughes, J. P., Whittington, W. L. H., Stoner, B., Garnett, G., Aral, S. O., & Holmes, K. K. (2010). Estimating duration in partnership studies: Issues, methods and examples. *Sex. Transm. Infect.*, *86*(2), 84–89. http://doi.org/10.1136/sti.2009.037960

Carnegie, N. B., Krivitsky, P. N., Hunter, D. R., & Goodreau, S. M. (2015). An Approximation Method for Improving Dynamic Network Model Fitting. *J. Comput. Graph. Stat.*, *24*(2), 502–519. http://doi.org/10.1080/10618600.2014.903087

Cavazos-Rehg, P. A., Krauss, M. J., Spitznagel, E. L., Schootman, M., Bucholz, K. K., Peipert, J. F., … Bierut, L. J. (2009). Age of sexual debut among US adolescents. *Contraception*, *80*(2), 158–162. http://doi.org/10.1016/j.contraception.2009.02.014

Cox, T. (2014). Testing the Equivalence of Survival Distributions using PP- and PPP-Plots. *Int. J. Stat. Med. Res.*, *3*(2), 161–173. http://doi.org/10.6000/1929-6029.2014.03.02.10

Davies, B., Anderson, S. J., Turner, K. M. E., & Ward, H. (2014, January). How robust are the natural history parameters used in chlamydia transmission dynamic models? A systematic review. BioMed Central. http://doi.org/10.1186/1742-4682-11-8

Fine, D., Dicker, L., Mosure, D., & Berman, S. (2008). Increasing chlamydia positivity in women screened in family planning clinics: Do we know why? *Sex. Transm. Dis.*, *35*(1), 47–52. http://doi.org/10.1097/OLQ.0b013e31813e0c26

Geisler, W. M. (2010).Duration of Untreated, Uncomplicated Chlamydia trachomatis Genital Infection and Factors Associated with Chlamydia Resolution: A Review of Human Studies . *J. Infect. Dis.*, *201*(S2), 104–113. http://doi.org/10.1086/652402

Geisler, W. M., Lensing, S. Y., Press, C. G., & Hook, E. W. (2013). Spontaneous resolution of genital chlamydia trachomatis infection in women and protection from reinfection. *J. Infect. Dis.*, *207*(12), 1850–1856. http://doi.org/10.1093/infdis/jit094

Gift, T. L., Kissinger, P., Mohammed, H., Leichliter, J. S., Hogben, M., & Golden, M. R. (2011). The cost and cost-effectiveness of expedited partner therapy compared with standard partner

referral for the treatment of chlamydia or gonorrhea. *Sex. Transm. Dis.*, *38*(11), 1067–1073. http://doi.org/10.1097/OLQ.0b013e31822e9192

Golden, M. R., Kerani, R. P., Stenger, M., Hughes, J. P., Aubin, M., Malinski, C., & Holmes, K. K. (2015). Uptake and Population-Level Impact of Expedited Partner Therapy (EPT) on Chlamydia trachomatis and Neisseria gonorrhoeae: The Washington State Community-Level Randomized Trial of EPT. *PLoS Med.*, *12*(1), 1–22. http://doi.org/10.1371/journal.pmed.1001777

Golden, M. R., Whittington, W. L. H., Handsfield, H. H., Hughes, J. P., Stamm, W. E., Hogben, M., ... Holmes, K. K. (2005). Effect of Expedited Treatment of Sex Partners on Recurrent or Persistent Gonorrhea or Chlamydial Infection. *N. Engl. J. Med.*, *352*(7), 676–685. http://doi.org/10.1056/NEJMoa041681

Goodreau, S. M., Rosenberg, E. S., Jenness, S. M., Luisi, N., Stansfield, S. E., Millett, G. A., & Sullivan, P. S. (2017). Sources of racial disparities in HIV prevalence in men who have sex with men in Atlanta, GA, USA: a modelling study. *Lancet HIV*, *4*(7), e311–e320. http://doi.org/10.1016/S2352-3018(17)30067-X

Hosenfeld, C. B., Workowski, K. A., Berman, S., Zaidi, A., Dyson, J., Mosure, D., ... Bauer, H. M. (2009). Repeat infection with chlamydia and gonorrhea among females: A systematic review of the literature. *Sex. Transm. Dis.*, *36*(8), 478–489. http://doi.org/10.1097/OLQ.0b013e3181a2a933

Jackson, C. H. (2016). Flexsurv: A platform for parametric survival modeling in R. *J. Stat. Softw.*, *70*(8), 1–33. http://doi.org/10.18637/jss.v070.i08

Jenness, S. M., Weiss, K. M., Goodreau, S. M., Gift, T., Chesson, H., Hoover, K. W., ... Rosenberg, E. S. (2017). Incidence of Gonorrhea and Chlamydia Following Human Immunodeficiency Virus Preexposure Prophylaxis Among Men Who Have Sex With Men: A Modeling Study. *Clin. Infect. Dis.*, *65*(5), 712–718. http://doi.org/10.1093/cid/cix439

Jolly, A. M., Muth, S. Q., Wylie, J. L., & Potterat, J. J. (2001). Sexual networks and sexually transmitted infections: A tale of two cities. *J. Urban Heal.*, *78*(3), 433–445. http://doi.org/10.1093/jurban/78.3.433

Kaestle, C. E., Morisky, D. E., & Wiley, D. J. (2002). Sexual Intercourse and the Age Difference between Adolescent Females and Their Romantic Partners. *Perspect. Sex. Reprod. Health*, *34*(6), 304. http://doi.org/10.2307/3097749

Kermack, W. o., & Mckendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London. Ser. A, Contain. Pap. A Math. Phys. Character*, *115*(772), 700–721. http://doi.org/10.1098/rspa.1927.0118

Kerns, J. L., Jones, H. E., Pressman, E. J., Fratarelli, L. A., Garth, J., & Westhoff, C. L. (2011). Implementation of expedited partner therapy among women with chlamydia infection at an urban family planning clinic. *Sex. Transm. Dis.*, *38*(8), 722–726. http://doi.org/10.1097/OLQ.0b013e318214bb83

Khosropour, C. M., Broad, J. M., Scholes, D., Saint-Johnson, J., Manhart, L. E., & Golden, M. R. (2014). Estimating chlamydia screening coverage: A comparison of self-report and health care effectiveness data and information set measures. *Sex. Transm. Dis.*, *41*(11), 665–670. http://doi.org/10.1097/OLQ.0000000000000186

Kong, F. Y. S., & Hocking, J. S. (2015, July). Treatment challenges for urogenital and anorectal Chlamydia trachomatis. BioMed Central. http://doi.org/10.1186/s12879-015-1030-9

Kretzschmar, M., Satterwhite, C., Leichliter, J., & Berman, S. (2012). Effects of screening and partner notification on chlamydia positivity in the united states: A modeling study. *Sex. Transm. Dis.*, *39*(5), 325–331. http://doi.org/10.1097/OLQ.0b013e31824e52c2

Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electron. J. Stat.*, *6*, 1100–1128. http://doi.org/10.1214/12-EJS696

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *76*(1), 29–46. http://doi.org/10.1111/rssb.12014

Krivitsky, P. N., Handcock, M. S., & Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Stat. Methodol.*, *8*(4), 319–339. http://doi.org/10.1016/j.stamet.2011.01.005

Krivitsky, P. N., & Morris, M. (2017). Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *Ann. Appl. Stat.*, *11*(1), 427–455. http://doi.org/10.1214/16-AOAS1010

Lara, L. A. S., & Abdo, C. H. N. (2016). Age at Time of Initial Sexual Intercourse and Health of Adolescent Girls. *J. Pediatr. Adolesc. Gynecol.*, *29*(5), 417–423. http://doi.org/10.1016/j.jpag.2015.11.012

Morris, M., & Kretzschmar, M. (1997). Concurrent partnerships and the spread of HIV. *AIDS*, *11*(5), 641–648. http://doi.org/10.1097/00002030-199705000-00012

Morris, M., Kurth, A. E., Hamilton, D. T., Moody, J., & Wakefield, S. (2009). Concurrent partnerships and HIV prevalence disparities by race: Linking science and public health practice. *Am. J. Public Health*, *99*(6), 1023–1031. http://doi.org/10.2105/AJPH.2008.147835

Niccolai, L. M., Rowhani-Rahbar, A., Jenkins, H., Green, S., & Dunne, D. W. (2005). Condom effectiveness for prevention of Chlamydia trochomotis infection. *Sex. Transm. Infect.*, *81*(4), 323–325. http://doi.org/10.1136/sti.2004.012799

Omori, R., Chemaitelly, H., Althaus, C. L., & Abu-Raddad, L. J. (2019). Does infection with Chlamydia trachomatis induce long-lasting partial immunity? Insights from mathematical modelling. *Sex. Transm. Infect.*, *95*(2), 115–121. http://doi.org/10.1136/sextrans-2018-053543

Peterman, T. A., Tian, L. H., Metcalf, C. A., Satterwhite, C. L., Malotte, C. K., DeAugustine, N., … Douglas, J. M. (2006). High incidence of new sexually transmitted infections in the year following a sexually transmitted infection: A case for rescreening. *Ann. Intern. Med.*, *145*(8), 564–572. http://doi.org/10.7326/0003-4819-145-8-200610170-00005

Satterwhite, C. L., Torrone, E., Meites, E., Dunne, E. F., Mahajan, R., Cheryl Bañez Ocfemia, M., … Weinstock, H. (2013). Sexually transmitted infections among US women and men: Prevalence and incidence estimates, 2008. *Sex. Transm. Dis.*, *40*(3), 187–193. http://doi.org/10.1097/OLQ.0b013e318286bb53

Shiely, F., Hayes, K., Thomas, K. K., Kerani, R. P., Hughes, J. P., Whittington, W. L. H., … Golden, M. R. (2010). Expedited partner therapy: A robust intervention. *Sex. Transm. Dis.*, *37*(10), 602–607. http://doi.org/10.1097/OLQ.0b013e3181e1a296

Taylor, M. M., Reilley, B., Yellowman, M., Anderson, L., Ravello, L. de, & Tulloch, S. (2013, May). Use of expedited partner therapy among chlamydia cases diagnosed at an urban Indian health centre, Arizona. http://doi.org/10.1177/0956462412472825

Torrone, E., Papp, J., Weinstock, H., & Centers for Disease Control and Prevention (CDC). (2014). Prevalence of Chlamydia trachomatis genital infection among persons aged 14-39 years–United States, 2007-2012. *MMWR. Morb. Mortal. Wkly. Rep.*, *63*(38), 834–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25254560%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4584673%20http://www.ncbi.nlm.nih.gov/pubmed/25254560%7B/%%7D0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4584673

Van Aar, F., Van Benthem, B. H. B., Van Den Broek, I. V. F., & Götz, H. M. (2018). STIs in sex partners notified for chlamydia exposure: Implications for expedited partner therapy. *Sex. Transm. Infect.*, *94*(8), 619–621. http://doi.org/10.1136/sextrans-2017-053364

Weiss, K. M., Goodreau, S. M., Morris, M., Prasad, P., Ramaraju, R., Sanchez, T., & Jenness, S. M. (2020). Egocentric sexual networks of men who have sex with men in the United States: Results from the ARTnet study. *Epidemics*, *30*, 100386. http://doi.org/10.1016/j.epidem.2020.100386

Weiss, K. M., Jones, J. S., Katz, D. A., Gift, T. L., Bernstein, K., Workowski, K., … Jenness, S. M. (2019). Epidemiological Impact of Expedited Partner Therapy for Men Who Have Sex with Men: A Modeling Study. *Sex. Transm. Dis.*, *46*(11), 697–705. http://doi.org/10.1097/OLQ.0000000000001058