Title
Hierarchical and semi-parametric Bayesian models for the study of longitudinal HIV behavior and tuberculosis incidence data

Author
Zhu, Yuda

Publication Date
2012

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Hierarchical and Semi-parametric Bayesian Models for the Study of Longitudinal HIV Behavior and Tuberculosis Incidence Data

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Biostatistics

by

Yuda Zhu

2012

Abstract of the Dissertation

Hierarchical and Semi-parametric Bayesian Models for the Study of Longitudinal HIV Behavior and Tuberculosis Incidence Data

by

Yuda Zhu

Doctor of Philosophy in Biostatistics

University of California, Los Angeles, 2012

Professor Robert E. Weiss, Chair

We propose and discuss two distinct and separate innovative Bayesian models. In the first model, we propose a replacement for standard statistical methodologies for longitudinal sexual behavior data. HIV intervention trials generally collect sexual behavior data repeatedly over time and involve multiple outcomes including the number of partners which are nested in subjects and the number of protected and unprotected sex acts with each partner which are inherently nested within partners. The data is further complicated by characteristics of both partners and acts. Partners can be HIV^+ or HIV^- while sex acts can be protected or unprotected. Properly modeling these outcomes and distinguishing these characteristics is critical. Here we use a multilevel multivariate Bayesian model for modeling sexual behavior outcomes. The proposed model accounts for the full complexity of sexual behavior allowing for simultaneous modeling of the number of partners and the number of sex acts with each partner, differentiation of behavior by partner serostatus, accounting for study eligibility criterions associated with the outcome of interest, and correlations between observations with the same subject, observations with the same partner, and observations across time. We further show that the proposed model can be used to quantify and draw inference on seroadaptive behaviors. Seroadaptive behaviors describe behaviors that vary based on the HIV status of partners with the goal of reducing the risk of transmission. The model is used to analyze data from the Healthy Living Project.

In the second half of this thesis, we explore a novel extension to the Dirichlet process mixture (DPM) model to accommodate longitudinal data. Longitudinal data is characterized by two features. First, the data are a function of time implying dependence between sampling densities across time. Second, the *same* subjects are repeatedly measured over time. The standard DPM model is a nonparametric Bayesian model that naturally clusters similar observations together and assigns a single value to each cluster. It can be used to model an unknown density but addresses neither of these two features in longitudinal data. A number of current extensions of the DPM model can accommodate dependent distributions which could be used to model the sampling distributions at each time point addressing the first feature. However, assumptions in these extensions imply these models do not take advantage of the second feature of longitudinal data where the *same* subjects are followed over time. To account for both features, we propose the cluster memory Dirichlet process mixture (cmDPM) model extending the DPM model to properly accommodate longitudinal data. In the cmDPM model, subjects are modeled as a DPM model at baseline. Cluster assignments at future time points depend on where the subject was previously clustered. Each subject may retain their cluster from the previous time point with some nonzero probability. This implies that at later times, subjects are no longer exchangeable and their observed values depend on their previous clustering history. Clusters that are retained over time evolve through a time dependent process. The cmDPM model extends the DPM to use both the information of where the subject was previously clustered and the value assigned to that cluster to model subject data at the current time point.

The dissertation of Yuda Zhu is approved.

William G. Cumberland

Donatello Telesca

Mary Jane Rotheram-Borum

Robert E. Weiss, Committee Chair

University of California, Los Angeles

2012

To my parents . . .

without whose sacrifices,

none of this would be possible.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Dr Robert E. Weiss, for his weekly counseling and advice. His encouragement and support gave me confidence to overcome the many stumbling blocks of research. His goading and insistence gave me the push to continue to search for the next challenge after reaching each goal. His patience in reading and correcting draft after draft of my papers, my proposal, and this thesis continues to amaze me. I would also like to thank my committee members: Dr Donatello Telesca, Dr William G. Cumberland, and Dr Mary Jane Rotheram-Borus for their helpful comments and suggestions. Dr Mary Jane Rotheram-Borus also provided and gave permission to use the HLP dataset without which this thesis would not have been possible. Finally, I would like to thank the NIH grant 5-T32-AI007370, Biostatistical Training in AIDS for financial support.

Vita

1978	Born, Beijing, China.
2000	B.S. (Bioengineering), University of California, Berkeley.
2000–2005	Engineer, Becton Dickinson Immunocytometry Systems.
2006-2008	M.S. (Biostatistics), University of California, Los Angeles.
2008–2012	Ph.D. (Biostatistics), University of California, Los Angeles.

Publications

Risa M. Hoffman, Beth D. Jamieson, Ronald J. Bosch, Judith Currier, Christina M. R. Kitchen, Ingrid Schmid, **Yuda Zhu**, Kara Bennett, and Ronald Mitsuyasu. (2011). Baseline Immune Phenotypes and $CD4^+$ T Lymphocyte Responses to Antiretroviral Therapy in Younger versus Older HIV-infected Individuals. *Journal of Clinical Immunology*, **31**(5): 873-881.

Sean D. Young and **Yuda Zhu**. (2011). Behavioral Evidence of HIV Testing Stigma. *AIDS and Behavior*, DOI: 10.1007/s10461-011-0018-8

CHAPTER 1

Introduction

1.1 Overview

In the behavioral and health sciences, there is often a need to model data with complex structures and patterns. In this dissertation, we present two distinct Bayesian models that address some of the specific problems faced when dealing with these kinds of data.

In Chapter 2, we examine data from the Healthy Living Project (HLP), an HIV behavioral intervention study conducted at four separate sites between 2000 and 2004. Standard statistical analysis for modeling sexual behavior and assessing HIV behavioral interventions use univariate longitudinal outcomes. These models fail to fully account for the complexity of sexual behavior data. We define a sexual behavior profile that is a multivariate multilevel summary of a participant's reported actions in the study. A hierarchical Bayesian longitudinal model is then proposed and we demonstrate the benefits of using this model. We utilize our proposed model to identify and quantify a specific form of risk management behavior in the HLP study known as *seroadaptation* where study participants change their behavior to reduce the risk of HIV transmission based on knowledge or perceived knowledge of their partner's HIV status.

In Chapter 3, we present the cluster memory Dirichlet process mixture (cmDPM) model, a novel non-parametric Bayesian model for longitudinal data. The cmDPM model clusters subjects at each time point based on both a Dirichlet process and where each subject was previously clustered at the previous time point. Longitudinal data is characterized by the *same* subjects observed repeatedly over time. My proposed model provides a flexible choice between the extremes of treating each time point as an independent Dirichlet process mixture

(DPM) model and treating the clustering structure at each time point as identical when modeling longitudinal data. We apply the cmDPM model to annual tuberculosis incidence rate data collected on 197 countries in the world from 1990-2010. Conclusions are drawn on the change in the distribution of tuberculosis incidence rates in the last 2 decades and individual predictions for 2011 are made.

The cmDPM model is also a valuable tool in many other applications of longitudinal data. One application is to address many potentially different and diverse groups of study subjects in longitudinal studies. This is often the case in large HIV behavioral studies like the HLP study that make use of multiple study sites. Participants recruited into these studies can also be from rather diverse backgrounds with varying behavior profiles. In particular, the HLP study recruited participants from four separate risk groups and four different races. In these situations, it may not be sufficient to simply include covariates for these differences. However, inclusion of each combination of location, risk group, and race at each time point would involve $4 \times 4 \times 4 \times 6 = 384$ covariates and would be excessive. The question of which combinations of participants behave similarly can be thought in statistical terms as a clustering problem for longitudinal data. In Chapter 4, we discuss future work that is required to enable this application of the model and additional future directions are also presented.

In the remainder of this chapter, we give a non-technical description of the multivariate multilevel longitudinal data we propose in Chapter 2 and the cmDPM model we present in Chapter 3. The non-technical descriptions outline the models presented and the benefits of using these models.

1.2 Non-technical description: A multilevel multivariate longitudinal model for sexual behavior data from the Healthy Living Project

Reduction in risky sexual behavior for HIV^+ individuals can result from reducing the number of HIV^-/unknown serostatus partners and/or from reducing the number of unprotected sex acts with each HIV^-/unknown partner. Subjects can also reduce their risk by being strategic in their behavioral choices. One example first studied in the population of men who have sex with men (MSM) is *seroadaptation*. We find that single measures of risky behavior fail to sufficiently describe the range and complexity of sexual behavior and instead define a sexual behavior profile that includes the number of HIV^+ partners, the number of HIV^-/unknown partners, the number of protected acts with each HIV^+ partner, the number of unprotected acts with each HIV^+ partner, the number of protected acts with each HIV^-/unknown partner, and the number of unprotected acts with each HIV^-/unknown partner. We propose a multilevel multivariate model that provides a more complete picture of sexual behavior and enables us to distinguish between reductions in numbers of partners and reductions in numbers of unprotected sex acts. With a more detailed model for sexual behavior than previous models, we can properly quantify behaviors that we previously were unable to including several forms of seroadaptive behaviors.

The HLP data was acquired from interviews with study participants. Participants were surveyed longitudinally at baseline and then once every five months in six equally spaced interviews over 25 months. A number of measures of sexual behavior from the previous three months were recorded including the serostatus of each partner and the total number of protected and unprotected sex acts with up to the five most recent partners. In addition, the number of protected and unprotected sex acts with each of the 5 most recent partners is recorded individually. A detailed description of the data collected is provided in Section 2.2.

In our analysis, the unknown serostatus partners are grouped with the HIV^- serostatus partners and we abbreviate this HIV^-/unknown serostatus group as HIV^- for short. We simultaneously model the number of HIV^+ partners, HIV^- partners, protected acts with

each HIV^+ partner, unprotected acts with each HIV^+ partner, protected acts with each HIV^- partner, and unprotected acts with each HIV^- partner longitudinally across 6 discrete time points. We refer to these six variables as the outcomes of the model.

Compared to the standard univariate longitudinal models used, use of our proposed model gives the following four advantages:

- The number of total reported protected and unprotected acts across all partners is broken down by the numbers per each individual partner. Modeling the number of protected and unprotected acts with each partner uses the partner specific information recorded for the five most recent partners and the total number of protected and unprotected acts recorded over all partners. Having outcomes for each partner distinguishes the scenario of a few acts with many partners from the scenario of many acts with a single partner which can have implications when designing new interventions. Having outcomes for each partner also allows us to easily incorporate covariates describing each partner. For example, in our model we distinguish partners who are considered main partners from those that are considered casual partners. We refer to this breakdown of total protected and unprotected sex acts by partner as disaggregation on a per partner basis.

- Modeling both the number of partners for a single subject and number of acts with each of several partners nested within the same subject creates a multilevel dataset. This separates variability attributed to different subjects, variability due to changes over time within the same subject, and variability across different partners nested within the same subject at the same time. Separating these sources of variation allows us to model data where the subject is followed longitudinally through time, the partners are followed longitudinally through time, or both. In modeling the HLP data, subjects are followed over time but partners were not identified across visits and therefore are assumed to be different at each interview. We use this distinction to justify creating a time dependent process specifically for observations across time with the same subject and a time independent process for nested partners within the same subject across

time.

- Joint analysis of sexual behaviors with HIV^+ and HIV^- partners in the model allows us to compare sexual behaviors with partners of different serostatuses within the same subject. We quantify differences in behavior with HIV^+ and HIV^- partners and use these differences to provide evidence for seroadaptation within our study population. Behaviors associated with seroadaptation are collectively called *seroadaptive behaviors*. We define and draw inference on three specific *seroadaptive behaviors* observable in the HLP data: (i) preferentially choosing fewer HIV^- partners, (ii) decreased level of sexual activity with HIV^- partners, and (iii) increased use of condoms with HIV^- partners.

- Using subjects' complete sexual behavior profiles allows us to correctly incorporate baseline eligibility criteria that were used to screen for a high risk population. In the HLP study, participants were eligible only if they (i) reported at least one partner at baseline and (ii) reported at least one unprotected sex act with an HIV^- partner or with an HIV^+ partner that was not categorized as a main partner. Since these criteria match with observed outcomes in our model, we can choose modeling distributions at baseline that exclude events which would fail the eligibility criterion. Failure to do this could falsely lead the model to infer the population is riskier than it actually is at baseline.

A number of covariates are used in the proposed model to model our outcomes. We include a different time effect at 5, 10, 15, 20, and 25 months and allow the intervention and control groups to differ at all time points including baseline. We also include covariate effects for the site location, risk group, race, and a number of other participant characteristics measured at the baseline interview. Main partners were defined as a partner specific covariate for each partner to distinguish them as more stable relationships and included to model the number of protected and unprotected acts with each partner. Covariate effects were allowed to be different across each outcome in the joint model with the exception of the main partner covariate which was assumed the same among both HIV^+ and HIV^- partners. A list of all variables other than time and intervention is found in Table 2.2 on page 37.

Latent effects are used to model differences among people. We make use of subject specific latent effects and partner specific latent effects in the analysis of the HLP data. A subject specific latent effect quantifies the idea that even subjects with the same covariates have inherently different tendencies in their behavior. For example, between two subjects who have the same covariates, one may have a tendency to consistently report greater numbers of HIV^- partners than the other. Similarly, a partner specific latent effect quantifies the idea that participants have inherently different tendencies in behavior with different partners. In our model, each outcome has a subject specific latent effect modeling that subject's inherent tendencies for that behavior. The number of protected and unprotected acts with each partner also includes partner specific latent effects. Subject and partner specific latent effects together help to explain additional variability in our observed data that our original modeling distribution may not otherwise account for.

When jointly modeling multiple outcomes, each outcome can provide information on all the other outcomes in the model. Outcomes related in this way are correlated. A model can make use of this by explicitly specifying a correlation structure between the outcomes. Our model consists of multiple outcomes observed over six discrete time points. In our model, I specify three different correlation structures between the outcomes.

Outcomes reported by the same subject at a given interview are all correlated. This is done by specifying a correlation structure between the subject specific latent effects for the number of HIV^+ partners, the number of HIV^- partners, the number of protected acts with an HIV^+ partner, the number of unprotected acts with an HIV^+ partner, the number of protected acts with an HIV^- partner, and the number of unprotected acts with an HIV^- partner. We collectively call these correlated subject specific latent effects a *multivariate subject specific latent effect*. This implies a subject's behavioral tendencies are all correlated. For example, a subject who has a tendency to consistently report greater numbers of HIV^+ partners may also consistently report greater numbers of HIV^- partners.

The number of protected and unprotected acts with the same partner has a different correlation structure than the number of protected and unprotected acts reported across different partners by the same subject. This is done by specifying a correlation structure

between the partner specific latent effects for the number of protected and unprotected acts. This implies that knowing the number of protected acts with a given partner should give you additional information about the number of unprotected acts with that partner and vice versa as compared to some other partner reported by the same subject. For example, a subject who reports larger numbers of protected acts with each partner may report fewer unprotected acts with each partner. If the same subject reports even larger numbers of protected acts with one specific partner, it is likely that the number of unprotected acts with that specific partner may be even lower.

Outcomes reported by the same subject across different interviews are correlated. This is done by specifying a correlation structure between the multivariate subject specific latent effects at each interview that depends on the multivariate subject specific latent effects at the previous interview. At each time point, the subject specific latent effects are specified to be similar to the subject specific latent effects at the previous time point. Latent effects close together in time are more correlated than those far apart in time. In this way, the subject specific latent effects at followup interview 4 will be more correlated to those at followup 3 than those at followup 2. The process specified in our model with which these multivariate subject specific latent effects change from one time point to the next is called a multivariate autoregressive process. The process is defined to be stationary, implying that the amount of variation in subject specific latent effects at each time point is specified in our model to be the same.

Our model uses a longitudinal multilevel multivariate outcome that provides a more complete picture of a subject's sexual behavior. We highlight some of the important assumptions underlying our model that should be kept in mind. The model makes the assumption that given covariates and latent effects, outcomes are independent from each other and follow a parametric Poisson distribution. Also, the model assumes subject and partner specific latent effects take on a Gaussian distribution. It is possible that the variability between subjects and partners may show more variability than accounted for with a Gaussian distribution and a different distribution may be more appropriate. It is critical that observed data actually matches these assumptions if the model is intended for use in prediction. One way to

check these assumptions is to make multiple predictions of the subjects in the study using the parameters estimated from the model and compare the resulting predicted outcomes to those observed in the data. This is shown for the HLP data in Figure 2.2 on page 28. When interpreting the effects of covariates, we assume that interactions do not exist between our modeled covariates. For example, we assume that the difference in effect between females and MSMs does not differ whether they are in Milwaukee or Los Angeles. This assumption is made due to the large number of potential combinations among covariates that would otherwise occur. In the following section on the cmDPM model, we present a novel nonparametric statistical method that could potentially be used if we do not wish to make this assumption.

1.3 Non-technical description - The cluster memory Dirichlet process mixture model

Nonparametric Bayesian methods are very flexible models that automatically adjust the model size and complexity based on the data. Parametric methods rely on a specified distribution to model our observed data. For instance, in the HLP data, we assume a Poisson distribution adequately describes our outcomes. Nonparametric Bayesian methods do not make this assumption and instead operate under the assumption that the density describing our outcomes is unknown.

One way to model an unknown density is to model it as a combination of several known densities. Observations drawn from this unknown density can be drawn from any one of these several known densities. This is known as a mixture model and each known density is one component of the mixture. The number of known densities used to describe the unknown density is the number of mixtures. Observations modeled under the same mixture of a mixture model are similar in value and considered *clustered*. When the number of mixtures is not predetermined and can grow to be as large as the number of observations being modeled, the model is known as an infinite mixture model.

The Dirichlet process mixture (DPM) model is a commonly used nonparametric Bayesian

method that models unknown densities as an infinite mixture model. In its simplest form, the DPM is parameterized by a cluster label for each observation describing the cluster in which that observation belongs and a cluster specific value for each cluster corresponding to the mean value of the observations in that cluster. Since the DPM results in a natural clustering structure of similar observations, it is often also be used in clustering applications. For example, the DPM model could be applied in the HLP data to each subgroup combination of race, site, and risk group. Subgroups that behaved similarly would naturally cluster together.

We propose a novel nonparametric Bayesian method, the *cmDPM model*, that extends the DPM model and is ideally suited for longitudinal data. In longitudinal data, subjects are repeatedly observed over time and each subject produces multiple observations over this given time frame. In the cmDPM model, the clustering of subjects can change from one time point to another. However, individualized history of where a subject was previously clustered plays a key role in determining the cluster that the subject belongs to at the current time point. The model is specified so that each subject is more likely to stay in the cluster they were labeled in at the previous time point. Dependence on previous cluster structure describes the concept that subjects who behaved similarly at a previous time point should be more likely to behave similarly at future time points as well. On the other hand, subjects may reevaluate their cluster membership and leave the cluster to which they were previously labeled at the new time point. How often subjects reevaluate their cluster membership over time can vary from dataset to dataset and application to application. We describe this characteristic in a dataset as *cluster mobility*.

Figure 3.1 on page 43 shows visually some simulated datasets with various levels of cluster mobility. Each dataset contains 75 subjects observed at 4 discrete time points. Observations are simulated from a mixture model with three clusters and correspondingly defined cluster specific values at each time point. Each observation is allowed to move between mixtures over time. The top three plots show data where the cluster specific values are constant across time for each cluster. Cluster mobility, the probability of reassessing cluster membership at each time point, is 0.2 on the top left, 0.1 in the top middle, and 0.05 on the top right. The

top right plot presents an example of a dataset where we can be fairly certain the cluster structure will be maintained over time and subjects will continue to behave similarly to other subjects in the cluster. The bottom three plots all have cluster mobility of 0.1 but show some variation in cluster behavior that can be modeled using the cmDPM model. The bottom left plot shows an example of data where cluster mobility is low but the cluster specific values are decreasing over time. This could be an example of a treatment that was effective but where subjects who initially were high continue to be higher than others. The bottom middle plot shows an example of data where cluster mobility is low but where cluster specific values switch between the top and bottom groups. The bottom right plot shows simulated data where variation in the observations is increasing over time. However, cluster membership would be accurately modeled from the earlier time points and maintained even when the increase in variability in the data at time point 4 makes it difficult to determine which cluster observations belong to.

The basic cmDPM model is parameterized by (i) cluster labels for each subject at each time point, (ii) a cluster specific value assigned to each cluster with observations in it at each time point, and (iii) an overall cluster stickiness parameter. A higher cluster stickiness parameter implies subjects tend to stay in the same cluster they were assigned to at previous time points and is indicative of a low cluster mobility dataset. On the other hand, a lower cluster stickiness parameter implies subjects leave the clusters they were previously in more frequently and is indicative of a high cluster mobility dataset.

Cluster assignment of subjects in the cmDPM model is based on a mix of both how close the observed outcomes between the subjects are at the given time point and which cluster these subjects were previously assigned at the previous time point. Higher values of the overall cluster stickiness parameter places more weight on the previous cluster assignment. Once clusters are assigned for each subject at each time point, cluster specific values are given for each cluster based on the observed values assigned to that cluster. If a cluster is maintained over time, we allow the values assigned to that cluster at each time point to be related by a defined functional relationship. For example, we may constrain the values of that cluster to be close to each other at neighboring time points.

Potential applications for the cmDPM model are fairly broad. The cmDPM model can be used for density estimations problems where we may be interested in dependent densities across time from longitudinal data. We can also apply the cmDPM model to mean effects of outcomes themselves as in the annual tuberculosis incidence rate dataset in Section 3.4. If fixed covariates exist in the model, the cmDPM model can be used as a model for the subject specific latent effects, relaxing the parametric assumption of the subject specific latent effects in longitudinal data. Finally, the cmDPM model can be applied to address large groups of interactions between categorical variables. For instance, we could apply the cmDPM model to all combinations of race/site/risk group in the HLP data and combinations that behaved closely would naturally cluster together.

CHAPTER 2

The Health Living Project

2.1 Introduction

Behavioral interventions designed for people living with HIV represent a targeted method of reducing the sexual transmission of HIV and can potentially play a critical role in controlling the human immunodeficiency virus (HIV) epidemic (Aral & Eterman 2002). The Health Living Project (HLP) (The Healthy Living Project Team 2007) is one behavioral intervention program in recent years that targets this population.

A successful intervention works to reduce risky behaviors that can lead to future transmission. This can come from reducing the number of partners who are HIV negative and/or reducing the number of unprotected sex acts with each partner who is HIV negative. In HLP and similar studies, measures of risky behavior are traditionally summarized into a single outcome. These outcomes typically involve summing behaviors across partners, for example the total number of unprotected sex acts with all partners who are either HIV negative or of whom the HIV status is uncertain. These measures are *aggregated outcomes*. While an aggregated outcome can be conveniently modeled using a univariate longitudinal model, it fails to accurately describe the complexity of risky sexual behavior. We show that reduction in risk can come from several different behavioral changes and argue that these changes require different interventions. *Seroadaptation* is one specific strategy where the subject uses knowledge or perceived knowledge of their potential partner's HIV serostatus to alter their behavior and limit their risk of transmission. The term HIV serostatus describes an individual's status of being positive or negative for the HIV antibody. We use the terms HIV^+ serostatus to describe an individual that is positive for the HIV antibody and HIV^-

serostatus for an individual who is negative for the HIV antibody. The behaviors associated with enacting strategies for risk reduction based on HIV serostatus of potential partners are called *seroadaptive behaviors*.

Our model has the ability to quantify seroadaptive behavior and therefore provides a critical contribution towards understanding risky sexual behavior. When measuring the effects of behavioral interventions, changes may occur through seroadaptive behaviors and/or as a change in frequency. An HIV^+ participant who changes from (5 HIV^+, 5 HIV^-) partners in a given unit of time to (8 HIV^+, 2 HIV^-) partners changes their behavior through seroadaptation. If that same participant instead changes from (5 HIV^+, 5 HIV^-) partners to (2 HIV^+, 2 HIV^-) partners, then the reduction in frequency is the source of risk reduction. Seroadaptation and changes in frequency are not mutually exclusive. The participant that changes from (5 HIV^+, 5 HIV^-) partners to (3 HIV^+, 1 HIV^-) partners decreases their risk from both seroadaptation and reduction in frequency. It is important that these two forms of risk reductions are distinguished from one another because seroadaptive behavior is contingent on accurate knowledge of partner serostatus but frequency changes are not. Seroadaptive behavior is not a guaranteed preventative strategy (Pinkerton 2008; Wilson et al. 2010; Butler & Smith 2007) and risks of infection with multiple strains of the virus exist (Poudel et al. 2007). However, seroadaptive behavior has been shown to be an effective form of risk reduction (Cassels et al. 2009; Jin et al. 2009; Golden et al. 2008). The study of seroadaptive behavior has been isolated almost exclusively in men who have sex with men (Parsons et al. 2005; Cox et al. 2004; Snowden et al. 2009 2011). The HLP data presents an opportunity to examine seroadaptive behavior in a variety of other risk groups. Understanding the role of seroadaptive behavior is important in other prevention studies as well. Reniers & Helleringer (2011) argue that the importance of HIV testing and counseling can not be accurately measured without accounting for seroadaptive behavior. A similar observation has been made for condom effectiveness studies (Warner et al. 2004).

In this chapter, we propose to jointly analyze the number of HIV^+ partners, the number of HIV^-/unknown serostatus partners, and the numbers of protected and unprotected acts with each partner. The number of partners of each serostatus are observed once at each time

point and the number of protected and unprotected acts with each partner form a bivariate outcome that is nested within both subject and time. We refer to treating the number of partners and behavior with each partner as a separate observation as *disaggregation on a per partner basis* and propose a Bayesian model for multivariate multilevel longitudinal count data for analysis of this type of outcome. While our methods are motivated by the HLP data, many other similar sexual behavior intervention studies involving people living with HIV have comparable data structures suitable for future analysis including Living in the Face of Trauma (LIFT) (Sikkema et al. 2007), Positive Choice: Interactive Video Doctor (Gilbert et al. 2008), and Healthy Relationships (Kalichman et al. 2001). In comparison with univariate models of longitudinal aggregated outcomes, our model offers the following advantages:

- Disaggregation of sex acts on a per partner basis allows us to model the full multivariate hierarchical nested structure of the data. This allows us to identify differences in behavior that previous analyses could not.
- Using a multivariate multilevel longitudinal model correctly models distinct levels of heterogeneity that were previously lumped into a single source of variation. Behavior variability exists both across different subjects as well as across partners within subjects. Previous analyses failed to account for variation across partners.
- Identification of seroadaptation in our population is critical to estimating the effects of behavioral interventions and can be determined directly from our model.
- Baseline eligibility criterions are correctly incorporated into the model.

We expand on each of these in turn.

Disaggregation on a per partner basis results in a separate bivariate outcome of the number of protected and unprotected sex acts for each partner. Modeling these outcomes separately differentiates a scenario of many partners with a few unprotected sex acts each from a scenario with a single partner and many unprotected acts. These different scenarios provide information for a more targeted counseling intervention focusing on reducing the

number of casual relationships with HIV$^-$ partners, increasing the use of protection, or both. In addition, these scenarios provide important information in deciding appropriate policy for prevention. The effectiveness of programs like PreExposure Prophylaxis (PrEP) (Grant et al. 2010) where HIV$^-$ people take antiretroviral medication daily to lower their chances of becoming infected may be viable and reasonable for the partners of our study population if only a few HIV$^-$ partners that engage in many unprotected acts with the subject were involved.

The multilevel multivariate outcome separately models subject level and partner level variation. Typical analyses in the field that analyze total protected acts in an aggregated fashion ignores this differentiation. Using multilevel outcomes facilitates proper modeling of the longitudinal process. Modeling separate sources of variation correctly also provides some estimation benefits since all observed information is used. Subjects in the HLP study are followed over time but unfortunately partners are not identified and are not traceable. This is reflected in our model through a time dependent process on subject specific latent effects leaving partner level variation independent of time. A simulation illustrating the benefits of modeling different sources of variation is given in Appendix A.1.

Our model enables us to quantify seroadaptive behavior. For the HLP data, we draw inference on 3 specific seroadaptive behaviors. Subjects can selectively choose (i) their partners, preferring HIV$^+$ partners over HIV$^-$/unknown serostatus partners, (ii) to have fewer sex acts with their HIV$^-$/unknown serostatus partners compared to their HIV$^+$ partners, and (iii) to use condoms a larger percentage of the time with partners who are HIV$^-$/unknown. Case (i) is behavior associated with partner selection while cases (ii) and (iii) describe sexual behavior during a relationship once it is formed. When seroadaptation is examined with an HIV$^-$ population, these seroadaptive behaviors still exist but in reverse form. The level of seroadaptation can change over time, differ between treatment and control groups, and differ across risk group, race, or by location.

Many studies, including HIV prevention studies, use eligibility criteria to initially screen for a high risk population. This recruitment method implies that some combination of observations that can occur at followup visits can not occur at baseline. For example, if only

participants who reported unprotected sex with an HIV$^-$ partner were recruited into the trial, we could have no zeros at baseline for this measure and a simple Poisson model would be inappropriate. If unaccounted for, the entire study population will appear more risky at baseline due to regression to the mean when no changes have actually occurred. The multivariate disaggregated data structure is necessary to model the complex eligibility criteria used in these trials. In our proposed model, we build this recruitment eligibility criterion directly into our conditional baseline distributions to remove bias due to the recruitment of a high risk population and to more accurately reflect the range of possible observations at baseline.

Previous multivariate count models (Chib & Winkelmann 2001) and (Tunaru 2002) model data without the multilevel structure using a single set of latent effects to model correlation between observations. A number of approaches extend modeling to multivariate multilevel data including Dunson (2000), Goldstein (2010), Rabe-Hesketh et al. (2005), and Browne & Draper (2006). Goldstein et al. (2009) specifically addresses multilevel models where data is observed on more than 1 level. This was later extended to model count data in Goldstein & Kounali (2009) by formulating a latent normal model where an extra step is inserted to sample an underlying latent normal variable from the count data. However, this method makes the fixed effect coefficients harder to interpret. None of these models examine the longitudinal component simultaneously with multilevel multivariate data. A common way to incorporate longitudinal dependence is through the inclusion of additive random effects that are correlated over time in the predictor. Alternatively, multivariate smoothing techniques such as Gaussian Markov random fields (Rue & Held 2005; Fahrmeir & Lang 2001) or dynamic Bayesian linear models (West et al. 1985; West & Harrison 1997) can be used.

The model we propose is for multilevel multivariate count data with observations on multiple levels and is followed through time. The introduction of a second level of latent effects properly accounts for heterogeneity across partners. Correlation across time is modeled through a generalized autoregressive process which can be applied to each level of latent effects as needed. In our model, subjects are followed through time while partners are not

so the process is applied only to the subject specific latent effects.

The remainder of the chapter is organized as follows. Section 2.2 provides a detailed description of the HLP data and outlines the multilevel structure. Terminology specific to HIV and sexual behavior is defined. Section 2.3 describes our model, specifies the prior structure, and outlines the algorithm for posterior computation. In Section 2.4, we outline how to draw inference on seroadaptive behavior with our model and define some specific terminology for the task. Section 2.5 presents the results of applying the presented model to the HLP dataset. Findings on seroadaptive behavior are explored in detail. Finally, we conclude with a discussion in Section 2.6.

2.2 Data Description

The Healthy Living Project is a multisite 2-group randomized controlled comprehensive behavioral intervention trial aimed at reducing risky sexual behavior among people living with HIV. After randomization, 469 eligible individuals were assigned to the control group and 467 were assigned to the behavioral intervention treatment group. HLP participants were surveyed longitudinally every 5 months in six equally spaced interviews over 25 months and a graphical representation of the collected data is shown in Figure 2.1. Participant demographic information was collected at baseline and a summary can be found in Appendix A.2. Subjects were categorized into 4 risk groups by transmission risk as intravenous drug users (IDU) and, provided they were not IDU, then by sexual orientation as men who have sex with men (MSM), heterosexual males (HTM), or females (FEM). The study was conducted in Los Angeles, Milwaukee, New York, and San Francisco. Other demographic information that was stored include education status, race, gender, and age.

At each interview including the baseline interview, outcomes pertaining to risky sexual behavior in the previous 3 months were recorded. This included the total number of partners, the serostatus of each partner, and the total number of protected and unprotected sex acts across all partners. Specific partner information on the number of protected and unprotected sex acts with up to the 5 most recent partners is also recorded. In the HLP study, a partner

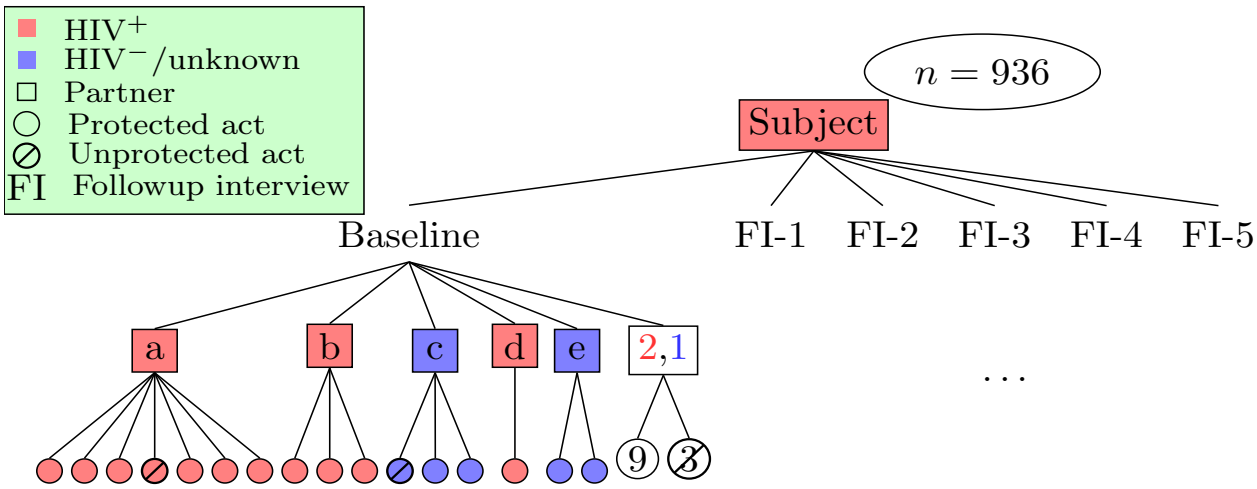

Figure 2.1: Graphical representation of longitudinally collected sexual behavior data in the HLP study.

was defined as someone with whom the participant had oral, vaginal, or anal sex. Sex acts in our analysis exclude oral sex and are defined as vaginal or anal intercourse only, due to the negligible transmission risk of HIV associated with oral sex leaving observations of 0 sex acts with a partner possible. The serostatus of each partner is categorized as HIV$^+$, HIV$^-$, or unknown. As usual in this field, we treat unknown serostatus partners as HIV$^-$ and for the remainder of the chapter, use of the description HIV$^-$ refers jointly to the HIV$^-$ and unknown serostatus population.

Recruitment was designed to screen for a high risk population. Study participants were only enrolled if they reported at least one unprotected sex act with an HIV$^-$ partner or with at least 1 HIV$^+$ partner who was not categorized as their primary partner at the baseline interview. Zero truncated joint distributions are used for the appropriate outcomes at baseline to reflect this enrollment criterion.

The number of HIV$^+$ partners, HIV$^-$ partners, and total protected and unprotected acts across all partners are subject level observations recorded longitudinally over time. Disaggregating the number of total protected and unprotected sex acts by partner gives the number of protected and unprotected sex acts with each specific partner which are bivariate partner level observations nested within subject and time. In the cases when a subject

reported more than 5 partners at an interview, the number of protected and unprotected sex acts with each of these additional partners are treated as random variables with a constrained sum to reflect the uncertainty in the data. After imputation of these outcomes, there was a total of 22,269 partner level bivariate observations spread over 4,695 separate interviews. Subject and partner level observations are jointly modeled as Poisson counts with mean parameters driven by participant and partner level covariates and random effects.

2.3 Model

2.3.1 Likelihood Specification

Let observations for subject $i = 1, \ldots, n$ occur at measurement times t_{ij} where $j = 1, \ldots, J_i$ and J_i varies by subject. At time t_{ij}, let V_{ij} denote the total number of partners associated with subject i in the previous 3 months, and let $\mathbf{W}_{ijk} = (P_{ijk}, U_{ijk})^T$ denote the bivariate outcome composed of the number of protected sex acts P_{ijk} and unprotected sex acts U_{ijk} nested in subject i and time t_{ij} with partner $k = 1, \ldots, V_{ij}$.

To model separate behavioral changes among HIV$^+$ and HIV$^-$ partners, we introduce an HIV serostatus indicator Z_{ijk} denoting the serostatus of partner k at time t_{ij} where $Z_{ijk} = 1$ indicates HIV$^+$ and $Z_{ijk} = 0$ indicates HIV$^-$/unknown status. Letting V_{ij}^+ be the number of HIV$^+$ partners and V_{ij}^- be the number of HIV$^-$ partners at time t_{ij}, then

$$V_{ij} = V_{ij}^+ + V_{ij}^- \tag{2.1}$$

$$V_{ij}^+ = \sum_{k=1}^{V_{ij}} Z_{ijk}. \tag{2.2}$$

The complete outcome vector, $\mathbf{Y}_{ij}$, for subject i at time t_{ij} is then modeled with a multivariate nested count model where

$$\mathbf{Y}_{ij} = (V_{ij}^+, V_{ij}^-, \mathbf{W}_{ij1}^T, \mathbf{W}_{ij2}^T, \ldots, \mathbf{W}_{ijV_{ij}}^T)^T \tag{2.3}$$

has length $2V_{ij} + 2$ that will vary across subjects and times. Outcomes are correlated both

between different outcomes and across time. We first present the cross sectional model and correlations between outcomes in $\mathbf{Y}_{ij}$. Correlations across time are then introduced through a multivariate stationary process.

We model each component of $\mathbf{Y}_{ij}$ as Poisson distributed $V_{ij}^{+} \sim \text{Po}(\lambda_{v,ij}^{+})$, $V_{ij}^{-} \sim \text{Po}(\lambda_{v,ij}^{-})$, $P_{ijk} \sim \text{Po}(\lambda_{p,ijk})$, and $U_{ijk} \sim \text{Po}(\lambda_{u,ijk})$ conditional on corresponding mean parameters $\lambda_{v,ij}^{+}$, $\lambda_{v,ij}^{-}$, $\lambda_{p,ijk}$, and $\lambda_{u,ijk}$ for $1 \leq i \leq n$, $1 \leq j \leq J_i$, and $1 \leq k \leq V_{ij}$. A log-linear random effects regression model characterizes these mean parameters

$$\begin{pmatrix} \lambda_{v,ij}^{+} \\ \lambda_{v,ij}^{-} \end{pmatrix} = \exp \begin{pmatrix} \mathbf{x}_{ij}' \boldsymbol{\alpha}_v^{+} + \beta_{v,ij}^{+} \\ \mathbf{x}_{ij}' \boldsymbol{\alpha}_v^{-} + \beta_{v,ij}^{-} \end{pmatrix} \tag{2.4}$$

and

$$\begin{pmatrix} \lambda_{p,ijk} \\ \lambda_{u,ijk} \end{pmatrix} = \begin{cases} \exp \begin{pmatrix} \mathbf{x}_{ijk}' \boldsymbol{\alpha}_p^{+} + \beta_{p,ij}^{+} + \delta_{p,ijk} \\ \mathbf{x}_{ijk}' \boldsymbol{\alpha}_u^{+} + \beta_{u,ij}^{+} + \delta_{u,ijk} \end{pmatrix}, \text{if } Z_{ijk} = 1, \\ \exp \begin{pmatrix} \mathbf{x}_{ijk}' \boldsymbol{\alpha}_p^{-} + \beta_{p,ij}^{-} + \delta_{p,ijk} \\ \mathbf{x}_{ijk}' \boldsymbol{\alpha}_u^{-} + \beta_{u,ij}^{-} + \delta_{u,ijk} \end{pmatrix}, \text{if } Z_{ijk} = 0 \end{cases} \tag{2.5}$$

where average number of protected and unprotected acts with partner k, $\lambda_{p,ijk}$ and $\lambda_{u,ijk}$, are parameterized differently depending on partner serostatus. Here, $\mathbf{x}_{ij}$ is the vector of subject level covariates and $\mathbf{x}_{ijk}$ appends partner level covariates to $\mathbf{x}_{ij}$. The set of fixed effects $\boldsymbol{\alpha}$ $= (\boldsymbol{\alpha}_v^{+}, \boldsymbol{\alpha}_v^{-}, \boldsymbol{\alpha}_p^{+}, \boldsymbol{\alpha}_u^{+}, \boldsymbol{\alpha}_p^{-}, \boldsymbol{\alpha}_u^{-})$ and subject specific latent effects $\boldsymbol{\beta}_{ij} = (\beta_{v,ij}^{+}, \beta_{v,ij}^{-}, \beta_{p,ij}^{+}, \beta_{u,ij}^{+}, \beta_{p,ij}^{-}, \beta_{u,ij}^{-})^T$ correspond to the number of HIV$^+$ partners, HIV$^-$ partners, protected acts with an HIV$^+$ partner, unprotected acts with an HIV$^+$ partner, protected acts with an HIV$^-$ partner, and unprotected acts with an HIV$^-$ partner respectively. Partner specific latent effects corresponding to partner k are denoted as $\boldsymbol{\delta}_{ijk} = (\delta_{p,ijk}, \delta_{u,ijk})$.

To complete specification of the hierarchical structure, random effects are modeled as Gaussian. Subject specific random effects $\boldsymbol{\beta}_{ij}$ evolve through time using a stationary multivariate autoregressive(1) process leading to correlation between observations within the same subject across time. At baseline $j = 1$, subject specific random effects $\boldsymbol{\beta}_{i1}$ are normal with

6×6 covariance matrix $\mathbf{L}$

$$\boldsymbol{\beta}_{i1}|\mathbf{L} \sim \mathrm{N}_6(\mathbf{0}, \mathbf{L}). \tag{2.6}$$

Future time points $j = 2, \ldots, J_i$ model subject specific random effects as conditional normal

$$\boldsymbol{\beta}_{ij}|\boldsymbol{\beta}_{ij-1}, \boldsymbol{\Sigma} \sim \mathrm{N}_6(\mathbf{A}\boldsymbol{\beta}_{ij-1}, \boldsymbol{\Sigma}) \tag{2.7}$$

where $\mathbf{A}$ is a diagonal matrix with each diagonal element taking values between -1 and 1, and $\boldsymbol{\Sigma}$ is a positive definite covariance matrix. We assume the process is stationary which implies

$$\mathbf{L} = \mathbf{A}\mathbf{L}\mathbf{A}^T + \boldsymbol{\Sigma}, \tag{2.8}$$

or equivalently elementwise

$$L_{st} = \frac{\Sigma_{st}}{1 - A_{ss}A_{tt}}, \tag{2.9}$$

where L_{st}, Σ_{st}, and A_{st} denote the s^{th} row t^{th} column element of matrix $\mathbf{L}$, $\boldsymbol{\Sigma}$, and $\mathbf{A}$ respectively. Partner specific bivariate random effects $\boldsymbol{\delta}_{ijk}$ are normal with covariance $\mathbf{D}$

$$\boldsymbol{\delta}_{ijk}|\mathbf{D} \sim \mathrm{N}_2(\mathbf{0}, \mathbf{D}). \tag{2.10}$$

Since partners are not followed over time and we can not determine if any of the same partners are kept between interviews, we assume no correlation between partner specific random effects.

Conditionally, observed outcomes for subject i, $\mathbf{Y}_i = (\mathbf{Y}_{i1}^T, \ldots, \mathbf{Y}_{iJ_i}^T)^T$ are multivariate Poisson-log normal as defined in Aitchison & Ho (1989). The expected value and correlation between outcomes after marginalizing over subject and partner specific random effects can be calculated using equations (2.4) and (2.5). Define conditional mean parameters $\mathrm{E}(\mathbf{Y}_i|\boldsymbol{\lambda}_i)$ $\equiv \boldsymbol{\lambda}_i = (\boldsymbol{\lambda}_{i1}^T, \ldots, \boldsymbol{\lambda}_{iJ_i}^T)^T$ where $\boldsymbol{\lambda}_{ij} = (\lambda_{v,ij}^+, \lambda_{v,ij}^-, \lambda_{p,ij1}, \lambda_{u,ij1}, \ldots, \lambda_{p,ijV_{ij}}, \lambda_{u,ijV_{ij}})^T$. From

our hierarchical model specification in equations (2.6), (2.7), and (2.10), it follows naturally

$$\log \boldsymbol{\lambda}_i \sim \mathrm{N}(\boldsymbol{\mu}_i, \boldsymbol{\Omega}) \tag{2.11}$$

where $\boldsymbol{\mu}_i$ with r^{th} element μ_{ir} is a function of fixed effects $\boldsymbol{\alpha}$ while covariance matrix $\boldsymbol{\Omega} = (\Omega_{rq})$ is a function of $\mathbf{A}$, $\mathbf{L}$, and $\mathbf{D}$. The mean and variance of the r^{th} component of $\mathbf{Y}_i$ are

$$\mathrm{E}(Y_{ir}) = \exp\left(\mu_{ir} + \frac{1}{2}\Omega_{rr}\right) \equiv \phi_{ir}, \tag{2.12}$$

$$\mathrm{var}(Y_{ir}) = \phi_{ir} + \phi_{ir}^2\{\exp\left(\Omega_{rr}\right) - 1\} \tag{2.13}$$

and the covariance between the q^{th} and r^{th} observations of $\mathbf{Y}_i$ is

$$\mathrm{cov}(Y_{iq}, Y_{ir}) = \phi_{iq}\phi_{ir}\{\exp\left(\Omega_{qr}\right) - 1\} \tag{2.14}$$

where $q \neq r$. Equation (2.14) shows that the sign and statistical significance of the covariance between observations can be determined directly from the $\boldsymbol{\Omega}$ matrix since $\mathrm{cov}(Y_{iq}, Y_{ir})$ is only positive when Ω_{qr} is positive and only negative when Ω_{qr} is negative. We define a parameter as statistically significant if the 95% posterior equal-tail credible interval of the parameter does not contain 0. The credible interval for Ω_{qr} is sufficient to determine if the credible interval for $\mathrm{cov}(Y_{iq}, Y_{ir})$ contains 0.

The eligibility criterion for the study can not be correctly modeled in a univariate framework but is readily incorporated into the joint baseline distributions of our model. The eligibility criterion excludes two distinct events at baseline $j = 1$. The first excluded event is the number of HIV$^+$ and HIV$^-$ partners both reported as 0. The second excluded event is the number of unprotected sex acts with all HIV$^-$/unknown serostatus partners and HIV$^+$ casual partners all reported as 0. To incorporate this information into the model, two joint zero-truncated Poisson distributions replace the conditionally independent Poisson assumption for partners and separately for acts. We define $(X_1, \ldots, X_G)$ to be distributed joint

zero-truncated Poisson if the joint density of $(X_1, \ldots, X_G)$ is

$$\left\{1 - \exp\left(-\sum_{g=1}^{G} \lambda_g\right)\right\}^{-1} \frac{\exp\left(-\sum_{g=1}^{G} \lambda_g\right) \prod_{g=1}^{G} \lambda_g^{X_g}}{\prod_{g=1}^{G} X_g!} \tag{2.15}$$

for $X_g = \{0, 1, 2, \ldots\}$ where $(X_1, \ldots, X_G)$ are not all equal to 0.

2.3.2 Prior Specification

Proper prior distributions are chosen for model parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\Sigma}$, $\mathbf{A}$, $\mathbf{D}$) to ensure the posterior is well defined. In selecting values for parameterizing our prior distributions, we chose values that are neutral favoring no positive or negative effect and non-influential thus letting the data drive our posterior inference. We consider priors to be non-influential if a twofold increase in the variance of our priors changes our posterior mean estimates by less than 1%. Normal priors for each fixed effect covariate $\alpha_l \sim \mathrm{N}(\mu_\alpha = 0,\ \sigma_\alpha^2 = 10)$ are assumed. Inference on the posterior mean number of partners and sex acts per partner involves exponentiating the diagonal terms of the covariance matrix as shown in equation (2.12). This results in a complication when the standard conjugate Inverse-Wishart density priors alone are used for covariance parameters $\boldsymbol{\Sigma}$ and $\mathbf{D}$ since the exponentiated diagonal terms have undefined posterior means due to long right tails. Details for the univariate case are shown in the Appendix. To correct this problem, priors for $\boldsymbol{\Sigma}$ and $\mathbf{D}$ are chosen to be proportional to the product of an Inverse-Wishart density distribution on the entire covariance matrix times independent left truncated normal density distributions on each of the diagonal elements truncated at 0. Letting $\Sigma_{s,s}$ and $D_{t,t}$ be the diagonal elements of $\boldsymbol{\Sigma}$ and $\mathbf{D}$, for $s = 1, \ldots 6$ and $t = 1, 2$, priors for $\boldsymbol{\Sigma}$ and $\mathbf{D}$ are

$$\pi(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{m_{\boldsymbol{\Sigma}}}{2}} \exp\left[-\frac{1}{2}\left\{\mathrm{tr}(\boldsymbol{\Psi}_\Sigma \boldsymbol{\Sigma}^{-1}) + \sum_{s=1}^{6}\left(\frac{\Sigma_{s,s} - \mu_\Sigma}{c_\Sigma}\right)^2\right\}\right], \tag{2.16}$$

$$\pi(\mathbf{D}) \propto |\mathbf{D}|^{\frac{m_D}{2}} \exp\left[-\frac{1}{2}\left\{\mathrm{tr}(\boldsymbol{\Psi}_D \mathbf{D}^{-1}) + \sum_{t=1}^{2}\left(\frac{D_{t,t} - \mu_D}{c_D}\right)^2\right\}\right], \tag{2.17}$$

where $\mathbf{\Sigma} > 0$ and $\mathbf{D} > 0$. This shortens the right tails resulting in less dispersion in the prior. Values for the parameters were chosen as $m_\Sigma = 10$, $\mathbf{\Psi}_\Sigma = I_6$, $m_D = 10$, $\mathbf{\Psi}_D = I_2$, $\mu_\Sigma = 2$, $\mu_D = 2$, $c_\Sigma = 100$, $c_D = 100$. Autoregressive factors $A_{l,l}$, the l^{th} diagonal element of $\mathbf{A}$ are given uniform priors $A_{l,l} \sim \text{Unif}(a_A = -1, b_A = 1)$.

2.3.3 Posterior Computation

Our posterior computations are sampled with Markov Chain Monte Carlo (MCMC) methods using the Metropolis algorithm (Metropolis et al. 1953; Hastings 1970; Gelfand & Smith 1990; Casella & George 1992). We provide a brief summary here with a more detailed step by step algorithm in Appendix A.4. For sampling $\boldsymbol{\alpha}$, a second-order Taylor expansion around the current state of the Markov chain (Rue & Held 2005) was used as the proposal function and this proposal substantially speeds convergence when compared to a random walk Gaussian proposal function. Sampling of subject specific random effects $\boldsymbol{\beta}_{ij}$, partner specific random effects $\boldsymbol{\delta}_{ijk}$, and autoregressive matrix A use random walk Gaussian proposal functions with the proposal function for the diagonal elements of $\mathbf{A}$ truncated at -1 and 1. Covariance parameters $\mathbf{\Sigma}$ and $\mathbf{D}$ are sampled using the Metropolis algorithm with Inverse-Wishart proposal functions that approximate the posterior density. Specifically, we use the proposal functions $q(\mathbf{\Sigma}|\boldsymbol{\beta}_{ij})$ and $q(\mathbf{D}|\boldsymbol{\delta}_{ijk})$ that are densities of Inverse-Wishart distributions $\text{IW}(m_{q1}, \mathbf{\Psi}_{q1})$ and $\text{IW}(m_{q2}, \mathbf{\Psi}_{q2})$ respectively where $m_{q1} = m_\Sigma + \sum_{i=1}^{n}\sum_{j=1}^{J_i} 1$, $\mathbf{\Psi}_{q1} = \sum_{i=1}^{n}\sum_{j=2}^{6}(\boldsymbol{\beta}_{ij} - \mathbf{A}\boldsymbol{\beta}_{ij-1})(\boldsymbol{\beta}_{ij} - \mathbf{A}\boldsymbol{\beta}_{ij-1})^T + \mathbf{\Psi}_\Sigma$, $m_{q2} = m_D + \sum_{i=1}^{n}\sum_{j=1}^{J_i} V_{ij}$, and $\mathbf{\Psi}_{q2} = \sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{V_{ij}} \boldsymbol{\delta}_{ijk}\boldsymbol{\delta}_{ijk}^T + \mathbf{\Psi}_D$.

Let S_{ij} be the set of partners k for which P_{ijk} and U_{ijk} are not observed in the data. Then conditional on total protected acts, total unprotected acts, and mean parameters $\boldsymbol{\lambda}_{ij}$, P_{ijk} and U_{ijk} are Multinomial distributed random variables and can be directly sampled.

The estimation procedure was implemented in R. A total of 100,000 iterations were collected after an initial 30,000 iterations were discarded as burn-in.

2.4 Seroadaptive behavior

Using our model, we make inference on 3 different forms of seroadaptive behavior to describe how the choice of (i) partner, (ii) level of sexual activity, and (iii) condom use differs depending on the serostatus of the partner. We define terms to help quantify these seroadaptive behaviors respectively as the (i) serodiscordant partner multiple (SPARM), (ii) serodiscordant activity multiple (SAM), and (iii) serodiscordant protection multiple (SPROM) and discuss them below.

The serodiscordant partner multiple measures whether partners of a specific serostatus are preferentially selected. To determine SPARM, we first calculate the expected probability that subjects with characteristics defined by $\mathbf{x}_{ij}$ would choose an HIV$^-$ partner as

$$\pi_{SPARM,ij} = \mathrm{E}_{\boldsymbol{\beta}_{ij}} \left(\frac{\lambda^-_{v,ij}}{\lambda^+_{v,ij} + \lambda^-_{v,ij}} \right) \tag{2.18}$$

where $\mathrm{E}_\theta(g(\theta))$ refers to the expectation of the function $g(\theta)$ with respect to the random variable θ. The corresponding serodiscordant partner multiple, SPARM_{ij} is then defined as the odds of picking an HIV$^-$ partner

$$\mathrm{SPARM}_{ij} = \frac{\pi_{SPARM,ij}}{1 - \pi_{SPARM,ij}} \tag{2.19}$$

and is interpreted to mean HIV$^-$ partners are picked SPARM_{ij} times as often as HIV$^+$ partners. The SPARM is a relative measure that depends on the prevalence and availability of HIV$^+$ partners in the area and should not be used by itself to assess the presence of seroadaptive behavior. It is however possible to compare SPARM between different groups or time periods to determine differences or changes in seroadaptive behavior. A ratio of SPARMs less than 1 implies a difference in seroadaptive partner selection with individuals characterized by the numerator showing greater preference towards selecting partners who are also HIV$^+$.

The serodiscordant activity multiple (SAM) measures differences in the level of sexual activity in relationships with partners of differing serostatus. Sexual activity level is mea-

sured by the total number of reported sex acts in the relationship regardless of protection status. The expected probability that subjects with characteristics defined by $\mathbf{x}_{ij}$ choose to engage in sex with an HIV$^-$ over an HIV$^+$ partner adjusting for differences in the number of partners of each serostatus is

$$\pi_{SAM,ij} = \mathrm{E}_{\boldsymbol{\beta}_{ij},\boldsymbol{\delta}_{ijk}} \left(\frac{\lambda_{u,ijk} + \lambda_{p,ijk}}{\lambda_{u,ijk} + \lambda_{p,ijk} + \lambda_{u,ijk'} + \lambda_{p,ijk'}} \right) \tag{2.20}$$

where partner k is HIV$^-$ and partner k' is HIV$^+$. Correspondingly, SAM$_{ij}$ is defined as the odds of engaging in sex with an HIV$^-$ partner,

$$\mathrm{SAM}_{ij} = \frac{\pi_{SAM,ij}}{1 - \pi_{SAM,ij}}. \tag{2.21}$$

A SAM value of 1 implies no difference in sexual activity level between partners of different serostatus while values less than 1 imply seroadaptation towards lower levels of sexual activity with HIV$^-$ partners.

The serodiscordant protection multiple (SPROM) measures selective use of protection based on partner serostatus. The SPROM measure describes differences in the tendency to use protection per sex act. The expected probability that a protected sex act is with an HIV$^-$ partner instead of an HIV$^+$ partner adjusting for differences in sexual activity level and the number of partners is

$$\pi_{SPROM,ij} = \mathrm{E}_{\boldsymbol{\beta}_{ij},\boldsymbol{\delta}_{ijk}} \left(\frac{w_{ijk}\lambda_{p,ijk}}{w_{ijk}\lambda_{p,ijk} + w_{ijk'}\lambda_{p,ijk'}} \right) \tag{2.22}$$

where partner k is HIV$^-$, partner k' is HIV$^+$, and $w_{ijk} = (\lambda_{p,ijk} + \lambda_{u,ijk})^{-1}$ are weights to adjust for the sexual activity level of partner k. The corresponding SPROM$_{ij}$ is defined as

$$\mathrm{SPROM}_{ij} = \frac{\pi_{SPROM,ij}}{1 - \pi_{SPROM,ij}}, \tag{2.23}$$

the odds a protected sex act is with an HIV$^-$ partner assuming equal numbers of HIV$^+$ and HIV$^-$ partners and equal levels of sexual activity. A SPROM value of 1 implies no

seroadaptive use of protection while values greater than 1 imply a greater tendency to use protection with HIV$^-$ partners.

2.5 Results

2.5.1 Model Validation

Accurate modeling of the high activity portions of the population is particularly important. These cases represent the greatest transmission risk and also test to see if our model adequately accounts for the high levels of variation in sexual behavior. To examine the accuracy of our model in this area, we compare the tail end of our posterior predictive distributions to the data for the number of HIV$^+$ partners, number of HIV$^-$ partners, total number of protected acts with all HIV$^+$ partners, total number of protected acts with all HIV$^-$ partners, total number of unprotected acts with all HIV$^+$ partners, and total number of unprotected acts with all HIV$^-$ partners at each time point. A total of 1000 datasets of new values for $\mathbf{Y}_{ij}$ are simulated using the sampled posteriors of parameters $\boldsymbol{\lambda}_{ij}$ for $1 \leq i \leq n$ and $1 \leq j \leq J_i$.

Figure 2.2 shows a summary of the tail portions of our posterior predictive distribution. For comparison, we also plot the percentage of observations in the HLP data that fit into each bin. In all these outcomes, our model reasonably predicts the percentage of high activity in the study population.

2.5.2 HLP Covariate Effects

Analysis of the HLP data does not find the counseling intervention to be efficacious. Subjects in the treatment group did not show evidence for (a) fewer HIV$^-$/unknown partners, (b) greater numbers of protected acts per HIV$^-$/unknown partner, or (c) fewer unprotected sex acts per HIV$^-$/unknown partner than their control group counterparts at any time point. The number of HIV$^+$ and HIV$^-$ partners does decrease over the course of the study across all study participants indicating a reduction in risky behavior as a result of being enrolled in the study regardless of treatment group assignment. Average number of unprotected acts

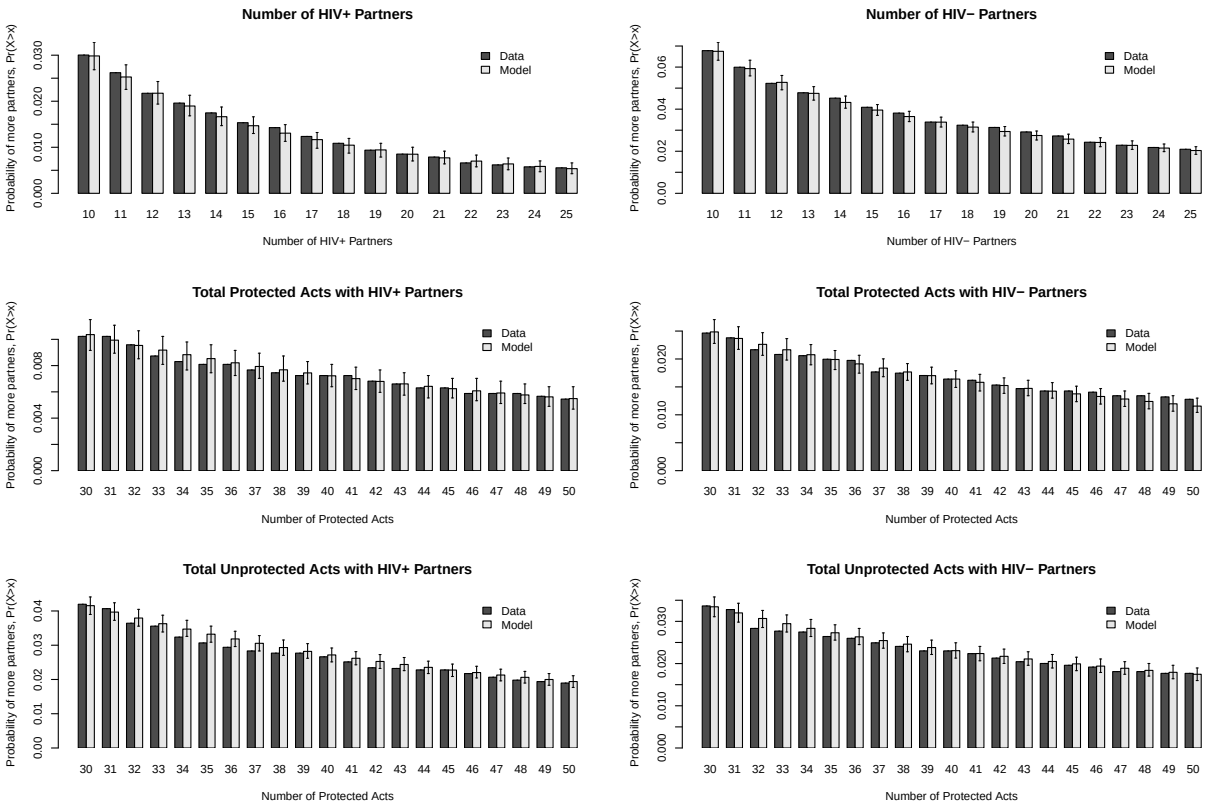

Figure 2.2: Comparison of tail probabilities from the model posterior predictive distributions and the percentage of HLP data fitting the bin criterion. Each bin reflects the probability of observing values greater than or equal to the outcome value. Outcomes include the number of HIV^+ partners, number of HIV^- partners, total number of protected acts with all HIV^+ partners, total number of protected acts with all HIV^- partners, total number of unprotected acts with all HIV^+ partners, and total number of unprotected acts with all HIV^- partners for all subjects over all time points.

per partner also decreases across both treatment and control groups while protected acts per partner stays fairly consistent throughout the duration of the study.

We find key differences in behavior across the different risk groups and summarize our findings in Figure 2.3. Females exhibited the riskiest sexual behavior, reporting the greatest numbers of unprotected sex acts with HIV^- partners. Even though the MSM group reported the largest numbers of HIV^- partners and the IDU group reported the largest numbers of total unprotected sex acts, both groups appear to mitigate their risk when their partner was HIV^- and reported fewer unprotected sex acts with those partners than the FEM group.

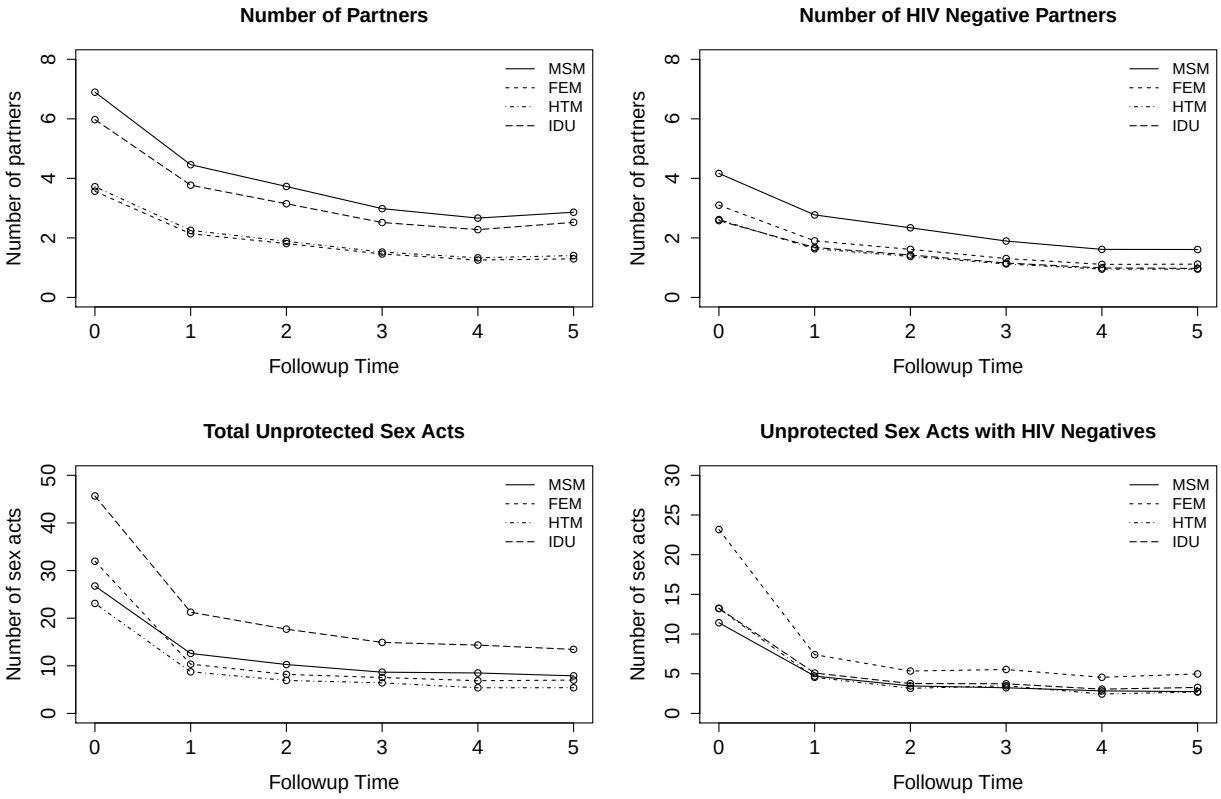

Figure 2.3: Summary of sexual behaviors across the 4 risk groups. All plots are average predicted outcomes with our model for participants who were white, assigned to the control group, less than 40 years old, from Los Angeles, did not graduate high school, were out of work, had no history of hard drug use, and reported only 1 main partner if any partners were reported.

The covariance parameters in the model do not directly affect interpretation of the covariates but show the extent and direction of correlation between the multiple outcomes in the model. Posterior estimates of the covariance parameters in the model can be found in Appendix A.5.

Covariate effects in our model are now explored in detail.

2.5.2.1 Treatment Over Time

We evaluated treatment efficacy of the HLP trial at each followup comparing the treatment group to the control group looking for evidence of any of the following scenarios:

Table 2.1: Difference between treatment and control groups at each followup time after adjusting for baseline. Values are multiplicative and reported as posterior mean PM followed by the 95% equal-tail credible interval (L_{CI}, U_{CI}).

Variable	Followup 1	Followup 2	Followup 3	Followup 4	Followup 5
Number of Partners					
HIV^+	1.04 (0.81, 1.33)	1.10 (0.83, 1.46)	1.14 (0.83, 1.55)	1.31 (0.92, 1.81)	0.89 (0.61, 1.24)
HIV^-/unknown	0.92 (0.74, 1.12)	0.87 (0.68, 1.11)	0.90 (0.68, 1.15)	0.92 (0.69, 1.21)	0.86 (0.63, 1.15)
Protected Acts					
HIV^+ partner	0.78 (0.38, 1.46)	0.74 (0.34, 1.47)	0.84 (0.33, 1.68)	1.29 (0.49, 2.75)	1.64 (0.62, 3.47)
HIV^-/unknown partner	0.90 (0.61, 1.26)	1.19 (0.77, 1.75)	1.39 (0.87, 2.10)	1.30 (0.78, 2.04)	1.19 (0.71, 1.88)
Unprotected Acts					
HIV^+ partner	1.08 (0.73, 1.55)	1.05 (0.68, 1.56)	1.16 (0.71, 1.79)	0.73 (0.43, 1.18)	1.13 (0.68, 1.79)
HIV^-/unknown partner	1.16 (0.81, 1.64)	1.21 (0.79, 1.80)	0.99 (0.59, 1.49)	0.87 (0.53, 1.38)	1.11 (0.65, 1.76)

- decrease in the number of HIV^-/unknown partners
- increase in the number of protected acts per HIV^-/unknown partner
- decrease in the number of unprotected acts per HIV^-/unknown partner.

A comparison of sexual behavior profiles for the treatment and control groups across time are shown in Figure 2.4. An overall decrease in the average number of both HIV^+ and HIV^-/unknown partners is observed across the entire study population. The average number of unprotected acts per partner also decreases across both treatment and control groups across partners of either serostatus while protected acts per partner stays fairly consistent throughout the study. However, the treatment group does not appear to behave differently from the control group at any time point. We show the difference between the treatment and control groups at each followup time period adjusting for baseline differences in Table 2.1. Values greater than 1 imply higher estimated counts in the treatment group than the control group after adjusting for initial differences in the two groups at baseline while values less than 1 imply lower estimated counts. For example, our estimates suggest the treatment group reported only 0.92 (0.74, 1.12) times as many HIV^-/unknown partners as the control group at followup 2 after adjusting for baseline differences.

2.5.2.2 Site, Risk Group, and Race

The HLP study was carried out across multiple geographical sites, contained multiple HIV transmission risk groups, and included multiple ethnicities of participants. Table 2.2 shows estimates for these covariate effects.

San Francisco appears to be the most risky population for HIV transmission followed by Los Angeles. New York and Milwaukee are less risky but for different reasons. Participants in Milwaukee behaved differently from those in Los Angeles, San Francisco, and New York, reporting significantly fewer HIV^+ partners likely due to the demographics of the population there. Participants at the New York site reported more protected sex with their partners than all other sites. Participants in San Francisco reported greater numbers of partners than the other locations. Most of the difference comes from the number of HIV^-/unknown serostatus partners.

Among the 4 risk groups, females report the largest numbers of unprotected sex acts with HIV^- partners. While MSMs have reported significantly greater numbers of HIV^-/unknown partners, their transmission risk is mitigated by their increased propensity to use protection with these partners. The IDU group also exhibits unsafe behavior, reporting the lowest number of protected sex acts per partner and the second highest number of unprotected sex acts among the 4 risk groups. However, their HIV transmission risk is partially mitigated because relative to other risk groups, a larger proportion of their partners are HIV^+ serostatus partners.

The race of the participant did not have as pronounced an effect on transmission risk as risk group and location. We found no significant differences between whites and others. The Hispanic group reported significantly larger numbers of HIV^-/unknown serostatus partners than all other groups but also report fewer unprotected sex acts with them. The African American group reported significantly fewer numbers of HIV^-/unknown serostatus partners than whites and Hispanics but also reported slightly higher numbers of unprotected sex acts per partner.

We spend the remainder of this section on presenting the new findings specific to seroad-

aptive behavior in this study population.

2.5.3 HLP Seroadaptation

We first examine seroadaptive behavior in white men who have sex with men (MSM) that are less than 40 years old from Los Angeles who did not graduate high school and were out of work with no history of hard drug use. Our model estimates these participants to have a SPARM of 1.50 (1.09, 1.99) entering relationships with HIV^- partners roughly 50% more often than with HIV^+ partners. We denote inference estimates as a_M (a_L, a_U) where a_M is the posterior mean, a_L and a_U are respectively the 2.5% quantile and 97.5% quantile estimates of the posterior distribution. Over the course of the study, the SPARM stays around this level with a slight dip to 1.32 (0.93, 1.78) by the end of the study. Lower levels of sexual activity were reported when the relationship was with an HIV^- partner resulting in a SAM of 0.78 (0.56, 1.07) though the value is not significant. The SAM stays fairly consistent throughout the study staying below 1 at each followup with values reported at followup 2 and followup 4 being marginally significant. The SPROM of 1.48 (1.22, 1.77) provides significant evidence that participants were almost 50% more likely to use protection during sex if their partner was HIV^-. This value increases somewhat during the course of the study to as much as 1.87 (1.44, 2.39) by followup 5 indicating participants may have further increased seroadaptive use of condoms while in the study although the difference in SPROMs at the two time points is not significant. Comparison with the treatment group shows that the treatment and control groups exhibit very similar behavior throughout the study. This shows that the counseling intervention did not significantly alter subjects' seroadaptive strategies. Overall, clear evidence exists for seroadaptive use of protection in our study group.

Seroadaptive behavior varies also across the 4 risk groups of the study population. Since most studies of seroadaptive behavior study MSM, we examine seroadaptive behavior across the risk groups using MSM as the comparison group. Our findings show the 4 risk groups do not significantly differ in SAM or SPROM values indicating similar seroadaptive behavior once relationships are formed. However, females reported a significantly higher SPARM

value of 4.26 (2.89, 5.93) compared to the SPARM value of 1.50 (1.09, 1.99) reported by the MSM population indicating that females chose HIV^- partners far more frequently over HIV^+ partners compared to MSM. The HTM population also report a relatively high SPARM value of 2.05 (1.37, 2.93) choosing HIV^- partners about twice as often as HIV^+ partners. Part of the observed difference in SPARM values between the heterosexual and MSM populations is likely a result of the higher percentage of HIV^+ individuals in the MSM population though this does not account for differences between the heterosexual males and females. Our findings suggest interventions targeting females and heterosexual males may benefit from encouraging seroadaptive selection of partners to reduce the SPARM value.

Our analysis also examined seroadaptive behavior differences across race and location. Among the race categories of white, black, Hispanic, and other, seroadaptive behavior strategies were fairly similar with Hispanics choosing the largest percentage of HIV^- partners though differences were not significant. HLP was a multisite study across Los Angeles, Milwaukee, New York, and San Francisco. Our findings indicate participants in Milwaukee reported a significantly higher SPARM value of 3.16 choosing HIV^- partners far more often than participants in the other locations. This could be a result of differences in HIV^+ partner availability in these respective areas. Nonetheless, it is important to recognize the increased percentage of HIV susceptible partners of people living with HIV in a location like Milwaukee and respond accordingly with the appropriate intervention.

A full summary of seroadaptive results is shown in Table 2.3.

2.6 Discussion

Our Bayesian model for multilevel multivariate longitudinal count data has distinct advantages in the modeling of sexual behavior data and allows a more comprehensive evaluation of covariate effects such as the effects of counseling intervention. Our proposed model differentiates between different forms of risk reduction in sexual behavior. HIV transmission risk can be lowered through decreasing the number of partners, decreasing the number of sex acts per partner, or increasing the proportion of sex acts where condoms are used. Each of these risk

reduction scenarios may also be selectively modified based on partner serostatus illustrating seroadaptation. The counseling to achieve and to respond to each of these scenarios can be very different. Identifying the specific areas of behavior that present the highest risk for transmission can highlight key points of interest for future intervention studies. For example, future intervention studies targeting females may prefer to focus on partner selection since that risk group chose a significantly larger proportion of HIV^- partners than any other risk group.

Our model is the first model to mathematically quantify different types of seroadaptive behavior providing a quantitative basis for future seroadaptation studies. In the HLP study, we find that subjects living with HIV already make some seroadaptive behavioral decisions that reduce the risk of transmission and that the current intervention under evaluation did not appear to alter these seroadaptive behaviors significantly. In addition, there is evidence that seroadaptive partner selection specifically varies based on the risk group and location. Individually tailored interventions focusing on enhancing these already existing seroadaptive behaviors may be another route of transmission risk reduction.

We also highlighted the importance of disaggregation on a per partner basis when analyzing sexual behavior data. Disaggregation of the number of protected and unprotected acts per partner treats behavior with each partner as separate events permitting partner level analysis and allowing the modeling of partner contributed variation, explicitly differentiating it from subject contributed variation. The situation of multiple unprotected sex acts with a single partner where only 1 possible HIV transmission could occur is now weighed differently than the situation of 1 unprotected sex act with multiple partners where multiple transmission acts could occur. The resulting multilevel model is ideal for sexual behavior data since we can separately model the effects of following subjects longitudinally through time even when their partners may be different at each followup. We make use of a generalized autoregressive stationary process on the subject specific latent effects accounting for heterogeneity of time trajectories across all outcomes.

Other intervention studies of sexual behavior may not record total number of acts across all partners. Instead, only partner specific information for the first few partners and the

total number of partners are recorded. In this scenario, imputation of the acts data for the remaining partners without the constraint on total number of acts is theoretically still possible with our model.

Although we have focused on modeling longitudinal sexual behavior profiles, our model can readily be applied to any application with nested longitudinal count data.

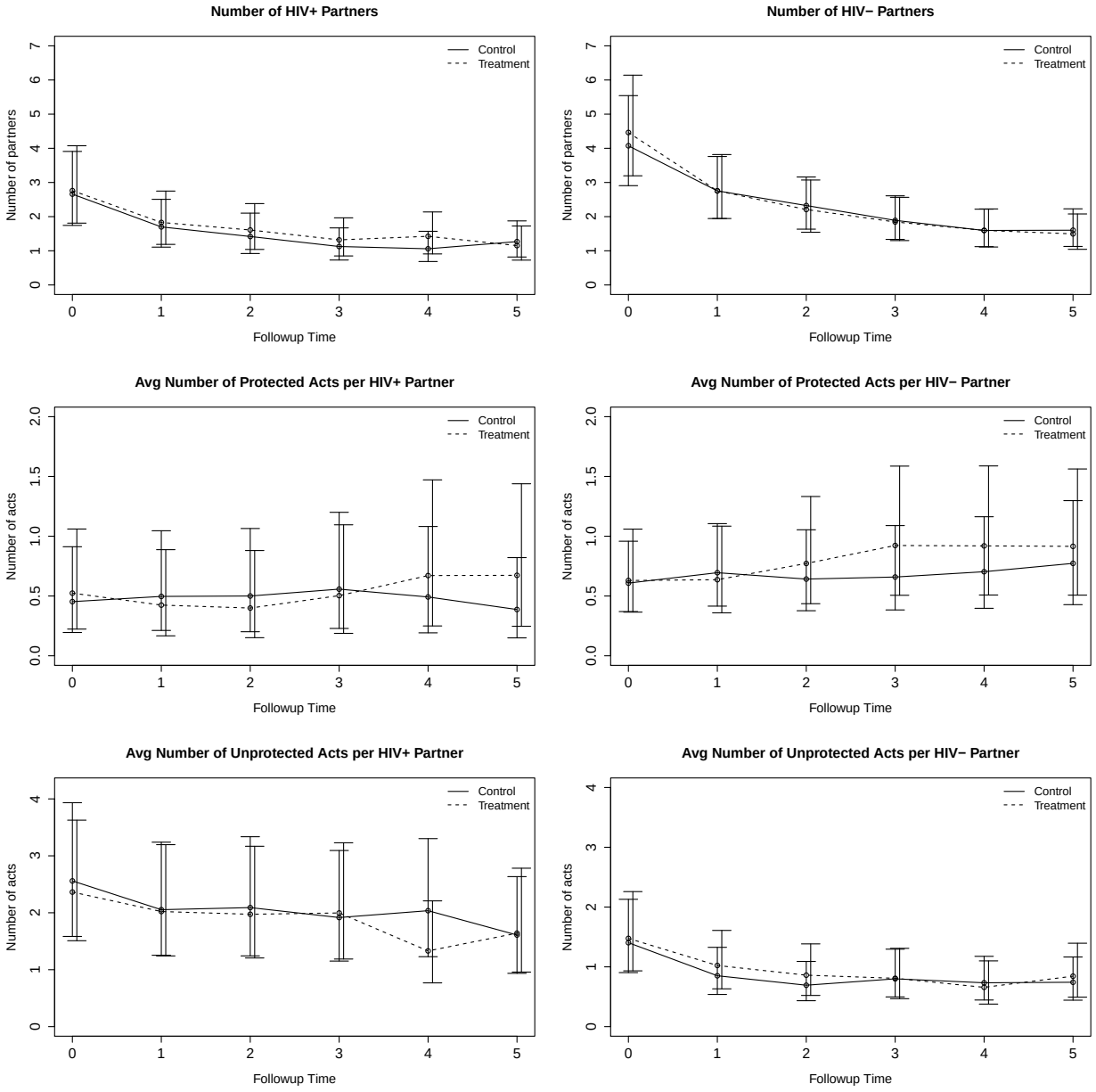

Figure 2.4: Sexual behavior profiles of treatment and control groups over time. Graphs represent averages for participants who are white, MSM, less than 40 years old, from Los Angeles, did not graduate high school, were out of work, and had no history of hard drug use

Table 2.2: The effects of covariates on sexual behavior. Multiplicative effects from the comparison group are reported as posterior mean PM followed by the 95% equal-tail credible interval (L_{CI}, U_{CI}). The comparison group are white MSMs assigned to the control group who are less than 40 years old, from Los Angeles, did not graduate high school, out of work, and had no history of hard drug use. Values with * indicate statistically significant evidence of difference from the comparison group.

Variable	Number of Partners	Protected Acts	Unprotected Acts
HIV^+ Partner			
Site			
Milwaukee	0.34 (0.23, 0.49)*	0.68 (0.31, 1.71)	1.29 (0.78, 2.12)
New York	1.10 (0.87, 1.38)	1.34 (0.85, 2.25)	0.94 (0.70, 1.23)
San Francisco	1.15 (0.93, 1.43)	0.78 (0.51, 1.19)	0.85 (0.66, 1.11)
Risk Group			
Female	0.14 (0.10, 0.19)*	2.06 (1.16, 3.76)*	1.98 (1.37, 2.94)*
HTM	0.36 (0.25, 0.51)*	3.35 (1.77, 6.13)*	1.12 (0.75, 1.66)
IDU	1.21 (0.94, 1.54)	0.50 (0.30, 0.85)*	1.65 (1.22, 2.18)*
Education			
HS grad or some college	0.92 (0.70, 1.18)	1.43 (0.94, 2.20)	1.00 (0.76, 1.30)
College graduate	1.34 (0.95, 1.87)	0.75 (0.39, 1.47)	0.98 (0.69, 1.45)
Age			
More than 40 yrs old	0.89 (0.75, 1.08)	1.07 (0.76, 1.52)	0.99 (0.82, 1.22)
Career			
Working	1.22 (1.08, 1.37)*	1.09 (0.82, 1.42)	1.14 (0.97, 1.36)
Race			
Black	0.96 (0.76, 1.22)	2.70 (1.81, 4.08)*	0.83 (0.65, 1.05)
Hispanic	0.83 (0.61, 1.12)	3.37 (1.98, 5.94)*	0.87 (0.64, 1.22)
Other	0.77 (0.53, 1.07)	1.70 (0.95, 3.15)	1.05 (0.72, 1.55)
Drug Use			
Hard drug usage (Lifetime)	1.31 (1.05, 1.66)*	0.94 (0.60, 1.42)	0.92 (0.71, 1.16)
Recent hard drug usage	1.09 (0.76, 1.56)	0.82 (0.43, 1.43)	1.13 (0.76, 1.64)
HIV^- or unknown Partner			
Site			
Milwaukee	1.08 (0.82, 1.41)	1.07 (0.73, 1.58)	1.07 (0.72, 1.56)
New York	0.95 (0.78, 1.16)	1.72 (1.28, 2.34)*	0.95 (0.74, 1.23)
San Francisco	1.50 (1.24, 1.80)*	0.93 (0.72, 1.22)	0.88 (0.67, 1.11)
Risk Group			
Female	0.69 (0.57, 0.84)*	1.70 (1.22, 2.30)*	1.88 (1.41, 2.54)*
HTM	0.58 (0.45, 0.77)*	2.39 (1.57, 3.56)*	1.29 (0.88, 1.87)
IDU	0.60 (0.47, 0.76)*	0.51 (0.35, 0.77)*	1.66 (1.19, 2.32)*
Education			
HS grad or some college	1.09 (0.90, 1.33)	1.05 (0.76, 1.44)	0.88 (0.68, 1.15)
College graduate	1.95 (1.53, 2.53)*	0.76 (0.49, 1.16)	0.84 (0.58, 1.23)
Age			
More than 40 yrs old	0.84 (0.73, 0.98)*	0.96 (0.79, 1.21)	0.79 (0.63, 0.96)*
Career			
Working	1.13 (1.02, 1.26)*	0.99 (0.83, 1.17)	1.15 (0.97, 1.37)
Race			
Black	0.83 (0.68, 0.99)*	1.26 (0.94, 1.73)	1.05 (0.80, 1.34)
Hispanic	1.36 (1.09, 1.69)*	1.90 (1.35, 2.73)*	0.78 (0.57, 1.06)
Other	0.86 (0.65, 1.14)	1.37 (0.89, 2.10)	0.94 (0.63, 1.40)
Drug Use			
Hard drug usage (Lifetime)	0.95 (0.79, 1.14)	0.90 (0.68, 1.17)	1.18 (0.90, 1.53)
Recent hard drug usage	0.98 (0.74, 1.34)	0.95 (0.64, 1.41)	1.41 (1.00, 1.96)*
Partner type			
Main Partner		3.70 (3.41, 4.00)*	

Table 2.3: Seroadaptive behavior comparison. Comparison group is white men who have sex with men (MSM) at baseline that are less than 40 years old from Los Angeles who did not graduate high school and were out of work with no history of hard drug use. For each group, we evaluated the serodiscordant partner selection (SPARM), serodiscordant activity multiple (SAM), and the serodiscordant protection multiple (SPROM). Values of SAM and SPROM with * indicate statistically significant evidence for seroadaptation. Values in **bold** are significantly different from the value in the comparison group.

Seroadaptive measures	SPARM	SAM	SPROM
Comparison group	1.50 (1.09, 1.99)	0.78 (0.56, 1.07)	1.48 (1.22, 1.77)*
Time			
Followup 1	1.55 (1.10, 2.09)	0.74 (0.52, 1.03)	1.71 (1.36, 2.14)*
Followup 2	1.56 (1.10, 2.12)	0.67 (0.46, 0.93)*	1.80 (1.41, 2.28)*
Followup 3	1.58 (1.11, 2.16)	0.73 (0.51, 1.02)	1.69 (1.30, 2.15)*
Followup 4	1.47 (1.05, 2.00)	0.72 (0.50, 1.00)*	1.82 (1.39, 2.32)*
Followup 5	1.32 (0.93, 1.78)	0.88 (0.60, 1.23)	1.87 (1.44, 2.39)*
Treatment (Trt)			
Trt Baseline	1.55 (1.10, 2.10)	0.82 (0.59, 1.13)	1.42 (1.16, 1.73)*
Trt Followup 1	1.48 (1.05, 2.03)	0.80 (0.56, 1.12)	1.62 (1.28, 2.03)*
Trt Followup 2	1.39 (0.98, 1.91)	0.82 (0.57, 1.15)	1.82 (1.42, 2.32)*
Trt Followup 3	1.41 (1.00, 1.94)	0.82 (0.56, 1.15)	1.91 (1.47, 2.49)*
Trt Followup 4	1.23 (0.87, 1.69)	0.90 (0.61, 1.27)	1.83 (1.33, 2.44)*
Trt Followup 5	1.35 (0.95, 1.87)	0.88 (0.58, 1.24)	1.75 (1.31, 2.30)*
Risk Group			
FEM	**4.26** (2.89, 5.93)	0.72 (0.49, 1.02)	1.44 (1.14, 1.73)*
HTM	2.05 (1.37, 2.93)	0.86 (0.56, 1.28)	1.48 (1.08, 1.95)*
IDU	0.96 (0.65, 1.34)	0.69 (0.45, 1.00)*	1.31 (1.10, 1.54)*
Race			
Black	1.35 (0.99, 1.79)	0.79 (0.58, 1.05)	1.27 (1.01, 1.57)*
Hispanic	2.02 (1.47, 2.72)	0.75 (0.55, 1.01)	1.52 (1.18, 1.92)*
Other	1.59 (1.07, 2.28)	0.75 (0.51, 1.06)	1.53 (1.18, 1.93)*
Location			
Milwaukee	**3.16**(2.09, 4.65)	0.76 (0.49, 1.11)	1.58 (1.26, 1.93)*
New York	1.36 (0.95, 1.87)	0.89 (0.63, 1.25)	1.64 (1.27, 2.02)*
San Francisco	1.77 (1.26, 2.37)	0.82 (0.58, 1.13)	1.51 (1.25, 1.81)*

CHAPTER 3

The cmDPM model

3.1 Introduction

The Dirichlet process mixture (DPM) model first introduced by Ferguson (1973) and Antoniak (1974) has been used extensively as a non-parametric Bayesian model. The DPM model can be used in clustering problems where the number of clusters are not known a priori. The standard DPM model assumes exchangeable observations from a single unknown distribution.

In the longitudinal data setting, Kleinman & Ibrahim (1998) has applied the DPM to the random effects model. In this model, multivariate subject specific random effects are distributed as from a single unknown multivariate distribution with a Dirichlet process prior. This model is restrictive since each subject is assumed to belong to the same cluster of the DPM model throughout the study. A more flexible approach specifies a separate distribution at each time point resulting in a collection of distributions.

Recent work has extended the DPM model to model a collection of distributions. These distributions are specified in the model so that they are not independent and identically distributed but instead share some common parameterization. MacEachern (1999) and MacEachern (2000) introduced the dependent Dirichlet process (DDP) that induces dependence between distributions by replacing the elements of the stick-breaking construction (Sethuraman 1994) of the Dirichlet process with stochastic processes. Along these lines, a number of models have been developed. De Iorio et al. (2004) uses a "single-p" DDP model where dependencies are introduced only through the point masses but not the weights of the stick breaking construction to create an ANOVA dependent Dirichlet process model. Teh

et al. (2006) introduces the hierarchical Dirichlet process (HDP) that models the baseline measure itself as a Dirichlet process. Dependence is induced in this case by sharing clusters and the corresponding values assigned to these clusters across all distributions. Rodríguez et al. (2008) introduced the nested Dirichlet process (nDP) which clusters similar distributions together. Dependent distributions across longitudinal data should also account for the ordering of time. In this area, Zhu et al. (2005) and Caron et al. (2007) introduced time dependent DP models based on varying the Polya urn-type scheme. Griffin & Steel (2006) developed a dependent Dirichlet process where the weights of the stick-breaking construction are conditional on an ordered covariate which could be time. Jensen & Shore (2011) extends the HDP model for longitudinal data allowing parameter values for each individual to either stay the same across time or be drawn from a time dependent distribution induced from the HDP.

With the exception of Jensen's Markovian hierarchical Dirichlet process, the other mentioned models model dependence on the distribution level but do not account for dependence resulting from the same subjects being followed over time. Parameter values at each time point ignore individual history and are only dependent on the past history of the population as a whole treating each subsequent time point as a new draw of subjects from the same general population instead of observations of the same subjects.

Jensen's Markovian hierarchical Dirichlet process accounts for individual history by specifying a nonzero probability to retain each subject's parameter value from the previous time point. Subjects whose parameter values were not retained are then modeled separately from the HDP model. However, this model only allows for flat trends over time for subjects who stay in the same cluster since the subject's parameter value is the same as the previous time point. More importantly, since the model only specifies whether values of subject parameters are retained, subjects can never join an existing cluster from the previous time point that was not their own cluster. A subject's parameter value drawn separately from the HDP model will never match that of a previously existing value. This implies all clusters eventually die out over time with new ones being created. We propose a model that approaches the longitudinal data analysis problem from the clustering perspective to address these two

restrictions.

In this chapter, we propose the cluster memory Dirichlet process mixture (cmDPM) model for the modeling of longitudinal data in a nonparametric Bayesian framework accounting for individual history through both previous cluster membership and previous parameter value. Subjects that behaved similarly at a given time point should continue to behave similarly at future time points as well. A probability parameter ρ is defined in the model to reflect how often this occurs. Alternatively, an individual may leave its current cluster with probability $1 - \rho$ giving it freedom to move to any cluster at the following time point. We refer to the characteristic describing how often observations reevaluate their cluster membership at each time point as *cluster mobility*. *Cluster mobility* is also an important statistic for describing the data. Low levels of cluster mobility indicate cluster membership that is "sticky" over time and corresponds to a high dependence of future behavior on historical behavior. In extreme cases, cluster mobility can be 0 implying no change in cluster membership at any time or 1 implying complete independence from previous cluster membership. When the cluster mobility is 0, a single standard Dirichlet model applied to the longitudinal vector of outcomes together (Kleinman & Ibrahim 1998) is appropriate to model the data. When the cluster mobility is 1, independent DPM models applied at each time point are appropriate to model the data. The cmDPM model with the ρ parameter is a flexible model that spans the range of possibilities between these two extremes.

Figure 3.1 shows six simulated datasets with varying levels of *cluster mobility* suitable for modeling with the cmDPM model. Each dataset contains 75 subjects recorded over 4 discrete time points. Subjects are assigned into one of three clusters with each cluster given a cluster specific value that may be different at each discrete time point. At baseline, 25 subjects are assigned into each cluster. At each following time point, each subject may reassess which cluster they are assigned to with probability equal to the dataset's cluster mobility. The top three plots show examples of data where the cluster specific values do not change over time but subjects occasionally switch clusters with the top left plot showing the highest level of cluster mobility at $1 - \rho = 0.2$, the top middle plot having $1 - \rho = 0.1$, and the top right plot showing the lowest level of cluster mobility at $1 - \rho = 0.05$. These are examples of

data where no overall change is occurring in the density at each time point but individuals are moving around between clusters. The information on cluster mobility that could be gained from the cmDPM model could help us determine the predictability of individuals in the dataset at future time points. Modeling this data without accounting for individual history would not show any differentiation between these three examples. The bottom three plots show other data examples that the cmDPM model would be well suited for. In the bottom left plot, cluster specific values are dropping over time in all three clusters. This could be a dataset where there are three unique groups of participants for which a treatment was effective in all groups. In the bottom middle plot, cluster specific values in two groups swap at time point 3. In this case, the treatment could have been effective in one group but counterproductive in the other group. In both of these datasets, we can only accurately estimate cluster mobility and the behavior in the data by modeling dependence on previous cluster membership of the participants using the cmDPM model since the cluster specific values are changing. The bottom right plot shows an example of simulated data that gets progressively noisier with increasing variance over time. Estimating cluster mobility in the dataset with the cmDPM model can help us better estimate the clustering structure at the last time point by borrowing clustering information for each subject from the earlier time points.

The cmDPM model forms clusters at future time points by carrying an existing cluster forward in time or creating a new cluster. When existing clusters are retained over time, the cluster specific value of that cluster at the previous time point can be useful to determine future cluster specific values. This information does not exist when new clusters are created. To make this distinction, and properly use individual parameter values at the previous time point when appropriate, we use two distinctly different distributions to model the cluster specific values at a new time point. When an existing cluster is carried forward, the cluster specific value is modeled under a cluster time specific distribution parameterized by the cluster specific value of that cluster at the previous time point. On the other hand, when a new cluster that was not present at the previous time is formed, the cluster specific value is modeled under a general time specific distribution to represent the lack of information. This

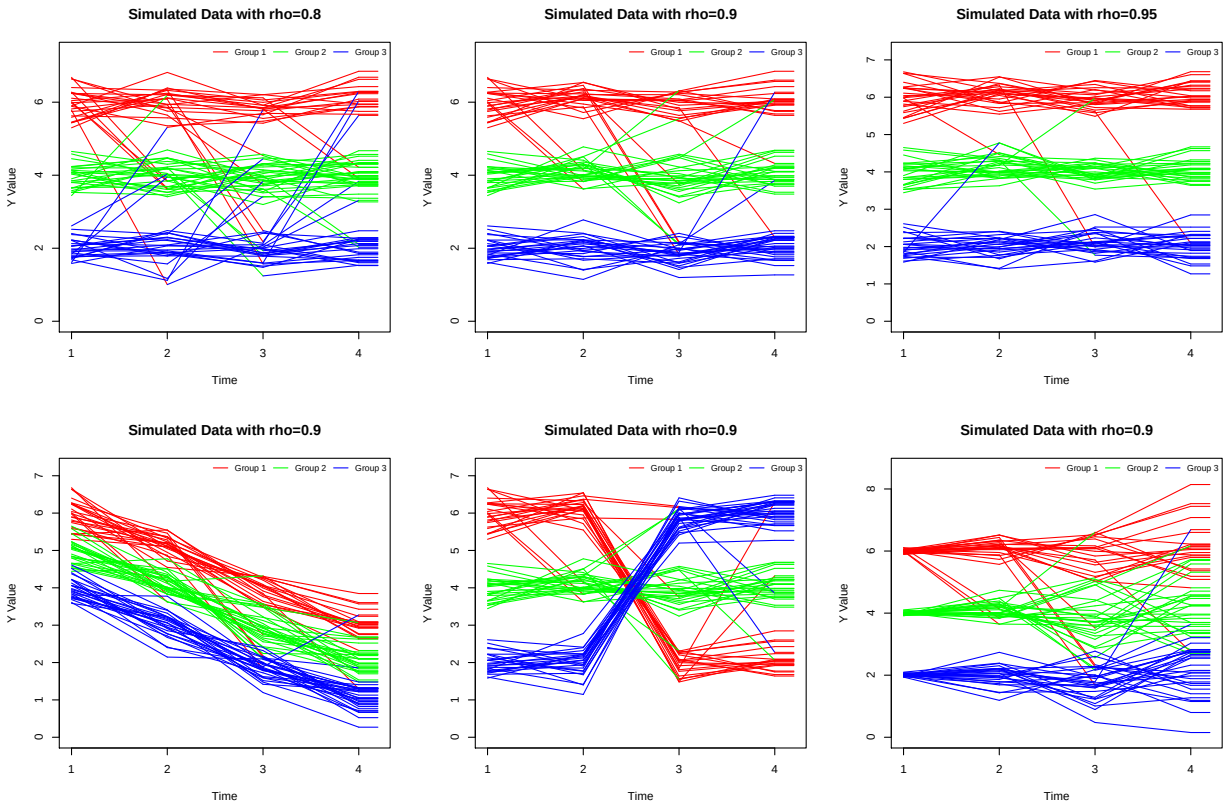

Figure 3.1: Simulated datasets with three clusters and four time points with various levels of cluster mobility, 1-ρ, that are suitable for the cmDPM model. Colors of lines indicate cluster membership at the previous time point.

distinction on how to model new cluster specific values makes use of the previous individual parameter values only from those individuals we believe are relevant to better predict the new cluster specific values at each time point.

The remainder of this chapter is organized as follows. We review the standard DPM model in Section 3.2. The cmDPM model is presented in Section 3.3. We develop the cmDPM model in three different ways. We first develop the model as an extended DPM model with two parts. We then develop the model by taking a finite mixture model and allowing the number of components to approach infinity. A persistent multi-day Chinese restaurant process is developed last which leads to the same limit. The cmDPM model is extended to include other parameters and hyperpriors. Finally, a summary of the posterior computation is given. In Section 3.4, we model tuberculosis incidence data from 197 countries around the

world over the last 21 years using our model. We conclude with a discussion in Section 3.5.

3.2 Dirichlet Process Mixture Models

Let $Y_1, \ldots, Y_n$ be the data assumed to be independently drawn from some unknown distribution U. The basic DPM model characterizes U by associating each observation Y_i with parameter θ_i and some specified likelihood density $f(Y_i|\theta_i)$. The parameter θ_i has a prior distribution G. To model uncertainty regarding the functional form of G, G is treated as random and modeled as a draw from a Dirichlet process characterized by strength parameter α and base distribution G_0. The full model can be summarized as

$$\begin{aligned} Y_i|\theta_i &\sim f(Y_i|\theta_i) \\ \theta_i|G &\sim G \\ G|\alpha, G_0 &\sim DP(\alpha, G_0) \end{aligned} \tag{3.1}$$

for i from 1, ..., n where $X \sim f(X|\theta)$ means that random variable X is distributed as some distribution F with density $f(X|\theta)$. Integrating over G gives the sequential conditional priors of θ_i (Blackwell & MacQueen 1973)

$$\theta_i|\theta_1, \ldots, \theta_{i-1} \sim \frac{1}{i-1+\alpha} \sum_{j=1}^{i-1} \delta(\theta_j) + \frac{\alpha}{i-1+\alpha} G_0, \tag{3.2}$$

where $\delta(\theta)$ is the distribution concentrated at a single point θ. Equation (3.2) shows that conditionally, each θ_i for $i = 1, \ldots, n$ has a nonzero probability of being equal to one of the previously given parameters $\theta_1, \ldots, \theta_{i-1}$ since the conditional distribution has a point mass at each of these values. The resulting draw of $\theta_1, \ldots, \theta_n$ from G is therefore clustered at some $K \leq n$ values.

Neal (2000) shows an equivalent model to (3.1) can be obtained from a finite mixture model with K components by taking the limit as K goes to infinity. The finite mixture

model has the form

$$\begin{aligned} Y_i|c_i, \boldsymbol{\phi} &\sim f(Y_i|\phi_{c_i}) \\ c_i|\boldsymbol{\pi} &\sim \text{Discrete}(\pi_1, \ldots, \pi_K) \\ \phi_c &\sim G_0 \\ \boldsymbol{\pi}|\alpha &\sim \text{Dirichlet}\left(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K}\right). \end{aligned} \tag{3.3}$$

Here, c_i is a cluster label that shows the cluster that observation Y_i belongs to. Each cluster $c \in \{a_1, \ldots a_K\}$ has a corresponding parameter ϕ_c that determines the distribution $f(Y_i|\phi_c)$ for that cluster. A discrete distributed random variable with mixing parameter $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ takes on values $a_1, \ldots, a_K$ with corresponding probabilities $(\pi_1, \ldots, \pi_K)$ where the values $a_1, \ldots, a_K$ are any assigned values. The mixing parameter $\boldsymbol{\pi}$ determines the probability of falling into each cluster $c \in \{a_1, \ldots a_K\}$ and is given a symmetric Dirichlet prior with concentration parameter α/K. In the limit as K goes to infinity, the two models in equations (3.1) and (3.3) are identical if we define $\theta_i = \phi_{c_i}$ for $i = 1, \ldots, n$.

3.3 Cluster Memory Dirichlet Process Mixture Model

We give three different elaborations of the cmDPM model in the following section which derive the model from three different perspectives. We first present the cmDPM model as a 2 part model where M_j observations remain in the clusters they were assigned to at the previous time point j and the remaining $n - M_j$ observations exhibit clustering properties like a standard DP with M_j already observed values. We then derive the cmDPM model directly from a longitudinal finite mixture model taking the number of components at each time point to infinity. In our final presentation, we give a process centered presentation of the cmDPM analogous to a Markovian Chinese restaurant process. We follow the specifications of the cmDPM model with a discussion of hyperpriors appropriate for the model. Posterior computation and an efficient estimation algorithm is presented in the final section.

3.3.1 Model Specification

3.3.1.1 DPM Extension

Let Y_{ij} denote longitudinal observations with $Y_{ij}|\theta_{ij} \sim f(Y_{ij}|\theta_{ij})$ for subject i $(i = 1, \ldots, n)$ at discrete time points j $(j = 1, \ldots, J)$. The parameter θ_{ij} is a subject specific parameter for subject i at time point j. At any given time point j, the parameters $\boldsymbol{\theta}_j = (\theta_{1j}, \ldots, \theta_{nj})$ are clustered into K_j clusters with all subject specific parameters in the same cluster assigned the same cluster specific value. A cluster label c_{ij} is used to denote which cluster subject i belongs to. Observations i and i' are in the same cluster at time j if cluster labels are equal, $c_{ij} = c_{i'j}$. An alternative parameterization of $\boldsymbol{\theta}_j$ assigns a value for each cluster instead of each subject with $\phi_{jc_{ij}} \equiv \theta_{ij}$. The parameter $\phi_{jc_{ij}}$ denotes the cluster specific value for the cluster with label c_{ij}.

At discrete time point $j = 1$, there is no prior history and observations at the first time point Y_{i1} are modeled using a standard DPM model,

$$Y_{i1}|\theta_{i1} \sim f(Y_{i1}|\theta_{i1}), \tag{3.4}$$

$$\theta_{i1}|G_1 \sim G_1, \tag{3.5}$$

$$G_1 \sim \mathrm{DP}(\alpha, G_{01}), \tag{3.6}$$

where α is the strength parameter and G_{01} is a time specific base distribution for $j = 1$.

At discrete time points $2 \geq j \geq J$, a cluster stickiness parameter ρ is used to determine how well clusters stay together over time. For each observation $1 \leq i \leq n$, we model a cluster memory indicator Z_{ij} as

$$Z_{ij}|\rho \sim \mathrm{Bernoulli}(\rho), \tag{3.7}$$

$$\rho \sim \mathrm{Beta}(a_\rho, b_\rho). \tag{3.8}$$

When $Z_{ij} = 1$, cluster label c_{ij} will retain the same value as the previous time point $c_{i(j-1)}$. When $Z_{ij} = 0$, cluster label c_{ij} is redrawn. Conditioned on $\mathbf{Z}_j = (Z_{1j}, \ldots Z_{nj})$, the cmDPM

model at time point j is a 2 part model which we now specify.

Let $S_j^1 = \{i|Z_{ij} = 1\}$ denote the set of observations $1 \leq i \leq n$ at time j with $Z_{ij} = 1$ and let $S_j^0 = \{i|Z_{ij} = 0\}$ denote the set of observations at time j with $Z_{ij} = 0$. For $i \in S_j^1$,

$$Y_{ij}|\theta_{ij} \sim f(Y_{ij}|\theta_{ij}), \tag{3.9}$$

$$\theta_{ij} \equiv \phi_{jc_{ij}}, \tag{3.10}$$

$$c_{ij}|Z_{ij} = c_{i(j-1)}, \tag{3.11}$$

$$\phi_{jc_{ij}}|\theta_{i(j-1)}, c_{ij} \sim G_{0jc_{ij}}(\theta_{i(j-1)}), \tag{3.12}$$

where cluster label $c_{ij} = c_{i(j-1)}$ because $Z_{ij} = 1$ and $G_{0jc_{ij}}(\theta_{i(j-1)})$ is the cluster time specific distribution parameterized by the value $\theta_{i(j-1)}$. This implies that when $Z_{ij} = 1$, subject i stays in the same cluster as the previous time point and the cluster specific value is allowed to vary as a function of the cluster specific value of that cluster at the previous time point. A total of $M_j = \sum_{i=1}^{n} Z_{ij}$ subjects stay in the same cluster as the previous time point in the first part of the model which provides some structure for the remaining observations.

The remaining observations $i \in S_j^0$ are now modeled as a DPM that has already observed M_j observations. Conditioned on $\boldsymbol{\theta}_j^1 = (\theta_{ij})$ for $i \in S_j^1$, observations $i \in S_j^0$ are modeled as

$$Y_{ij}|\theta_{ij} \sim f(Y_{ij}|\theta_{ij}), \tag{3.13}$$

$$\theta_{ij}|G_j \sim G_j, \tag{3.14}$$

$$G_j|\boldsymbol{\theta}_j^1 \sim \text{DP}\left(\alpha + M_j, \frac{\alpha G_{0j} + \sum_{i \in S_j^1} \delta_{\theta_{ij}}}{\alpha + M_j}\right), \tag{3.15}$$

where G_{0j} is a time specific base distribution for time j.

We define an abbreviated notation for equations (3.7), (3.8), (3.10), (3.11), (3.12), (3.14), and (3.15) for $1 \leq i \leq n$ and $2 \leq j \leq J$ as

$$\theta_{ij}|G_j \sim G_j, \tag{3.16}$$

$$G_j \sim \text{cmDP}(\alpha, G_{0jc}, \mathbf{c}_{j-1}, \boldsymbol{\theta}_{j-1}, \rho), \tag{3.17}$$

where $\mathbf{c}_j = (c_{1j}, \ldots, c_{nj})$ and cluster time specific base distributions G_{0jc} are

$$G_{0jc} = \begin{cases} G_{0jc}(\phi_{(j-1)c}), & \text{if } c \in \mathbf{c}_{j-1}, \\ G_{0j}, & \text{otherwise.} \end{cases} \tag{3.18}$$

Parameterizing the cluster specific base distribution G_{0jc} for clusters carried over from the previous time point and new clusters formed at the current time point differently in our model reflects the belief that additional information is gained when we know where the cluster had previously been. When a new cluster is formed, the cluster specific value is drawn from the entire range of possible cluster specific values distributed as G_{0j}. However, a carried over cluster c already contains a cluster specific value $\phi_{(j-1)c}$ from the previous time point $j-1$ which helps us identify where the value at the current time point might be. Therefore, $G_{0jc}(\phi_{(j-1)c})$ will generally be a distribution with much smaller variance than G_{0j}. For example, both G_{0j} and $G_{0jc}(\phi_{(j-1)c})$ may be Gaussian distributions but cluster time specific distribution $G_{0jc}(\phi_{(j-1)c})$ would be informatively centered with mean $\phi_{(j-1)c}$ and a much smaller variance.

3.3.1.2 Finite Mixture

The cmDPM model can also be presented by extending the finite mixture model (3.3). In the finite mixture model, the likelihood of observations $Y_{ij}|\phi_{jc_{ij}}$ is denoted as $Y_{ij}|\phi_{jc_{ij}} \sim f(Y_{ij}|\phi_{jc_{ij}})$ where cluster labels c_{ij} define the mixture components. The cluster labels of all the observations at time j, $\mathbf{c}_j = (c_{1j}, \ldots, c_{nj})$, map each subject to the appropriate mixture component. Vector $\mathbf{c}_j$ defines the cluster structure at time j and has K_j unique values. At time point $j = 1$ when no clustering history is available for the subjects, the standard DPM model is used. For time points $j \geq 2$, specification of the cluster structure $\mathbf{c}_j$ depends on the cluster structure at the previous time point $\mathbf{c}_{j-1}$. Observations retain the same cluster label value as the previous time point with probability cluster stickiness parameter ρ. Otherwise, they are assigned a cluster label value drawn from a discrete distribution determined by time specific mixing parameter $\boldsymbol{\pi}_j$. As the number of components goes to infinity, we show that

cluster assignment of the later scenario behaves the same as a DPM updated with the M_j observations retained from the previous time point. At the end of this section, we marginalize out Z_{ij} to show how cluster stickiness parameter ρ changes the conditional cluster assignment probabilities from those in the standard DPM model.

For observation i at time point $j = 1$,

$$c_{i1}|\boldsymbol{\pi}_1 \sim \text{Discrete}(\pi_{11}, \ldots, \pi_{1K_1}), \tag{3.19}$$

$$\phi_{1c} \sim G_{01}, \tag{3.20}$$

$$\boldsymbol{\pi}_1|\alpha \sim \text{Dirichlet}\left(\frac{\alpha}{K_1}, \ldots, \frac{\alpha}{K_1}\right), \tag{3.21}$$

where c is the value of c_{i1} drawn from (3.19). This specifies the model for the initial cluster structure $\mathbf{c}_1$. At times $2 \leq j \leq J$,

$$c_{ij}|Z_{ij}, c_{i(j-1)}, \boldsymbol{\pi}_j \sim \begin{cases} \delta(c_{i(j-1)}), & \text{if } Z_{ij} = 1, \\ \text{Discrete}(\pi_{j1}, \ldots, \pi_{jK_j}), & \text{if } Z_{ij} = 0, \end{cases} \tag{3.22}$$

$$\phi_{jc} \sim G_{0jc}, \tag{3.23}$$

$$Z_{ij}|\rho \sim \text{Bernoulli}(\rho), \tag{3.24}$$

$$\rho \sim \text{Beta}(a_\rho, b_\rho) \tag{3.25}$$

$$\boldsymbol{\pi}_j|\alpha, \mathbf{c}_j^1 \sim \text{Dirichlet}\left(n_{j1}^1 + \frac{\alpha}{K_j}, \ldots, n_{jK_j}^1 + \frac{\alpha}{K_j}\right). \tag{3.26}$$

The time specific mixing parameter $\boldsymbol{\pi}_j$ has a non-symmetric Dirichlet prior (3.26), given α and the cluster structure $\mathbf{c}_j^1$ where $\mathbf{c}_j^1 = (c_{ij})$ for all $i \in S_j^1$ is the vector of all c_{ij} with $Z_{ij} = 1$. Let $|S|$ be the number of elements in set S. The values $n_{jc}^1 = |\{c_{ij}|(Z_{ij} = 1 \text{ and } c_{ij} = c)\}|$ are the number of c_{ij} in $\mathbf{c}_j^1$ equal to c.

The joint conditional density of $\mathbf{c}_j$ is

$$\begin{aligned} p(\mathbf{c}_j|\alpha, \mathbf{Z}_j, \mathbf{c}_{j-1}) &= p(\mathbf{c}_j^1|\mathbf{Z}_j, \mathbf{c}_{j-1}) p(\mathbf{c}_j^0|\mathbf{c}_j^1, \alpha, \mathbf{Z}_j) \\ &= \left[\prod_{i \in S_j^1} \delta_{c_{ij}}(c_{i(j-1)}) \right] p(\mathbf{c}_j^0|\mathbf{c}_j^1, \alpha, \mathbf{Z}_j). \end{aligned} \tag{3.27}$$

where $\mathbf{c}_j^0 = (c_{ij})$ for all $i \in S_j^0$ and $\delta_x(c)$ is the point mass of random variable x with value of 1 at $x = c$ and 0 everywhere else. The joint conditional density $p(\mathbf{c}_j^0|\mathbf{c}_j^1, \alpha, \mathbf{Z}_j)$ can also be specified as the product of individual conditional densities

$$p(\mathbf{c}_j^0|\mathbf{c}_j^1, \alpha, \mathbf{Z}_j) = \prod_{i \in S_j^0} p(c_{ij}|\mathbf{c}_j^1, \mathbf{c}_{<ij}^0, \alpha, \mathbf{Z}_j) \tag{3.28}$$

where $\mathbf{c}_{<ij}^0 = \{c_{rj}|(r < i \text{ and } r \in S_j^0)\}$ is the set of $\mathbf{c}_j^0$ with all observations $r < i$.

The conditional probability mass function of $c_{ij} \in \mathbf{c}_j^0$ at $c_{ij} = c$ is

$$\begin{aligned} &P(c_{ij} = c|\mathbf{c}_j^1, \mathbf{c}_{<ij}^0, \alpha, \mathbf{Z}_j) \\ &= \frac{P(\mathbf{c}_{<ij}^0, c_{ij} = c|n_{j1}^1, \ldots, n_{jK_j}^1)}{P(\mathbf{c}_{<ij}^0|n_{j1}^1, \ldots, n_{jK_j}^1)} \\ &= \frac{\displaystyle\int \left(\prod_{c_{rj} \in \mathbf{c}_{<ij}^0} \pi_{jc_{rj}} \right) \pi_{jc} \frac{\Gamma(\alpha + \sum_{k=1}^{K_j} n_{jk}^1)}{\prod_{k=1}^{K_j} \Gamma(n_{jk}^1 + \alpha/K_j)} \pi_{j1}^{n_{j1}^1 + (\alpha/K_j) - 1} \ldots \pi_{jK_j}^{n_{jK_j}^1 + (\alpha/K_j) - 1} d\boldsymbol{\pi}_j}{\displaystyle\int \left(\prod_{c_{rj} \in \mathbf{c}_{<ij}^0} \pi_{jc_{rj}} \right) \frac{\Gamma(\alpha + \sum_{k=1}^{K_j} n_{jk}^1)}{\prod_{k=1}^{K_j} \Gamma(n_{jk}^1 + \alpha/K_j)} \pi_{j1}^{n_{j1}^1 + (\alpha/K_j) - 1} \ldots \pi_{jK_j}^{n_{jK_j}^1 + (\alpha/K_j) - 1} d\boldsymbol{\pi}_j} \\ &= \frac{n_{jc}^1 + n_{<ijc}^0 + (\alpha/K_j)}{\alpha + M_j + i^1 - 1} \end{aligned} \tag{3.29}$$

where $n_{<ijc}^0 = |\{c_{rj}|(c_{rj} = c \text{ and } c_{rj} \in \mathbf{c}_j^0)\}|$ is the number of $c_{rj} = c$ for all $c_{rj} \in \mathbf{c}_{<ij}^0$ and $i^1 = |\mathbf{c}_{<ij}^0|$ is the number of observations $k < i$ that are in S_j^0, the cardinality of $\mathbf{c}_{<ij}^0$. Let $n_{<ijc} = n_{jc}^1 + n_{<ijc}^0$ be the total number of observations r with either $Z_{rj} = 1$ or with $r < i$ in cluster c. We can rewrite equation (3.29) as

$$P(c_{ij} = c|\mathbf{c}_j^1, \mathbf{c}_{<ij}^0, \alpha, \mathbf{Z}_j) = \frac{n_{<ijc} + (\alpha/K_j)}{\alpha + M_j + i^1 - 1}. \tag{3.30}$$

Allow K_j to go to infinity; the conditional probability that cluster label $c_{ij} \in \mathbf{c}_j^0$ takes on the value of an existing cluster c then approaches

$$P(c_{ij} = c|\mathbf{c}_j^1, \mathbf{c}_{<ij}^0, \alpha, \mathbf{Z}_j) \to \frac{n_{<ijc}}{\alpha + M_j + i^1 - 1}, \tag{3.31}$$

while the conditional probability that c_{ij} takes on a new cluster approaches

$$P(c_{ij} = c|\mathbf{c}_j^1, \mathbf{c}_{<ij}^0, \alpha, \mathbf{Z}_j) \to \frac{\alpha}{\alpha + M_j + i^1 - 1}, \tag{3.32}$$

which is the standard DPM model after having first observed all M_j observations for $i \in S_j^1$. This implies that for $2 \leq j \leq J$, conditional on cluster memory indicators $\mathbf{Z}_j$, observations in the model either belong to an assigned cluster defined by previous cluster membership as in (3.9), (3.10), and (3.11) or they belong in a DPM model with observations carried over from the previous time point as in (3.13), (3.14), and (3.15).

The conditional probabilities for $c_{ij} \in S_j^0$ in (3.31) and (3.32) can also be written conditional on all other cluster labels by treating observation i as the n^{th} observation.

Marginalizing out Z_{ij} from the conditional probabilities of c_{ij} shows that the cluster stickiness parameter ρ puts additional weight on the probability of keeping the same cluster label as the previous time point $c_{i(j-1)}$ and gives insight on what happens to the model at the extreme conditions of $\rho = 0$ and $\rho = 1$. First, we define some needed notation. Let $\mathbf{c}_{-ij}$ denote the vector of parameters $\mathbf{c}_j$ without observation i and let $\mathbf{c}_{-(ij)}$ denote the vector of parameters $(\mathbf{c}_1, \ldots, \mathbf{c}_{-ij}, \ldots, \mathbf{c}_J)$. Further let n_{-ijc} be the total number of observations in cluster c at time j other than observation i, n_{-ijc}^1 be the total number of observations in cluster c that were carried over from the previous time point excluding observation i, and $M_{-ij} = \sum_{r=1}^n Z_{rj} - Z_{ij}$ be the total number of observations carried over from the previous time point excluding observation i. We can then specify the conditional probabilities of

$p(c_{ij}|\mathbf{c}_{-ij}, \rho, \alpha)$ marginalizing out Z_{ij} as 3 mutually exclusive and exhaustive cases

$$P(c_{ij} = c|\mathbf{c}_{-ij}, \mathbf{c}_{j-1}, \rho, \alpha) = \begin{cases} b\left[(1-\rho)n_{-ijc} + \rho(\alpha + M_{-ij})\frac{n_{-ijc}}{n^1_{-ijc}}\mathrm{I}(c = c_{i(j-1)})\right], \\ b\rho(\alpha + M_{-ij}), \\ b(1-\rho)\alpha, \end{cases} \tag{3.33}$$

with corresponding values of c in each case defined as

$$\begin{cases} c = c_{rj} \text{ for some } r \neq i, \\ [c = c_{i(j-1)}] \text{ and } [c \neq c_{rj} \text{ for every } (c_{rj} \in \mathbf{c}^0_j, r \neq i)], \\ c \neq \text{any value in } \mathbf{c}_{-(ij)}, \end{cases} \tag{3.34}$$

where b in (3.33) is the normalizing constant

$$b^{-1} = \sum_c P(c_{ij} = c|\mathbf{c}_{-ij}, \mathbf{c}_{j-1}, \rho, \alpha). \tag{3.35}$$

The standard DPM model conditional probabilities for cluster label c_{ij} are down weighted by a factor of $1 - \rho$ and the cluster with $c = c_{i(j-1)}$ is given additional weight to account for retained memory. Cluster $c = c_{i(j-1)}$ is given weight $b\rho(\alpha + M_{-ij})$ if not in $\mathbf{c}_{-ij}$ and given additional weight $b\rho(\alpha + M_{-ij})n_{-ijc}/n^1_{-ijc}$ if already in $\mathbf{c}_{-ij}$. When $\rho = 0$, the effect of the cluster structure at the previous time point $\mathbf{c}_{i(j-1)}$ becomes null and the model is equivalent to the standard Dirichlet model applied independently to each time point. When $\rho = 1$, all observations maintain their previous cluster structure and the model is equivalent to a single standard Dirichlet model applied to the longitudinal vector of outcomes together. The introduction of cluster stickiness parameter ρ maintains flexibility and presents a model with a clustering structure somewhere between the 2 extremes.

3.3.1.3 Markovian Chinese Restaurant Process

An alternative way to visualize the cluster memory Dirichlet process $G \sim \text{cmDP}(\alpha, G_{0c}, \mathbf{c}_1, \boldsymbol{\theta}_1, \rho)$ is by examining the conditional distributions of $Y_i|Y_1, \ldots, Y_{(i-1)}$ through using the Markovian Chinese restaurant process (MCRP). The MCRP is a novel variant of the Chinese restaurant process (CRP) (Aldous 1985) used to visualize the Dirichlet process that allows partial retention of the cluster structure. We first describe the CRP then present the modified MRCP. The CRP in simplest terms describes a distribution on clusters. In the CRP metaphor, the Chinese restaurant has infinitely many tables where each table can seat infinitely many customers. The tables are ordered and customers coming in can sit at any occupied table or the next unoccupied table. Specifically, customer i sits at an already occupied table k with probability $n_k/(\alpha+i-1)$ where n_k is the number of customers already at table k or sits at the next unoccupied table with probability $\alpha/(\alpha+i-1)$. Each table is then served a dish ϕ_k taken from a distribution G_0.

The MCRP starts with an existing seating arrangement $\mathbf{c}_1$ describing a previous cluster structure and a set of corresponding observation specific values $\boldsymbol{\theta}_1$ for observations $1, \ldots, n$. This existing seating arrangement could be obtained through an exchangeable process like the CRP, another MCRP, or be simply assigned based on prior belief. The customers represent subjects in longitudinal data and the tables represent clusters. All customers seated at the same table as customer i are served a dish $\phi_{1c_{i1}} = \theta_{i1}$. After tasting the dish, customers leave the table with probability ρ with a total of M customers leaving their tables to return to the front of the restaurant. A partially retained cluster structure remains from those customers who did not leave. If a table is left empty, that table is removed. The customers who left their tables and returned to the front of the restaurant then get seated sequentially at an already occupied table k with probability $n_k/(\alpha + M + i^1 - 1)$ or the next unoccupied table with probability $\alpha/(\alpha+M+i^1-1)$ where $M+i^1-1$ is the total number of customers already seated. This portion of the seating process is the same as a CRP that had already seated $n - M$ customers exactly the same way as they were seated in the previous seating arrangement. The seating arrangement of customers are therefore dependent on the previously existing seating

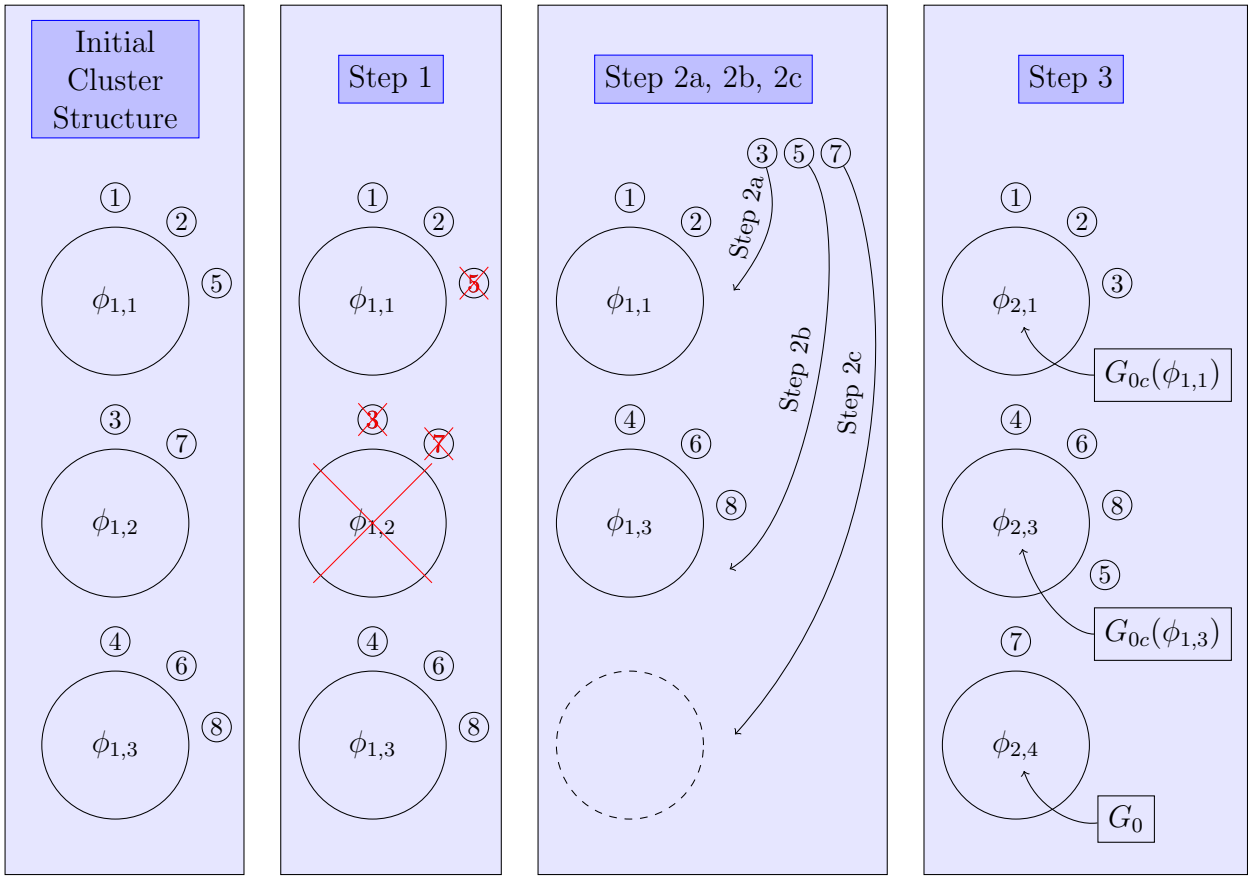

Figure 3.2: Graphical summary of the Markovian Chinese restaurant process. We start with a predefined seating arrangement for n customers. In step 1, each customer is removed with probability ρ. If an entire table is left empty, the table is removed as well. In step 2, removed customers are assigned seating with probability defined in the CRP. In step 3, each table is served a dish $\phi_{2,c}$. For each table c, if the table existed in the original seating arrangement, customers are served a similar dish from $G_{0c}(\phi_{1,c})$. If the table is new, customers are served a new dish from a non-specific base distribution G_0 instead.

arrangement $\mathbf{c}_1$ as inferred by the name Markovian Chinese restaurant process. Each table c is then served dishes $\phi_{2,c}$ depending on whether that table existed in the original seating arrangement. Tables that existed in the original seating arrangement are served a dish similar to their original dish from a table specific distribution $G_{0c}(\phi_{1c})$. Tables that did not exist are served a dish from a non-specific base distribution G_0. In this manner, each time point $j > 1$ after the initial time point $j = 1$ of our cmDPM model is a Markovian Chinese restaurant process with the cluster structure based on the previous time point's cluster structure. A graphical summary of the MCRP is shown in Figure 3.2.

3.3.2 Hyperpriors

The presented model can be extended with hyperpriors to incorporate uncertainty about some parameters that were considered constants in the previous section.

Rather than a fixed strength parameter α, a hyperprior for α can be specified to incorporate uncertainty. The posterior of strength parameter α is only dependent on the number of new unique clusters at each time point. We define K_j^0 as the number of new values of cluster labels c_{ij} for $1 \leq i \leq n$ that appears in $\mathbf{c}_j$ but did not previously appear at any time $l < j$ and the number of retained cluster labels is M_j where at the initial time point $j = 1$, we have $M_1 = 0$. The posterior of α given K_j^0 across all time points $1 \leq j \leq J$ is

$$\begin{aligned} f(K_1^0, \ldots, K_J^0|\alpha) &\propto \prod_{j=1}^{J} \alpha^{K_j^0} \frac{\Gamma(\alpha + M_j)}{\Gamma(\alpha + n)}, \\ f(\alpha|K_1^0, \ldots, K_J^0) &\propto f(\alpha) f(K_1^0, \ldots, K_J^0|\alpha). \end{aligned} \tag{3.36}$$

A reasonable hyperprior for α could be based on prior belief of the number of unique clusters across all time points. We use an extension of Antoniak (1974) to establish an approximate relationship between α and the total number of unique clusters. Let W_{ij} be 1 when a new cluster c_{ij} is chosen for subject i at time j and 0 otherwise. Then from (3.32) in the cmDPM model,

$$\begin{aligned} Pr(W_{ij} = 1) &= \frac{\alpha}{\alpha + M_j + i^1 - 1}, && \text{for } i \in S_j^0, \\ &= 0, && \text{for } i \in S_j^1. \end{aligned} \tag{3.37}$$

Let $Z_n = \sum_{j=1}^{J} \sum_{i=1}^{n} W_{ij}$ be the total number of unique clusters over all time points. Then the expected number of unique clusters across all time points can be approximated as

$$E(Z_n|M_2, \ldots M_J, \alpha) = \sum_{j=1}^{J} \sum_{m=M_j+1}^{n} \frac{\alpha}{\alpha + m - 1} \approx \alpha \sum_{j=1}^{J} \log\left(\frac{n + \alpha}{\alpha + M_j}\right) \tag{3.38}$$

since there are $n - M_j$ observations total in S_j^0 at each time point j. A prior distribution

for α can be established by estimating some reasonable number of unique clusters and using (3.38) to approximate reasonable α values. To faciliate computation of the expected number of clusters $E(Z_n)$, we can estimate a possibly subjective value for ρ and substitute $n\rho$ in place of M_j.

The cluster time specific base distributions G_{0jc} is parameterized with additional cluster time specific base distribution parameters $\boldsymbol{\eta}_j = (\eta_j^1, \eta_{j0})$ such that

$$G_{0jc} = \begin{cases} G_{0jc}(\phi_{(j-1)c}, \eta_j^1), & \text{if } c \in \mathbf{c}_{j-1}, \\ G_{0j}(\eta_{j0}), & \text{otherwise,} \end{cases} \tag{3.39}$$

where η_j^1 parameterizes G_{0jc} for clusters $c \in \mathbf{c}_{j-1}$ carried over from the previous time point and η_{j0} parameterizes G_{0jc} for new clusters $c \notin \mathbf{c}_{j-1}$. Hyperpriors on η_j^1 and η_{j0} further incorporate uncertainty regarding the base distribution G_{0jc}.

Similarly, the conditional distribution describing our outcome Y_{ij} can also include additional likelihood distribution parameter λ so that $Y_{ij}|\theta_{ij}, \lambda \sim f(Y_{ij}|\theta_{ij}, \lambda)$. A hyper prior on λ incorporates uncertainty regarding this distribution.

3.3.3 Posterior Computation

Our posterior computations are sampled with Markov Chain Monte Carlo (MCMC) methods using the Gibbs sampler and the Metropolis algorithm. Neal (2000) provides a basis for sampling the standard DPM model which we extend. In theory, jointly sampling cluster memory indicator Z_{ij} and cluster label c_{ij} for each subject $1 \leq i \leq n$ at each time $1 \leq j \leq J$ and sampling a cluster specific value ϕ_{jc} only when cluster c first appears at time j is sufficient for sampling each cluster and the values of each cluster. In practice, this results in an extremely slow convergence rate and we implement two additional sampling steps to speed convergence. The other parameters in the model are sampled one at a time from their full conditional distributions.

The full cmDPM model is defined by strength parameter α, cluster stickiness parameter

ρ, additional likelihood distribution parameter λ, vectors of cluster memory indicators $\mathbf{Z}_j$ for $j = 2, \ldots J$, vectors of subject specific parameters $\boldsymbol{\theta}_j$, vectors of cluster labels $\mathbf{c}_j$, and cluster time specific base distribution parameters $\boldsymbol{\eta}_j$ for $j = 1, \ldots, J$. We use parameters $\phi_{jc_{ij}}$ and θ_{ij} interchangeably based on the equivalence relationship $\phi_{jc_{ij}} = \theta_{ij}$. To speed convergence,we first determine at each iteration sets of cluster labels for each subject that will take on the same value staying in the same cluster over several time points conditioned on the cluster memory indicator at those time points. These sets of *bound* cluster labels which we define more rigorously below are sampled jointly. Second, we resample cluster specific values ϕ_{jc} for each cluster c at each time point j.

When conditionally conjugate priors are used, conditional posteriors of each are sampled in turn using a Gibbs sampling step. We present a brief summary of the posterior computation algorithm here. A detailed description of the computation algorithm is included in the Appendix.

Let $G(\phi; \eta)$ represents the density of the distribution G at ϕ parameterized by η.

1. For $1 \leq i \leq n$ and $j = 1$, sample cluster label $c_{i1}|\alpha, \mathbf{c}_{-i1}, Z_{i2}, Y_{i1}, \lambda$. When a new cluster c is formed with observation Y_{i1}, a value of ϕ_{1c} is sampled from $f(\phi_{1c}|Y_{i1}, \eta_{10}, \lambda) \propto f(Y_{i1}|\phi_{1c}, \lambda)G_{01}(\phi_{1c}; \eta_{10})$.

2. For $1 \leq i \leq n$ and $2 \leq j \leq J$ jointly sample $(c_{ij}, Z_{ij})|\alpha, \rho, \mathbf{c}_{-(ij)}, Y_{ij}, \lambda$. When a new cluster c is formed with observation Y_{ij}, a value of ϕ_{jc} is sampled from $f(\phi_{jc}|Y_{ij}, \eta_{j0}, \lambda) \propto f(Y_{ij}|\phi_{jc}, \lambda)G_{0j}(\phi_{jc}; \eta_{j0})$. When a cluster c is retained from the previous time point $j-1$ for observation Y_{ij} with no current cluster specific value ϕ_{jc} at time point j, a new ϕ_{jc} is sampled from $f(\phi_{jc}|Y_{ij}, \phi_{(j-1)c}, \eta_j^1, \lambda) \propto f(Y_{ij}|\phi_{jc}, \lambda)G_{0jc}(\phi_{jc}; \phi_{(j-1)c}, \eta_j^1)$.

3. When cluster stickiness parameter ρ is high, cluster labels can mix very slowly since multiple cluster labels across time are set equal. We introduce a cluster mixing step across time for each observation $1 \leq i \leq n$ to speed up convergence. Given $\mathbf{Z}_i$, define a set of consecutive cluster labels $(c_{ip}, \ldots, c_{iq})$ as *bound* if (i) $Z_{i(p+1)} = Z_{i(p+2)} = \cdots = Z_{iq} = 1$, (ii) $p = 1$ or $Z_{ip} = 0$, and (iii) $q = J$ or $Z_{i(q+1)} = 0$. These *bound* cluster labels (i) are all equal to each other because the value is retained from the previous

time point, (ii) begin either at time point $j = 1$ or at a time p where the cluster label was not carried over from time $p-1$ and the cluster labels at these two times are therefore not guaranteed to be the same, and (iii) ends at either time point $j = J$ or at a time q where the cluster label was not carried over to time point $q+1$ and the cluster labels at these two times are therefore not guaranteed to be the same. From our model, this implies $c_{ip} = c_{i(p+1)} = \cdots = c_{iq}$ and $q - p$ is as large as possible with as many observations in each bound set as possible. Each observation i has a total of $J - \sum_{j=2}^{J} Z_{ij}$ *bound* clusters. We jointly sample sets of *bound* cluster labels $(c_{ip}, \ldots, c_{iq})|\alpha, \mathbf{Z}_i, \mathbf{c}_{-ip}, \ldots, \mathbf{c}_{-iq}, Y_{ip}, \ldots, Y_{iq}$ for each subject $1 \leq i \leq n$. If a new cluster c is formed at any time point $p \leq j \leq q$, sample a new ϕ_{jc}.

4. When clusters do not disappear very often, the cluster specific values will mix very slowly if they are not resampled. A remixing step for cluster specific values $\boldsymbol{\phi}_{jc}$ significantly speeds up convergence. This is used in the standard DPM model as well. Sample ϕ_{jc} from $f(\phi_{jc}|\mathbf{Y}_j, \boldsymbol{\eta}_j, \lambda) \propto \prod_{i \text{ s.t. } c_i=c} f(Y_{ij}|\phi_{jc}, \lambda) G_{0jc}(\phi_{jc}; \phi_{(j-1),c}, \boldsymbol{\eta}_j)$ $G_{0jc}(\phi_{(j+1)c}; \phi_{jc}, \boldsymbol{\eta}_j)$. Repeat for $1 \leq j \leq J$ and all unique values of ϕ_{jc}.

5. Sample ρ from $f(\rho|\mathbf{Z}_2, \ldots, \mathbf{Z}_J)$.

6. Sample additional hyperparameters α, $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_J$, and λ.

3.4 Applications

In this section, we apply the cmDPM model to data on annual tuberculosis (TB) incidence rates around the world over the last 21 years. The cmDPM model clusters countries that behave similarly together each year. The clustering takes into account their currently observed incidence rates and where they were clustered previously. We summarize the change in world TB incidence rates over the last 21 years as modeled by the cmDPM model showing how the overall density and clustering structures have evolved over time and we make individual country predictions for 2011, the following year. A comparison to the standard time independent DPM model is also given.

3.4.1 Tuberculosis incidence by country from 1990-2010

Tuberculosis (TB) is a common infectious disease that in humans typically result in an asymptomatic, latent infection. However, those co-infected with HIV are at substantially increased risk of developing active disease compared to those who are not. Gibson et al. (2008) states that among people who are infected with tuberculosis, roughly 30% of people who are HIV^+ go on to develop active disease during their lifetime as opposed to only 5-10% of those who are HIV^-. Glynn (1998) compares the relative risk of tuberculosis between individuals infected with HIV compared to those not infected in a number of studies and while these studies show large variation in estimates of relative risk, it is consistent that those infected with HIV have greater risk. Due in part to the impact of HIV on tuberculosis at the population level, the World Health Organization (WHO) declared TB to be "a global health emergency" in 1993.

We examine annual tuberculosis incidence per 100k population across 197 countries in the world over the last 21 years from 1990-2010. The dataset is available publicly at `http://data.worldbank.org`. Let X_{ij} be the annual tuberculosis incidence rate per 100k for country i at time j for $1 \leq i \leq 197$ and $1 \leq j \leq 21$ and let $Y_{ij} = \log(X_{ij} + 1)$ be the corresponding annual log tuberculosis incidence rate.

A profile plot of the annual tuberculosis incidence rate data across all 21 years is shown in Figure 3.3. Annual tuberculosis incidence rates for each country are connected by a single line. Most countries show a slow but steady drop in tuberculosis incidence over the last 2 decades. It is possible to visually identify clusters of countries that behave similarly over time as some countries report tuberculosis incidence rates that are very close to each other year after year. A few countries also show mobility between clusters moving over the years to join different clusters of countries. The overall level of *cluster mobility* is of interest when studying tuberculosis incidence on a global scale as it is an indication of the cluster predictability across time. Low levels of *cluster mobility* indicates that most countries do not vary behavior from their peers over time and cluster membership tends to stay the same. Longitudinal trends of entire clusters tend to be a result of global changes. Identifying specific

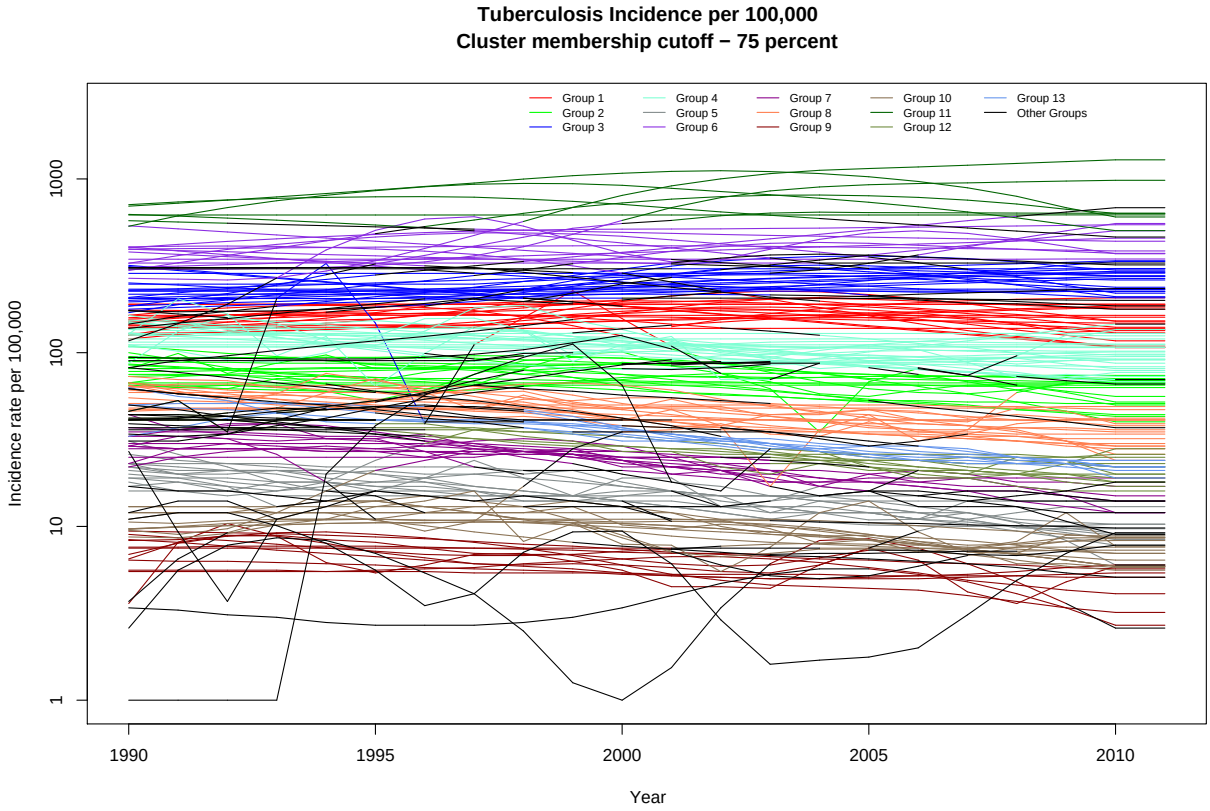

Figure 3.3: Longitudinal data of tuberculosis incidence per 100k population for 197 countries in the world. Each line plots the reported tuberculosis incidence rates for a single country over 21 years. Posterior clustering of countries based on the cmDPM model is shown. Countries i and i' are considered in the same cluster at a given time j if the cluster labels c_{ij} and $c_{i'j}$ are equivalent at least 75% of the time.

countries that do change cluster memberships often can be important in understanding how large changes in TB incidence can occur.

3.4.2 The cmDPM model for tuberculosis incidence

We model log tuberculosis incidence as

$$\begin{aligned}
Y_{ij}|\theta_{ij},\sigma^2 &\sim \mathrm{N}(\theta_{ij},\sigma^2),\\
\theta_{ij} &\sim G_j,\\
G_1 &\sim \mathrm{DP}(\alpha,\mathrm{N}(\mu_G,\sigma_G^2)),\\
G_j &\sim \mathrm{cmDP}(\alpha,G_{0jc},\mathbf{c}_{j-1},\rho) \qquad \text{for } j\geq 2,\\
\rho|a_\rho,b_\rho &\sim \mathrm{Beta}(a_\rho,b_\rho)\\
G_{0jc} &= \begin{cases} \mathrm{N}(\phi_{(j-1)c},\sigma_c^2), & \text{if } c\in\mathbf{c}_{j-1},\\ \mathrm{N}(\mu_G,\sigma_G^2), & \text{otherwise,}\end{cases}
\end{aligned} \tag{3.40}$$

with hyperpriors

$$\begin{aligned}
\sigma^2|a_{\sigma^2},b_{\sigma^2} &\sim \mathrm{InvGamma}(a_{\sigma^2},b_{\sigma^2}),\\
\mu_G|\sigma_G^2,\mu_G,v &\sim \mathrm{N}(\mu_0,\sigma_G^2/v),\\
\sigma_G^2|a_{\sigma_G^2},b_{\sigma_G^2} &\sim \mathrm{InvGamma}(a_{\sigma_G^2},b_{\sigma_G^2}),\\
\sigma_c^2|a_{\sigma_c^2},b_{\sigma_c^2} &\sim \mathrm{InvGamma}(a_{\sigma_c^2},b_{\sigma_c^2}),\\
\alpha|a_\alpha,b_\alpha &\sim \mathrm{Uniform}(a_\alpha,b_\alpha)
\end{aligned} \tag{3.41}$$

where $X\sim\mathrm{InvGamma}(a,b)$ denotes a random variable X is distributed inverse-gamma with mean $b/(a-1)$ and $X\sim\mathrm{Beta}(a,b)$ denotes a random variable X drawn from a beta distribution with mean $a/(a+b)$. From Section 3.3.2, $\sigma^2\equiv\lambda$ is the additional likelihood distribution parameter; parameters $(\mu_G,\sigma_G^2)\equiv\eta_{j0}$ and $\sigma_c^2\equiv\eta_j^1$ are the cluster time specific base distribution parameters in this model.

Parameters for the hyperpriors are selected to reflect constraints that X_{ij} tuberculosis incidence per 100,000 must be between (0, 100,000) and correspondingly, Y_{ij} must be between (0, 11.5). We estimate incidences of around 55 per 100,000 to be the median and choose $\mu_0=4\approx\log(55)$. We select parameters for a_{σ^2}, b_{σ^2}, and v so that the expected standard

Table 3.1: Parameters of hyperpriors chosen in (3.40). Selection of parameters reflect prior belief that median annual tuberculosis incidence per 100k should be around 150.

Parameter Name	Parameter	Value
Hyperprior parameter 1 for σ^2	a_{σ^2}	4
Hyperprior parameter 2 for σ^2	b_{σ^2}	2
Mean of hyperprior for μ_G	μ_0	4
Inverse variance scale parameter for μ_G	v	2
Hyperprior parameter 1 for σ_G^2	$a_{\sigma_G^2}$	6
Hyperprior parameter 2 for σ_G^2	$b_{\sigma_G^2}$	10
Hyperprior parameter 1 for σ_c^2	$a_{\sigma_c^2}$	21
Hyperprior parameter 2 for σ_c^2	$b_{\sigma_c^2}$	2
Cluster stickiness hyperprior parameter 1	a_ρ	1
Cluster stickiness hyperprior parameter 2	b_ρ	1
Cluster strength hyperprior parameter 1	a_α	0.5
Cluster strength hyperprior parameter 2	b_α	10

deviation in μ_G is approximately 0.35 allowing flexible overall population average values of incidence per 100,000 with typical values ranging from 20 to 150. A priori, we believe variation of 1 log incidence rate for observations within the same cluster would not be uncommon. We chose values of a_{σ^2} and b_{σ^2} so that the mean of the prior for σ^2 was 0.66 corresponding to an expected standard deviation of approximately 0.82 around the cluster specific value for each cluster to reflect this belief. Similarly, values of $a_{\sigma_c^2}$ and $b_{\sigma_c^2}$ for the prior of σ_c^2 were chosen so that the mean of the prior was 0.1. This reflects our belief that variations of the mean value of the same cluster from year to year should mostly stay within $\sqrt{0.1} \approx 0.3$ log incidence rate from one year to the next. The upper and lower limits of strength parameter α was selected based on (3.38) and assuming a value for $\rho = 0.9$ to reflect our prior belief that total number of different clusters across all time points should be between 4 to 46. Table 3.1 provides a full list of chosen values characterizing all priors in the model.

Posterior calculations are done using the MCMC algorithm discussed in Section 3.3.3. After an initial 5,000 iterations used for burn-in, we ran the algorithm for 40,000 iterations and thin the resulting MCMC chain to save on every 4^{th} iteration. Estimation of the subject time specific variables θ_{ij} is of primary interest in this model as this directly reflects estimates of the annual log TB incidence rates for country i at time j. Trace plots for select θ_{ij} parameters are shown in Appendix B.2 providing a visual cue for convergence. Some posterior

estimates of θ_{ij} are multimodal and do jump between multiple modes reflecting uncertainty of cluster assignments for these countries.

Table 3.2 shows select results for the model. The tuberculosis incidence dataset shows low levels of cluster mobility. The parameter ρ has a posterior mean of 0.97 implying 97% of countries retained cluster membership without any reevaluation. When cluster mobility is low, information on cluster membership is borrowed over time and the distribution on the partitioning of the countries is closely related at each time point. Figure 3.3 is color coded to show posterior clustering of the TB incidence data across all 197 countries. Countries i and i' are considered to be in the same cluster in this Figure if their cluster labels c_{ij} and $c_{i'j}$ are the same a posteriori at least 75% of the time. Lines in black were used to represent clusters with fewer than 3 members. This color coding shows cluster membership with a single cutoff value. A more detailed graphical representation of the cluster structure at select time points is shown in Appendix B.3. Figure B.2 shows a heat map of how cluster memberships evolve over time. Country IDs are ordered by the value of the observed tuberculosis incidence rate in 1990. Each individual graph shows the posterior probabilities of two countries sharing the same cluster in a given year. Over time, countries tend to join neighboring clusters but not those further away.

Comparison of values between the across time cluster variance σ_c^2 and the new cluster parameter variance σ_G^2 gives us an idea of how much information is gained in determining cluster specific value ϕ_{jc} when cluster c is carried over time from $j-1$ to j. If a country is retained in the same cluster from year to year, variation in annual log incidence rate only has a standard deviation of 0.08 with 95% credible interval (0.04, 0.13) whereas a country moving into a new unknown cluster would have a standard deviation of 2.95 with a 95% credible interval of (1.70, 5.00). The combination of having a low level of cluster mobility as determined by ρ and a high level of retained information on cluster specific values as determined by a comparison between σ_c^2 and σ_G^2 implies the cmDPM model should allow for considerably better modeling of the posterior density functions than a time independent method. Further proof is provided by a comparison of the kernel density estimates of Y_{ij} in the data using a Gaussian smoothing kernel compared to the posterior predictive densities

Table 3.2: Posterior summaries of select parameters for TB data with the cmDPM model.

Parameter Name	Parameter	Posterior Mean	Credible Interval (2.5%, 97.5%)
Cluster strength parameter	α	3.48	(1.84, 5.67)
Cluster stickiness parameter	ρ	0.97	(0.96, 0.98)
New cluster parameter mean	μ_G	3.33	(2.52, 4.16)
Within cluster variance	σ^2	0.04	(0.02, 0.08)
New cluster parameter variance	σ^2_G	2.95	(1.70, 5.00)
Across time cluster variance	σ^2_c	0.08	(0.04, 0.13)

of Y_{ij} using both the cmDPM model and the time independent DPM models at select years is shown in Appendix B.4.

Figure 3.4 provides a summary of how the posterior predictive densities of annual log tuberculosis incidence rate, Y_{ij}, changes with time. Overall, Y_{ij} has only been decreasing over the past 2 decades in the countries lower than the median while the countries reporting tuberculosis incidence rates higher than the median has roughly stayed the same or even increased slightly. This implies that the countries managing the rate of TB disease well in 2010 are doing even better than the countries who managed the rate of TB disease well in 1990. However, the countries managing the rate TB disease poorly in 2010 are managing it just as badly or worse than the countries who managed the rate of TB disease poorly in 1990. Alarmingly as of 2010, we continue to have roughly the same percentage of countries reporting annual log incidence rates higher than 100 per 100k population. Furthermore, given the low cluster mobility estimated from our model, these countries are also unlikely to move out of this situation in the near future.

We can make predictions for individual countries in future years and of other unobserved countries. Posterior predictions are done by first simulating cluster memory indicator Z_{ij} for the desired observation i at time j. Cluster label c_{ij} is then simulated based on the value of Z_{ij} and all other cluster labels $\mathbf{c}_{-(ij)}$. Finally the value of ϕ_{jc} is sampled given which cluster c observation i belongs to. This is repeated to obtain a posterior predictive density for observation i at time j. Posterior predictive densities for 3 select countries are shown in Figure 3.5 as an example. The United States is slated to improve while Uganda and Argentina are slated to get a little worse if all trends continue.

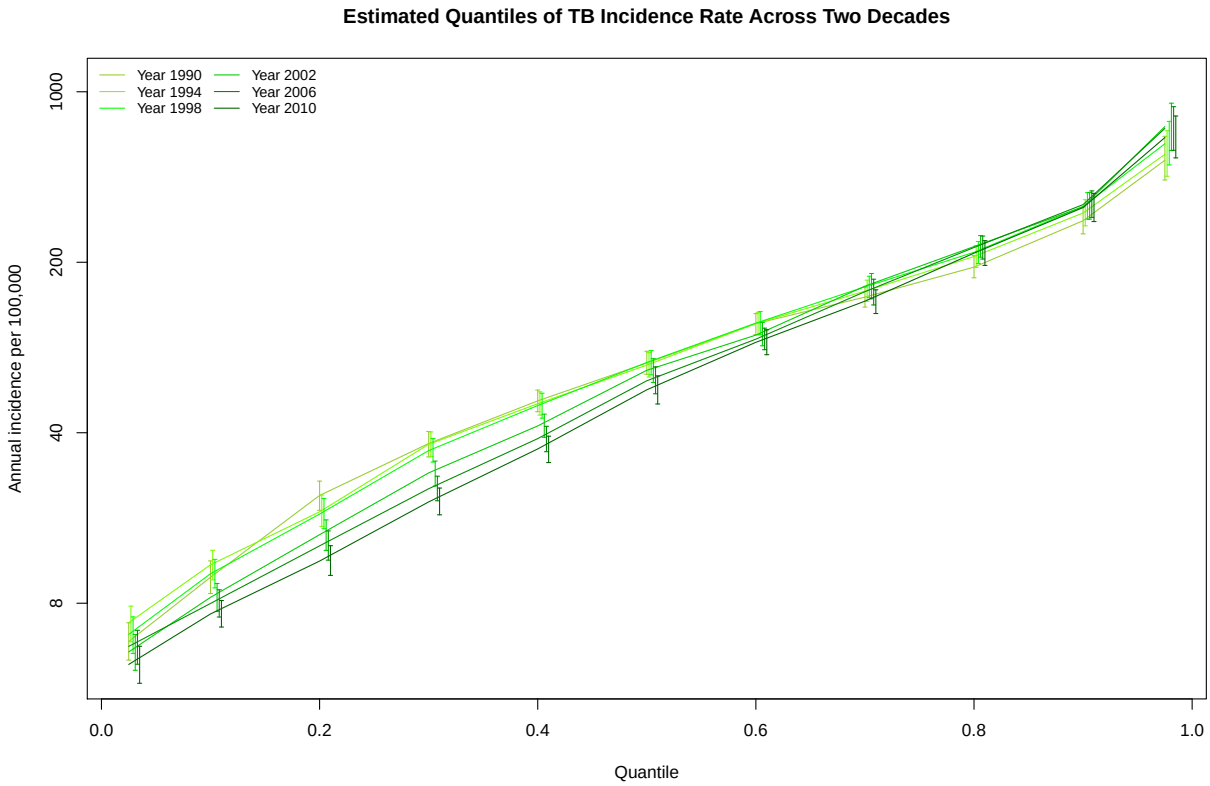

Figure 3.4: Quantiles of the posterior predicted densities of annual TB incidence rate at specific time points. Error bars shown are the 95% credible intervals at selected quantile measures. Overlapping quantile lines across different years at a specific TB incidence rate value imply that there has been no change in the total percentage of countries reporting TB incidence rates at or below that value.

3.5 Discussion

The cmDPM model is ideally suited for modeling longitudinal data in a nonparametric Bayesian framework. Our model is the first model to account for observation level clustering history. Clusters are given greater tendency to stay together while parameter values assigned to these clusters are still allowed to change at each time point. We show that density estimation can be significantly improved from the independent DPM case by borrowing clustering information across time. We also introduce the concept of *cluster mobility* in datasets. This is modeled directly in the cmDPM model as the parameter $1 - \rho$ describing how often subjects change their cluster assignments from the previous time point. This

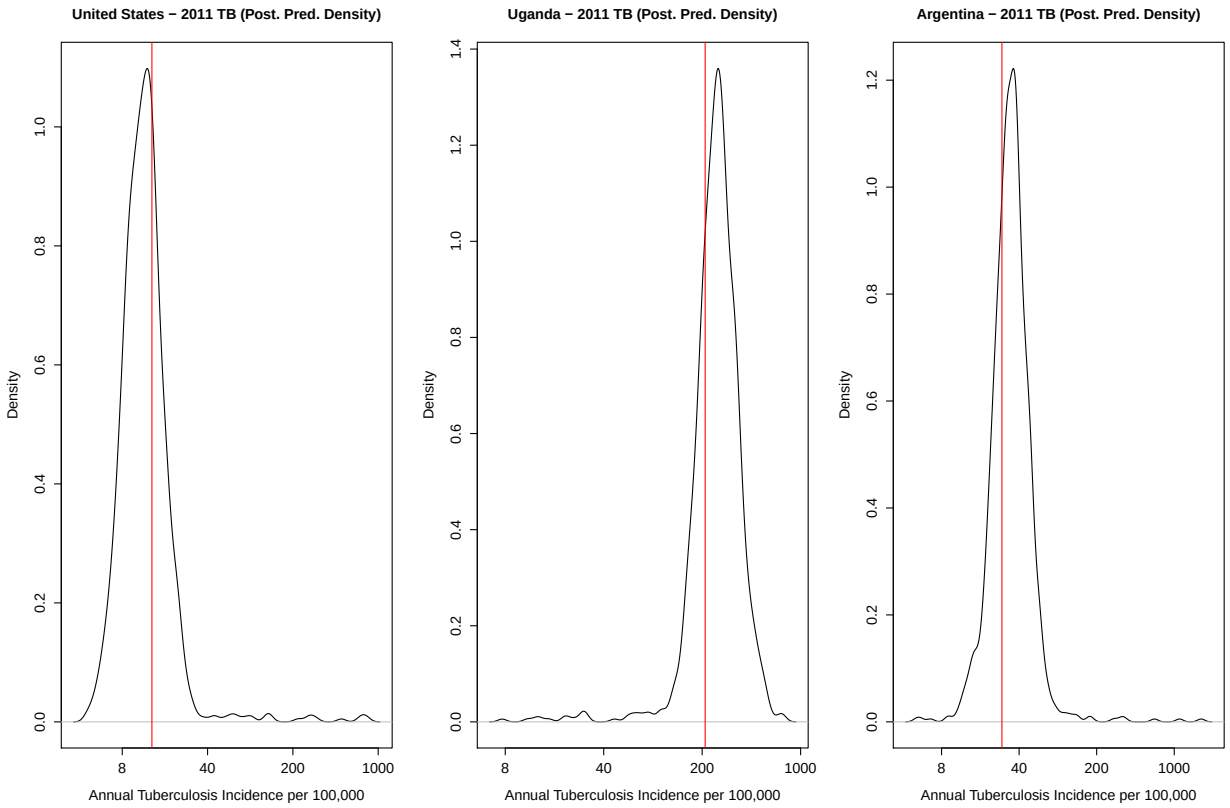

Figure 3.5: Posterior predictive densities of annual tuberculosis incidence rate per 100,000 in 2011 for three select countries (United States, Uganda, Argentina). Lines in red represent observed tuberculosis incidence rates in 2010. The tails of these distributions show small modes where our model predicts there is a possibility of the country moving to a different cluster.

parameter can be used to quantify the durability of current clusters in the data at future time points.

CHAPTER 4

Future Work

In this chapter, I discuss plans for future work. The proposed work is divided into near term and long term goals based on a mix of importance and ease of implementation.

4.1 Near term goals

In the near term, I will submit a paper on the cmDPM model with application to tuberculosis to an appropriate statistical journal. I plan to extend the cmDPM model to incorporate subject level covariates $\mathbf{x}'_{ij}\boldsymbol{\beta}$ where $\mathbf{x}_{ij}$ are the observed subject level covariates at time j and $\boldsymbol{\beta}$ are a set of fixed covariate effects with specified prior distributions. For example, in the tuberculosis incidence data, we could modify (3.40) with

$$Y_{ij}|\mathbf{x}_{ij}, \boldsymbol{\beta}, \theta_{ij}, \sigma^2 \sim \mathrm{N}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \theta_{ij}, \sigma^2), \tag{4.1}$$

$$\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\mu}_\beta, \Sigma_{\boldsymbol{\beta}}), \tag{4.2}$$

where $\boldsymbol{\mu}_\beta$ and $\Sigma_{\boldsymbol{\beta}}$ are specified parameters for the prior of $\boldsymbol{\beta}$ and specification of the remaining parameters remain the same. Covariate effects $\boldsymbol{\beta}$ are part of the set of λ parameters defined at the end of Section 3.3.2 and standard estimation techniques can be used when conditioned on all other parameters. Implementation of subject level covariates should significantly tighten prediction ranges for individual outcomes at future time points. An R package will be released to facilitate use of the cmDPM model in this form.

4.2 Long term goals

I wish to extend the cmDPM model to accommodate for a variety of likelihood densities $f(Y_{ij};\phi)$ and allow for the selection of a variety of non-conjugate cluster time specific base distributions G_{0jc}. Both these extensions significantly broaden the applications of the cmDPM model. Using different likelihood densities $f(Y_{ij};\phi)$ allows the cmDPM model to be applied to a larger variety of data types. Selection of different distributions of G_{0jc} allows us to alter how we specify the evolution of cluster specific values over time. For example, cluster specific values may have a greater probability of decreasing than increasing from one time point to the next. Cluster specific values over multiple previous time points may also also be incorporated in parameterizing G_{0jc}.

Both these extensions can be difficult with the existing estimation algorithm. Currently, the estimation algorithm for the cmDPM model to determine the probability of assigning observation i at time j to a cluster c that exists at time point $j-1$ but does not currently exist at time j relies on the value of $\int f(Y_{ij};\phi)dG_{0jc}(\phi;\phi_{(j-1)c})$ having a closed form solution. When this is not the case, the algorithm becomes significantly more cumbersome requiring numerical estimation of the integral. An efficient solution to this problem is required to make these two proposed extensions feasible.

One interesting application of the cmDPM model is to cluster groups of participants in the HLP dataset. Subjects can be cross-classified by all possible combinations of race, site, and risk group; we can then see which of these groups behave similarly across time. This can be done by assigning a group specific effect to each combination of groups present in the dataset and applying the cmDPM model to those specific effects. This will cluster groups of participants that behave similarly together yet still allow changes in cluster membership across time. Implementation of this model requires adding subject level covariates, assigning a distribution to the outcomes that suitably accounts for count data, and finding an appropriate G_{0jc} density.

The cmDPM model can also be modified with a number of other extensions that I propose from a mechanistic perspective. The cluster stickiness parameter ρ is currently constant

across all subjects. Incorporation of subject or time specific ρ_{ij} parameters based on observed covariates could provide additional information about how these covariates affect movement between clusters. In the context of data like the annual tuberculosis incidence data, this is particularly interesting as it could allow us to identify characteristics that promote changes in tuberculosis incidence.

Current assignment of clusters for each subject is Markovian and depends on the cluster location of that subject only through the previous time point. An argument could be made that subjects that have stayed in the same cluster for multiple years should be less likely to leave that cluster than a subject which has just been in a cluster for a single year. Furthermore, a subject that has just left a cluster may be more likely to return to that cluster if not too much time has passed. I can extend the cmDPM model to incorporate information on subject specific cluster membership over multiple previous years. Currently, a subject will stay in the same cluster as the previous year with probability ρ and reevaluate cluster membership with probability $1 - \rho$. One option is to assign a weighting for the previous r years such that $w_1 \geq w_2 \geq \cdots \geq w_r \geq 0$ and $\sum_{q=1}^{r} w_q = 1$. The probability of subject i being assigned to some previous cluster $c_{i(j-q)}$ at time j could then be specified as $w_q\rho$ and the probability of reevaluating cluster membership would stay as $1 - \rho$.

The current version of the cmDPM model does not account for the ordinal nature of the cluster specific parameters $\phi_{jc_{ij}}$ at time j for all unique values of c_{ij}. In a dataset like the tuberculosis incidence data, it is reasonable to assume a priori that countries may be more likely to move to an adjacent cluster where the value of the adjacent cluster is similar than to jump to a cluster that is far away. Adding this information to the model is another potential improvement that can be made.

Finally, it will be useful to extend the cmDPM model to accommodate continuous time data.

APPENDIX A

Health Living Project

A.1 Estimation benefits of modeling separate sources of variation

Our multilevel model introduces separate sources of variation for estimating the number of protected and unprotected acts. Not only do subjects behave differently but the same subject varies their behavior with different partners bringing another level of heterogeneity to observed outcomes. In the Poisson model, when the variation comes from multiple sources, some estimation benefits can be gained by correctly modeling these sources of variation. We examine the effects of failing to account for this level of heterogeneity through simulation of univariate random variables with 2 levels of heterogeneity. This is similar to our study data limited to a single outcome at a single time point but repeatedly observed for multiple partners.

Simulation data is generated assuming $Y_{ik} \sim \text{Po}(\lambda_{ik})$ for participant $i = 1, \ldots, 12$, and partner $k = 1, \ldots K_i$ where K_i is itself a zero-truncated Poisson distributed variable with parameter $\lambda_{K_i} = 10$. Mean parameters λ_{ij} follow a log linear function

$$\log \lambda_{ik} = \mu + \beta_i + \delta_{ik} \tag{A.1}$$

with Gaussian distributed latent effects $\beta_i \sim \text{N}(0, \sigma^2)$ and $\delta_{ik} \sim \text{N}(0, d^2)$. Values of μ, σ^2, d^2 are set to 1, 1.5, and 2 respectively with the resulting expected value $\text{E}(Y_{ik}) = \exp\{\mu + 0.5(\sigma^2 + d^2)\} = 15.64$.

We consider 3 separate analyses for making inference on the expected value of Y_{ik}. Analysis 1 uses the full disaggregated data and the correct 2 level heterogeneity model as presented

in (A.1). This is analogous to the ideal situation where all partner level data in the HLP study is observed. Analysis 2 also uses the correct model but assumes we only observe Y_{ik} for $k \leq 5$ and the aggregated totals Y_{iT}=$\sum_{k=1}^{K_i} Y_{ik}$. Unobserved disaggregated values $\mathbf{Y}_{i,miss} = (Y_{i6}, \ldots Y_{iK_i})$ for $K_i \geq 6$ are imputed from observed info as

$$\mathbf{Y}_{i,miss}|Y_{iT}, Y_{i1}, \ldots, Y_{i5} \sim \text{Multinomial}(N_i, \mathbf{p}_i) \tag{A.2}$$

where $N_i = Y_{iT} - \sum_{k=1}^{5} Y_{ik}$, $\mathbf{p}_i = (\lambda_{i6}/\lambda_{iT}, \ldots, \lambda_{iK_i}/\lambda_{iT})$, and $\lambda_{iT} = \sum_{k=6}^{K_i} \lambda_{ik}$. This analysis mimics the actual data from the HLP study. Analysis 3 uses only aggregated totals, $Y_{iT} \sim \text{Po}(\lambda_i)$, with mean parameters

$$\log \lambda_i = \mu + \beta_i \tag{A.3}$$

using a single level of Gaussian distributed latent effect $\beta_i \sim \text{N}(0, \sigma^2)$. This mimics the traditional analysis of total sex acts in studies like HLP. Inference in analysis 3 is made from $\text{E}(Y_{iT}) = \text{E}_{K_i}\text{E}(Y_{iT}|K_i) = \text{E}_{K_i}\text{E}(\sum_{k=1}^{K_i} Y_{ik}|K_i) = \text{E}(K_i)\text{E}(Y_{ik})$.

Table A.1 shows summary results from 600 simulated datasets. Use of aggregated totals in analysis 3 results in longer intervals, an average width of 7.8 as compared to 7.3 from analysis 2 and 7.2 from analysis 1. A bigger mean square error was also found in analysis 3, an average of 8.4 as compared to 6.8 from analysis 2 and 6.5 from analysis 1. This shows that failing to account for subject and partner level variation correctly can result in posterior estimates with larger credible intervals.

Table A.1: A summary of the average estimated posterior median ($PMed$), posterior mean (PM), lower limit (L_{CI}) and upper limit (U_{CI}) of the 95% equal-tail credible interval (CI), coverage probability of the CI (CP), length of the CI (G_{CI}), and mean square error MSE from 3 analysis methods across 600 simulated datasets with 2 levels of heterogeneity to examine the effects of disaggregation. Analysis 1 uses the true model with complete disaggregated observations. Analysis 2 uses the true model with partial disaggregated observations. Analysis 3 looks at only aggregated totals and assumes data with only a single level of heterogeneity.

Analysis	$PMed$	PM	L_{CI}	U_{CI}	CP	G_{CI}	MSE
1	16.0	16.1	13.0	20.2	97.3	7.2	6.5
2	16.0	16.2	13.0	20.3	96.9	7.3	6.8
3	16.2	16.3	13.0	20.8	95.4	7.8	8.4

A.2 HLP Demographics

A summary of participants' demographics information at baseline is given in Table A.2.

Table A.2: General demographics of $n = 936$ subjects at baseline with stratification by intervention assigment.

Variable	Treatment (n=467)	Control (n=469)	Total (n=936)
Site, n (%)			
Los Angeles	163 (34.9)	170 (36.2)	333 (35.6)
Milwaukee	43 (9.2)	44 (9.4)	87 (9.3)
New York	127 (27.2)	118 (25.2)	245 (26.2)
San Francisco	134 (28.7)	137 (29.2)	271 (29.0)
Risk Group, n (%)			
MSM	256 (54.8)	278 (59.3)	534 (57.1)
IDU	57 (12.2)	50 (10.7)	107 (11.4)
FEM	103 (22.1)	93 (19.8)	196 (20.9)
HTM	51 (10.9)	48 (10.2)	99 (10.6)
Education, n (%)			
Less than HS	88 (18.8)	97 (20.7)	185 (19.8)
HS Grad	126 (27.0)	99 (21.1)	225 (24.0)
Some College	176 (37.7)	183 (39.0)	359 (38.4)
College Grad	77 (16.5)	90 (19.2)	167 (17.8)
Race, n (%)			
White	143 (30.6)	157 (33.5)	300 (32.1)
Black	231 (49.5)	190 (40.5)	421 (45.0)
Hispanic	61 (13.1)	82 (17.5)	143 (15.3)
Other	32 (6.9)	40 (8.5)	72 (7.7)
Gender, n (%)			
Male	364 (77.9)	376 (80.2)	740 (79.1)
Female	103 (22.1)	93 (19.8)	196 (20.9)
Mean Age (sd)			
Age	39.57 (7.15)	40.11 (7.68)	39.84 (7.42)

A.3 Choice of priors on variance hyperparameters

We consider the hierarchical Poisson model with random effects

$$Y_i \sim \text{Po}(\lambda_i) \tag{A.4}$$

$$\lambda_i = \exp(\mu + \beta_i) \tag{A.5}$$

$$\beta_i \sim \text{N}(0, \sigma^2) \tag{A.6}$$

for $i = 1, \ldots, n$. When prior independence is assumed between μ and σ^2, the sampling mean, $\text{E}(Y_i)$, can be found using iterated conditional expections

$$\begin{aligned}\text{E}(Y_i) = \text{E}(\text{E}(Y_i|\lambda_i)) = \text{E}(\lambda_i) = \text{E}(\text{E}(\lambda_i|\mu, \sigma^2)) &= \text{E}(\exp(\mu + \frac{1}{2}\sigma^2)) \\ &= \text{E}(\exp(\mu))\text{E}(\exp(\frac{1}{2}\sigma^2)).\end{aligned} \tag{A.7}$$

Use of the conjugate inverse gamma prior for $\sigma^2 \sim \text{IG}(a, b)$ results in a posterior density

$$f(\sigma^2|\beta_i, Y_i) \propto \exp\left(-\frac{b^*}{\sigma^2}\right)(\sigma^2)^{-a^*-1} \tag{A.8}$$

where $a^* = a + n/2$ and $b^* = b + \sum_{i=1}^{n} \beta_i^2/2$ for $\sigma^2 \in (0, \infty)$. Letting $v = \exp(\sigma^2/2)$, the corresponding posterior density

$$f(v|\beta_i, Y_i) \propto \frac{1}{v}\exp\left(-\frac{b^*}{2\log(v)}\right)(\log v)^{-a^*-1} \tag{A.9}$$

has undefined mean $\text{E}(v)$. As a direct result, $\text{E}(Y_i)$ is also undefined. The use of a zero left truncated normal prior for $\sigma^2 \sim \text{truncN}(c, d^2)$ where c and d^2 are respectively the mean and

variance of the untruncated normal density results in posterior densities

$$f(\sigma^2|\beta_i, Y_i) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2}\left(\frac{(\sigma^2 - c)^2}{d^2} + \frac{\sum_{i=1}^n \beta_i}{\sigma^2}\right)\right) \tag{A.10}$$

$$f(v|\beta_i, Y_i) \propto \frac{1}{v}(\log v)^{-n/2} \exp\left(-\frac{1}{2}\left(\frac{(2\log v - c)^2}{d^2} + \frac{\sum_{i=1}^n \beta_i}{2\log v}\right)\right) \tag{A.11}$$

$$\propto \frac{1}{v}(\log v)^{-n/2} v^{(2c-2\log v)/d^2} \exp\left(-\frac{\sum_{i=1}^n \beta_i}{4\log v}\right) \tag{A.12}$$

that avoid this complication and has a finite $\mathrm{E}(Y_i)$. In the multivariate case, use of zero left truncated normal priors on each diagonal element of the covariance matrix has the same effect.

A.4 Posterior sampling

Posterior sampling of model parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}_{ij}$, $\boldsymbol{\delta}_{ijk}$, $\boldsymbol{\Sigma}$, $\mathbf{D}$, $\mathbf{A}$) for $1 \leq i \leq n$, $1 \leq j \leq J_i$, $1 \leq k \leq V_{ij}$ uses Markov Chain Monte Carlo (MCMC) methods (Metropolis et al. 1953; Hastings 1970; Gelfand & Smith 1990; Casella & George 1992). We also simultaneously sample from unobserved partner level outcomes (P_{ijk}, U_{ijk}) for $k \in S_{ij}$ where S_{ij} denotes the set of partners for subject i at time t_{ij} for which partner specific act information was not observed. Posterior sampling of the conditional distributions of $\boldsymbol{\alpha}$, $\boldsymbol{\Sigma}$, $\mathbf{D}$, and $\mathbf{A}$ use the Metropolis-Hastings algorithm with a non-symmetric proposal function. Detailed sampling algorithms are given here.

In this section, we use the notation $\widehat{r}$ to denote the current iteration of parameter r. A proposal function $q(r_{prop}|\widehat{r}) \sim G(h(\widehat{r}))$ where G is a specified distribution denotes that $q(r_{prop}|\widehat{r})$ is the density of the distribution G with parameters defined by $h(\widehat{r})$ evaluated at r_{prop}.

1. Sampling from the posterior distributions of $\boldsymbol{\alpha}_v^+$ and $\boldsymbol{\alpha}_v^-$ takes on the same form. We

only present the sampling density $f(\boldsymbol{\alpha}_v^+|.)$ here. Sample $\boldsymbol{\alpha}_v^+$ from

$$f(\boldsymbol{\alpha}_v^+|.) \propto \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{J_i}\left\{V_{ij}^+\mathbf{x}_{ij}'\boldsymbol{\alpha}_v^+ - \lambda_{v,ij}^+\right\} - G(\boldsymbol{\alpha}_v^+)\right]\{1-\exp(-\lambda_{v,i1}^+ - \lambda_{v,i1}^-)\}^{-1} \tag{A.13}$$

where $\lambda_{v,ij}^+ = \exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_v^+ + \beta_{v,ij}^+)$, $\lambda_{v,ij}^- = \exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_v^- + \beta_{v,ij}^-)$ and $G(\boldsymbol{\alpha}_v^+) = \frac{1}{2}(\boldsymbol{\alpha}_v^+ - \boldsymbol{\mu}_{\alpha_v^+})'\boldsymbol{\Sigma}_{\alpha_v^+}(\boldsymbol{\alpha}_v^+ - \boldsymbol{\mu}_{\alpha_v^+})$ comes from the prior of $\boldsymbol{\alpha}_v^+$. Prior parameters $\boldsymbol{\mu}_{\alpha_v^+}$ and $\boldsymbol{\Sigma}_{\alpha_v^+}$ are correspondingly a vector with each element set to $\mu_\alpha = 0$ and a diagonal matrix with each diagonal element set to $\sigma_\alpha^2 = 10$. We use a second-order Taylor approximation of equation (A.13) as the adaptive proposal function. When $\mu_\alpha = 0$ and $\boldsymbol{\Sigma}_{\alpha_v^+}$ is diagonal, we use a multivariate normal proposal function $q(\boldsymbol{\alpha}_{v,prop}^+|\widehat{\boldsymbol{\alpha}}_v^+) \sim \text{MVN}(\mathbf{T}_{v^+}^{-1}\mathbf{M}_{v^+}', \mathbf{T}_{v^+}^{-1})$ with

$$\mathbf{T}_{v^+} = \Sigma_{\alpha_v^+}^{-1} + \sum_{i=1}^{n}\sum_{j=1}^{J_i}\widehat{\lambda}_{v,ij}^+\mathbf{x}_{ij}\mathbf{x}_{ij}' - \sum_{i=1}^{n}R_i\mathbf{x}_{i1}\mathbf{x}_{i1}', \tag{A.14}$$

$$\mathbf{M}_{v^+} = \sum_{i=1}^{n}\sum_{j=1}^{J_i}(\widehat{\lambda}_{v,ij}^+\widehat{\boldsymbol{\alpha}}_v^{+\prime}\mathbf{x}_{ij} - \widehat{\lambda}_{v,ij}^+ + V_{ij}^+)\mathbf{x}_{ij}' - \sum_{i=1}^{n}(R_i\widehat{\boldsymbol{\alpha}}_v^{+\prime}\mathbf{x}_{i1} - H_i)\mathbf{x}_{i1}', \tag{A.15}$$

where $\widehat{\lambda}_{v,ij}^+ = \exp(\mathbf{x}_{ij}'\widehat{\boldsymbol{\alpha}}_v^+ + \beta_{v,ij}^+)$, $H_i = -\widehat{\lambda}_{v,i1}^+\{1-\exp(-\widehat{\lambda}_{v,i1}^+ - \lambda_{v,i1}^-)\}^{-1}$ $\exp(-\widehat{\lambda}_{v,i1}^+ - \lambda_{v,i1}^-)$, and $R_i = H_i^2 - H_i\widehat{\lambda}_{v,i1}^+ + H_i$.

2. Sampling the posterior distributions of $\boldsymbol{\alpha}_u^+, \boldsymbol{\alpha}_u^-, \boldsymbol{\alpha}_p^+$, and $\boldsymbol{\alpha}_p^-$ is similar in form to sampling $\boldsymbol{\alpha}_v^+$ but uses the partner level observations. We present here only the posterior sampling algorithm for $\boldsymbol{\alpha}_u^+$. To sample from posterior distributions of $\boldsymbol{\alpha}_u^+$, we first define a set Φ_i such that $k \in \Phi_i$ denotes all partners nested within subject i who at baseline were defined as an HIV$^-$ partner or an HIV$^+$ partner who was not categorized as a primary partner. We also reorder partner observations for each subject i at each time point j such that partners $1, \ldots, V_{ij}^+$ are HIV$^+$ and the rest are HIV$^-$ for

notational convenience. We sample $\boldsymbol{\alpha}_u^+$ from

$$f(\boldsymbol{\alpha}_u^+|.) \propto \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{V_{ij}^+}\left\{U_{ijk}\mathbf{x}'_{ijk}\boldsymbol{\alpha}_u^+ - \lambda_{u,ijk}\right\} - G(\boldsymbol{\alpha}_u^+)\right]\{1-\exp(-\sum_{k\in\Phi_i}\lambda_{u,i1k})\}^{-1} \tag{A.16}$$

where $\lambda_{u,ijk} = \exp\left(\mathbf{x}'_{ijk}\boldsymbol{\alpha}_u^+ + \beta_{u,ij}^+ + \delta_{u,ijk}\right)$ for $k \le V_{ij}^+$, $\lambda_{u,ijk} =$ $\exp\left(\mathbf{x}'_{ijk}\boldsymbol{\alpha}_u^- + \beta_{u,ij}^- + \delta_{u,ijk}\right)$ for $k > V_{ij}^+$, and $G(\boldsymbol{\alpha}_u^+) = \frac{1}{2}(\boldsymbol{\alpha}_u^+ - \boldsymbol{\mu}_{\alpha_u^+})'\boldsymbol{\Sigma}_{\alpha_u^+}(\boldsymbol{\alpha}_u^+ - \boldsymbol{\mu}_{\alpha_u^+})$ comes from the prior of $\boldsymbol{\alpha}_u^+$. Prior parameters $\boldsymbol{\mu}_{\alpha_u^+}$ and $\boldsymbol{\Sigma}_{\alpha_u^+}$ are correspondingly a vector with each element set to $\mu_\alpha = 0$ and a diagonal matrix with each diagonal element set to $\sigma_\alpha^2 = 10$. Similar to the proposal function $q(\boldsymbol{\alpha}_{v,prop}^+|\widehat{\boldsymbol{\alpha}}_v^+)$, we use a second-order Taylor approximation of equation (A.16) as the adaptive proposal function $q(\boldsymbol{\alpha}_{u,prop}^+|\widehat{\boldsymbol{\alpha}}_u^+)$. When $\mu_\alpha = 0$ and $\boldsymbol{\Sigma}_{\alpha_u^+}$ is diagonal, $q(\boldsymbol{\alpha}_{u,prop}^+|\widehat{\boldsymbol{\alpha}}_u^+) \sim$ MVN$(\mathbf{T}_{u^+}^{-1}\mathbf{M}'_{u^+}, \mathbf{T}_{u^+}^{-1})$ with

$$\mathbf{T}_{u^+} = \Sigma_{\alpha_v^+}^{-1} + \sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{V_{ij}^+}\widehat{\lambda}_{u,ijk}\mathbf{x}_{ijk}\mathbf{x}'_{ijk} - \sum_{i=1}^{n}\sum_{k\in\Phi_i}R_{i,u^+}\mathbf{x}_{i1k}\mathbf{x}'_{i1k}, \tag{A.17}$$

$$\begin{aligned}\mathbf{M}_{u^+} =& \sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{V_{ij}^+}(\widehat{\lambda}_{u,ijk}\widehat{\boldsymbol{\alpha}}_u^{+\prime}\mathbf{x}_{ijk} - \widehat{\lambda}_{u,ijk}^+ + U_{ijk})\mathbf{x}'_{ijk} - \\ & \sum_{i=1}^{n}\sum_{k\in\Phi_i}(R_{i,u^+}\widehat{\boldsymbol{\alpha}}_u^{+\prime}\mathbf{x}_{i1k} - H_{i,u^+})\mathbf{x}'_{i1k},\end{aligned} \tag{A.18}$$

where

$$\widehat{\lambda}_{u,ijk} = \exp\left(\mathbf{x}'_{ijk}\widehat{\boldsymbol{\alpha}}_u^+ + \beta_{u,ij}^+ + \delta_{u,ijk}\right) \text{ for } k \le V_{ij}^+, \tag{A.19}$$

$$H_{i,u^+} = -\widehat{\lambda}_{u,i1T}^+\{1-\exp\left(-\widehat{\lambda}_{u,i1T}^+ - \lambda_{u,i1T}^-\right)\}^{-1}\exp\left(-\widehat{\lambda}_{u,i1T}^+ - \lambda_{u,i1T}^-\right), \tag{A.20}$$

$$\widehat{\lambda}_{u,i1T}^+ = \sum_k \widehat{\lambda}_{u,i1k} \text{ for } (k\in\Phi_i)\cap(k\le V_{ij}^+), \tag{A.21}$$

$$\lambda_{u,i1T}^- = \sum_k \lambda_{u,i1k} \text{ for } (k\in\Phi_i)\cap(k> V_{ij}^+), \text{ and} \tag{A.22}$$

$$R_{i,u^+} = H_{i,u^+}^2 - H_{i,u^+}\widehat{\lambda}_{u,i1T}^+ + H_{i,u^+}. \tag{A.23}$$

3. Sample $\boldsymbol{\beta}_{i1}, \ldots, \boldsymbol{\beta}_{iJ}$ for $1 \leq i \leq n$ from

$$f(\boldsymbol{\beta}_{i1}|.) \propto f(\mathbf{Y}_{i1}|\boldsymbol{\beta}_{i1}) f(\boldsymbol{\beta}_{i1}|\mathbf{L}) f(\boldsymbol{\beta}_{i2}|\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}_{i1}), \tag{A.24}$$

$$f(\boldsymbol{\beta}_{ij}|.) \propto f(\mathbf{Y}_{ij}|\boldsymbol{\beta}_{ij}) f(\boldsymbol{\beta}_{ij}|\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}_{i(j-1)}) f(\boldsymbol{\beta}_{i(j+1)}|\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}_{ij}), \quad \text{for } 2 \leq j < J, \tag{A.25}$$

$$f(\boldsymbol{\beta}_{iJ}|.) \propto f(\mathbf{Y}_{iJ}|\boldsymbol{\beta}_{iJ}) f(\boldsymbol{\beta}_{iJ}|\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}_{i(J-1)}), \tag{A.26}$$

where $f(Y_{ij}|\boldsymbol{\beta}_{ij}) = f(V_{ij}^+|\lambda_{v,ij}^+) f(V_{ij}^-|\lambda_{v,ij}^-) \prod_{k=1}^{V_{ij}} f(U_{ijk}|\lambda_{u,ijk}) f(P_{ijk}|\lambda_{p,ijk})$, $\lambda_{p,ijk} = \exp\left(\mathbf{x}_{ijk}'\boldsymbol{\alpha}_p^+ + \beta_{p,ij}^+ + \delta_{p,ijk}\right)$ for $k \leq V_{ij}^+$, $\lambda_{p,ijk} = \exp\left(\mathbf{x}_{ijk}'\boldsymbol{\alpha}_p^- + \beta_{p,ij}^- + \delta_{p,ijk}\right)$ for $k > V_{ij}^+$. We use a random walk Gaussian proposal function $q(\boldsymbol{\beta}_{ij,prop}|\widehat{\boldsymbol{\beta}}_{ij}) \sim \text{MVN}(\widehat{\boldsymbol{\beta}}_{ij}, \boldsymbol{\Sigma}_{\beta,q})$ where $\boldsymbol{\Sigma}_{\beta,q}$ is a diagonal matrix with diagonal elements chosen to obtain an acceptable rate of acceptance for $\boldsymbol{\beta}_{ij,prop}$.

4. Sample $\boldsymbol{\delta}_{ijk} = (\delta_{p,ijk}, \delta_{u,ijk})^T$ for $1 \leq i \leq n$, $1 \leq j \leq J_i$, $1 \leq k \leq V_{ij}$ from

$$f(\boldsymbol{\delta}_{ijk}|.) \propto f(P_{ijk}|\lambda_{p,ijk}) f(U_{ijk}|\lambda_{u,ijk}) f(\boldsymbol{\delta}_{ijk}|\mathbf{D}) \tag{A.27}$$

using the Metropolis algorithm with a random walk Gaussian proposal function $q(\boldsymbol{\delta}_{ijk,prop}|\widehat{\boldsymbol{\delta}}_{ijk}) \sim \text{MVN}(\widehat{\boldsymbol{\delta}}_{ijk}, \boldsymbol{\Sigma}_{\delta,q})$ where $\boldsymbol{\Sigma}_{\delta,q}$ is a diagonal matrix with diagonal elements chosen to obtain an acceptable rate of acceptance for $\boldsymbol{\delta}_{ijk,prop}$.

5. Sample $\boldsymbol{\Sigma}$ from

$$f(\boldsymbol{\Sigma}|.) \propto \prod_{i=1}^{n} \prod_{j=1}^{J_i} f(\boldsymbol{\beta}_{ij}|\boldsymbol{\Sigma}) \pi(\boldsymbol{\Sigma}) \tag{A.28}$$

where the prior $\pi(\boldsymbol{\Sigma})$ is defined in Section 2.3.2. A proposal function, $q(\boldsymbol{\Sigma}_{prop}|\widehat{\boldsymbol{\Sigma}}) \sim \text{IW}(\boldsymbol{\Psi}_{\Sigma,q}, m_{\Sigma,q})$ is used to approximate $f(\boldsymbol{\Sigma}|.)$ where $\boldsymbol{\Psi}_{\Sigma,prop} = \sum_{i=1}^{n} \sum_{j=2}^{J_i} (\boldsymbol{\beta}_{ij} - \mathbf{A}\boldsymbol{\beta}_{i(j-1)})$ $(\boldsymbol{\beta}_{ij} - \mathbf{A}\boldsymbol{\beta}_{i(j-1)})' + \boldsymbol{\Psi}_{\Sigma}$ and $m_{\Sigma,prop} = m_{\Sigma} + \sum_{i=1}^{n} \sum_{j=2}^{J_i} 1$. The proposal function $q(\boldsymbol{\Sigma}_{prop}|\boldsymbol{\Sigma}_{current})$ is modestly overdispersed compared to $f(\boldsymbol{\Sigma}|.)$ due to the normal prior terms in the diagonal elements.

6. Sample $\mathbf{D}$ from

$$f(\mathbf{D}|.) \propto \prod_{i=1}^{n} \prod_{j=1}^{J_i} \prod_{k=1}^{V_{ij}} f(\boldsymbol{\delta}_{ijk}|\mathbf{D})\pi(\mathbf{D}) \tag{A.29}$$

where the prior $\pi(\mathbf{D})$ is defined in Section 2.3.2. A proposal function, $q(\mathbf{D}_{prop}|\widehat{\mathbf{D}}) \sim \text{IW}(\boldsymbol{\Psi}_{D,q}, m_{D,q})$ is used to approximate $f(\mathbf{D}|.)$ where $\boldsymbol{\Psi}_{D,q} = \sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{V_{ij}} \boldsymbol{\delta}_{ijk}\boldsymbol{\delta}'_{ijk} + \Psi_D$ and $m_{D,q} = m_D + \sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{V_{ij}} 1$.

7. Since parameter $\mathbf{A}$ is diagonal with diagonal elements $A_{l,l}$, we can sample $\mathbf{A}$ from

$$f(\mathbf{A}|.) \propto \prod_{i=1}^{n} \left\{ f(\boldsymbol{\beta}_{i1}|\mathbf{L}) \prod_{j=2}^{J_i} f(\boldsymbol{\beta}_{ij}|\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}_{i(j-1)}) \right\} \prod_{l=1}^{6} \pi(A_{l,l}) \tag{A.30}$$

where $\pi(A_{l,l}) = 1/2\ \mathbf{1}(-1 \leq A_{l,l} \leq 1)$ is a uniform distribution from -1 to 1. A random walk truncated Gaussian proposal function is used for $A_{l,l}$

$$q(A_{(l,l),prop}|\widehat{A}_{l,l}) \sim \text{truncN}(\widehat{A}_{l,l}, \sigma_A^2) \tag{A.31}$$

where $A_{(l,l),prop}$ is between -1 to 1 and joined to create the new proposed A.

8. Let $P_{ijT} = \sum_{k=1}^{V_{ij}} P_{ijk}$ and $U_{ijT} = \sum_{k=1}^{V_{ij}} U_{ijk}$ respectively be the total protected and unprotected acts for subject i at time t_{ij}. Also let S_{ij} be the set of partners k with subject i at time t_{ij} for which partner specific protected and unprotected acts is not recorded and let $\mathbf{P}_{ijS_{ij}}$ and $\mathbf{U}_{ijS_{ij}}$ be the set of protected and unprotected acts corresponding to these partners. We sample $\mathbf{P}_{ijS_{ij}}$ and $\mathbf{U}_{ijS_{ij}}$ for $1 \leq i \leq n$, $1 \leq j \leq J_i$ from

$$\mathbf{P}_{ijS_{ij}}|\boldsymbol{\lambda}_{ij}, P_{ijT}, (P_{ijk} \text{ for } k \notin S_{ij}) \sim \text{Multinomial}(N_{P,ij}, \boldsymbol{\pi}_{P_{ij}}), \tag{A.32}$$

$$\mathbf{U}_{ijS_{ij}}|\boldsymbol{\lambda}_{ij}, U_{ijT}, (P_{ijk} \text{ for } k \notin S_{ij}) \sim \text{Multinomial}(N_{U,ij}, \boldsymbol{\pi}_{U_{ij}}), \tag{A.33}$$

where $\boldsymbol{\lambda}_{ij} = (\lambda^+_{v,ij},\ \lambda^-_{v,ij},\ \lambda_{p,ij1},\ \lambda_{u,ij1},\ \ldots,\ \lambda_{p,ijV_{ij}},\ \lambda_{u,ijV_{ij}})^T$, $N_{P,ij} = P_{ijT} - \sum_{k\notin S_{ij}} P_{ijk}$, $N_{U,ij} = U_{ijT} - \sum_{k\notin S_{ij}} U_{ijk}$, $\boldsymbol{\pi}_{P_{ij}} = (\pi_{P_{ijk}}) = (\lambda_{p,ijk}/\lambda_{p,ijT})$ for every $k \in S_{ij}$, $\lambda_{p,ijT} = \sum_{k\in S_{ij}} \lambda_{p,ijk}$, $\boldsymbol{\pi}_{U_{ij}} = (\pi_{U_{ijk}}) = (\lambda_{u,ijk}/\lambda_{u,ijT})$ for every $k \in S_{ij}$, and $\lambda_{u,ijT} = \sum_{k\in S_{ij}} \lambda_{u,ijk}$.

We repeat Steps 1 through 8 until convergence and to collect a sample from the posterior.

A.5 Summary of HLP covariance estimates

Table A.3 shows estimated covariance between log mean parameters of the corresponding outcomes, (V_{ij}^{+}, V_{ij}^{-}, P_{ijk}^{+}, U_{ijk}^{+}, P_{ijk}^{-}, U_{ijk}^{-}), which are a function of parameters $\mathbf{L}$ and $\mathbf{D}$ in the model. Significant positive (negative) covariance estimates represent significant positive (negative) correlation between corresponding outcomes. Estimates and 95% posterior intervals for the autoregressive parameter $\mathbf{A}$ are also shown.

We find significant negative correlation between the number of protected acts per partner and the number of unprotected acts per partner with the same subject regardless of partner serostatus. Our results also suggest significant positive correlation between both the number of protected acts per partner across different partners with the same subject and the number of unprotected acts per partner across different partners with the same subject regardless of partner serostatus. These findings suggest that subjects who tend to report greater number of protected acts per partner tend to report fewer unprotected acts per partner and vice versa. In addition, if the subject reports greater numbers of protected or unprotected acts for a single partner, they tend to consistently report greater numbers of the same type of act for other partners. There is marginally significant positive correlation between the number of HIV^{+} and number of HIV^{-} partners though the correlation is very low. Finally, fairly high and consistent values of across time correlation A at each time point suggests that subject behavior was highly correlated from interview to interview and that the strong correlation remained over the course of the study.

Table A.3: Summary of variance and covariance for the log mean parameters of the outcomes and the autoregressive parameter A. Values are reported as posterior mean PM on the first line and the 95% equal-tail credible interval (L_{CI}, U_{CI}) on the second.

Outcomes	V_{ij}^{+}	V_{ij}^{-}	P_{ijk}^{+}	U_{ijk}^{+}	P_{ijk}^{-}	U_{ijk}^{-}
Different Partners						
V_{ij}^{+}	2.05 (1.84, 2.28)	0.11 (0.01, 0.21)	-0.38 (-0.66, -0.12)	-0.47 (-0.65, -0.31)	-0.14 (-0.29, 0.03)	0.06 (-0.10, 0.22)
V_{ij}^{-}	– –	1.68 (1.55, 1.84)	-0.13 (-0.34, 0.07)	-0.22 (-0.35, -0.10)	-0.61 (-0.78, -0.46)	-0.60 (-0.74, -0.44)
$P_{ijk'}^{+}$	– –	– –	5.09 (4.34, 5.79)	-1.17 (-1.45, -0.91)	2.08 (1.72, 2.55)	-0.82 (-1.17, -0.43)
$U_{ijk'}^{+}$	– –	– –	– –	1.84 (1.58, 2.10)	-0.35 (-0.57, -0.13)	0.98 (0.74, 1.21)
$P_{ijk'}^{-}$	– –	– –	– –	– –	2.84 (2.52, 3.20)	-0.46 (-0.64, -0.27)
$U_{ijk'}^{-}$	– –	– –	– –	– –	– –	2.46 (2.18, 2.75)
Same Partner						
P_{ijk}^{+}	– –	– –	6.71 (5.96, 7.41)	-1.44 (-1.72, -1.17)	– –	– –
U_{ijk}^{+}	– –	– –	– –	3.45 (3.19, 3.72)	– –	– –
P_{ijk}^{-}	– –	– –	– –	– –	4.46 (4.11, 4.86)	-0.72 (-0.91, -0.54)
U_{ijk}^{-}	– –	– –	– –	– –	– –	4.07 (3.79, 4.36)
Across time correlation						
A	0.77 (0.74, 0.80)	0.72 (0.69, 0.75)	0.64 (0.59, 0.70)	0.67 (0.62, 0.72)	0.75 (0.72, 0.79)	0.67 (0.63, 0.72)

APPENDIX B

The cmDPM model

B.1 Posterior computation for the cmDPM model

The cmDPM model is first specified in Section 3.3.1.1 with strength parameter α, cluster stickiness parameter ρ, vectors of cluster memory indicators $\mathbf{Z}_j$ for $j = 2, \ldots, J$, vectors of subject specific parameters $\boldsymbol{\theta}_j$, and vectors of cluster labels $\mathbf{c}_j$ for $j = 1, \ldots, J$. Hyperpriors for the model are specified in Section 3.3.2 with cluster time specific base distribution parameters $\boldsymbol{\eta}_j$ for $j = 1, \ldots, J$ and additional likelihood distribution parameter λ. Posterior sampling of these model parameters uses Markov Chain Monte Carlo (MCMC) methods and the detailed posterior sampling algorithm is provided here.

1. For $1 \leq i \leq n$, we sample the cluster labels c_{ij} for subject i at time point 1 from the conditional density $f(c_{i1}|\mathbf{c}_{-(i1)}, Y_{i1}, \alpha, \boldsymbol{\phi}_1, \boldsymbol{\eta}_1, \lambda)$. The conditional density of c_{i1} at c can be broken down into two discrete cases.

   (a) If $c = c_{k1}$ for some $k \neq i$:
   $Pr(c_{i1} = c|\mathbf{c}_{-i1}, \alpha, Y_{i1}, \boldsymbol{\phi}_{-i1}, \lambda) = b(n - 1 + \alpha)^{-1} n_{-i1,c} f(Y_{i1}|\phi_{1c}, \lambda)$.

   (b) If $c \neq$ any current value in $\mathbf{c}_{-(ij)}$
   $Pr(c_{i1} = c|\mathbf{c}_{-(i1)}, \alpha, Y_{i1}, \boldsymbol{\phi}_{-i1}, \lambda) = b(n - 1 + \alpha)^{-1} \alpha \int f(Y_{i1}|\phi, \lambda) dG_{01}(\phi; \boldsymbol{\eta}_1)$,

   where b is the normalizing constant.

   When a new cluster c is formed with observation Y_{i1}, a value of ϕ_{1c} is sampled from $f(\phi_{1c}|Y_{i1}, \eta_{10}, \lambda) \propto f(Y_{i1}|\phi_{1c}, \lambda) G_{01}(\phi_{1c}; \eta_{10})$.

2. For $1 \leq i \leq n$ at each time point $2 \leq j \leq J$, we jointly sample both the cluster label c_{ij} and the cluster memory indicator Z_{ij} from the conditional density

$f(c_{ij}, Z_{ij}|\mathbf{c}_{-(ij)}, Y_{ij}, \alpha, \boldsymbol{\phi}_j, Z_{i(j+1)}, \boldsymbol{\eta}_j, \boldsymbol{\eta}_{j+1}, \lambda)$. For this step, we define values of $Z_{i(J+1)} = 0$ for $1 \leq i \leq n$. The joint conditional density of of (c_{ij}, Z_{ij}) can be broken down into three discrete cases.

(a) If $Z_{i(j+1)} = 1$, $c_{i(j+1)} = c_{i(j-1)}$, and $n^1_{-ijc} \neq 0$,

$$Pr(c_{ij}, Z_{ij}|.) = \begin{cases} (c, 1) = b\rho(\alpha + M_{-ij})(n^1_{-ijc})^{-1}(n_{-ijc} + n^1_{-ijc}) \\ (c, 0) = b(1-\rho)n_{-ijc} \end{cases}$$

where $c = c_{i(j-1)}$ and b is the normalizing constant. Otherewise, if $n_{-ijc} = 0$ or $n^1_{-ijc} = 0$, $(c_{ij}, Z_{ij}) = (c, 1)$ where $c = c_{i(j-1)}$.

(b) If $Z_{i(j+1)} = 1$ and $c_{i(j+1)} \neq c_{i(j-1)}$

$(c_{ij}, Z_{ij}) = (c, 0)$ where $c = c_{i(j-1)}$.

(c) If $Z_{i(j+1)} = 0$, let $c_1 \neq c_{i(j-1)}$ and $c_1 = c_{kj}$ for some $k \neq i$, $c_2 = c_{i(j-1)}$ and $c_2 = c_{kj}$ for some $k \neq i$, $c_3 = c_{i(j-1)}$ and $c_3 \neq c_{kj}$ for any $k \neq i$, $c_4 \neq$ any current value in $\mathbf{c}_{-(ij)}$. The joint probability can be sampled as

$$Pr(c_{ij}, Z_{ij}|.) = \begin{cases} (c_1, 0) = b\left[(1-\rho)n_{-ijc}f(Y_{ij}|\phi_{jc}, \lambda)\right] \\ (c_2, 0) = b\left[(1-\rho)n_{-ijc}f(Y_{ij}|\phi_{jc}, \lambda)\right] \\ (c_2, 1) = b\rho(\alpha + M_{-ij})(n^1_{-ijc})^{-1}n_{-ijc}f(Y_{ij}|\phi_{jc}, \lambda) \\ (c_3, 1) = b\rho(\alpha + M_{-ij}) \int f(Y_{ij}; \phi, \lambda) \\ \qquad\qquad dG_{0jc}(\phi; \phi_{(j-1)c}, \boldsymbol{\eta}_j)G_{0(j+1)c}(\phi_{(j+1)c}; \phi, \boldsymbol{\eta}_{j+1}) \\ (c_4, 0) = b(1-\rho)\alpha \int f(Y_{ij}; \phi, \lambda)dG_{0j}(\phi, \eta_{j0}) \end{cases}$$

if $n^1_{-ijc} \neq 0$. If $n_{-ijc} \neq 0$ and $n^1_{-ijc} = 0$, then $(c_{ij}, Z_{ij}) = (c_2, 1)$. The constant b is used as a normalizing constant.

When a new cluster c is formed with observation Y_{ij}, a value of ϕ_{jc} is sampled from $f(\phi_{jc}|Y_{ij}, \eta_{j0}, \lambda) \propto f(Y_{ij}|\phi_{jc}, \lambda)G_{0j}(\phi_{jc}; \eta_{j0})$. When a cluster c is retained from the previous time point $j-1$ for observation Y_{ij} with no current cluster specific value ϕ_{jc} at time point j, a new ϕ_{jc} is sampled from $f(\phi_{jc}|Y_{ij}, \phi_{(j-1)c}, \eta^1_j, \lambda) \propto f(Y_{ij}|\phi_{jc}, \lambda)$ $G_{0jc}(\phi_{jc}; \phi_{(j-1)c}, \eta^1_j)$.

3. For $1 \leq j \leq J$, sample a new cluster specific value ϕ_{jc} from $f(\phi_{jc}|Y_{1j}, \ldots, Y_{nj}) \propto \prod_{i \text{ s.t. } c_{ij}=c} f(Y_{ij}|\phi_{jc}, \lambda) G_{0jc}(\phi_{jc}; \phi_{(j-1)c}, \boldsymbol{\eta}_j) G_{0(j+1)c}(\phi_{(j+1)c}; \phi_{jc}, \boldsymbol{\eta}_{j+1})$ for every unique value of ϕ_{jc}. This resampling of cluster specific values significantly speeds convergence.

4. For $1 \leq i \leq n$, jointly sample each set of *bound* cluster labels $(c_{ip}, \ldots, c_{iq})$.

   (a) Let $c_1 = c_{rp}$ for some $r \neq i$ and $c_2 \neq$ any current value in $\mathbf{c}_{-(ip)}$. Also let A_{pq} be the set of times $p \leq j \leq q$ for which cluster label $c_{ij} \neq c_{kj}$ for every $k \neq i$ and B_{pq} be the set of times $(p, \ldots, q) \notin A_{pq}$. Then the *bound* cluster labels $(c_{ip}, \ldots, c_{iq})$ can be sampled as

$$Pr(c_{ip}, \ldots, c_{iq}) = \begin{cases} (c_1, \ldots, c_1) = b(n - 1 + \alpha)^{-1} n_{-ip,c} L_{c1} \\ (c_2, \ldots, c_2) = b(n - 1 + \alpha)^{-1} \alpha L_{c2} \end{cases}$$

   where $L_c = \prod_{j \in B_{pq}} f(Y_{ij}; \phi_{jc}, \lambda) \prod_{j \in A_{pq}} \int f(Y_{ij}; \phi, \lambda) G_{0jc}(\phi; \phi_{(j-1)c}, \boldsymbol{\eta}_j)$ $G_{0(j+1)c}(\phi_{(j+1)c}; \phi, \boldsymbol{\eta}_{j+1}) d\phi$. This joint sampling of bound cluster labels is another remixing step that significantly speeds convergence.

   Resampling of bound cluster labels can lead to clusters that previously did not exist at some of these time points. For these clusters, we then sample new cluster specific values. When a new cluster label c is formed with observation Y_{ip}, a value of ϕ_{pc} is sampled from $f(\phi_{pc}|Y_{ip}, \eta_{p0}, \lambda) \propto f(Y_{ip}|\phi_{pc}, \lambda) G_{0p}(\phi_{pc}; \eta_{p0})$. For $j = p+1, \ldots, q$, when a cluster c is retained from the previous time point $j-1$ for observation Y_{ij} with no current cluster specific value ϕ_{jc} at time point j, a new ϕ_{jc} is sampled from $f(\phi_{jc}|Y_{ij}, \phi_{(j-1)c}, \eta_j^1, \lambda) \propto f(Y_{ij}|\phi_{jc}, \lambda) G_{0jc}(\phi_{jc}; \phi_{(j-1)c}, \eta_j^1)$.

   (b) For $j \in A_{pq}$, sample new ϕ_{jc} from $f(\phi_{jc}|Y_{ij}) \propto f(Y_{ij}|\phi, \lambda) G_{0jc}(\phi; \phi_{(j-1)c}, \boldsymbol{\eta}_j)$ $G_{0(j+1)c}(\phi_{(j+1)c}; \phi, \boldsymbol{\eta}_{j+1})$.

5. Sample cluster stickiness parameter ρ from Beta$(a_{post,\rho}, b_{post,\rho})$ where $a_{post,\rho} = a_\rho + \sum_{j=2}^{J} Z_{ij}$ and $b_{post,\rho} = b_\rho + n(J-1) - \sum_{j=2}^{J} Z_{ij}$.

6. Sample cluster time specific base distribution parameters $\boldsymbol{\eta}_j = (\eta_j^1, \eta_{j0})$ from $f(\eta_j^1|\boldsymbol{\phi}_j) \propto$

$\prod_{k \in \mathbf{c}_j^1} f(\phi_{jk}|\eta_j^1) H_1(\eta_j^1)$ and $f(\eta_{j0}|\boldsymbol{\phi}_j) \propto \prod_{c \in U_j} f(\phi_{jc}|\eta_{j0}) H_2(\eta_{j0})$ where U_j is the set of all spontaneously created unique clusters c that appears in $\mathbf{c}_j$ but not at any prior time point, $H_1(\eta_j^1)$ is the prior specified for η_j^1, and $H_2(\eta_{j0})$ is the prior specified for η_{j0}.

7. Sample cluster strength parameter α from $f(\alpha|K_1, \ldots K_j)$ from (3.36).

8. Let $\mathbf{Y} = (Y_{11}, \ldots, Y_{nj})^T$ be the vector of all observed outcomes for $1 \leq i \leq n$ at times $1 \leq j \leq J$. Sample additional likelihood distribution parameter λ from $f(\lambda|\mathbf{Y}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \propto \prod_{i=1}^n \prod_{j=1}^J f(Y_{ij}|\theta_{ij}, \lambda) H_3(\lambda)$ where $H_3(\lambda)$ is the prior specified for λ.

Repeat steps 1 through 8 until convergence.

B.2 Trace plots of posterior estimates for subject specific parameters θ_{ij}

Trace plots of posterior estimates of a few subject specific parameters at different time points are shown in Figure B.1 to provide a visual assessment of convergence. Subject specific parameters θ_{ij} may have multiple modes in their posterior densities. This reflects uncertainty of which cluster that particular country i may belong to at a given time j.

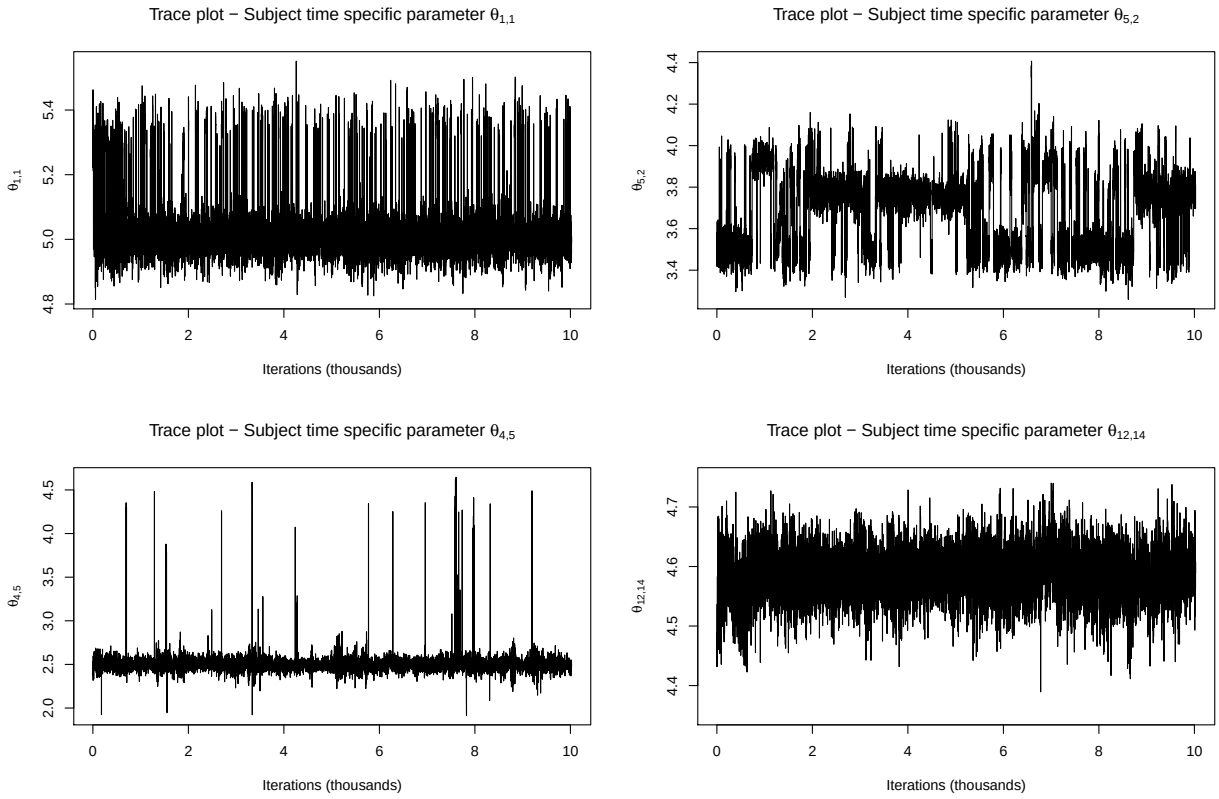

Figure B.1: Trace plots of posterior estimates for subject specific parameters θ_{ij}. Iterations have been thinned to every 4^{th} value in the MCMC chain.

B.3 Evolution of cluster memberships for the countries over time

We show how cluster membership evolves over time for all 197 countries in Figure B.2. Darker shades of blue represent higher posterior probabilities of countries sharing the same cluster. A heat map legend matching the shade of blue to the posterior probabilities is given at the right of each plot. Cluster membership is mostly maintained over time. Countries that do move between clusters still move between clusters with similar cluster specific values. This indicates we should be able to make fairly accurate predictions of cluster membership even several years in the future. Countries are also fairly distinctly clustered in the model. Most posterior probabilities between two countries sharing the same cluster is either close to 1 (dark blue) or close to 0 (white).

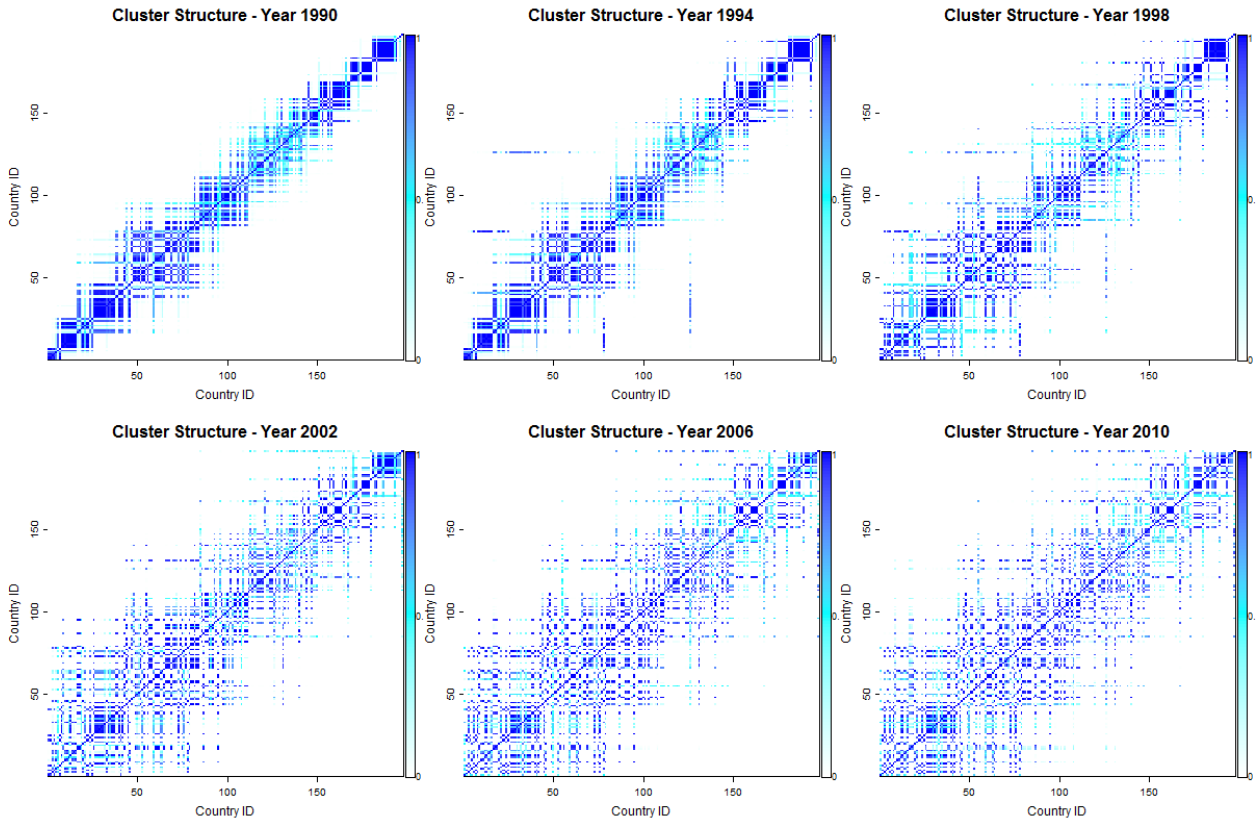

Figure B.2: Posterior probabilities of shared cluster membership between countries at select years. Darker shades of blue represent higher posterior probabilities with a legend given at the right of each plot. Country IDs are ordered by observed values of annual log TB incidence rate in 1990 from lowest to highest.

B.4 Comparison of posterior predictive densities for cmDPM and DPM models

In Figure B.3, we show for select years, the (1) estimated densities of the data, (2) the posterior predictive densities of the annual TB incidence rate in the cmDPM model, and (3) the posterior posterior predictive densities of the annual TB incidence rate in the DPM model. It is evident that the cmDPM model results in significantly better approximations of the density of the data at each time point by borrowing clustering information across time.

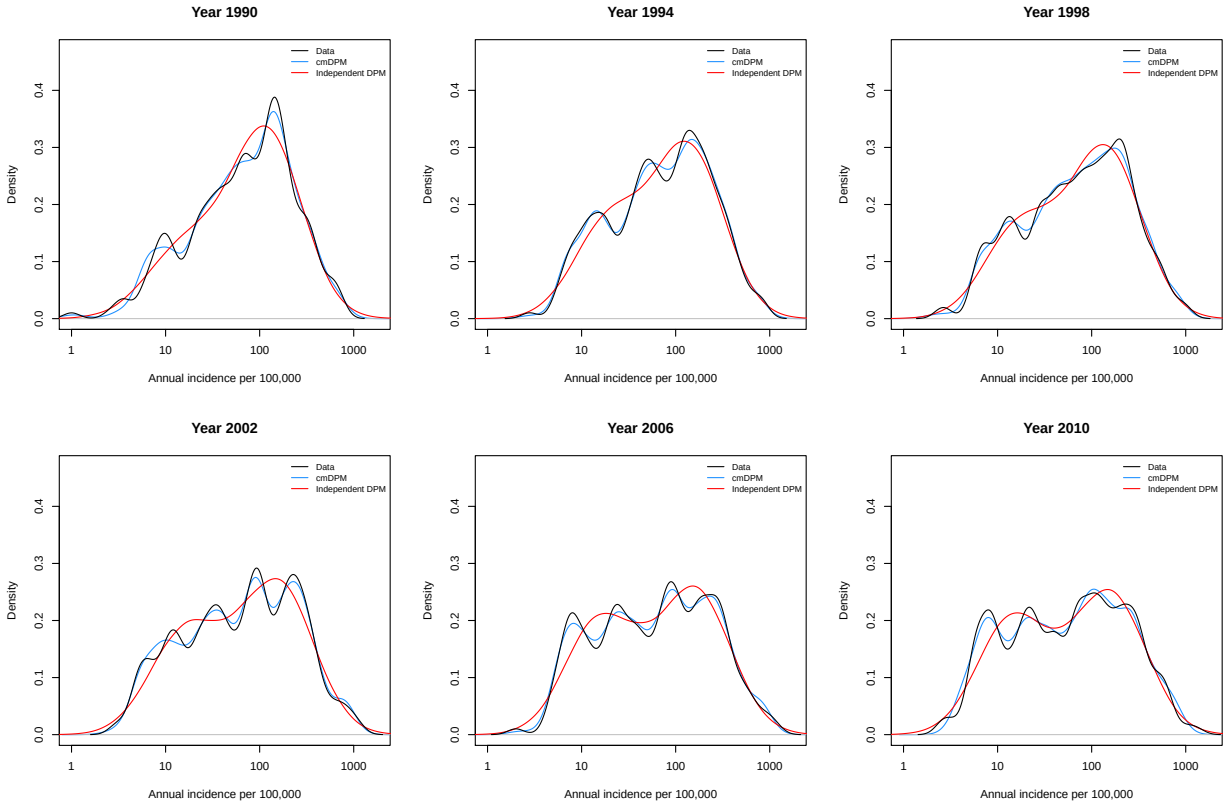

Figure B.3: A comparison of estimated densities at select years for the data Y_{ij}, the posterior predictive densities of Y_{ij} in the cmDPM model, and the posterior predictive densities of Y_{ij} in the time independent DPM model.

Bibliography

Aitchison, J. & Ho, C. H. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4), 643–653.

Aldous, D. (1985). Exchangeability and related topics. In *Ecole d'Ete de Probabilities de Saint-Flour XIII 1983*, volume 1117 of *Lecture Notes in Mathematics* (pp. 1–198). Berlin: Springer.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6), 1152–1174.

Aral, S. O. & Eterman, T. A. (2002). A stratified approach to untangling the behavioral/biomedical outcomes conundrum. *Sexually Transmitted Diseases*, 29(9), 530–532.

Blackwell, D. & MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes. *The Annals of Statistics*, 1(2), 353–355.

Browne, W. J. & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1, 473–514.

Butler, D. M. & Smith, D. M. (2007). Serosorting can potentially increase HIV transmissions. *AIDS*, 21(9), 1218–1220.

Caron, F., Davy, M., & Doucet, A. (2007). Generalized Polya urn for time-varying Dirichlet process mixtures. In *Proceedings of the Twenty-Third Annual Conference on Uncertainty in Artificial Intelligence*: AUAI Press.

Casella, G. & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), 167–174.

Cassels, S., Menza, T. W., Goodreau, S. M., & Golden, M. R. (2009). HIV serosorting as a harm reduction strategy: evidence from Seattle, Washington. *AIDS*, 23(18), 2497–2506.

Chib, S. & Winkelmann, R. (2001). Markov chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4), 428–435.

Cox, J., Beauchemin, J., & Allard, R. (2004). HIV status of sexual partners is more important than antiretroviral treatment related perceptions for risk taking by HIV positive MSM in Montreal, Canada. *Sexually Transmitted Infections*, 80(6), 518–523.

De Iorio, M., Muller, P., Rosner, G. L., & MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association*, 99, 205–215.

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2), 355–366.

Fahrmeir, L. & Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(2), 201–220.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2), 209–230.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

Gibson, P., Abramson, M., Wood-Baker, R., Volmink, J., Hensley, M., & Costabel, U. (2008). *Evidence-Based Respiratory Medicine with CD-ROM*. Evidence-Based Medicine. John Wiley & Sons.

Gilbert, P., Ciccarone, D., Gansky, S. A., Bangsberg, D. R., Clanon, K., McPhee, S. J., CalderÃşn, S. H., Bogetz, A., & Gerbert, B. (2008). Interactive "Video Doctor" counseling reduces drug and sexual risk behaviors among HIV-positive patients in diverse outpatient settings. *PLoS ONE*, 3(4), 1–10.

Glynn, J. R. (1998). Resurgence of tuberculosis and the impact of HIV infection. *British Medical Bulletin*, 54(3), 579–593.

Golden, M. R., Stekler, J., Hughes, J. P., & Wood, R. W. (2008). HIV serosorting in men who have sex with men: is it safe? *Journal of Acquired Immune Deficiency Syndromes*, 49(2), 212–218.

Goldstein, H. (2010). *Multilevel Statistical Models*. John Wiley & Sons, Ltd.

Goldstein, H., Carpenter, J., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling*, 9(3), 173–197.

Goldstein, H. & Kounali, D. (2009). Multilevel multivariate modelling of childhood growth, numbers of growth measurements and adult characteristics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3), 599–613.

Grant, R. M., Lama, J. R., Anderson, P. L., McMahan, V., Liu, A. Y., Vargas, L., Goicochea, P., Casapía, M., Guanira-Carranza, J. V., Ramirez-Cardich, M. E., Montoya-Herrera, O., Fernández, T., Veloso, V. G., Buchbinder, S. P., Chariyalertsak, S., Schechter, M., Bekker, L.-G., Mayer, K. H., Kallás, E. G., Amico, K. R., Mulligan, K., Bushman, L. R., Hance, R. J., Ganoza, C., Defechereux, P., Postle, B., Wang, F., McConnell, J. J., Zheng, J.-H., Lee, J., Rooney, J. F., Jaffe, H. S., Martinez, A. I., Burns, D. N., & Glidden, D. V. (2010). Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *New England Journal of Medicine*, 363(27), 2587–2599.

Griffin, J. E. & Steel, M. F. J. (2006). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, 101(473), 179–194.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.

Jensen, S. T. & Shore, S. H. (2011). Semiparametric Bayesian modeling of income volatility heterogeneity. *Journal of the American Statistical Association*, 106, 1280–1290.

Jin, F., Crawford, J., Prestage, G. P., Zablotska, I., Imrie, J., Kippax, S. C., Kaldor, J. M., & Grulich, A. E. (2009). Unprotected anal intercourse, risk reduction behaviours, and subsequent HIV infection in a cohort of homosexual men. *AIDS*, 23(2), 243–252.

Kalichman, S. C., Rompa, D., Cage, M., DiFonzo, K., Simpson, D., Austin, J., Luke, W., Buckles, J., Kyomugisha, F., Benotsch, E., Pinkerton, S., & Graham, J. (2001). Effective-

ness of an intervention to reduce HIV transmission risks in HIV-positive people. *American Journal of Preventive Medicine*, 21(2), 84–92.

Kleinman, K. P. & Ibrahim, J. G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics*, 54(3), pp. 921–938.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science* (pp. 50–55). Alexandria, VA: American Statistical Association.

MacEachern, S. N. (2000). *Dependent Dirichlet Processes*. Technical report, Department of Statistics, The Ohio State University.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9, 249–265.

Parsons, J. T., Schrimshaw, E. W., Wolitski, R. J., Halkitis, P. N., Purcell, D. W., Hoff, C. C., & Gómez, C. A. (2005). Sexual harm reduction practices of HIV-seropositive gay and bisexual men: serosorting, strategic positioning, and withdrawal before ejaculation. *AIDS*, 19(S1), S13–S25.

Pinkerton, S. D. (2008). Acute HIV infection increases the dangers of serosorting. *American Journal of Preventive Medicine*, 35(2), 184.

Poudel, K. C., Poudel-Tandukar, K., Yasuoka, J., & Jimba, M. (2007). HIV superinfection: another reason to avoid serosorting practice. *Lancet*, 370(9581), 23.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128(2), 301–323.

Reniers, G. & Helleringer, S. (2011). Serosorting and the evaluation of HIV testing and counseling for HIV prevention in generalized epidemics. *AIDS and Behavior*, 15(1), 1–8.

Rodríguez, A., Dunson, D. B., & Gelfand, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association*, 103(483), 1131–1154.

Rue, H. & Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman & Hall.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639–650.

Sikkema, K., Hansen, N., Kochman, A., Tarakeshwar, N., Neufeld, S., Meade, C., & Fox, A. (2007). Outcomes from a group intervention for coping with HIV/AIDS and childhood sexual abuse: Reductions in traumatic stress. *AIDS and Behavior*, 11, 49–60.

Snowden, J. M., Raymond, H. F., & McFarland, W. (2009). Prevalence of seroadaptive behaviours of men who have sex with men, San Francisco, 2004. *Sexually Transmitted Infections*, 85(6), 469–476.

Snowden, J. M., Raymond, H. F., & McFarland, W. (2011). Seroadaptive behaviours among men who have sex with men in San Francisco: the situation in 2008. *Sexually Transmitted Infections*, 87(2), 162–164.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476), 1566–1581.

The Healthy Living Project Team (2007). Effects of a behavioral intervention to reduce risk of transmission among people living with HIV: The Healthy Living Project randomized controlled study. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 44(2), 213–221.

Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. *Austrian Journal of Statistics*, 31(2-3), 221–229.

Warner, L., Newman, D. R., Austin, H. D., Kamb, M. L., Douglas, J. M., Malotte, C. K., Zenilman, J. M., Rogers, J., Bolan, G., Fishbein, M., Kleinbaum, D. G., Macaluso, M., Peterman, T. A., & for the Project RESPECT Study Group (2004). Condom effectiveness for reducing transmission of gonorrhea and chlamydia: The importance of assessing partner infection status. *American Journal of Epidemiology*, 159(3), 242–251.

West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (2nd ed.)*. New York: Springer-Verlag.

West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80(389), 73–83.

Wilson, D. P., Regan, D. G., Heymer, K.-J., Jin, F., Prestage, G. P., & Grulich, A. E. (2010). Serosorting may increase the risk of HIV acquisition among men who have sex with men. *Sexually Transmitted Diseases*, 37(1), 13–17.

Zhu, X., Ghahramani, Z., & Lafferty, J. (2005). *Time-Sensitive Dirichlet Process Mixture Models*. Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.